



RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Kointegration in Theorie und Praxis:
Statistische Analyse gemeinsamer Entwicklungstrends in
psychologischen Zeitreihensystemen

Dissertation

zur Erlangung des akademischen Grades
doctor philosophiae (Dr. phil.)

vorgelegt von

Dipl.-Math.

Antje Birgit Gruber

geboren am 07. November 1983 in Schwetzingen

Heidelberg, den 7. März 2011

Betreuer und Gutachter: Prof. Dr. Joachim Werner
(Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg)
Zweiter Gutachter: Prof. Dr. Andreas Voß
(Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg)

Zusammenfassung

Die vorliegende Dissertation *Kointegration in Theorie und Praxis: Statistische Analyse gemeinsamer Entwicklungstrends in psychologischen Zeitreihensystemen* besteht aus drei Teilen.

Die ersten beiden Kapitel dienen der theoretischen Darstellung der Kointegrationsanalyse. Hier werden die notwendigen statistischen Kenntnisse zur Zeitreihenanalyse und dem Phänomen der Kointegration vermittelt, um den psychologischen Anwender in die Lage zu versetzen die vorgestellten Verfahren bei empirischen Daten einzusetzen. Die typische Einführung in die Kointegrationstheorie wendet sich an Mathematiker oder Anwender mit entsprechendem Expertenwissen. So ist die Darstellung meist sehr verkürzt und Zusammenhänge zwischen einzelnen Methoden sind nicht ersichtlicht.

Ziel dieser Arbeit ist es, die Methode für Sozialwissenschaftler zugänglich zu machen und durch die didaktische Auf- und Überarbeitung die Verbindung der einzelnen Ansätze herauszustellen. So unterscheidet sich diese Darstellung durch Anschaulichkeit und Ausführlichkeit von den üblichen Einführungen in die Zeitreihen- und Kointegrationsanalyse, ohne dabei an mathematischer Exaktheit zu verlieren. Sie kann sowohl als Einstieg in die Theorie als auch für den mit der Methode vertrauten Leser als Nachschlagewerk dienen.

Im dritten Kapitel wird in ausführlichen Simulationsstudien untersucht, ob sich die im ersten Teil eingeführten Analysemethoden auch bei einer Datenqualität einsetzen lassen, die in der psychologischen Forschung typisch ist. Hier werden Fragen zur notwendigen Länge der Systeme, zur Anwendbarkeit auf verschiedenen Skalenniveaus oder dem Umgang mit fehlenden Messwerten beantwortet.

Ein mathematischer Anhang rundet die Arbeit ab und ermöglicht es dem interessierten Leser mit grundlegenden Kenntnissen in Wahrscheinlichkeitstheorie und Regressionsanalyse zu jedem Kapitel weitere mathematisch komplexere Zusammenhänge zu ergänzen, ohne weitere Hilfsmittel hinzunehmen zu müssen.

Inhaltsverzeichnis

Einleitung	9
1 Zeitreihenanalyse: Theoretischer Hintergrund	15
1.1 Zeitreihen zur Beschreibung von Verlaufsdaten	15
1.2 Stationarität: stabile Prozesse	19
1.3 Moving Average Darstellung: Einfluss externer Innovationen	22
1.4 Autoregressive Darstellung: Analyse der internen Struktur	25
1.5 Lag Operationen	26
1.6 Wolddarstellung	29
1.7 Stationaritätsbedingung für autoregressive Prozesse	31
1.8 Random Walk: ein erstes dynamisches Prozessmodell	33
1.9 Autokorrelationsfunktion	35
1.10 Partielle Autokorrelationsfunktion	39
1.11 Anzahl der Lags: Modellwahl	41
1.12 Unit Root Tests zur Erkennung von Stationarität	42
1.13 Multivariate Zeitreihen zur Beschreibung von Zeitreihensystemen	47
1.14 Lagpolynommatrizen	50
1.15 Impulsantwortfunktionen: langfristige Wirkung externer Schocks	52
2 Kointegration: gemeinsame Trends in Zeitreihensystemen	55
2.1 Integrierte Prozesse: Zeitreihen mit stochastischem Trend	59
2.1.1 Stationäre Prozesse	59
2.1.2 Trendstationäre Prozesse	60
2.1.3 Differenzenstationäre Prozesse	61
2.1.4 Integrierte Prozesse	63
2.1.5 Stationäre und differenzenstationäre Prozesse	64
2.2 Kointegrierte Prozesse: Systeme mit gemeinsamen Trends	66
2.2.1 Der Begriff der Kointegration	66
2.2.2 Kointegrationsbeziehungen	68
2.2.3 Gemeinsamer Trend - Anschauung	72
2.2.4 Spurious Regression	79
2.3 Darstellungsformen kointegrierter Systeme	81
2.3.1 Granger Repräsentationstheorem	81
2.3.2 Moving-Average- und Common-Trend-Darstellung	85
2.3.3 Vektorautoregressive und Fehlerkorrekturdarstellung	90
2.3.4 Weitere Darstellungen eines kointegrierten Systems	94

2.4	A Drunk and her Dog: Fehlerkorrekturmechanismen am konstruierten Beispiel	95
2.4.1	Random Walk or Drunkard's Walk	95
2.4.2	A Drunk and her Dog	96
2.4.3	A Drunk, her Dog and a Boyfriend	98
2.5	Fehlerkorrekturmodell: langfristiges Gleichgewicht und kurzfristige Dynamik	100
2.5.1	Der allgemeine bivariate Fall	101
2.5.2	Fehlerkorrektur im bivariaten System	103
2.5.3	Der multivariate Fall am dreidimensionalen Beispiel	112
2.6	Deterministische Trends in kointegrierten Prozessen	114
2.6.1	Übersicht der Modelle	115
2.6.2	Berechnungen	118
2.7	Kointegrationstests	124
2.7.1	Tests bei Kointegrationsrang Eins	124
2.7.2	Johansen-Test	127
2.8	Couples in Collusion - Kointegrationsanalyse in psychologischer Praxis	138
2.8.1	Kollusionstheorie	138
2.8.2	Kollusion in der Psychosomatik	139
2.8.3	Interpretation durch Fehlerkorrekturdarstellung	140
3	Simulationsstudien: Kointegration in psychologischen Daten	145
3.1	Datenbasis und Auswertungskriterien	148
3.1.1	Datenbasis	149
3.1.2	Auswertungskriterien	151
3.2	Die Länge einer Zeitreihe	160
3.2.1	Bekannter Kointegrationsvektor	160
3.2.2	Johansen-Test: Klassifikation der Systeme verschiedener Länge	169
3.2.3	Ergebnis: Länge einer Zeitreihe	181
3.3	Messung auf Intervallskalenniveau	182
3.3.1	Vorstudie: diskretisierte Daten	184
3.3.2	10-stufige Intervallskala	184
3.3.3	7-stufige Intervallskala	193
3.3.4	5-stufige Intervallskala	200
3.3.5	3-stufige Intervallskala	205
3.3.6	Ergebnis: Messung auf Intervallskalenniveau	211
3.4	Ceiling- und Flooreffekte	212
3.4.1	Von einer Seite beschränkte Skalen	212
3.4.2	Bis zur Hälfte beschränkte Skala	219
3.4.3	Von beiden Seiten beschränkt	226
3.4.4	Ergebnis: Messung mit Ceiling- und Flooreffekten	231
3.5	Messung auf Ordinalskalenniveau	231
3.5.1	Ordinalskala 1: „symmetrisch um Median gewichtet“	231

3.5.2	Ordinalskala 2: „symmetrisch mit Differenzierung der Randbereiche“	237
3.5.3	Ordinalskala 3: „wachsende Stufenabstände“	243
3.5.4	Ordinalskala 4: „unsystematisch variiertes Stufenabstände“	248
3.5.5	Ergebnis: Messung auf Ordinalskalenniveau	254
3.6	Missing Values	255
3.6.1	Regelmäßiges Entfernen jeder 10. Messung	255
3.6.2	Regelmäßiges Entfernen jeder 7. Messung	261
3.6.3	Regelmäßiges Entfernen jeder 5. Messung	266
3.6.4	Regelmäßiges Entfernen jeder 3. Messung	270
3.6.5	Zufälliges Entfernen eines Zehntels der Messungen	275
3.6.6	Zufälliges Entfernen eines Siebtels der Messungen	281
3.6.7	Zufälliges Entfernen eines Fünftels der Messungen	286
3.6.8	Zufälliges Entfernen eines Drittels der Messungen	290
3.6.9	Ergebnis: Messung mit Missing Values	294
3.7	Ergebnisse der Simulationsstudien	295
 Ausblick		 301
 A Mathematischer Anhang		 305
A.1	Algebraische Strukturen	305
A.2	Einführung in die komplexen Zahlen	317
A.3	Verallgemeinerte Binomialkoeffizienten	324
 B Zeitreihenanalyse		 331
B.1	Stationarität	331
B.2	Lag Operationen	332
B.3	Wolddarstellung	334
B.4	Impulsantwortfunktion	335
 C Kointegrationsanalyse		 337
C.1	Integrierte Prozesse	337
C.2	Darstellungsformen	341
C.3	Kointegrationstests	351
 D Simulationsstudien		 357
D.1	Die Länge einer Zeitreihe	357
D.2	Messung auf Intervallskalenniveau	360
D.3	Boxplot-Statistiken für die Schätzung von β	364
D.4	Erstellung der Testsets	386
D.5	Allgemeine Auswertung	390
D.6	Funktionen zur Datentransformation	394
 Literaturverzeichnis		 403

Einleitung

Zeitreihen sind allgegenwärtig. Denn viele verschiedene Phänomene oder Kennzahlen werden in regelmäßigen Abständen gemessen und notiert. Das wohl prominenteste Beispiel sind die Kurse von Aktien, wie sie täglich in den Zeitungen abgebildet werden. Außer diesen Börsenkursen zeichnet die deutsche Bundesbank weitere Zeitreihendaten über Kredithöhen, Zinsen, Renditen, Goldpreise, Außenwirtschaft oder Konjunkturdaten bereits über einen Zeitraum von Jahren oder gar Jahrzehnten auf (Deutsche Bundesbank, 2011).

Doch auch in vielen psychologischen oder der Psychologie nahen Bereichen werden Zeitreihendaten erhoben und entsprechend analysiert (z.B. Umweltpsychologie Bullinger, 1989; Musiktherapie, Ferstl, 2006). Zur Unterscheidung verschiedener Schlafphasen werden Zeitreihenmuster verwendet (Gottman, 1981, S.3), man kann die Bewertung von Schmerzen durch den Patienten (z.B. bei Migräne Gottman, 1981, S.34 Schmitz & Otto, 1984 oder Neurodermitis, auch Werner, 2005) mit Hilfe eines Schmerztagebuches als Zeitreihe erheben und anschließend auf die interne Struktur, Perioden oder enthaltene Trends untersuchen. Das Interaktionsverhalten zwischen Mutter und Kind (Lester, Hoffman & Brazelton, 1985) oder Gleichaltrigen (Gottman, 1981, S.38ff) kann genauso analysiert werden wie die Wort-Assoziationen bei einem schizophrenen Patienten während des Therapieprozesses (Moseley, 1963) oder die Entwicklung des Selbstwerts (z.B. Delignières, Fortes & Ninot, 2004).

Zur Wettervorhersage werden zum Teil bereits seit Jahrhunderten Klimadaten aufgezeichnet (Deutscher Wetterdienst, 2011). Daraus lassen sich auch mögliche Ursachen für den aktuellen CO_2 -Anstieg in der Atmosphäre bestimmen (Frank et al., 2010). In der Geophysik werden Zeitreihen zur Vorhersage von Erdbeben verwendet (Southern California Earthquake Data Center, 2011). Des Weiteren gibt es Anwendungen von Zeitreihen bei der Berechnung des Ruinrisikos einer Versicherungsgesellschaft (Beard, Pentikainen & Pesonen, 1977; Bühlmann, 1970) oder zur Optimierung von Lagerhaltungskosten (Hochstädter, 1969; Girlich, 1973).

Monatlich werden die aktuellen Arbeitslosenzahlen berichtet, stets im Vergleich mit den Werten des Vormonats und desselben Monats im Vorjahr (Bundesagentur für Arbeit, 2011). Auf den Webseiten des Bundeskriminalamtes (Bundeskriminalamt, 2011) sind Kriminalstatistiken für die Jahre 1987-2009 einsehbar und das Statistische Bundesamt hält viele Zeitreihendaten über Bevölkerungswachstum, Geburten- und Sterberaten oder Verbraucherpreise bereit (Statistisches Bundesamt Deutschland, 2011).

Während viele dieser Daten vor allem deskriptive Verwendung finden, besteht bei zahlreichen ökonomischen Zeitreihendaten ein wirtschaftlich-politisches Interesse an der Prognose des zukünftigen Verlaufs. Deshalb hat sich die Zeitreihenanalyse in den Wirtschaftswissenschaften zu einem Standardverfahren entwickelt (z.B. Taylor, 2005; Wagatha, 2005; Mills & Markellos, 2008). So werden viele ökonomische Größen regelmäßig und über lange Zeiträume

zuverlässig aufgezeichnet. Es gibt Daten zum Bruttoinlandsprodukt verschiedener Länder, zum Verlauf des Ölpreises, Wechselkursen, Geldmarktzinsen, Inflationsraten und vielem mehr (z.B. Schröder & Buscher, 2002; Deutsche Bundesbank, 2011).

Da man in der psychologischen Forschung häufig an kausalen Zusammenhängen oder Gruppenvergleichen interessiert ist, werden vor allem statistische Verfahren wie die Regressions- oder Faktorenanalyse zur Auswertung der Ergebnisse verwendet (siehe u.v. Fahrmeir, Kneib & Lang, 2007; Werner, 1997). Die meisten dieser Verfahren setzen die Unabhängigkeit der vorliegenden Daten voraus. Zeitreihendaten besitzen jedoch meist eine innere Abhängigkeitsstruktur, so dass die üblichen Verfahren nicht eingesetzt werden können (Kratochwill, 1978; Wagenmakers, Farrell & Ratcliff, 2005). Die Auswertung von Verlaufsdaten mittels der Inferenzstatistik kann so zu Widersprüchen oder gar falschen Ergebnissen führen (Petermann, 1978, 1989), die herkömmlichen kritischen Werte in der ANOVA können nicht mehr verwendet werden (Glass, Willson & Gottman, 1975; Sharpley & Alavosius, 1988). Häufig ist man in der Psychologie jedoch an der Untersuchung und Erklärung der interindividuellen Variabilität diverser Merkmale interessiert, so dass die *intraindividuelle* Variabilität als Störgröße betrachtet wird (z.B. Lienert & Raatz, 1998) oder zu Lasten der Testreliabilität geht (Petermann, 1978).

Zeitreihen enthalten einen wichtigen Informationsaspekt, der in Stichprobendaten nicht vorkommt - die Zeit. Bürgy (1998) bemerkt, dass durch die Aggregation von Verlaufsdaten zu Mittelwerten und anschließender inferenzstatistischer Analyse (wie z.B. bei Cooney, Clarke & Morris, 1986, oder Fahrenberg, 1983) wichtige Informationen wie Trends, Periodizität und serielle Abhängigkeiten verloren gehen, so dass man die Interpretierbarkeit der erhaltenen Kenngrößen in Frage stellen muss.

Dadurch, dass die zeitliche Reihenfolge der einzelnen Messungen vorgegeben ist, ergeben sich komplexe Abhängigkeitsmuster, die mit den Standardverfahren nicht greifbar sind. Auch die Ergebnisse einer Regressionsanalyse mit Zeitreihendaten können stark verfälscht sein. So kann es durch bestimmte dynamische Aspekte in den Reihen zu einer Scheinregression (Granger & Newbold, 1974) kommen, obwohl die Reihen unabhängig voneinander sind. Wir werden dieses Phänomen in Kapitel 2.2 genauer betrachten.

Man braucht folglich spezielle statistische Verfahren, um Zeitreihendaten adäquat analysieren zu können. Seit Jahrzehnten gibt es intensive Forschungsarbeit in diesem Bereich, der vor allem in den Wirtschaftswissenschaften (Box & Jenkins, 1976), aber auch zunehmende in anderen Disziplinen Anwendung findet. Durch die Werke von Glass et al. (1975) und Gottman (1981) wurde die Zeitreihenanalyse auch in die Psychologie eingeführt und findet dort zunehmend Verwendung. Schon Lasswell (1937) betrachtete den Therapieverlauf bereits als Zeitreihe. Deshalb verwundert es nicht, dass es vor allem in der Therapieforschung zahlreiche Beiträge mit zeitreihenanalytischer Auswertung gibt (u.v. Elkins & Moore, 2011; Dugas, Francis & Bouchard, 2009; Tschacher & Ramseyer, 2009; Hull, Clarkin & Alexopoulos, 1993).

Da bei psychologischen Messungen in relativ kurzer Zeit an vielen Messzeitpunkten erhoben werden kann, eignen sich auch Fragestellungen im Zusammenhang mit Daten wie Pulsfrequenz, Hautwiderstand und -temperatur oder aus den Neurowissenschaften zur zeitreihenanalytischen Auswertung (z.B. Bürgy, 1998; Link, Endt, Oeff & Trahms, 2000).

Häufig ist man nicht nur an den Abhängigkeiten innerhalb eines zeitlichen Verlaufs interessiert, sondern möchte vielmehr Systeme von Zeitreihen analysieren. Ähnlich wie bei der Regressionsanalyse sind Aussagen über mögliche Korrelationen notwendig, um Zusammenhänge zwischen verschiedenen Phänomenen zu entdecken und erklären zu können. Ein besonderes multivariates Verfahren, die *Kointegrationsanalyse*, ermöglicht es eine langfristige Gleichgewichtsbeziehung zwischen mehreren Zeitreihen zu erkennen und bietet somit eine Möglichkeit, Scheinregression von wahren Zusammenhängen zu unterscheiden (C. W. J. Granger & Lee, 1990).

Im Jahr 2003 erhielt Sir Clive William John Granger „for methods of analyzing economic time series with common trends (cointegration)“¹ den Nobelpreis für Wirtschaftswissenschaften (Nobelstiftung, 2003). Ausgehend von sogenannten Fehlerkorrekturmodellen konnte er zeigen, dass viele Zeitreihen, die bei ersten grafischen Betrachtungen einen unregelmäßigen Verlauf zeigen, als System verschiedene gemeinsame Komponenten und dadurch ein stabiles langfristiges Gleichgewicht (*Kointegrationsbeziehung*) besitzen.

Mittlerweile gibt es eine unüberschaubare Menge an Beiträgen zum Phänomen der Kointegration² (z.B. Engle, 1991). Vor allem in den Wirtschaftswissenschaften ist man an der Anwendung und Weiterentwicklung der Verfahren der Kointegrationsanalyse interessiert (Herrmann, 2005). So wird mittels Kointegrationsanalyse die Beziehung zwischen verschiedenen europäischen Aktienmärkten untersucht (u.a. Kasa, 1997; Cai, Chou & Li, 2009). Teilweise werden in Sonderausgaben einiger Fachzeitschriften Artikel zur Kointegration gesammelt (z.B. Aoki, 1988, *Econometrics*, 2010).

Zunehmend finden die Verfahren auch in den Sozialwissenschaften Verwendung. So gibt es diverse Studien, die einen Zusammenhang der weiblichen Arbeitskraftbeteiligung mit der Scheidungsrate (Bremmer & Kesselring, 2004) oder einer altersabhängigen Geburtenrate (McNown, 2003) anhand von Kointegrationsbeziehungen ermittelt haben. Lin und Brannigan (2003) untersuchten den Zusammenhang zwischen Kriminalität und Immigration, während Luiz (2001) die Kriminalitätsdynamik in Südafrika anhand von ökonomischen Determinanten mit Hilfe der Kointegrationsanalyse erklären konnte. Weitere Ergebnisse in der kriminologischen Forschung konnten mittels aufgedeckten oder widerlegten Kointegrationsbeziehungen gewonnen werden (Witt & Witte, 2000; O'Brien, 1999) und Darrat und Al-Yousif (1999) konnten zeigen, dass das Bevölkerungswachstum in Entwicklungsländern in einem positiven Zusammenhang mit der dortigen wirtschaftlichen Entwicklung steht. Andrés und Halicioğlu

¹deutsch etwa: „für seine Arbeit an Methoden zur Analyse ökonomischer Zeitreihen mit gemeinsam veränderlichen Trends (Kointegration)“

²Suche zum Begriff „cointegration“ in der wirtschaftswissenschaftlichen Datenbank *EconLit* ergibt über 6000 Beiträge

(2010) zeigten mittels Kointegrationsanalyse Verbindungen zwischen der Selbstmordrate in Dänemark und ökonomischen Faktoren oder der Scheidungsrate.

Durch die Veröffentlichungen der Arbeitsgruppe „Zeitreihenanalyse“ am Psychologischen Institut der Ruprecht-Karls-Universität in Heidelberg unter Leitung von Herrn Prof. Dr. Werner und insbesondere im Rahmen der Promotion von Frau Dr. Stroe-Kunold (Stroe-Kunold, 2010, siehe auch Stroe-Kunold, 2005) wurde das Verfahren verstärkt in die psychologische Forschung eingeführt.

In einem ersten Überblicksartikel (Stroe-Kunold & Werner, 2007) wurde die Idee der Kointegration dargestellt und die Anwendung der Methode für psychologische Fragestellungen motiviert. In Stroe-Kunold und Werner (2008) wurde eine Anleitung zur Verwendung der Kointegrationsanalyse mittels des Fehlerkorrekturmodells präsentiert. Basierend auf Monte Carlo Simulationen wurden erste Überlegungen zur Bedeutung der beiden wichtigsten Parameter des Modells getroffen. Schließlich wurde hier erstmals ein psychologischer Datensatz mittels Kointegrationsanalyse ausgewertet. In einem weiteren Artikel greifen die Autoren die erwähnte Problematik der Scheinregression in Abgrenzung zu Kointegration auf (Stroe-Kunold & Werner, 2009).

Aufbauend auf diesen Erkenntnissen setzt die erste gemeinsame Arbeit „Cointegration Methodology for Psychological Researchers: An Introduction to the Analysis of Dynamic Process Systems“ (Stroe-Kunold, Gruber, Stadnytska, Werner & Brosig, 2010) an. Darin wird versucht das Fehlerkorrektur-Modell, das für die Kointegrationsanalyse besonders geeignet ist, möglichst detailliert und für den Anwender anschaulich zu beschreiben und dabei auch das mathematische Modell ausführlich zu analysieren, um dem Leser eine Interpretation der geschätzten Fehlerkorrektur-Darstellung im Anwendungsfall zu ermöglichen.

In ähnlicher Weise soll die vorliegende Arbeit dem interessierten Anwender das notwendige mathematische Verständnis und ausreichende methodische Kenntnisse vermitteln, um eigene als Zeitreihen erhobene Daten mit diesem Verfahren auszuwerten. Nur mit dem notwendigen (mathematischen) Hintergrundwissen kann ein Anwender genug Vertrauen in eine statistische Methode entwickeln, um sie mit ausreichender Sicherheit für eigene Auswertungen zu nutzen. Außerdem muss die Frage beantwortet werden, ob die analytisch gewonnenen Erkenntnisse auf psychologische Daten anwendbar sind und so die Psychologie von der Kointegrationsanalyse profitieren kann.

Im Gegensatz zu ökonomischen Daten weisen psychologische Daten einige Besonderheiten auf. Der methodisch entscheidende Faktor ist die notwendige Länge einer Zeitreihe. Während sich bestimmte Daten wie der Wert einer Aktie problemlos täglich bestimmen lassen, ist man in psychologischen Untersuchungen häufig von dem Verhalten der Versuchspersonen abhängig. Mit Hilfe der Methoden im Rahmen des *Amulanten Assessments* (Fahrenberg, Myrtek, Pawlik & Perrez, 2007) können viele Daten zum Beispiel mittels elektronischer Tagebücher bereits regelmäßig im Alltag erfasst werden, wodurch die Erhebung vieler Merkmale über mehrere Tage und Wochen vereinfacht wird. Je länger der Messzeitraum, desto höher

ist jedoch das Risiko eines vorzeitigen Abbruchs. Deshalb ist die wichtigste Frage, wie lang ein System von Zeitreihen sein muss, damit man die Ergebnisse einer Kointegrationsanalyse sinnvoll verwerten kann.

Des Weiteren lässt es die Operationalisierung der meisten psychologischen Merkmale nicht zu, sie beliebig genau zu bestimmen, wie das beispielsweise bei einer in Geldeinheiten gemessenen Größe der Fall ist. Vielmehr werden häufig Messinstrumente wie Fragebögen mit wenigen, diskreten Antwortmöglichkeiten verwendet. Hier ist zu klären, wie viele Punkte auf der Werteskala notwendig sind, um einen stochastischen Trend und damit Kointegration in den Daten erkennen zu können. Zudem ist bei vielen dieser Instrumente die Skalenqualität zu beachten. Während Geldbeträge auf einer Verhältnisskala gemessen werden, hat man bei psychologischen Daten häufig nur Ordinalskalenqualität. Kann bei diesen Voraussetzungen eine Kointegrationsanalyse sinnvoll eingesetzt werden?

Die Gefahr eines vorzeitigen Abbruchs wurde bereits erwähnt. Häufig sind es jedoch nur einzelne Messungen, die aus den verschiedensten Gründen nicht erhoben werden können. Wie sollen diese *missing values* in der Auswertung berücksichtigt werden? Bedeutet ein ausgebliebener Wert automatisch das Ende der erhobenen Zeitreihe und den Neubeginn zum nächsten Messzeitpunkt?

Alle diese Fragen stellen sich dem psychologischen Anwender, wenn er die Kointegrationsanalyse auf seine Daten anwenden möchte. Die Simulationsstudien dieser Arbeit geben erste Antworten.

Im ersten Kapitel dieser Dissertation werden die Grundkenntnisse der Zeitreihenanalyse vermittelt, auf denen die Kointegrationsanalyse beruht. Es wird nur auf jene Methoden eingegangen, die später verwendet werden. Dazu werden Grundkenntnisse in Statistik und insbesondere Regressionsanalyse vorausgesetzt. Nachdem zunächst verschiedene Methoden zur univariaten Analyse *einer* Zeitreihe beschrieben werden, folgen in den Kapitel 2.1 bis 2.7 Verfahren zur multivariaten Auswertungen von *Zeitreihensystemen*.

An die Einführung in die allgemeine Zeitreihenanalyse schließt sich die theoretische Darstellung der Kointegrationsanalyse in Kapitel 2 an. Aufbauend auf der Darstellung *integrierter* Prozesse in Kapitel 2.1 folgt die anschauliche Einführung des Begriffs der *kointegrierten* Prozesse und die Darstellung des Problems der Scheinregression in Kapitel 2.2.

Nachdem die verschiedenen Darstellungsformen kointegrierter Systeme, die in ihnen enthaltene Information und die Umrechnung ineinander vorgestellt wurden (Kapitel 2.3), wird das gebräuchlichste Modell, das vektorielle Fehlerkorrekturmodell, an dem von Murray (1994) konstruierten Beispiel einer „Betrunkenen und ihres Hundes“ verdeutlicht, das als Leitfaden für die psychologische Interpretation dienen kann (Kapitel 2.4) und anschließend in Kapitel 2.5 genau analysiert (vgl. auch Stroe-Kunold, Gruber et.al., 2010).

Des Weiteren wird in Kapitel 2.6 gezeigt, wie man Mittelwertsunterschiede zwischen den einzelnen Reihen des Systems oder auch lineare Trends im Kointegrationsmodell berück-

sichtigen kann. Der theoretische Teil wird mit einer ausführlichen Erläuterung verschiedener Kointegrationstests beendet (Kapitel 2.7). Dabei liegt der Schwerpunkt auf der Darstellung des Johansenstests (z.B. Johansen, 1988), der in der Praxis die Grundlage für die Kointegrationsanalyse bildet. Abschließend wird die Anwendung dieses Verfahrens an einem empirischen Datensatz demonstriert, wodurch Zusammenhänge zwischen den psychologischen Merkmalen und Krankheitssymptomen eines Ehepaars aufgedeckt werden können.

Im dritten Kapitel wird mit Hilfe von Simulationsstudien untersucht, welche Merkmale die vorliegenden Daten erfüllen müssen, damit eine Kointegrationsanalyse mit dem Johansen-Test sinnvoll angewendet werden kann. Beginnend mit der notwendigen Länge des Systems (Kapitel 3.2) werden auch die erwähnten Fragen nach der Skalenbeschaffenheit (Anzahl der Stufen und Skalenqualität), sowie nach dem Umgang mit *missing values* geklärt.

Da im Umfang dieser Arbeit viele weitere Aspekte der Kointegrationsanalyse unberücksichtigt bleiben müssen, soll abschließend ein Ausblick auf einige Erweiterungen der Kointegration, z.B. bei saisonalen Daten, gegeben und zu weiteren Untersuchungen im Bezug auf psychologische Daten motiviert werden.

1 Zeitreihenanalyse: Theoretischer Hintergrund

Mit der Veröffentlichung von Box und Jenkins (1976) hatte die Zeitreihenanalyse ihren Durchbruch in den Wirtschaftswissenschaften. Kurze Zeit später erfolgte durch die Einführung von vektoriiell autoregressiven Prozessen die Erweiterung in den multivariaten Bereich (Sims, 1980). Mittlerweile findet man zunehmend Verwendungen in den Sozialwissenschaften wie der Psychologie. So nicht nur in den Neurowissenschaften (z.B. Mezer, Yovel, Pasternak, Gorfine & Assaf, 2009), sondern auch zur Dokumentation eines Therapieverlaufs (z.B. J. D. Smith, Handler & Nash, 2010; Kupfer, Brosig & Brähler, 2005).

Da die Kointegrationsanalyse zur Analyse gemeinsamer Dynamiken in (psychologischen) Zeitreihensystemen auf den Erkenntnissen dieser historischen Entwicklung aufbaut, ist es unerlässlich, dass der Anwender mit Grundkenntnissen der univariaten und multivariaten Zeitreihenanalyse vertraut ist, bevor er Kointegrationsmethoden zur Datenanalyse verwendet. In diesem einleitenden Kapitel soll ein ausreichendes Wissen vermittelt werden, das zum Verständnis der Konzepte der Kointegrationsanalyse notwendig ist.

1.1 Zeitreihen zur Beschreibung von Verlaufsdaten

Will man Aussagen über den zeitlichen Verlauf eines Phänomens machen, muss man die Variable, durch die das Phänomen operationalisiert wird, in regelmäßigen zeitlichen Abständen (z.B. täglich, stündlich) messen. Eine Folge von Messungen einer Variable über die Zeit nennt man *Zeitreihe* (z.B. Schlittgen & Streitberg, 1999; Assenmacher, 2002).

Definition 1.1.1 (Zeitreihe) *Als Zeitreihe bezeichnet man eine mit $t \in \mathbb{T}$ indizierte Folge von Zufallsvariablen*

$$(Y_t)_{t \in \mathbb{T}} = Y_1, Y_2, Y_3, \dots,$$

die alle Werte in einer Menge E annehmen. Die Indexmenge \mathbb{T} wird in der Zeitreihenanalyse stets als Zeitbereich behandelt.

Man unterscheidet Zeitreihen nun nach ihrem Parameterraum \mathbb{T} und ihrem Zustandsraum E (z.B. Neusser, 2006, S.7). In dieser Arbeit werden wir nur *diskrete* Zeitreihen betrachten. Der Zeitbereich ist also stets abzählbar $\mathbb{T} \subseteq \mathbb{Z}$, meist sogar, wie in der Praxis immer, endlich $\mathbb{T} = \{1, 2, \dots, T\}$. Die Theorie stetiger stochastischer Prozesse ($\mathbb{T} = \mathbb{R}$) und damit stetiger

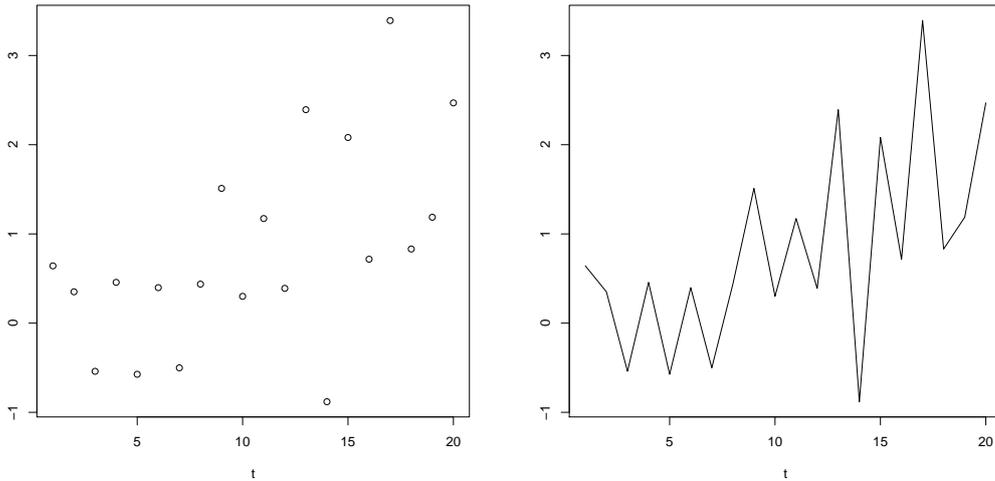


Abbildung 1.1: Eine Zeitreihe, diskret und verbunden dargestellt

Zeitreihen ist mathematisch komplexer und soll hier nicht behandelt werden. Eine Einführung ist bei Klenke (2006) zu finden. Dennoch werden die meisten Zeitreihen grafisch nicht als diskrete Punkte, sondern verbunden als stetige Funktion der Zeit dargestellt (vgl. Abbildung 1.1), wodurch die Verlaufsbetrachtung erleichtert wird.

Die Zufallsvariablen haben alle Werte in demselben Zustandsraum E . Während in den Wirtschaftswissenschaften meist reellwertige Zeitreihen mit $E = \mathbb{R}$ betrachtet werden (vgl. Schröder & Buscher, 2002), nehmen psychologische Messwerte häufig Werte auf einer Skala mit endlich vielen Stufen an. Für das Phänomen der Kointegration werden wir diesen Unterschied in Kapitel 3 genau untersuchen. In Kapitel 1.13 werden wir die Definition auf vektorielle Zeitreihen erweitern, E also als einen mehrdimensionalen Zustandsraum annehmen. An dieser Stelle konzentrieren wir uns auf eine eindimensionale, aber beliebige Zustandsmenge.

Formal ist jede Zufallsvariable einer Zeitreihe eine Funktion auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) (Stier, 2001, V.2 Stochastische Prozesse). Meist wird der zugrunde liegende Wahrscheinlichkeitsraum, bestehend aus der Menge der Elementarereignisse Ω , einer auf ihm definierten Sigmaalgebra \mathcal{F} und einem Wahrscheinlichkeitsmaß P gar nicht erwähnt, so dass in diesem Fall angenommen wird, dass alle Zufallsvariablen auf demselben Wahrscheinlichkeitsraum definiert sind (Lütkepohl, 2005, 1.2 Some Basics). Wir werden unter einer Zeitreihe also die Realisierung eines (diskreten) stochastischen Prozesses verstehen, der auch als *Daten generierender Prozess (DGP)* bezeichnet wird (z.B. Brockwell & Davis, 1996; Lütkepohl, 2005; Klenke, 2006, Definition 9.1).

Definition 1.1.2 (Stochastischer Prozess) *Sei $\mathbb{T} \subset \mathbb{R}$. Eine Folge von Zufallsvariablen $(Y_t)_{t \in \mathbb{T}}$ auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit Werten in der Menge E heißt stochastischer Prozess mit Parameterraum \mathbb{T} und Zustandsraum E^3 .*

³Exakterweise fordert man noch, dass die Menge E zusammen mit der Borelschen Sigmaalgebra gewisse topologische Eigenschaften erfüllt, so dass (E, \mathcal{E}) ein sogenannter *polnischer Raum* ist (Klenke, 2006,

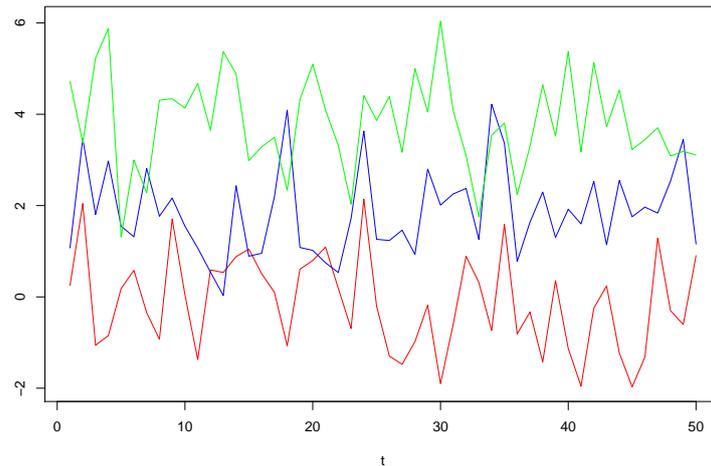


Abbildung 1.2: Drei Realisierungen eines stochastischen Prozesses

Ein stochastischer Prozess ist also eine Funktion, die sowohl von einem Wert t aus dem Parameterraum \mathbb{T} , als auch einem Elementarereignis ω aus dem zugrundeliegenden Wahrscheinlichkeitsraum Ω abhängt

$$Y : \mathbb{T} \times \Omega \rightarrow E.$$

Für einen festen Wert t erhält man die Zufallsvariable $Y_t(\omega)$ an diesem Zeitpunkt, die formal vom Elementarereignis ω abhängt. Da in der Regel in der Praxis weder der Wahrscheinlichkeitsraum noch das Elementarereignis bekannt sind, schreibt man einfach Y_t .

Hält man nun das Elementarereignis ω in seiner Betrachtung fest, ergibt sich eine Funktion von \mathbb{T} in den Wertebereich E . Für jedes ω erhält man theoretisch eine eigene Funktion, die jedem Zeitpunkt einen bestimmten Wert zuordnet. Diese Funktion von $\mathbb{T} \rightarrow E$ nennt man eine *Trajektorie* oder eine *Realisierung* des stochastischen Prozesses (Rozanov, 1975; Fahrmeir, Kaufmann & Ost, 1981; Bhattacharya & Waymire, 1990). In Abbildung 1.2 sind drei Realisierungen (für $\omega_1, \omega_2, \omega_3$) eines stochastischen Prozesses dargestellt. Durch das Zufallsergebnis wird eine Realisierung ausgewählt (Rozanov, 1975, S.140). Man kann sich vorstellen, dass man aus drei Versuchspersonen zufällig eine bestimmt, deren Reaktionsvermögen o.Ä. getestet wird. Je nachdem, welche Versuchsperson am Experiment teilnimmt, erhält man einen unterschiedlichen Verlauf.

Das Besondere in der Zeitreihenanalyse ist nun, dass man stets nur genau eine Realisierung beobachten kann, da es nicht möglich ist, die Betrachtung am selben Zeitpunkt zu wiederholen. Man hat zu jedem Zeitpunkt nur genau eine Ausprägung der Zufallsvariablen Y_t . Die Realisierung (oder ein endlicher Teil einer Realisierung) eines stochastischen Prozesses ist eine Zeitreihe in oben definiertem Sinn. Wie in der Literatur üblich (vgl. Neusser, 2006; Lütkepohl, 2005, 1.2 Some Basics) werden wir eine Zeitreihe im obigen Sinn und gegebenenfalls

S.180 und S.183). Da alle von uns verwendeten Zahlenmengen dies erfüllen, beschränken wir uns auf die Betrachtung der Menge E .

auch den stochastischen Prozess ebenfalls mit Y_t bezeichnen. Aus dem Zusammenhang wird stets eindeutig hervorgehen, ob es sich um eine einzelne Variable zum Zeitpunkt t , um den Prozess in seiner Gesamtheit oder eine einzelne Trajektorie handelt. In der Literatur wird der Begriff *Zeitreihe* sowohl für eine Realisierung als auch für den stochastischen Prozess verwendet (Neusser, 2006; Brockwell & Davis, 1996, Remark 2).

Ein stochastischer Prozess wird durch die gemeinsamen Verteilungen aller endlichen Teilmengen $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$ beschrieben (z.B. Rozanov, 1975), falls diese eine Konsistenzbedingung erfüllen (Fahrmeir et al., 1981, Definition 3.2 konsistent und Satz 3.4 Satz von Kolmogorov) (siehe Anhang B.1, S.331). Wenn man also die Verteilungen der einzelnen Zufallsvariablen ($n = 1$), aller möglichen Paare (Y_{t_1}, Y_{t_2}) , Tripel usw. in konsistenter Weise festlegt, hat man den stochastischen Prozess eindeutig definiert (Brockwell & Davis, 1996). In der Praxis ist dieses System der endlich-dimensionalen Verteilungen jedoch unbekannt, so dass man sich auf die ersten zwei Momente der Verteilungen beschränkt, also die Erwartungswerte, Varianzen und Kovarianzen (Lütkepohl, 2005, 1.2 Some Basics), die man mit einfachen Methoden schätzen kann.

Da sie theoretisch von einem Zufallsereignis abhängt, hat jede Zufallsvariable Y_t einen Erwartungswert und eine Varianz. Für die Zufallsvariablen an je zwei Zeitpunkten kann man die Kovarianz wie üblich berechnen. Betrachtet man diese Momente in Abhängigkeit von der Zeit, erhält man die *Erwartungswertfunktion* $\mu_t = \mathbb{E}[Y_t]$ und die *Varianzfunktion* $\sigma_t^2 = \text{Var}[Y_t]$, die zum Zeitpunkt t dem Erwartungswert bzw. der Varianz der Zufallsvariablen an diesem Messpunkt Y_t entspricht (z.B. Beichelt, 2003, S. 110). Um auszudrücken, dass die Kovarianzfunktion eher von den Zeitpunkten $s, t \in \mathbb{T}$, als von den eigentlichen Zufallsvariablen Y_s, Y_t abhängt, schreibt man $\gamma(s, t) = \text{Cov}[Y_s, Y_t] = \mathbb{E}[(Y_s - \mu_s)(Y_t - \mu_t)]$. Gleiches gilt für die Korrelationsfunktion $\rho(s, t) = \frac{\gamma(s, t)}{(\text{Var}[Y_s]\text{Var}[Y_t])^{1/2}}$ (z.B. Fahrmeir et al., 1981, S. 210 Def. 1.2). Häufig spricht man statt von Erwartungswertfunktion auch von Erwartungswertreihe oder Erwartungswertprozess, um die zeitliche Abhängigkeit zum Ausdruck zu bringen.

Eine Zeitreihe ist als Folge von Zufallsvariablen definiert. Dennoch wird es manchmal hilfreich oder notwendig sein, auch deterministische Entwicklungen zu betrachten. Dazu führen wir nun den Begriff der deterministischen Zeitreihe ein (Brockwell & Davis, 2006, S.180 Remark1).

Definition 1.1.3 (Deterministische Zeitreihe) *Eine Zeitreihe, die sich nicht zufällig verändert, deren Verlauf vielmehr genau durch eine Funktion f vorgeschrieben ist*

$$Y_t = f(t),$$

nennt man *deterministisch*.

Beispiel 1.1.4 (Einige deterministische Zeitreihen) *Das einfachste Beispiel einer deterministischen Reihe ist ein konstanter Prozess $Y_t = c$, der zu jedem Zeitpunkt den konstanten*

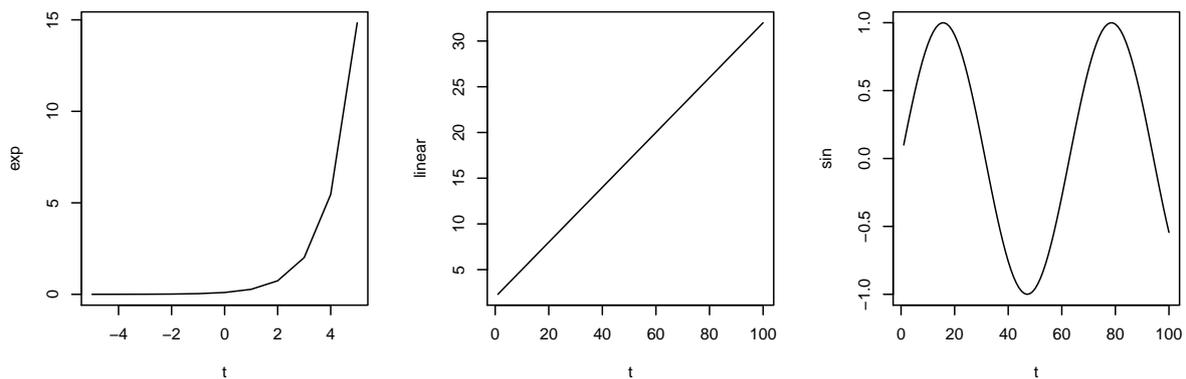


Abbildung 1.3: Beispiele deterministischer Zeitreihen: exponentieller Verlauf $Y_t = \exp(t)$, linearer Verlauf $Y_t = at + b$, periodischer Verlauf $Y_t = \sin(t)$

Wert $c \in \mathbb{R}$ annimmt. Weitere deterministische Prozesse wären

$$Y_{1t} = \exp(t)$$

$$Y_{2t} = at + b$$

$$Y_{3t} = \sin(t).$$

Diese Beispiele sind in Abbildung 1.3 (stetig) dargestellt.

Wir werden deterministische Reihen als Komponenten in Zeitreihen verwenden, um regelmäßige Entwicklungen zu berücksichtigen. Dies gibt uns die Möglichkeit, regelmäßige und zufällige Verläufe gemeinsam zu betrachten.

1.2 Stationarität: stabile Prozesse

Es gibt viele Phänomene in der Psychologie, die über die Zeit stabil sind. So nimmt man an, dass sich die (kristalline) Intelligenz eines Menschen über die Lebensspanne nicht verändert (Cattell, 1987). Ähnliches gilt für Persönlichkeitsmerkmale, sogenannte *Traits*, wie zum Beispiel die „Big Five“ (Costa & McCrae, 1992). Dennoch erhält man nicht bei jeder Messung denselben Wert. Dies liegt vor allem an Messfehlern, aber auch an nicht kontrollierbaren Störfaktoren wie z.B. tagesabhängigen Bedingungen. Typischerweise schwanken die Messwerte dann um einen bestimmten Wert, über- oder unterschreiten ihn, liegen aber selten weit von ihm entfernt. Einen Prozess, dessen Erwartungswert und Varianz konstant über die Zeit sind, nennt man (*schwach*) *stationär* (z.B. Gottman, 1981; Kirchgässner & Wolters, 2007; Brockwell & Davis, 1996, Def. 1.3.2).

Definition 1.2.1 (Stationär) *Eine Zeitreihe Y_t heißt (schwach) stationär, wenn ihr Erwartungswert und ihre Varianz unabhängig von der Zeit sind, also wenn an allen Messpunkten*

gilt

$$\begin{aligned}\mu_t &= \mu, \text{ also konstant,} \\ \sigma_t^2 &= \sigma^2, \text{ also konstant.}\end{aligned}$$

Außerdem sei die Kovarianzfunktion nur vom Abstand h der beiden Zeitpunkte, nicht jedoch von deren zeitlicher Lage abhängig

$$\forall t \in \mathbb{T}, \forall h > 0 \text{ gilt } \text{Cov}[Y_t, Y_{t+h}] = \gamma(t, t+h) = \gamma(h).$$

Eine stationäre Zeitreihe hat zu jedem Zeitpunkt denselben Erwartungswert und dieselbe Varianz. Die Kovarianz bleibt immer gleich, wenn man zwei Variablen mit demselben Abstand betrachtet, und ist unabhängig von der zeitlichen Lage der beiden Variablen. So haben die fünfte und siebte Variable dieselbe Autokovarianz wie die Variablen an den Zeitpunkten 62 und 64. Denn $\text{Cov}[Y_5, Y_7] = \gamma(5, 7) = \gamma(2)$ und $\text{Cov}[Y_{62}, Y_{64}] = \gamma(62, 64) = \gamma(2)$.

Definition 1.2.2 (Stark/streng stationär) *Ein stochastischer Prozess $(Y_t)_{t \in \mathbb{T}}$ heißt stark stationär (z.B. Fahrmeir et al., 1981, S.209, Def. 1.1), falls die gemeinsamen Verteilungen⁴ aller möglichen Tupel $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$ zeitinvariant sind.*

Stark stationäre Prozesse sind also zeittranslationsinvariant (Stier, 2001), d.h. alle gemeinsamen Verteilungen von Zufallsvariablen sind unabhängig vom Betrachtungspunkt t . Für schwach stationäre Prozesse fordert man dies nur für die ersten zwei Momente, da man in der Praxis die Zeitunabhängigkeit der gemeinsamen Verteilungen nicht gewähren kann und bei normalverteilten Variablen aus der schwachen Stationarität die starke direkt folgt (Maddala & Kim, 2007, S. 10). Ein stark stationärer Prozess ist trivialerweise auch schwach stationär, da die Momente durch die eindimensionale Verteilung festgelegt werden. Wir werden im Folgenden unter einem stationären Prozess stets einen *schwach stationären* Prozess verstehen, wie es in der Zeitreihenliteratur üblich ist (z.B. Schröder & Buscher, 2002, S.15).

Bei einem schwach stationären Prozess sind die ersten zwei Momente konstant über die Zeit, also für alle Zufallsvariablen Y_t gleich. Der einfachste stationäre Prozess ist dadurch charakterisiert, dass die Zufallsvariablen an den einzelnen Messzeitpunkten nicht nur denselben Erwartungswert (meist normiert gleich Null) (1.1) und dieselbe Varianz (1.2) besitzen, sondern untereinander noch paarweise unkorreliert (1.3) sind. Einen solchen Prozess nennt man *Weißes Rauschen* oder *White Noise* (z.B. Neidhardt, 1964; Fan & Yao, 2003; Kuo, 2006).

⁴siehe Definition B.1.1, S.331

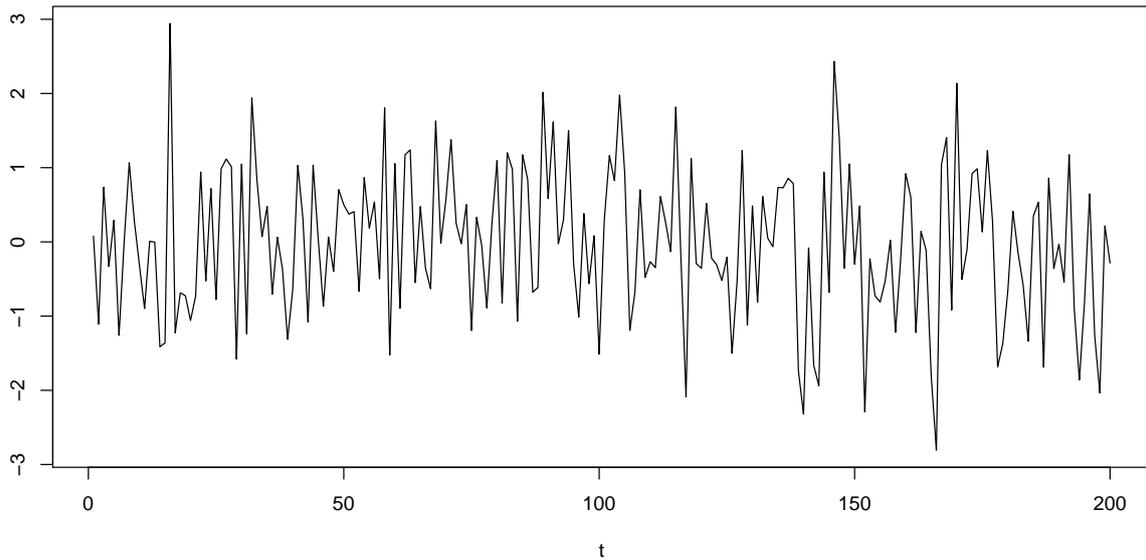


Abbildung 1.4: Gaußsches Weißes Rauschen mit Erwartungswert Null und Varianz Eins

Definition 1.2.3 (Weißes Rauschen) *Eine Zeitreihe ϵ_t heißt Weißes Rauschen⁵, wenn gilt*

$$\mu_t = c, \quad (\text{meist } = 0), \quad (1.1)$$

$$\sigma_t^2 = \sigma^2, \quad (1.2)$$

$$\gamma_\epsilon(h) = 0 \text{ für } (h \neq 0). \quad (1.3)$$

Weißes Rauschen ist stets stationär. Für $h = 0$ entspricht die Kovarianzfunktion der Varianz, also $\gamma(0) = \sigma^2$. Alle anderen Werte der Kovarianzfunktion sind wegen der Unkorreliertheit gleich Null.

Nimmt man zusätzlich an, dass alle Variablen normalverteilt sind, erhält man das sogenannte *Gaußsche Weiße Rauschen* (Rozanov, 1975; Fan & Yao, 2003). In Abbildung 1.4 ist eine Trajektorie mit Mittelwert Null und Varianz Eins dargestellt. Durch die Normalverteilung wird die Eigenschaft der Unkorreliertheit (1.3) sogar zur stärkeren Unabhängigkeit (Klenke, 2006, Satz 5.4 S. 98).

Oft wird die Zeitreihe der (Mess-)Fehler ϵ_t als Weißes Rauschen angenommen, da sich eben die Messfehler untereinander nicht beeinflussen sollten, aber durch denselben Vorgang entstehen, also dieselbe Verteilung haben. Da die Messfehler wiederum viele mögliche Quellen haben können, werden sie als normalverteilt angenommen.

Da die Variablen eines Weißen Rauschens paarweise unkorreliert sind, spricht man auch von einer Zeitreihe „ohne Gedächtnis“. Der aktuelle Wert des Prozesses ist völlig unabhängig von seinen vergangenen Zuständen. Gibt es jedoch eine Abhängigkeit zwischen dem aktuellen Wert Y_t und einem oder mehreren vergangenen Werten, spricht man von einer Zeitreihe „mit

⁵Untersucht man einen solchen Prozess spektralanalytisch, stellt man fest, dass er alle Frequenzen enthält, wie dies bei weißem Licht der Fall ist (Neidhardt, 1964, S.90).

Gedächtnis“. Je nachdem wie weit diese zeitliche Abhängigkeit zurückreicht, spricht man auch von „langem“ oder „kurzem“ Gedächtnis (Beran, 1998). Ein Beispiel für ein Prozessmodell mit kurzem Gedächtnis ist die Moving Average Darstellung.

1.3 Moving Average Darstellung: Einfluss externer Innovationen

Der Wert einer Zeitreihe in *Moving-Average-Darstellung* (z.B. Stier, 2001, 5.4.3; Kirchgässner & Wolters, 2007, Kap. 2.2.2) berechnet sich von einem Startwert $Y_0 = \mu$ ausgehend durch Aufsummieren einer bestimmten Anzahl von zufälligen Schocks, die auch *Innovationen* (Fan & Yao, 2003; Alho & Spencer, 2005) genannt werden.

Definition 1.3.1 (Moving Average Prozess) *Ein Prozess, der sich wie folgt darstellen lässt, heißt Moving Average Prozess der Ordnung q (z.B. Fahrmeir et al., 1981, S. 214, Def. 2.1)*

$$\begin{aligned} Y_t &= \mu + b_0\epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q} \\ &= \mu + \sum_{j=0}^q b_j\epsilon_{t-j}, \end{aligned}$$

wobei ϵ_t Weißes Rauschen ist.

Der Zustand des Prozesses Y_t zum Zeitpunkt t ist also abhängig von den Werten des Weißen Rauschens ϵ_t zu den letzten q Zeitpunkten. Sein Wert ist gegeben durch die gewichtete Summe (Mittel = Average) der Werte $\epsilon_{t-q}, \dots, \epsilon_t$. Da sich das zeitliche „Fenster“, aus dem die Werte zur Summation verwendet werden, mit jedem Zeitpunkt verschiebt, spricht man von einem *Moving Average* Prozess. Zur Analyse betrachtet man meist eine normierte Variante des Prozesses, so dass man in der Theorie meist $b_0 = 1$ setzt⁶.

Definition 1.3.2 (Lag) *Die einzelnen Variablen Y_t einer Zeitreihe bezeichnet man auch als (Time-)Lags der Zeitreihe (z.B. Hamilton, 1994).*

Meist verwendet man den Begriff „Lag“ für vergangene Werte. In Kapitel 1.5 werden wir den Lag-Operator (Definition 1.5.1, S.26) einführen, der eine Zufallsvariable der Zeitreihe (in die Vergangenheit) „verschiebt“.

Je nachdem, auf wie viele Lags der Prozess zu jedem Zeitpunkt zurückgreift, spricht man von einem Moving Average Prozess der Ordnung q . Ein MA(1)-Prozess wird (neben einem Startwert $Y_0 = \mu$) als Summe des aktuellen Wertes eines White Noise Prozesses und seines ersten Timelags dargestellt

$$Y_t = \mu + \epsilon_t + b_1\epsilon_{t-1}.$$

⁶Die anderen Koeffizienten werden bei Normierung zu $\frac{b_i}{b_0}$

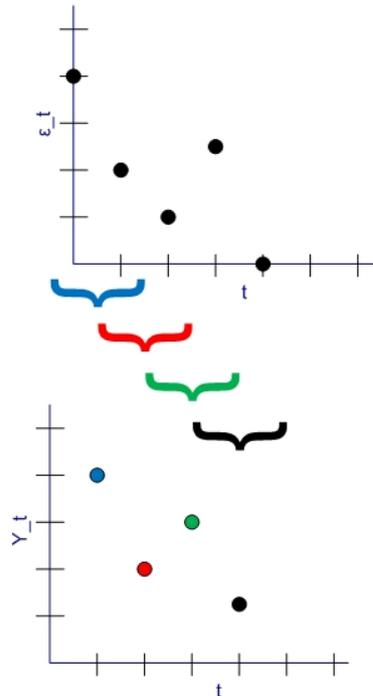


Abbildung 1.5: Entstehung des Moving Average Prozesses erster Ordnung $Y_t = \epsilon_t + 0.5\epsilon_{t-1}$ aus Weißem Rauschen ϵ_t . Das „Fenster“, aus dem die Werte gewichtet werden, bewegt sich mit der Zeit jeweils um einen Zeitschritt nach rechts.

Man nimmt also außer ϵ_t den Wert ϵ_{t-1} zum vorigen Zeitpunkt mit hinzu. Ohne Lags wäre dies $Y_t = \mu + \epsilon_t$, ein White Noise Prozess mit Erwartungswert μ . In Abbildung 1.5 ist das Prinzip, wie der MA(1)-Prozess $Y_t = \epsilon_t + 0.5\epsilon_{t-1}$ aus Weißem Rauschen entsteht, dargestellt.

Unabhängig von der Ordnung oder der Wahl der Gewichte b_i hat ein Moving Average Prozess immer den Erwartungswert $\mathbb{E}[Y_t] = \mu$. Auch seine Varianz ist konstant (Hamilton, 1994, S.48; Kirchgässner & Wolters, 2007, S.64).

Bemerkung 1.3.3 (Erwartungswert und Varianz eines Moving Average Prozesses) *Um den Erwartungswert und die Varianz eines Moving Average Prozesses zu berechnen, benötigt man vor allem die Information, dass der Erwartungswert von Weißem Rauschen Null und die Varianz konstant σ_ϵ^2 ist. Dann ergibt sich die Formel aus den Summeneigenschaften von Erwartungswert⁷ und Varianz⁸*

$$\begin{aligned}
 \mathbb{E}[Y_t] &= \mathbb{E}[\mu + \epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q}] \\
 &= \mathbb{E}[\mu] + \mathbb{E}[\epsilon_t] + \mathbb{E}[b_1\epsilon_{t-1}] + \dots + \mathbb{E}[b_q\epsilon_{t-q}] \\
 &= \mu + 0 + 0 + \dots + 0 \\
 &= \mu
 \end{aligned}$$

⁷Der Erwartungswert ist linear $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.

⁸Die Varianz ist additiv, aber nicht homogen. Ein Faktor kann quadratisch aus der Varianz gezogen werden $\text{Var}[aX + bY] = a^2\text{Var}[X] + b^2\text{Var}[Y]$.

$$\begin{aligned}
 \text{Var}[Y_t] &= \text{Var}[\mu + \epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q}] \\
 &= \text{Var}[\mu] + \text{Var}[\epsilon_t] + \text{Var}[b_1\epsilon_{t-1}] + \dots + \text{Var}[b_q\epsilon_{t-q}] \\
 &= 0 + \sigma_\epsilon^2 + b_1^2\sigma_\epsilon^2 + \dots + b_q^2\sigma_\epsilon^2 \\
 &= \sigma_\epsilon^2 \sum_{j=0}^q b_j^2, \text{ mit } b_0 = 1.
 \end{aligned}$$

Wie wir gleich sehen werden, hängt der Kovarianzprozess eines MA-Prozesses nur vom Abstand der betrachteten beiden Variablen, nicht jedoch von deren zeitlicher Lage ab. So können wir festhalten (Fahrmeir et al., 1981; Rinne & Specht, 2002):

Bemerkung 1.3.4 (Stationarität eines Moving Average Prozesses) *Ein Moving Average Prozess ist stets stationär.*

Betrachtet man die Darstellungen zweier Werte, die einen Abstand q haben

$$\begin{aligned}
 Y_{t-q} &= \mu + b_0\epsilon_{t-q} + b_1\epsilon_{(t-q)-1} + \dots + b_q\epsilon_{(t-q)-q} \\
 &\vdots \\
 Y_t &= \mu + b_0\epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q},
 \end{aligned}$$

so sieht man, dass das Lag ϵ_{t-q} in beiden (und in allen zwischen liegenden) Y_i enthalten ist; bei Y_{t-q} gewichtet mit b_q , bei Y_t mit b_0 . Hingegen haben Y_t und Y_{t-q-1} keinen gemeinsamen Bestandteil, da Y_{t-q-1} auf weiter zurückliegenden Lags basiert, so dass sie unkorreliert sind. Die Korrelation erhält man durch Normierung aus der Kovarianz, die für einen Moving Average Prozess ähnlich wie die Varianz berechnet wird

$$\text{Cov}[Y_t, Y_{t+h}] = \gamma(h) = \sigma_\epsilon^2 \sum_{j=0}^{q-h} b_j b_{j+h}.$$

Diese Formel gilt allerdings nur für $h \leq q$. Für $h > 0$ gilt wegen der Unkorreliertheit $\gamma(h) = 0$.

Da die (Auto-)Kovarianz bis zu einem exakt definierten Lag in die Vergangenheit zurückreicht, sagt man der MA-Prozess habe ein kurzes Gedächtnis (Fahrmeir et al., 1981; Rinne & Specht, 2002).

Wir werden die Autokovarianz im Kapitel 1.9 noch genauer betrachten und diese Eigenschaft für die Schätzung der Prozessordnung q nutzen.

Während ein Moving Average Prozess sich aus den Lags eines Weißen Rauschens zusammensetzt, greift ein autoregressiver Prozess auf die eigene Vergangenheit zurück und hat theoretisch ein unendlich langes Gedächtnis.

1.4 Autoregressive Darstellung: Analyse der internen Struktur

Während die Moving Average Darstellung vor allem die Abhängigkeit des Prozesses von den Innovationen betont, erleichtert die autoregressive Darstellung die Betrachtung der internen Struktur und inneren Abhängigkeiten des Prozesses.

Im einfachsten Fall ist ein autoregressiver Prozess nur von seinem Zustand am vorigen Zeitpunkt abhängig (z.B. Kirchgässner & Wolters, 2007, Kap. 2.1.1; Schröder & Buscher, 2002, Def.6; Alho & Spencer, 2005, S. 203). Bei täglicher Messung lässt sich der Wert des Prozesses am besten durch den Zustand am Vortag erklären.

Definition 1.4.1 (Autoregressiver Prozess der Ordnung 1) *Ein autoregressiver Prozess der Ordnung 1, kurz AR(1)-Prozess, ist ein Prozess, der sich wie folgt darstellen lässt*

$$Y_t = c + aY_{t-1} + \epsilon_t,$$

wobei $a, c \in \mathbb{R}$ Konstanten sind und ϵ_t Weißes Rauschen darstellt.

Diese Darstellung erinnert an die einfache Regression mit Y_t als abhängige und Y_{t-1} als unabhängige Variable. Der Zustand des Prozesses zum Zeitpunkt t lässt sich also (zu großen Teilen) durch seinen Zustand zum Zeitpunkt $t - 1$ erklären. Er „regressiert“ auf sich selbst, weshalb er *autoregressiv* genannt wird. Wie stark diese Abhängigkeit ist, beschreibt der Parameter a . Zusätzlich wird der Prozess von der Konstanten c und einem Zufallsprozess ϵ_t beeinflusst.

Man sieht leicht, dass sich aus der Konstanten c der Mittelwert des Prozesses berechnen lässt (und umgekehrt), falls der Prozess stationär ist⁹. Es gilt dann

$$\mu = \frac{c}{1 - a},$$

wobei man häufig (oBdA¹⁰) $c = 0$ und damit auch $\mu = 0$ setzt (Kirchgässner & Wolters, 2007).

Ist der Wert des Prozesses nicht nur abhängig von seinem Zustand am letzten Zeitpunkt, sondern greift er noch weiter in seine eigene Vergangenheit zurück, dann spricht man von einem autoregressiven Prozess der Ordnung p . Man braucht als erklärende Variablen also die letzten p beobachteten Zustände des Prozesses (Fahrmeir et al., 1981, S.215, Def.2.2; Schröder & Buscher, 2002, S.155).

Definition 1.4.2 (Autoregressiver Prozess der Ordnung p) *Ein Prozess, der sich in Abhän-*

⁹Die Stationarität ist abhängig vom Parameter a , wie wir in Kapitel 1.7 noch sehen werden. Man sieht bereits, dass man bei $a = 1$ den Erwartungswert nicht mehr berechnen kann.

¹⁰„ohne Beschränkung der Allgemeinheit“, also ohne die Betrachtung auf Prozesse mit $c = 0$ zu beschränken, da man einen solchen Prozess stets erhält, indem man statt Y_t dann $Y_t - c$ betrachtet.

gigkeit seiner Lags in folgender Form darstellen lässt

$$Y_t = c + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + \epsilon_t,$$

heißt autoregressiver Prozess der Ordnung p , kurz $AR(p)$. Dabei ist $c \in \mathbb{R}$ eine Konstante, die $a_i \in \mathbb{R}$ beschreiben die Gewichtung des Einflusses der einzelnen Lags Y_{t-i} und ϵ_t ist Weißes Rauschen.

Der Erwartungswert eines allgemeinen stationären autoregressiven Prozesses berechnet sich durch (z.B. Kirchgässner & Wolters, 2007, S.50)

$$\mu = \frac{c}{1 - a_1 - a_2 - \dots - a_p}. \quad (1.4)$$

Daraus ergibt sich die Frage, wann ein AR-Prozess stationär ist. Wir wissen bereits, dass ein Prozess dann stationär ist, wenn man ihn als MA-Prozess darstellen kann. Damit wir einen vorgegebenen AR-Prozess in eine MA-Darstellung bringen können, brauchen wir einige mathematische Hilfsmittel wie z.B. den Lag-Operator für die Verwendung von Lagpolynomen.

1.5 Lag Operationen

Wir haben bereits gesehen, dass man häufig vergangene Zustände benötigt, um Prozesse darzustellen. So greifen autoregressive Prozesse auf eigene Lags zurück, während Moving Average Prozesse aus der gewichteten Summe der Lags eines Weißen Rauschens entstehen. Es erweist sich als nützlich, für diese Zeitverschiebung eines Prozesses einen speziellen Operator einzuführen, mit dessen Hilfe man Lagpolynome definieren kann, die für die Analyse von Zeitreihen sehr hilfreich sind (Fahrmeir et al., 1981; Kirchgässner & Wolters, 2007; Schlittgen & Streitberg, 1999).

Definition 1.5.1 (Lag Operator) *Der Lag Operator (auch Backshift Operator genannt) führt einen Linksshift aus. Dadurch wird der Prozess Y_t um eine Zeiteinheit (ein Lag) in die Vergangenheit, also auf der Zeitachse nach links, verschoben*

$$LY_t = Y_{t-1}.$$

Natürlich kann man einen Prozess auch um mehrere Lags in die Vergangenheit „schieben“. Dies geschieht durch mehrfache Anwendung des Lag-Operators. $L(LY_t) = LY_{t-1} = Y_{t-2}$. Eine solche mehrfache Anwendung schreibt man auch als Potenz

$$L^2 Y_t = L(LY_t) = Y_{t-2}$$

oder allgemein

$$L^k Y_t = Y_{t-k}.$$

Wird der Lag-Operator nicht angewandt, bleibt man bei dem Wert des Prozesses zum aktuellen Zeitpunkt $L^0 Y_t = Y_t$, weshalb für den „neutralen“ Lagoperator $L^0 = 1$ gilt. Da ein konstanter Prozess $Y_t = c$ zu jedem Zeitpunkt denselben Wert hat, bewirkt die Anwendung des Lag-Operators nichts. Es gilt also

$$L^k c = c \text{ für alle } k.$$

Für das Verbinden von Lag-Operatoren gelten die üblichen Rechenregeln für Potenzen. Im Anhang B.2 (S.332) werden weitere Eigenschaften des Lagoperators beschrieben. Außerdem ist der Lagoperator linear (siehe Bemerkung B.2.2, S.332), aus diesem Grund ist es möglich Polynome im Lagoperator zu bilden (Kirchgässner & Wolters, 2007, S.11).

Definition 1.5.2 (Lagpolynom) *Allgemein kann ein Lagpolynom der Ordnung p so dargestellt werden:*

$$P(L) = p_0 + p_1 L + p_2 L^2 + \dots + p_p L^p.$$

Dabei sind $p_k, k = 0, 1, \dots, p$ reelle Koeffizienten.

Für Lagpolynome gelten alle Rechenregeln, die für Polynome bekannt sind (siehe Anhang A.1.4, S.313). Häufig wird das Lagpolynom normiert mit $p_0 = 1$ betrachtet.

Die Anwendung eines Lagpolynoms erfolgt komponentenweise

$$\begin{aligned} P(L)X_t &= (p_0 + p_1 L + p_2 L^2 + \dots + p_p L^p)X_t \\ &= p_0 X_t + p_1 L X_t + p_2 L^2 X_t + \dots + p_p L^p X_t \\ &= p_0 X_t + p_1 X_{t-1} + p_2 X_{t-2} + \dots + p_p X_{t-p}. \end{aligned}$$

Bei der Anwendung eines Lagpolynoms auf einen konstanten Prozess bewirken die einzelnen Lagoperationen keine Verzögerung. Deshalb gilt

$$P(L)c = (p_0 + p_1 + p_2 + \dots + p_p) \cdot c.$$

Damit ergibt sich der Wert des Lagpolynoms für einen konstanten Prozess aus der Summe aller seiner Koeffizienten multipliziert mit der Konstanten c . Man erhält die Summe der Koeffizienten auch, indem man L durch 1 ersetzt

$$P(1) = \sum_{i=0}^p p_i.$$

Deshalb kann man die Anwendung eines Lagpolynoms auf eine Konstante wie folgt vereinfachen:

Bemerkung 1.5.3 Für ein Lagpolynom $P(L)$ und einen konstanten Prozess c gilt:

$$P(L)c = P(1)c.$$

Ein Lagpolynom ist ein Polynom im Lagoperator. Man kann jedoch dasselbe Polynom auch auf andere Objekte, zum Beispiel auf reelle oder komplexe¹¹ Zahlen bzw. Variablen anwenden. Dazu braucht man die sogenannte z -Transformierte des Lagpolynoms (Haßler, 2007; Schröder & Buscher, 2002, S.207).

Definition 1.5.4 (z -Transformierte) Die z -Transformierte $P(z)$ zu einem Lagpolynom $P(L)$ ist ein Polynom in einer (komplexwertigen) Variablen z mit denselben Koeffizienten p_k wie $P(L)$

$$P(z) = p_0 + p_1z + p_2z^2 + \dots + p_pz^p.$$

Durch die Verwendung von Lag-Polynomen erhält man eine neue Schreibweise für die verschiedenen Darstellungsformen eines Prozesses.

Bemerkung 1.5.5 (MA-Polynom) Man kann einen Moving Average Prozess auch mit Hilfe eines Lagpolynoms darstellen (Kirchgässner & Wolters, 2007; Rinne & Specht, 2002) über

$$Y_t = \mu + B(L)\epsilon_t.$$

Dabei ist $B(L) = 1 + b_1L + b_2L^2 + \dots + b_qL^q$. Denn es gilt

$$\begin{aligned} B(L)\epsilon_t &= (1 + b_1L + b_2L^2 + \dots + b_qL^q)\epsilon_t \\ &= 1 \cdot \epsilon_t + b_1L\epsilon_t + b_2L^2\epsilon_t + \dots + b_qL^q\epsilon_t \\ &= \epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} + \dots + b_q\epsilon_{t-q}. \end{aligned}$$

Bemerkung 1.5.6 (AR-Polynom) Auch einen autoregressiven Prozess kann man über ein Lagpolynom darstellen (Kirchgässner & Wolters, 2007; Fahrmeir et al., 1981, S.216). Aus der bekannten Darstellung

$$Y_t = c + a_1Y_{t-1} + a_2Y_{t-2} + \dots + a_pY_{t-p} + \epsilon_t$$

gelangt man zum AR-Polynom, indem man zunächst alle Ausdrücke in Y_t auf eine Seite der

¹¹Eine Einführung in die komplexen Zahlen findet sich im Anhang A.2 ab Seite 317.

Gleichung bringt und dann die Lags in Operatorenschreibweise überführt

$$\begin{aligned} \underbrace{Y_t}_{1 \cdot Y_t} - a_1 \underbrace{Y_{t-1}}_{LY_t} - a_2 \underbrace{Y_{t-2}}_{L^2 Y_t} - \dots - a_p \underbrace{Y_{t-p}}_{L^p Y_t} &= c + \epsilon_t \\ (1 - a_1 L - a_2 L^2 - \dots - a_p L^p) Y_t &= c + \epsilon_t \\ A(L) Y_t &= c + \epsilon_t. \end{aligned}$$

Das AR-Polynom ist also $A(L) = 1 - a_1 L - a_2 L^2 - \dots - a_p L^p$.

Da gilt $A(1) = 1 - a_1 - a_2 - \dots - a_p$, kann man den Erwartungswert eines stationären AR(p)-Prozesses auch mit dem Lagpolynom darstellen (vgl. Gleichung (1.4), S.26)

$$\mu = \frac{c}{A(1)}.$$

1.6 Wolddarstellung

Mit Hilfe der Darstellung über das Lagpolynom können wir die Wolddarstellung eines AR(1)-Prozesses herleiten. Der Satz von Wold besagt, dass man jeden *stationären* Prozess nach Abzug seiner deterministischen Bestandteile als einen Moving Average Prozess darstellen kann. Dies wollen wir verwenden, um eine Stationaritätsbedingung für autoregressive Prozesse abzuleiten.

Betrachten wir nun einen AR(1)-Prozess und zwar in der neuen Darstellung mit AR-Lag-Polynom, wobei wir zur Vereinfachung noch (oBdA¹²) $c = 0$ setzen.

$$\begin{aligned} A(L) Y_t &= \epsilon_t \\ \text{also } (1 - aL) Y_t &= \epsilon_t \end{aligned}$$

Wenn wir das Lagpolynom invertieren und sein Inverses $(1 - aL)^{-1}$, das in Anhang B.2 (S.332ff) berechnet wird, auf den Prozess ϵ_t anwenden, erhalten wir eine Darstellung des Prozesses über die Lags eines Weißen Rauschens (Schröder & Buscher, 2002, S. 171; Gottman, 1981, S.123ff)

$$\begin{aligned} Y_t &= (1 - aL)^{-1} \epsilon_t \\ &= (1 + aL + a^2 L^2 + a^3 L^3 \dots) \epsilon_t \\ &= \epsilon_t + a \epsilon_{t-1} + a^2 \epsilon_{t-2} + a^3 \epsilon_{t-3} + \dots \\ &= \sum_{j=0}^{\infty} a^j \epsilon_{t-j}. \end{aligned} \tag{1.5}$$

¹²„ohne Beschränkung der Allgemeinheit“, also ohne die Betrachtung auf Prozesse mit $c = 0$ zu beschränken, da man einen solchen Prozess stets erhält, indem man statt Y_t dann $Y_t - c$ betrachtet.

Diese Darstellung erinnert schon sehr an die MA-Darstellung, die wir im Kapitel 1.3 (S.22) kennengelernt hatten. Analog wie bei einer endlichen Summe spricht man hier deshalb auch von einer $MA(\infty)$ -Darstellung. Bei unendlichen Summen muss man jedoch immer etwas vorsichtiger sein als bei endlichen Summen. Denn durch die Gleichung (1.5) auf Seite 29 wird nur dann ein Prozess dargestellt, wenn die Reihe konvergiert, also wenn der Wert der unendlichen Summe endlich ist¹³. Dies ist genau dann der Fall, wenn $|a| < 1$. Denn dann handelt es sich um eine geometrische Reihe, die stets konvergiert (siehe Anhang B.3, S.334).

Da er sich als die gewichtete unendliche Summe eines Weißen Rauschens darstellen lässt, ist jeder $AR(1)$ -Prozess mit $|a| < 1$ auch ein $MA(\infty)$ -Prozess (Neusser, 2006, Definition 2.5; Fahrmeir et al., 1981; Fan & Yao, 2003; Brockwell & Davis, 1996, Def.3.2.1).

Definition 1.6.1 ($MA(\infty)$ -Prozess) *Ein stochastischer Prozess Y_t heißt ein $MA(\infty)$ -Prozess, falls eine Folge von absolut summierbaren Zahlen $\psi_j, j = 0, 1, 2, \dots$ mit $\sum_{j=0}^{\infty} |\psi_j| < \infty$ existiert, so dass gilt*

$$\begin{aligned} Y_t &= \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \\ &= \Psi(L)\epsilon_t \end{aligned}$$

mit Weißem Rauschen ϵ_t .

Die Definition ist analog zu der Definition von Moving Average Prozessen endlicher Ordnung (Definition 1.3.1, S.22). Der Prozess lässt sich als unendliche Summe von Lags eines Weißen Rauschens darstellen. Die zusätzliche Bedingung über die Koeffizienten ψ_j sichert nur die Konvergenz der Reihe, die für die Existenz des Prozesses notwendig ist.

Beispiel 1.6.2 ($MA(\infty)$ -Prozesse) *$MA(\infty)$ -Prozesse unterscheiden sich in der Ausprägung ihrer Gewichte b_i . Natürlich müssen nicht alle Koeffizienten in einer $MA(\infty)$ -Darstellung von Null verschieden sein. Wir kennen bereits einige $MA(\infty)$ -Prozesse, nämlich*

- *$MA(q)$ -Prozesse: alle Gewichte b_i über der Prozessordnung $i > q$ sind gleich Null*
- *$AR(1)$ -Prozesse mit $|a| < 1, b_i = a^i$*

Wie für MA -Prozesse endlicher Ordnung gilt auch im unendlichen Fall (Fahrmeir et al., 1981):

Theorem 1.6.3 *Ein $MA(\infty)$ -Prozess ist immer stationär.*

Umgekehrt gilt für alle stationären Prozesse, dass sie sich in einer Form darstellen lassen, in der ihr stochastischer Anteil als $MA(\infty)$ -Prozess auftaucht. Dies ist der Satz von Wold (Wold, 1938; siehe auch Neusser, 2006, Theorem 4.1; Fahrmeir et al. (1981); Kirchgässner und Wolters (2007); C. W. Granger & Newbold, 1977, Kap.1.11).

Theorem 1.6.4 (Satz von Wold) *Jeder stationäre stochastische Prozess Y_t mit Mittelwert*

¹³Eine Summe aus unendlich vielen Summanden hat nur im Fall der Konvergenz keinen unendlichen Wert.

Null und positiver endlicher Varianz kann als Summe eines $MA(\infty)$ -Prozesses und eines rein deterministischen Prozesses δ_t dargestellt werden:

$$\begin{aligned} Y_t &= \delta_t + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \\ &= \delta_t + \Psi(L)\epsilon_t. \end{aligned}$$

Diese Darstellung nennt man auch *Wolddarstellung* eines Prozesses.

Dass im Satz von Wold ein Erwartungswert gleich Null gefordert wird, ist keine echte Einschränkung. Hat der stationäre Prozess einen von Null verschiedenen Erwartungswert, wird in seiner Wolddarstellung einfach ein konstanter Prozess addiert¹⁴. Für uns ist vor allem folgender Sonderfall interessant, bei dem sich obiger Satz stark vereinfacht (Neusser, 2006, S.66; Neidhardt, 1964, S.56).

Definition 1.6.5 (rein stochastischer Prozess) *Prozesse, bei denen die rein deterministische Komponente δ_t Null ist, heißen rein nicht-deterministisch, rein stochastisch oder regulär.*

Korollar 1.6.6 (Wolddarstellung eines rein stochastischen Prozesses) *Zu jedem stationären rein stochastischen Prozess Y_t (mit $\delta_t = 0$) gibt es ein Weißes Rauschen ϵ_t , so dass Y_t ein $MA(\infty)$ -Prozess bezüglich ϵ_t ist (Fan & Yao, 2003, Kap.2.1.3).*

Alle (rein stochastischen) stationären Prozesse haben also eine Darstellung als $MA(\infty)$ -Prozess. Wir wenden uns nun der MA -Darstellung von allgemeinen autoregressiven Prozessen zu.

1.7 Stationaritätsbedingung für autoregressive Prozesse

Wir haben also gesehen, dass man einen $AR(1)$ -Prozess genau dann als MA -Prozess darstellen kann, wenn $|a| < 1$ gilt. Diese Erkenntnis wollen wir nun auf allgemeinere autoregressive Prozesse erweitern. Dabei hilft uns die Betrachtungsweise über Lagpolynome.

Das Lagpolynom eines $AR(1)$ -Prozesses lautet $A(L) = 1 - aL$. Betrachtet man die Nullstellen seiner z -Transformierten, erhält man ein interessantes Ergebnis

$$\begin{aligned} A(z) = 1 - az &= 0 \\ \Leftrightarrow 1 &= az \\ \Leftrightarrow \frac{1}{a} &= z. \end{aligned}$$

Man nennt die Gleichung $A(z) = 0$ auch *charakteristische Gleichung* (Johansen, 1995, S.14 Assumption 1; Schröder & Buscher, 2002, S. 162). Obige Rechnung zeigt nun, dass die Nullstellen der charakteristischen Gleichung vom Parameter a des Lagpolynoms abhängig

¹⁴Ein konstanter Prozess ist trivialerweise deterministisch.

sind. Da $|a| < 1$ gelten muss, damit der AR(1)-Prozess stationär ist, muss für die Nullstelle z gelten

$$|z| = \left| \frac{1}{a} \right| > 1.$$

Das bedeutet aber umgekehrt auch, dass ein AR(1)-Prozess nur genau dann stationär ist, wenn die Nullstelle der charakteristischen Gleichung vom Betrag größer als Eins ist.

Während dieses Ergebnis für einen AR(1)-Prozess noch trivial erscheint, wird die Notwendigkeit der charakteristischen Gleichung erst im Fall autoregressiver Prozesse allgemeiner Ordnung einsichtig.

Dazu muss man den möglichen Wertebereich der Nullstellen vergrößern und mit komplexen Zahlen arbeiten. Eine Einführung in dieses Gebiet, die sich nur auf die hier notwendigen Kenntnisse bezieht, findet sich im Anhang A.2 ab Seite 317 dieser Arbeit. Ein wichtiger Satz in diesem Zusammenhang ist der *Fundamentalsatz der Algebra* (z.B. Meyberg, 1999), der besagt, dass jedes Polynom vom Grad p in den komplexen Zahlen genau p Nullstellen hat. Des Weiteren kann man sich komplexe Zahlen als Punkte in der Ebene vorstellen. Ihr Betrag berechnet sich analog wie der von Vektoren im \mathbb{R}^2 (siehe Anhang A.2.1, S.319). Für komplexe Zahlen, deren Betrag größer als Eins ist, kann man also sagen, dass sie vom Nullpunkt $(0, 0)$ einen Abstand haben, der größer als Eins ist. Sie liegen also außerhalb eines Kreises mit Mittelpunkt Null und Radius Eins, dem sogenannten Einheitskreis. Deshalb wird die Stationaritätsbedingung für autoregressive Prozesse meist wie folgt formuliert (z.B. Fahrmeir et al., 1981, S.221):

Theorem 1.7.1 (Stationaritätsbedingung für autoregressive Prozesse) *Ein autoregressiver Prozess der Ordnung p ist genau dann stationär, wenn alle p Lösungen seiner charakteristischen Gleichung $A(L) = 0$ außerhalb des Einheitskreises liegen, also vom Betrag größer als Eins sind.*

Das Lagpolynom eines AR(1)-Prozesses hat den Grad Eins und deshalb genau eine Nullstelle. Bei reellen Koeffizienten des Polynoms kann die Nullstelle auch nur reellwertige Nullstellen annehmen. Auf dem Einheitskreis (Abbildung A.5 auf Seite 324 im Anhang A.2.2) gibt es nur zwei reelle Werte, nämlich 1 und -1 . Der Prozess ist also genau dann stationär, wenn die Nullstelle seines charakteristischen Polynoms nicht im (reellen) Intervall $[-1, 1]$ liegt¹⁵.

Bei autoregressiven Prozessen höherer Ordnung hat das Lagpolynom einen Grad größer/gleich Zwei. Nach dem Fundamentalsatz der Algebra gibt es dann zwei bzw. p komplexwertige Nullstellen. Liegt mindestens eine Nullstelle innerhalb des Einheitskreises, so konvergiert die MA-Darstellung nicht mehr, der Prozess ist also nicht stationär und das AR-Polynom nicht invertierbar. Wir werden uns in dieser Arbeit wie in der Zeitreihenliteratur allgemein üblich vor allem auf den Fall einer reellwertigen Einheitswurzel im Prozess konzentrieren.

¹⁵Das Intervall $[-1, 1]$ ist der reelle Teil des Einheitskreises.

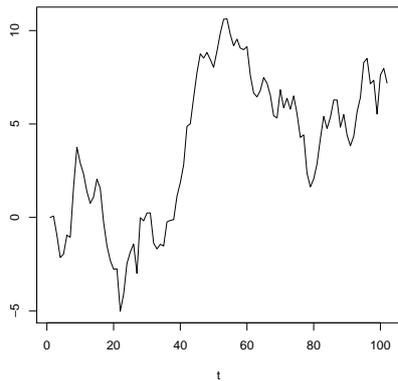


Abbildung 1.6: Random Walk

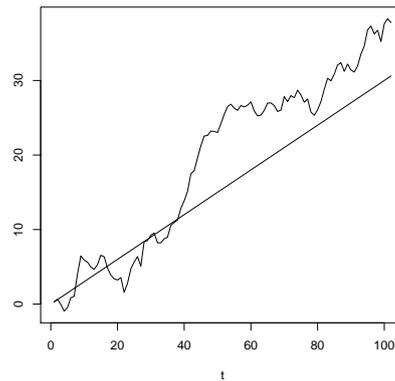


Abbildung 1.7: Random Walk mit Drift

Man kann die Stationaritätsbedingung für einen Test nutzen. Nämlich, indem man prüft, ob ein Prozess in seiner AR-Darstellung Lösungen innerhalb des Einheitskreises hat. Verschiedene Arten dieses Tests werden wir in Kapitel 1.12 einführen. Die dort vorhandene Abbildung 1.9 auf Seite 42 zeigt den Vergleich zwischen Prozessen mit Nullstellen innerhalb, außerhalb und auf dem Einheitskreis.

1.8 Random Walk: ein erstes dynamisches Prozessmodell

Liegen die Lösungen der charakteristischen Gleichung eines AR-Prozesses außerhalb des Einheitskreises, so ist er stationär. Gibt es eine Lösung, die innerhalb liegt, so explodiert der Prozess. Ein bislang nicht betrachteter Sonderfall ist der, in dem Lösungen der charakteristischen Gleichung *genau auf* dem Einheitskreis liegen. Diese Punkte nennt man auch *Einheitswurzeln*, englisch *Unit Roots* (siehe Anhang A.2.2, S.323).

Der einfachste Prozess mit einer Einheitswurzel ist ein AR(1)-Prozess mit Erwartungswert Null, bei dem $a = 1$ ist

$$Y_t = Y_{t-1} + \epsilon_t. \quad (1.6)$$

Sein Lagpolynom ist $A(L) = 1 - L$, so dass die Lösung des charakteristischen Polynoms $A(z) = 1 - z = 0$ gleich $z = 1$ ist, die natürlich eine Einheitswurzel ist. Eine Realisierung dieses Random Walks ist in Abbildung 1.6 zu sehen. Man kann deutlich erkennen, dass seine Varianz sich mit der Zeit vergrößert. Der Prozess besitzt jedoch keinen klaren (deterministischer) Verlauf. Vielmehr folgt der Prozess einem zufälligen (stochastischen) Trend.

Ein stationärer AR(1)-Prozess hat eine Darstellung als MA(∞)-Prozess, was für den nicht-stationären Prozess Y_t aus Gleichung (1.6) natürlich nicht gilt. Dennoch ist es interessant, ihn in eine ähnliche Darstellung zu überführen.

Die Gleichung (1.6) gilt für jede Variable Y_t des Prozesses, also auch für $Y_{t-1} = Y_{t-2} + \epsilon_{t-1}$. Dies kann man nun in obige Gleichung (1.6) einsetzen und erhält

$$Y_t = \underbrace{Y_{t-2} + \epsilon_{t-1}}_{Y_{t-1}} + \epsilon_t.$$

Setzt man dieses Einsetzen rekursiv fort, stellt also Y_{t-2} durch $Y_{t-3} + \epsilon_{t-2}$ dar usw., erhält man eine Darstellung des Prozesses über seinen Startwert Y_0 als Summe¹⁶ von Weißem Rauschen

$$\begin{aligned} Y_t &= Y_0 + \epsilon_1 + \epsilon_2 + \dots + \epsilon_{t-1} + \epsilon_t \\ &= Y_0 + \sum_{j=1}^t \epsilon_j. \end{aligned}$$

Der Prozess „bewegt“ sich also von seinem Startwert Y_0 mit jedem Zeitpunkt um eine zufällige Größe ϵ_t weiter. Deshalb wird dieser Prozess auch *Random Walk* (z.B. Alho & Spencer, 2005; Bhattacharya & Waymire, 1990, S.4 Def.2.1) oder auch *Drunkard's Walk* (Gottman, 1981, S.118) genannt. Wir werden seine Eigenschaften im Kapitel 2.1 noch genauer betrachten. Wir können jedoch schon sehen, dass keine Innovation ϵ_t jemals verloren geht. Man spricht deshalb auch von einem Prozess mit *unendlich langem Gedächtnis*.

Entscheidend ist hier jedoch vor allem die Nicht-Stationarität. Für seinen Erwartungswert gilt zwar

$$\begin{aligned} \mathbb{E}[Y_t] &= \mathbb{E}[Y_0] + \sum_{j=1}^t \underbrace{\mathbb{E}[\epsilon_j]}_{=0} \\ &= Y_0, \end{aligned}$$

wodurch die erste Bedingung für Stationarität erfüllt ist. Doch für seinen Varianzprozess gilt

$$\sigma_t^2 = \text{Var}[Y_t] = \underbrace{\text{Var}[Y_0]}_{=0} + \sum_{j=1}^t \underbrace{\text{Var}[\epsilon_j]}_{=\sigma_\epsilon^2},$$

die Varianz ist also zum Zeitpunkt t abhängig von t Summanden, die jeweils denselben Wert, nämlich die konstante Varianz des Weißes Rauschens σ_ϵ^2 , haben. Es gilt folglich

$$\sigma_t^2 = t \cdot \sigma_\epsilon^2.$$

Die Varianz ist also zeitabhängig, der Prozess somit nicht stationär.

Eine Abwandlung dieses Prozesses erhält man, indem man eine Konstante in die Prozess-

¹⁶Wegen der Einheitswurzel ist die zugehörige Reihe nicht konvergent, der Prozess - wie wir bereits wissen - also nicht stationär.

gleichung aufnimmt (z.B. Alho & Spencer, 2005)

$$Y_t = d + Y_{t-1} + \epsilon_t. \quad (1.7)$$

Dieser deterministische Anteil sorgt dafür, dass der Prozess eine „Bewegungsrichtung“ bekommt, weshalb er auch *Random Walk mit Drift* genannt wird (Kirchgässner & Wolters, 2007, S.156). Sein Erwartungswert ist $\mathbb{E}[Y_t] = Y_0 + dt$, so dass der Prozess eine lineare Grundrichtung bekommt. Für den Varianzprozess gilt weiterhin $\text{Var}[Y_t] = t\sigma_\epsilon^2$. Die Abbildung 1.7 auf Seite 33 zeigt die Realisierung eines Random Walks wie Abbildung 1.6 mit einem zusätzlichen Drift $d = 0.3$.

Es ist möglich, einen Random Walk mit oder ohne Driftkonstante ($d = 0$) durch geeignete Transformation in einen stationären Prozess zu verwandeln, der mit den entsprechenden Hilfsmitteln analysiert werden kann. Und zwar folgt aus der Prozessgleichung (1.6) bzw. (1.7), dass der Differenzenprozess $Y_t - Y_{t-1} = d + \epsilon_t$ stationär ist.

Um die Voraussetzungen für eine Kointegrationsanalyse zu prüfen, wird es notwendig sein einen Test auf Stationarität und speziell auf Weißes Rauschen zu haben. Solche Tests basieren auf der Autokorrelationsfunktion.

1.9 Autokorrelationsfunktion

Wir haben bereits gesehen, dass sich MA- und AR-Prozesse endlicher Ordnung darin unterscheiden, „wie lang ihr Gedächtnis“ ist. Dabei haben MA-Prozesse ein endliches und AR-Prozesse theoretisch ein unendlich langes Gedächtnis, das jedoch, wie wir noch sehen werden, in der Realität exponentiell abklingt, formal bleibt jede Innovation immer im Prozess, ihr Einfluss schwindet jedoch mit fortschreitender Zeit.

Diese Gedächtniseigenschaft lässt sich formal über die Autokorrelationsfunktion eines Prozesses beschreiben. In Kapitel 1.3 (S.22) haben wir gezeigt, dass die Kovarianzfunktion eines MA-Prozesses nach q Lags abbricht. Dies kann man nun in einer vorhandenen Zeitreihe dazu nutzen, die Prozessordnung q eines darstellenden MA-Prozesses zu schätzen. Ein White Noise Prozess ist als Folge von unkorrelierten Zufallsvariablen definiert, seine Autokorrelationsfunktion ist also bis auf $\rho(0)$ identisch Null(funktion)¹⁷. Die Schätzungen aller Autokorrelationskoeffizienten sollten also für $h \geq 1$ nicht signifikant von Null abweichen.

Betrachten wir nun also den Fall, dass uns eine Zeitreihe $(Y_t)_{t \in \mathbb{T}}$ bestehend aus T Beobachtungen vorliegt, wir jedoch den Daten erzeugenden Prozess nicht kennen. Neben dem Erwartungswert ist es möglich aus diesen Daten auch die (Auto-)Kovarianzfunktion zu schätzen. Da bei einem stationären Prozess Erwartungswert und Varianz unabhängig von der Zeit

¹⁷Natürlich ist der erste Wert $\rho(0)$, der die Korrelation der unverschobenen Zeitreihe mit sich selbst misst, stets exakt Eins.

sind, betrachten wir dazu die beobachteten Zeitpunkte Y_1, Y_2, \dots, Y_T als „ganz normale“ Stichprobe und berechnen wie üblich (Bortz, 2005, S.188f)

$$\bar{Y}_T := \frac{1}{T} \sum_{t=1}^T Y_t$$

und

$$\hat{\gamma}(h) := \frac{1}{T} \sum_{t=1}^{T-h} (Y_t - \bar{Y}_T)(Y_{t+h} - \bar{Y}_T).$$

Der Schätzer $\hat{\gamma}(h)$ ist verzerrt („biased“, also nicht erwartungstreu), da mit T statt mit $T - h$ normiert wird (Fahrmeir et al., 1981). Er hat jedoch den Vorteil, dass es mit seiner Hilfe möglich ist, die Autokovarianzmatrix Γ_T so zu schätzen, dass sie stets positiv definit und für $\hat{\gamma}(0) > 0$ nicht-singulär ist (Brockwell & Davis, 2006, Theorem 1.5.1).

$$\hat{\Gamma}_T = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \dots & \hat{\gamma}(T-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \dots & \hat{\gamma}(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(T-1) & \hat{\gamma}(T-2) & \dots & \hat{\gamma}(0) \end{pmatrix}$$

Box und Jenkins (1976) zeigen noch, dass die Schätzungen für $\gamma(h)$ nur dann ausreichend gut sind, wenn $T \geq 50$ und $h \leq \frac{T}{4}$ ist. Wir werden uns in Kapitel 3.2 noch mit der Länge einer Zeitreihe beschäftigen.

Definition 1.9.1 (Autokorrelationskoeffizienten) *Aus der geschätzten Autokovarianzfunktion werden die geschätzten Autokorrelationskoeffizienten wie üblich durch*

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

abgeleitet (Gottman, 1981, S.69; Fahrmeir et al., 1981, S.212 Bem.1.5).

Es kann gezeigt werden, dass der Vektor $(\hat{\rho}(1), \dots, \hat{\rho}(h))$ asymptotisch multivariat normalverteilt um die wahren Werte $(\rho(1), \dots, \rho(h))$ mit aus den Autokorrelationen und T berechenbarer Kovarianzmatrix ist (Neusser, 2006, Theorem 3.3). Diese Verteilungseigenschaft kann man nun für die erwähnten Tests nutzen. Wenn es sich bei dem vorliegenden Prozess um ein Weißes Rauschen handelt, sollten die Schätzungen aller Autokorrelationskoeffizienten nicht signifikant von Null abweichen, da die Lags untereinander nicht korreliert sind.

Für große T konvergieren diese Schätzungen gegen normalverteilte Zufallsvariablen mit Mittelwert Null und Varianz $\frac{1}{T}$, so dass ein 95%- Konfidenzintervall um Null gegeben ist durch $\pm 1,96T^{-\frac{1}{2}}$. So kann man für jeden Autokorrelationskoeffizienten feststellen, ob er in diesem Konfidenzintervall liegt oder eine signifikante Korrelation an diesem Lag vorliegt. Auf diese Weise lässt sich aus der grafischen Darstellung der Autokorrelationsschätzungen sagen, ob

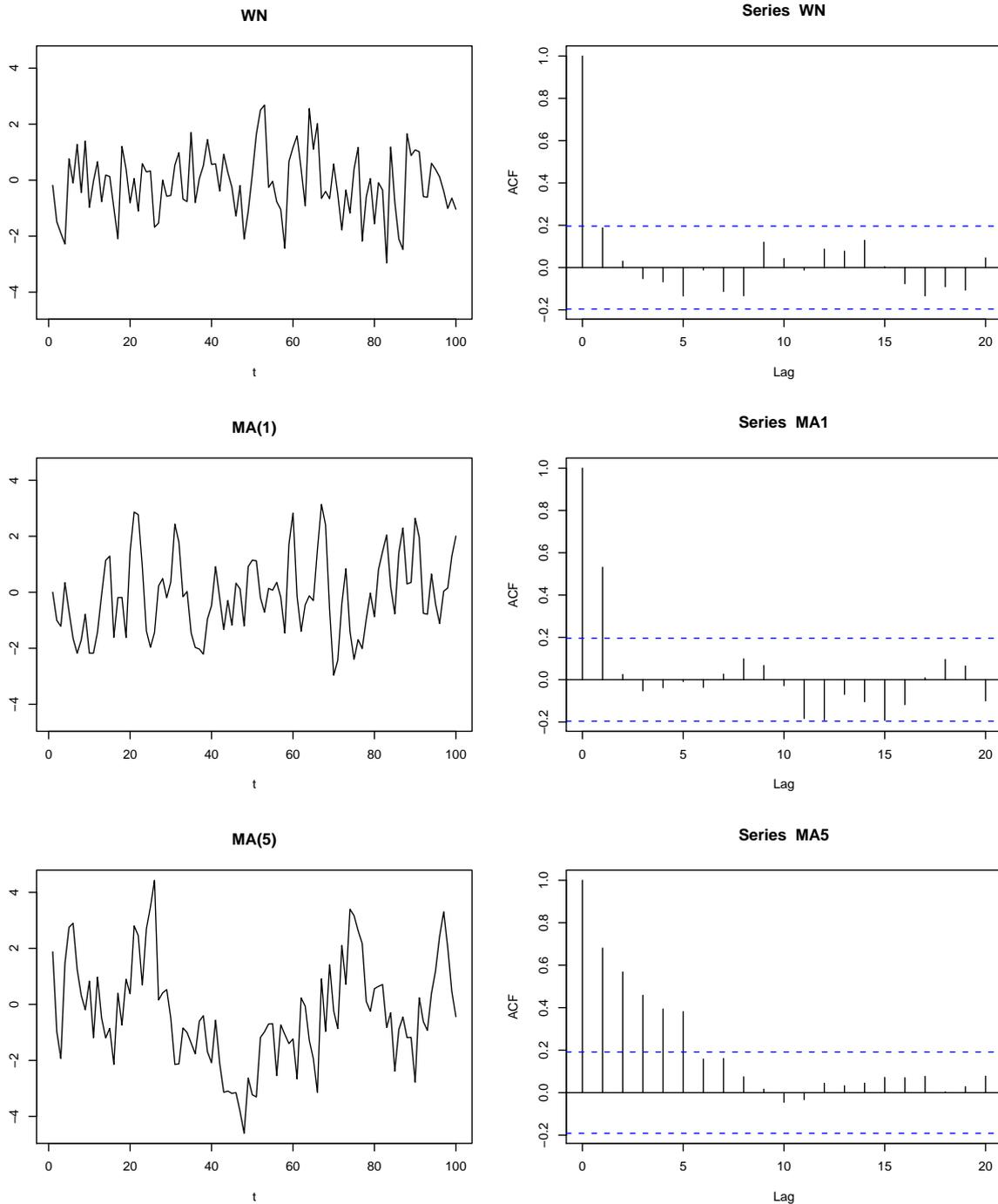


Abbildung 1.8: Drei Zeitreihen und ihre Autokorrelationsfunktion: Weißes Rauschen (oben), MA(1)-Prozess (Mitte) und MA(5)-Prozess (unten): Bis zu dem Lag der Prozessordnung sind die Autokorrelationskoeffizienten signifikant von Null verschieden, so dass man die Prozessordnung durch die Autokorrelationsfunktion bestimmen kann.

es sich um Weißes Rauschen handelt.

In der Abbildung 1.8 sind die Autokorrelationsfunktionen eines White Noise-, eines MA(1)- und eines MA(5)-Prozesses mit zugehörigen Konfidenzintervallen dargestellt. Die Autokorrelation am Lag Null ist trivialerweise in allen drei Abbildungen gleich Eins, da hier die Kovarianz zwischen Y_t und sich selbst berechnet wird. In der ersten Grafik für Weißes Rauschen liegen alle Schätzwerte der Autokorrelationskoeffizienten innerhalb des Konfidenzintervalls, sind folglich statisch nicht von Null verschieden. Daran erkennt man, dass die Lags des Weißes Rauschens unkorreliert sind. In der zweiten Abbildung sind ein MA(1)-Prozess und seine Autokorrelationsfunktion dargestellt. Nur am Lag Eins übersteigt die Schätzung den kritischen Wert des Konfidenzintervalls deutlich. In der Darstellung des MA(5)-Prozesses ist dies bis zum Lag Fünf richtig. Allein aus den Plots (links in den Abbildungen) kann man nicht auf die Prozessordnung schließen, während man sie in der geschätzten Autokorrelationsfunktion (rechts in den Abbildungen) direkt ablesen kann.

Um Weißes Rauschen handelt es sich eigentlich nur dann, wenn *alle* Autokorrelationskoeffizienten gemeinsam in einem Konfidenzband um Null liegen. Diese verbundene Hypothese lässt sich eigentlich nicht durch die Annahme aller Einzelhypothesen $\rho(h) = 0, \forall h$ verifizieren. Das Testen dieser gemeinsamen Hypothese

$$\rho(1) = \rho(2) = \dots = \rho(N) = 0, N = 1, 2, \dots,$$

übernimmt der sogenannte Portemanteau-Test (z.B. Brockwell & Davis, 1996, S.36).

Da der Vektor $(\hat{\rho}(1), \dots, \hat{\rho}(h))$ für große T standardnormalverteilt sind Neusser (2006, Theorem 3.3), folgen die beiden Statistiken (Kirchgässner & Wolters, 2007, S.17 Gleichung (1.11) und (1.12))

$$Q_1 := T \sum_{h=1}^N \hat{\rho}^2(h) \text{ Box-Pierce-Statistik}$$

$$Q_2 := T(T+2) \sum_{h=1}^N \frac{\hat{\rho}^2(h)}{T-h} \text{ Ljung-Box-Statistik}$$

einer χ_N^2 -Verteilung. Die Ljung-Box-Statistik berücksichtigt als eine Verfeinerung der Box-Pierce-Statistik, dass die Schätzung für große h auf weniger Summanden beruht als für kleine (Kirchgässner & Wolters, 2007, S.17). Man testet die Nullhypothese, dass es sich bei dem Prozess um Weißes Rauschen handelt. Unter ihrer Gültigkeit liegen also keine Autokorrelationen vor. Bei großen Werten für die beiden Statistiken Q_1 und Q_2 muss sie verworfen werden. Die kritischen Werte für diesen Fall kann man der χ^2 -Verteilung für N entnehmen (z.B. Bowerman & O'Connell, 1993).

Mit Hilfe der Autokorrelationsfunktion hat man also einen Test auf Weißes Rauschen erhalten und kann auch die Prozessordnung q eines MA-Prozesses schätzen. Da sich autoregressive

Prozesse stets als $MA(\infty)$ -Prozess darstellen lassen, kann man ihre Prozessordnung p jedoch so nicht ermitteln. Dies kann man durch die Verwendung von partiellen Autokorrelationskoeffizienten erreichen.

1.10 Partielle Autokorrelationsfunktion

Autoregressive Prozesse haben theoretisch ein unendlich langes Gedächtnis, da keine Innovation im Prozess verloren geht. Dabei hält der Einfluss einer einzelnen Innovation jedoch nicht beständig an, sondern nimmt mit der Zeit ab. Um den direkten Einfluss eines Lags auf ein anderes zu untersuchen, muss man zunächst die Abhängigkeiten von dazwischen liegenden Lags aus der Korrelation entfernen. Bei der *partiellen* Autokorrelationsfunktion (z.B. Schröder & Buscher, 2002) handelt es sich um die von indirekten Einflüssen bereinigte Korrelation.

Beispiel 1.10.1 (Beispiel ohne Zeitreihe) *Betrachtet man zwei Variablen X, Y , die jeweils mit einer dritten Variable Z korreliert sind, dann wirken sich die Korrelationen von X, Y mit Z auf die Korrelation von X und Y aus.*

Will man nun den linearen Zusammenhang von X und Y ohne den Einfluss von Z bestimmen, muss man den partiellen Korrelationskoeffizienten berechnen Bortz (2005, Kap.13)

$$\rho(X, Y | Z) = \frac{\rho(X, Y) - \rho(X, Z)\rho(Y, Z)}{(1 - \rho(X, Z)^2)(1 - \rho(Y, Z)^2)}.$$

Man sieht, dass alle drei Korrelationen in die Berechnung eingehen.

Wir haben gesehen, dass bei einer autoregressiven Zeitreihe jede Variable mit jeder anderen korreliert ist. Das liegt daran, dass in einem autoregressiven Prozess mindestens auf das letzte Lag zurückgriffen wird, das wiederum vom vorletzten abhängt usw. Um den direkten Einfluss eines Lags Y_{t-h} auf Y_t zu bestimmen, müssen die Einflüssen dazwischenliegender Variablen $Y_{t-1}, \dots, Y_{t-h+1}$ herauspartialisiert werden.

Beispiel 1.10.2 (Partieller Autokorrelationskoeffizient für $h = 2$) *Überträgt man das obige Beispiel auf eine autoregressive Zeitreihe, ist etwa $Y = Y_t$, $Z = Y_{t-1}$ und $X = Y_{t-2}$. Man will nun die partielle Autokorrelation von Y_{t-2} auf Y_t berechnen. Da Y_t stationär ist, gilt:*

$$\begin{aligned} \rho(X, Y) &= \rho(Y_{t-2}, Y_t) = \rho(2), & \text{da der Abstand } h = 2 \text{ ist,} \\ \rho(X, Z) &= \rho(Y_{t-2}, Y_{t-1}) = \rho(1), & \text{da der Abstand } h = 1 \text{ ist,} \\ \rho(Y, Z) &= \rho(Y_t, Y_{t-1}) = \rho(1), & \text{da der Abstand } h = 1 \text{ ist.} \end{aligned}$$

Mit obiger Formel ergibt sich (z.B. Gottman, 1981, S.142, Gleichung (13.2))

$$\rho(Y_t, Y_{t-2} | Y_{t-1}) = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)}. \quad (1.8)$$

Aus dem Regressionsmodell $Y = aX + \epsilon$ ist bekannt, dass

$$\rho(Y, X) = \frac{\sqrt{\text{Var}[X]}}{\sqrt{\text{Var}[Y]}} \cdot a \text{ gilt.}$$

Angewandt auf ein AR(1)-Modell $Y_t = aY_{t-1} + \epsilon_t$ ergibt sich

$$\rho(Y_t, Y_{t-1}) = \frac{\sqrt{\text{Var}[Y_{t-1}]}}{\sqrt{\text{Var}[Y_t]}} \cdot a = a,$$

da in einem stationären Prozess alle Varianzen gleich sind. Hier gilt, dass die Autokorrelation gleich der partiellen Autokorrelation ist, da es keinen Zeitpunkt zwischen den beiden Variablen gibt.

Um die partiellen Autokorrelationskoeffizienten schätzen zu können, bedient man sich der Tatsache, dass sie aus den Koeffizienten der Prognosefunktion bestehen, falls man Y_t anhand vergangener Lags vorhersagen will (Neusser, 2006). Da diese Koeffizienten wiederum aus den Autokorrelationskoeffizienten berechnet werden, kann man die partiellen Autokorrelationskoeffizienten π_i wie in Gleichung (1.8) im Fall $h = 2$ über die Autokorrelationskoeffizienten $\rho(h)$ gewinnen (Neusser, 2006).

Die sogenannten Yule-Walker-Gleichungen (z.B. Fan & Yao, 2003; Schröder & Buscher, 2002; Stier, 2001) stellen die Verbindung zwischen den Autokorrelationskoeffizienten $\rho(h)$ und den partiellen Autokorrelationskoeffizienten π_h her. Dabei ist π_h der partielle Korrelationskoeffizient zum Lag h .

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(h) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \dots & \dots & \dots & \dots \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \dots \\ \pi_h \end{pmatrix} \quad (1.9)$$

Löst man das Gleichungssystem (1.9) bei bekannten (oder bereits geschätzten) Autokorrelationskoeffizienten $\rho(1), \dots, \rho(h)$, so erhält man Schätzwerte für π_1, \dots, π_h . Eine weitere Möglichkeit besteht darin, die Koeffizienten rekursiv durch den sogenannten Durbin-Algorithmus (Durbin, 1960) zu bestimmen.

Man kann zeigen, dass die partiellen Autokorrelationsfunktion für einen Moving Average Prozess theoretisch unendlich lang ist, dass die partiellen Autokorrelationskoeffizienten in der Praxis jedoch exponentiell abnehmen (z.B. Stier, 2001).

Für einen autoregressiven Prozess der Ordnung p bricht die Autokorrelationsfunktion jedoch nach dem p -ten Lag ab, so dass man eine Möglichkeit gefunden hat, auch die Prozessordnung eines AR-Prozesses zu bestimmen. Da wir im nächsten Kapitel jedoch ein weiteres effizienteres Verfahren kennenlernen werden, gehen wir auf partielle Autokorrelation hier nicht weiter ein. Eine ausführliche Beschreibung der Schätzverfahren findet sich in Kirchgässner und

Wolters (2007, Kap.2.1.4).

1.11 Anzahl der Lags: Modellwahl

Jeder stationäre Prozess kann durch einen autoregressiven Prozess beliebig genau angenähert werden. Zur Modellierung eines Prozesses ist die Anzahl der einzubeziehenden Lags entscheidend. Die Anpassungsgüte eines Modells (=“Fit“) wird umso besser, je kleiner die geschätzte Varianz der Residuen ist. Je mehr der Varianz durch die Lags des Prozesses erklärt werden kann, desto besser ist das Modell. Dabei kann es jedoch zum „Overfitting“ kommen, so dass p zu groß geschätzt wird (Neusser, 2006). Auch wird der Schätzaufwand mit jedem zusätzlichen Parameter ständig größer.

Um einen Kompromiss zwischen möglichst guter Anpassung und möglichst wenigen Modellparametern zu finden, versucht man ein sogenanntes *Informationskriterium* zu minimieren. Neben der Residualvarianz enthält das Informationskriterium einen Strafterm, der das Hin-zunehmen weiterer Parameter „bestraft“. Die wichtigsten Informationskriterien sind nach folgendem Schema aufgebaut:

$$IC(p) = \ln \hat{\sigma}_{\epsilon(p)}^2 + p + \frac{C(T)}{T},$$

wobei $C(T)$ eine monoton fallende¹⁸ Funktion ist. Für den Strafterm $p + \frac{C(T)}{T}$ gibt es unterschiedliche Möglichkeiten. Die bekanntesten Informationskriterien sind das Akaike Informationskriterium (Akaike, 1978)

$$AIC(p) = \ln \hat{\sigma}_{\epsilon(p)}^2 + p + \frac{2}{T},$$

das Bayes-Informationkriterium (Schwarz, 1978)

$$BIC(p) = \ln \hat{\sigma}_{\epsilon(p)}^2 + p + \frac{\ln(T)}{T}$$

und das Hannan-Quinn-Kriterium (Hannan & Quinn, 1979)

$$HQ(p) = \ln \hat{\sigma}_{\epsilon(p)}^2 + p + \frac{\ln(\ln(T))}{T}.$$

Das AIC wird zwar am häufigsten verwendet, Simulationsstudien haben jedoch ergeben, dass es die Modellordnung eher überschätzt (Schlittgen & Streitberg, 1999). Die beiden anderen führen zu konsistenten Schätzungen.

Im Anwendungsfall berechnet man die Werte des Informationskriteriums für $p = 0, 1, 2, \dots$ und wählt dasjenige p als Prozessordnung für die autoregressive Darstellung, bei dem der

¹⁸Also nicht-steigend. Eine konstante Funktion ist auch monoton fallend.

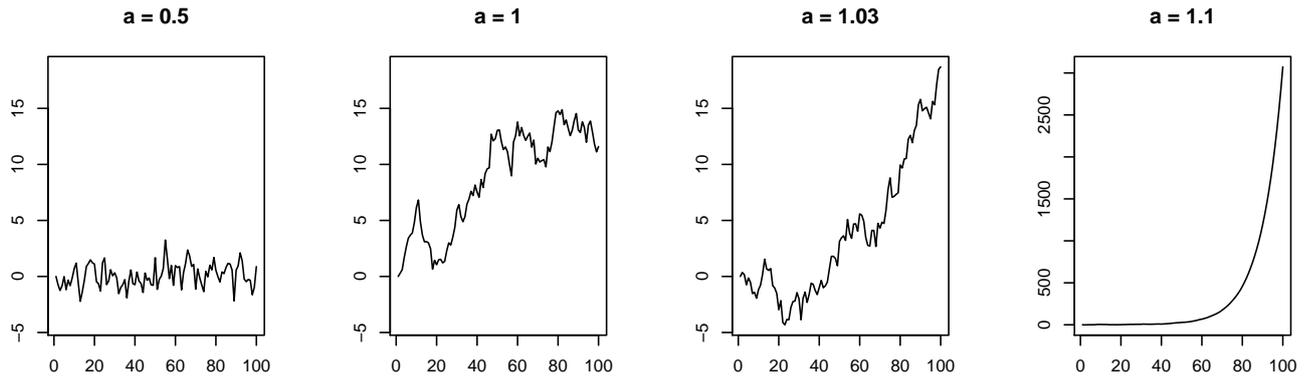


Abbildung 1.9: Der Prozess $Y_t = a \cdot Y_{t-1} + \epsilon_t$ mit verschiedenen Werten für a : stationärer Prozess, Random Walk und zwei Zeitreihen, bei denen der Verlauf „explodiert“. (Beachte: die Abbildung für $a = 1.1$ unterscheidet sich in der Skalierung)

Wert am niedrigsten war.

Wir haben nun einen speziellen Test für Weißes Rauschen und solche Tests kennen gelernt, mit denen man die Prozessordnung in der MA- oder AR-Darstellung aus den vorhandenen Daten ermitteln kann. Uns fehlt nun noch ein Verfahren, um festzustellen, ob ein Prozess stationär ist.

1.12 Unit Root Tests zur Erkennung von Stationarität

Wie haben bereits gesehen, dass eine autoregressive Darstellung nicht nur für stationäre Prozesse, sondern auch für einen Random Walk Prozess existiert. Als Stationaritätskriterium haben wir eine Aussage über die Lösungen der charakteristischen Gleichung verwendet und festgestellt, dass dies für einen AR(1)-Prozess gleichbedeutend damit ist, dass in der Darstellung

$$Y_t = a \cdot Y_{t-1} + \epsilon_t, \tag{1.10}$$

der Koeffizient a vom Betrag kleiner als Eins ist. Für $a = 1$ handelt es sich um einen nicht-stationären Random Walk. Im Fall $|a| > 1$ explodiert der Prozess, so dass er nur von theoretischem Interesse ist, da er in der Praxis nicht auftauchen sollte oder falls doch bereits im Plot eindeutig zu identifizieren ist, wie man in Abbildung 1.9 klar erkennen kann (Man beachte, dass der Plot für $a = 1.1$ auf einer anderen Skala aufgetragen ist). Eine naheliegende Idee ist es nun, für einen Stationaritätstest zu prüfen, ob die Schätzung von a signifikant von Eins abweicht. Da die Nullhypothese stets $a = 1$ lautet, handelt es sich vielmehr um einen Einheitswurzeltest (Unit Root Test) als um einen Stationaritätstest. Die einzig plausible Alternativhypothese ist jedoch die des stationären Prozesses (Neusser, 2006).

Um zu testen, ob eine beobachtete Realisierung Y_1, \dots, Y_T ¹⁹ einem einfachen Random Walk

$$Y_t = Y_{t-1} + \epsilon_t$$

entspricht, schätzt man zunächst in einem AR(1)-Modell wie (1.10) den Parameter a . Dazu kann man den aus der Regressionsanalyse bekannten Kleinste-Quadrate-Schätzer verwenden, der hier mit dem Maximum-Likelihood-Schätzer identisch ist (Seber & Lee, 2003),

$$\hat{a}_T = \frac{\sum_{t=1}^T Y_{t-1} Y_t}{\sum_{t=1}^T Y_{t-1}^2}.$$

Nun kann man zur Überprüfung der Hypothese jedoch den üblichen t-Test nicht benutzen, da die der Regressionsanalyse bekannte t-Statistik in diesem Fall keiner t-Verteilung folgt. Betrachtet man die asymptotische Verteilung der Statistik, so konvergiert sie zwar für $|a| < 1$

$$\sqrt{T} (\hat{a}_T - a) \xrightarrow{d} \mathcal{N}(0, 1 - a^2),$$

im Fall einer gültigen Nullhypothese $a = 1$ degeneriert die Grenzverteilung jedoch zu einer „Ein-Punkt-Verteilung“ mit Varianz Null, $\sigma^2 = 1 - a^2 = 1 - 1^2 = 0$, um den Erwartungswert Null.

Durch Multiplikation mit \sqrt{T} kann man die Statistik jedoch so verändern

$$T (\hat{a}_T - a),$$

dass sie in jedem Fall gegen eine existierende Grenzverteilung konvergiert. Da es sich jedoch nicht um eine Standardverteilung handelt, müssen die kritischen Werte in Simulationsstudien gewonnen werden (Fuller, 1996).

Durch die Skalierung mit T statt mit \sqrt{T} konvergiert der geschätzte Koeffizient im Random Walk Fall schneller gegen den wahren Wert 1, als im stationären Fall $|a| < 1$. Diese Eigenschaft des Schätzers wird als *Superkonsistenz* (Hamilton, 1994, S.476) bezeichnet.

Die Herleitung der Grenzverteilung geschieht mit Hilfe der Darstellung über eine Brownsche Bewegung. Eine gute Beschreibung findet sich bei Stier (2001, Kap. 18.4) oder Maddala und Kim (2007, Kap.3.3.2). Für uns ist jedoch nur die Existenz dieser Grenzverteilung bedeutend. Ahtola und Tiao (1987) leiten die Grenzverteilung im Falle von komplexen Einheitswurzeln auf dem Einheitskreis ab, wie sie bei AR-Prozessen höherer Ordnung vorkommen können.

Wie im obigen Beispiel ist der Ausgangspunkt für Einheitswurzeltests zunächst die Regression von Y_t auf Y_{t-1} (siehe Gleichung (1.10), S.42). Meist werden jedoch auch deterministische Terme in die dann sogenannte Dickey-Fuller-Regression (Hamilton, 1994) aufgenommen.

Bemerkung 1.12.1 (Dickey-Fuller-Regression) *Die Dickey-Fuller-Regression wird als Grund-*

¹⁹Wir nehmen an, dass $Y_0 = 0$.

lage für die bekanntesten Unit-Root-Tests verwendet

$$Y_t = \text{deterministische Terme} + aY_{t-1} + \epsilon_t. \quad (1.11)$$

Häufig findet man auch die Schreibweise dieser Regression über ihre Differenz

$$\begin{aligned} Y_t - Y_{t-1} &= \text{deterministische Terme} + aY_{t-1} - Y_{t-1} + \epsilon_t \\ \Delta Y_t &= \text{deterministische Terme} + \underbrace{(a-1)}_{\rho} Y_{t-1} + \epsilon_t \\ \Delta Y_t &= \text{deterministische Terme} + \rho Y_{t-1} + \epsilon_t \end{aligned} \quad (1.12)$$

Die beiden Formulierungen (1.11) und (1.12) sind numerisch identisch. Die Nullhypothese ist stets „Der Prozess hat eine Einheitswurzel“, also

$$\begin{aligned} H_0 : a = 1 \text{ gegen } H_1 : -1 < a < 1 \\ \text{bzw. } H_0 : \rho = 0 \text{ gegen } H_1 : -2 < \rho < 0 \end{aligned} \quad (1.13)$$

In der zweiten Darstellung (1.13) sieht man besonders gut, dass es sich um einen einseitigen Test handelt²⁰.

Für den Dickey-Fuller-Test (Dickey & Fuller, 1979) gibt es zwei Teststatistiken, wobei die oben betrachtete Statistik $T(\hat{a} - 1)$ ρ -Statistik genannt wird. Sie ist nicht normalverteilt, doch gut zu simulieren, so dass für ihre Verteilung Tabellen vorliegen (Fuller, 1996; Hamilton, 1994). Die zweite ist die t-Statistik

$$\begin{aligned} t_{\hat{a}} &= \frac{\hat{a} - 1}{\hat{\sigma}_{\hat{a}}} \\ \text{mit } \hat{\sigma}_{\hat{a}} &= \sqrt{\frac{s_T^2}{\sum_{t=1}^T Y_{t-1}^2}} \\ \text{und } s_T^2 &= \frac{1}{T-2} \sum_{t=2}^T (Y_t - \hat{a}_T Y_{t-1})^2. \end{aligned}$$

Auch ihre Verteilung liegt nur tabellarisch vor. MacKinnon (1981) bietet eine Möglichkeit an, aus den Tabellen kritische Werte für beliebige Längen T zu interpolieren.

Erschwerend kommt noch hinzu, dass die Grenzverteilungen der beiden Statistiken sowohl von den deterministischen Termen, die man in die Dickey-Fuller-Regression aufnimmt, als auch von dem (unbekannten) wahren Prozess abhängen, der die Daten tatsächlich erzeugt.

Dabei unterscheidet man vier Fälle (siehe Tabelle 1.1) (Hamilton, 1994), die sich vor allem darin unterscheiden, ob man die Dickey-Fuller-Regression mit einem linearen Trend betrachtet, und den Daten erzeugenden Prozess als Random Walk mit oder ohne Drift (siehe

²⁰Wie bereits erwähnt, ist es nicht üblich den Fall $|a| > 1$ bzw. $\rho < -2$ zu testen.

Tabelle 1.1: Vier Fälle bei Dickey-Fuller-Regression

	Daten generierender Prozess	Dickey-Fuller-Regression
Fall 1	einfacher Random Walk	ohne deterministischen Anteil
Fall 2	einfacher Random Walk	konstanter Mittelwert
Fall 3	Random Walk mit Drift	konstanter Mittelwert
Fall 4	Random Walk mit Drift	linearer Trend

Gleichung (1.7), S.35) annimmt.

Im einfachsten Fall (Fall 1) nimmt man einen einfachen Random Walk an und verwendet die Dickey-Fuller-Regression ganz ohne deterministischen Anteil. Diesen Fall haben wir zu Beginn des Kapitels bereits betrachtet. Wird die Nullhypothese „Prozess ist einfacher Random Walk“ abgelehnt, handelt es sich um einen stationären Prozess mit Erwartungswert Null. Da man einen von Null verschiedenen Mittelwert in der Anwendung jedoch selten wirklich ausschließen kann, ist dieser Fall eher von theoretischem Interesse (Neusser, 2006), so dass man in die Regressionsgleichung noch eine Konstante aufnehmen sollte. Dadurch ändert sich jedoch wider der Erwartung die Grenzverteilung (Stier, 2001; Hamilton, 1994), so dass man folgenden Fall 2 betrachtet.

Auch hier wird der Daten generierende Prozess als einfacher Random Walk angenommen, in der Dickey-Fuller-Regression wird jedoch zusätzlich ein konstanter Mittelwert berücksichtigt.

$$\begin{aligned} \text{DGP}^{22} : Y_t &= Y_{t-1} + \epsilon_t \\ \text{DFR}^{22} : Y_t &= c + aY_{t-1} + \epsilon_t \end{aligned}$$

Der dritte Fall (Fall 3) unterscheidet sich vom zweiten darin, dass man nun einen Random Walk mit Drift betrachtet.

$$\begin{aligned} \text{DGP} : Y_t &= c + Y_{t-1} + \epsilon_t \\ \text{DFR} : Y_t &= c + aY_{t-1} + \epsilon_t \end{aligned}$$

In diesem Fall ist die t-Statistik asymptotisch normalverteilt (Neusser, 2006). Bei genauer Betrachtung sieht man jedoch, dass unter der Nullhypothese (DGP) ein Prozess mit zeitlichem Trend erwartet wird, in der Alternativhypothese (DFR) dies jedoch ausgeschlossen wird. In der Praxis sollte man jedoch wissen, ob ein linearer Trend vorliegt, und nur noch testen, ob der Prozess eine Einheitswurzel besitzt oder stationär um diesen Trend verläuft. Wir werden im Abschnitt 2.1.2 noch genauer auf trendstationäre Prozesse eingehen.

Interessant ist also vor allem der Fall 4, bei dem in beiden Hypothesen ein linearer Trend

²¹Daten generierender Prozess

²²Dickey-Fuller-Regression

hinzugenommen wird.

$$\begin{aligned} \text{DGP: } Y_t &= c + Y_{t-1} + \epsilon_t \\ \text{DFR: } Y_t &= c + d \cdot t a Y_{t-1} + \epsilon_t \end{aligned}$$

Die Verteilungen der Statistiken sind nach diesen vier Fällen getrennt in (Hamilton, 1994) tabellarisiert.

Bevor man einen Unit-Root-Test anwendet, sollte man also zunächst den Prozess plotten und anhand der Grafik entscheiden, ob man einen linearen Trend ins Modell aufnehmen will (Fall 4) oder ob eine Konstante ausreichend ist (Fall 2), weil kein zeitlicher Trend zu erkennen ist (Neusser, 2006). In seltenen Fällen kann man auch auf die Konstante verzichten (Fall 1), wenn man einen von Null verschiedenen Mittelwert ausschließen kann.

Auch die F-Statistik für die verbundene Hypothese $c = 0, a = 1$ (Fall 2) bzw. $d = 0, a = 1$ (Fall 4) folgt keiner bekannten Standardverteilung und liegt deshalb nur in Form von Tabellen vor, wird jedoch von den meisten Software-Paketen angeboten.

Die bisherigen Ausführungen beschränkten sich auf die Regression auf Y_{t-1} . Weitere mögliche Autokorrelationen im Prozess wurden bislang nicht berücksichtigt. Die beiden bekanntesten Unit-Root-Tests, der Augmented Dickey-Fuller-Test (Dickey & Fuller, 1979) und der Phillips-Perron-Test (P. C. B. Phillips & Perron, 1988), haben unterschiedliche Strategien dieses Problem zu lösen.

Für den **Augmented Dickey-Fuller-Test** werden Differenzen weiterer Lags in die Regression mit aufgenommen und so einer möglichen Autokorrelation Rechnung getragen.

$$Y_t = \text{deterministische Terme} + aY_{t-1} + b_1\Delta Y_{t-1} + \dots + b_{p-1}\Delta Y_{t-p+1} + \epsilon_t$$

Durch diesen Vorgang ändern sich die besprochenen Grenzverteilungen nicht, so dass die Tabellen weiter verwendet werden können. Die Ordnung p sollte so groß gewählt werden, dass ϵ_t ein Weißes Rauschen ist. Die Wahl von p erfolgt in den Software-Lösungen automatisch nach bewährten Algorithmen (Trapletti & Hornik, 2009). Wir haben in Kapitel 1.9 den Portemanteau-Test als einen Test auf Weißes Rauschen kennengelernt.

Beim **Phillips-Perron-Test** wird die Idee verfolgt, die Statistik nachträglich (nichtparametrisch) zu korrigieren (Neusser, 2006). Dazu muss V_t in der Dickey-Fuller-Regression

$$Y_t = \text{deterministische Terme} + aY_{t-1} + V_t$$

kein Weißes Rauschen sein, sondern kann einen beliebigen stationären Prozess mit Erwartungswert Null darstellen. Die Korrektur geschieht dann mit Hilfe der sogenannten langfristigen Varianz (Neusser, 2006). Der PP-Test lässt auch Heteroskedastizität zu (Stier, 2001; Perron, 1988; Banerjee, 2000).

Im Vergleich hat der Phillips-Perron-Test eine höhere Macht als der ADF-Test, so dass die Wahrscheinlichkeit eine falsche Nullhypothese abzulehnen größer ist. Allerdings ist er auch stärker verzerrt, so dass die Nullhypothese zu häufig abgelehnt wird (Neusser, 2006). Diese Eigenschaften bestätigen sich auch in den in Kapitel 3 durchgeführten Simulationsstudien.

Die Literatur zu Unit-Root-Tests ist unermesslich groß (Diebold & Nerlove, 1990, S.4). Wir haben hier nur die Grundlagen beschrieben, die wir im weiteren Verlauf der Arbeit verwenden werden. Eine ausführliche Darstellung weiterer Einheitswurzeltests findet sich bei Stier (2001) oder Maddala und Kim (2007). Elder und Kennedy (2001) beschreiben, wie man vorgehen kann, wenn aus dem Plot nicht eindeutig hervorgeht, ob man einen Trend hinzunehmen sollte.

1.13 Multivariate Zeitreihen zur Beschreibung von Zeitreihensystemen

Während man in der univariaten Zeitreihenanalyse nur eine einzige Reihe betrachtet, dient die multivariate Zeitreihenanalyse dazu, Systeme von Zeitreihen und ihre Wirkungsweise zu erkennen (z.B. Brockwell & Davis, 2006, Kap.11 Multivariate Time Series). Da es sich bei Kointegration um ein multivariates Phänomen handelt, müssen wir unsere Ergebnisse auf multivariate Zeitreihen übertragen, um sie im nächsten Kapitel für die Kointegrationsanalyse verwenden zu können (z.B. Schröder & Buscher, 2002, S.221).

Definition 1.13.1 (Multivariate Zeitreihe, multivariater Prozess) *Unter einem multivariaten (diskreten) Prozess oder einer multivariaten Zeitreihe versteht man eine Folge von vektoriellen Zufallsvariablen Y_1, Y_2, \dots , die fortlaufend in der Zeit indiziert sind.*

Diese Definition ist vergleichbar mit der univariaten Definition (Definition 1.1.1, S.15). Nur dass es sich im multivariaten Fall um *vektorielle* Zufallsvariablen bzw. Zufallsvektoren handelt, so dass sich der Prozess wie folgt darstellt

$$Y_t = \left\{ \left(\begin{array}{c} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{Kt} \end{array} \right) = \left(\begin{array}{c} Y_{1,0} \\ Y_{2,0} \\ \vdots \\ Y_{K,0} \end{array} \right), \left(\begin{array}{c} Y_{1,1} \\ Y_{2,1} \\ \vdots \\ Y_{K,1} \end{array} \right), \dots, \left(\begin{array}{c} Y_{1,t} \\ Y_{2,t} \\ \vdots \\ Y_{K,t} \end{array} \right) \right\}.$$

Ein vektorieller Zufallsprozess hat zu jedem Zeitpunkt einen vektoriellen Wert. Also je nach Dimension des Prozesses $K = 1, 2, 3, \dots$ Komponenten. Eine univariate Zeitreihe ist folglich ein Spezialfall mit $K = 1$. Mit diesem Wissen im Hintergrund werden wir auch multivariate Prozesse weiterhin mit Y_t bezeichnen, ohne ihre Mehrdimensionalität durch etwa Fettschreibung zu betonen.

Die Betrachtung als *vektorieller* Zufallsprozess dient häufig nur der einfacheren Schreibweise.

Viele Ausdrücke vereinfachen sich, wenn man sie in Vektoren- und Matrizenschreibweise angeben kann, da sie dann sehr an den univariaten Fall erinnern und leicht zu überblicken sind.

Für die Vorstellung ist es jedoch passender, einen multivariaten Prozess als System von univariaten Zeitreihen zu betrachten. So hat ein multivariater Prozess in der einen Sichtweise zu jedem Zeitpunkt einen vektoriellen K -dimensionalen Wert oder er besteht in der anderen Perspektive aus K eindimensionalen Reihen, die jeweils einen reellen (skalaren²³) Wert haben.

$$\begin{aligned} Y_{1t} &= Y_{1,0}, Y_{1,1}, Y_{1,2}, \dots, Y_{1,t} \\ Y_{2t} &= Y_{2,0}, Y_{2,1}, Y_{2,2}, \dots, Y_{2,t} \\ &\vdots \\ Y_{Kt} &= Y_{K,0}, Y_{K,1}, Y_{K,2}, \dots, Y_{K,t} \end{aligned}$$

So kann man nun in diesem multivariaten System zunächst die K Einzelreihen z.B. mit den bisher kennengelernten Methoden analysieren und anschließend das gemeinsame System betrachten. Denn ein multivariater Prozess ist mehr als „die Summe seiner Teile“. Es handelt sich nicht einfach um die Zusammenfassung mehrerer Reihen zu einem vektoriellen Prozess, sondern um ein *System* von mehreren Reihen, in dem es verschiedene Wechselwirkungen der Einzelreihen geben kann. Dabei können die Einflüsse einseitig (kausal) sein oder es können Feedbacksysteme vorliegen, in denen sich mehrere Einzelprozesse zeitgleich gegenseitig beeinflussen (C. W. Granger, 1969). Für die Kointegrationsanalyse ist dieser Systemcharakter entscheidend.

Analog zum univariaten Prozess hat jeder multivariate Prozess einen (vektoriellen) Erwartungswertprozess, der als K -dimensionaler Erwartungsvektor zum Zeitpunkt t aus den K Erwartungswerten der Komponentenreihen besteht (Fuller, 1996, Kap.1.7)

$$\mu_t = \mathbb{E} \left[\begin{pmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{Kt} \end{pmatrix} \right] = \begin{pmatrix} \mathbb{E}[Y_{1,t}] \\ \mathbb{E}[Y_{2,t}] \\ \vdots \\ \mathbb{E}[Y_{K,t}] \end{pmatrix}.$$

Der multivariate Erwartungswertprozess μ entsteht aus den Einzelprozessen einfach durch gemeinsame Betrachtung der Einzelerwartungswerte zu einem gegebenen Zeitpunkt t . Für die zweiten Momente des Prozesses ergibt sich durch die mögliche Wechselwirkung der Einzelprozesse folgende Erweiterung. Während man bei einer univariaten Zeitreihe ausschließlich die Kovarianzen der einzelnen Lags betrachten kann, kommen im multivariaten Fall noch Kovarianzen zwischen den Reihen im System dazu. Man kann also für je zwei Zeitpunkte $K \cdot K$ Kovarianzen betrachten. Denn das Lag jeder der K Einzelreihen kann mit dem Lag der

²³in Abgrenzung zu vektoriell

eigenen Reihe zum zweiten Zeitpunkt und mit dem entsprechenden Lag jeder der $K - 1$ anderen Einzelreihen korrelieren. So erhält man als Kovarianzfunktion zu je zwei Zeitpunkten eine Matrix, in der alle (Kreuz-)Kovarianzen der Vektoren Y_t und Y_{t-h} aufgetragen sind

$$\Gamma(t, t - h) = \mathbb{E}[Y_t Y_{t-h}] = \begin{pmatrix} \gamma_{11}(t, t - h) & \dots & \gamma_{1K}(t, t - h) \\ & \dots & \\ \gamma_{K1}(t, t - h) & \dots & \gamma_{KK}(t, t - h) \end{pmatrix}.$$

Dabei sind die einzelnen Kovarianzen wie üblich durch $\gamma_{ij}(t, t - h) = \text{Cov}[Y_{i,t}, Y_{j,t-h}] = \mathbb{E}[Y_{i,t} Y_{j,t-h}]$ definiert.

Beispiel 1.13.2 (Bivariater Fall) *Im bivariaten Fall gibt es nun also zu je zwei Zeitpunkten vier Kovarianzen. Wir notieren sich hier für t und $t - 1$*

$$\Gamma(t, t - 1) = \begin{pmatrix} \gamma_{11}(t, t - 1) & \gamma_{12}(t, t - 1) \\ \gamma_{21}(t, t - 1) & \gamma_{22}(t, t - 1) \end{pmatrix} = \begin{pmatrix} \text{Cov}[Y_{1,t}, Y_{1,t-1}] & \text{Cov}[Y_{1,t}, Y_{2,t-1}] \\ \text{Cov}[Y_{2,t}, Y_{1,t-1}] & \text{Cov}[Y_{2,t}, Y_{2,t-1}] \end{pmatrix}.$$

Für $h = 0$ sind dies die gleichzeitigen Einflüsse. Auf der Diagonalen von $\Gamma(t, t)$ stehen also die Varianzen der Einzelprozesse. Auch ein multivariater Prozess ist stationär, wenn seine Momente zeitunabhängig sind (Friz & Victoir, 2010; Brockwell & Davis, 2006, Def.11.1.1):

Definition 1.13.3 (stationärer multivariater Prozess) *Ein multivariater Prozess ist stationär, wenn sein Erwartungswertvektor konstant ist und seine Kovarianzfunktion nur vom Lag h und nicht von der Zeit abhängig ist.*

Dann kann man die Kovarianzmatrix für $t, t - h$ unabhängig von dem konkreten Zeitpunkt t schreiben als (Stier, 2001)

$$\Gamma(h) = \begin{pmatrix} \gamma_{11}(h) & \dots & \gamma_{1K}(h) \\ & \dots & \\ \gamma_{K1}(h) & \dots & \gamma_{KK}(h) \end{pmatrix}.$$

In einem stationären multivariaten Prozess sind auch alle Einzelprozesse stationär. Umgekehrt muss ein System von stationären Einzelprozessen nicht zu einem stationären multivariaten Prozess führen, da es Kovarianzen zwischen den Einzelreihen geben könnte, die zeitabhängig sind (Stier, 2001, S.66) .

Der einfachste multivariate stationäre Prozess ergibt sich aus einem System von Weißem Rauschen, das wir als einfachsten univariaten stationären Prozess kennengelernt hatten (Brockwell & Davis, 2006, Def.11.1.2).

Definition 1.13.4 (vektorielles Rauschen) *Ein Prozess ϵ_t heißt multivariates Weißes Rauschen, falls sein Erwartungswertvektor gleich dem Nullvektor ist und für seine Kovarianz-*

matrixfunktion gilt

$$\mathbb{E}[\epsilon_t \epsilon_{t-h}] = \begin{cases} \Sigma & h = 0 \\ 0 & h \neq 0. \end{cases}$$

Die Lags der White Noise-Prozesse sind unkorreliert, allerdings kann es Korrelationen zwischen den Lags verschiedener Einzelprozesse zum gleichen Zeitpunkt geben. Hat Σ Diagonalgestalt sind auch gleichzeitige (kontemporäre) Korrelationen ausgeschlossen.

1.14 Lagpolynommatrizen

Wir hatten univariate Moving Average oder autoregressive Prozesse mit Hilfe von Lagpolynomen dargestellt. Um vektorielle Prozesse analysieren zu können, müssen wir diesen Ansatz für den multivariaten Fall erweitern. Unsere Zufallsvariablen Y_t sind zu K -dimensionalen Zufallsvektoren geworden. Ebenso werden aus den Konstanten $c \in \mathbb{R}$ nun konstante Vektoren $c \in \mathbb{R}^K$. Die Parameter der Lagpolynome werden nun zu *Parametermatrizen*. Diese $K \times K$ -Matrizen machen deutlich, dass jedes Lag jeder Einzelreihe auch den Zustand aller anderen Einzelreihen beeinflussen kann. Ein zweidimensionales Beispiel soll dies verdeutlichen.

Beispiel 1.14.1 (Bivariater Fall; VAR(2)) *Ein bivariater vektorieller autoregressiver Prozess der Ordnung zwei, kurz VAR(2), lässt sich wie folgt darstellen*

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \epsilon_t. \quad (1.14)$$

Dabei sind Y_t , Y_{t-1} und Y_{t-2} Zufallsvektoren, A_1 und A_2 Parametermatrizen und ϵ_t vektorielles Weißes Rauschen. In ausführlicher Schreibweise stellt sich ein bivariater VAR(2)-Prozess also wie folgt dar

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} a_{1,11} & a_{1,12} \\ a_{1,21} & a_{1,22} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} a_{2,11} & a_{2,12} \\ a_{2,21} & a_{2,22} \end{pmatrix} \begin{pmatrix} Y_{1,t-2} \\ Y_{2,t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Schreibt man diesen vektoriellen Prozess nun als multivariates System der einzelnen univariaten Prozesse ergibt sich

$$\begin{aligned} Y_{1t} &= a_{1,11} Y_{1,t-1} + a_{1,12} Y_{2,t-1} + a_{2,11} Y_{1,t-2} + a_{2,12} Y_{2,t-2} + \epsilon_{1t} \\ Y_{2t} &= a_{1,21} Y_{1,t-1} + a_{1,22} Y_{2,t-1} + a_{2,21} Y_{1,t-2} + a_{2,22} Y_{2,t-2} + \epsilon_{2t}. \end{aligned}$$

Jede Einzelreihe hängt also sowohl von ihren eigenen Lags als auch von denen der anderen Reihe ab. In dieser Systemschreibweise als univariate Prozesse kann man nun einige

Ausdrücke über den Lagoperator darstellen, denn es gilt $Y_{t-1} = LY_t$ usw.

$$\begin{aligned} Y_{1t} &= a_{1,11}LY_{1,t} + a_{1,12}LY_{2,t} + a_{2,11}L^2Y_{1,t} + a_{2,12}L^2Y_{2,t} + \epsilon_{1t} \\ Y_{2t} &= a_{1,21}LY_{1,t} + a_{1,22}LY_{2,t} + a_{2,21}L^2Y_{1,t} + a_{2,22}L^2Y_{2,t} + \epsilon_{2t} \end{aligned}$$

und anschließend zu Lagpolynomen zusammenfassen

$$\begin{aligned} Y_{1t} &= (a_{1,11}L + a_{2,11}L^2) Y_{1,t} + (a_{1,12}L + a_{2,12}L^2) Y_{2,t} + \epsilon_{1t} \\ Y_{2t} &= (a_{1,21}L + a_{2,21}L^2) Y_{1,t} + (a_{1,22}L + a_{2,22}L^2) Y_{2,t} + \epsilon_{2t}. \end{aligned}$$

Bringt man diese nun auf die linke Seite der Gleichung und nimmt $Y_{1t} = 1 \cdot Y_{1t}$ mit auf, dann erhält man

$$\begin{aligned} (1 - a_{1,11}L - a_{2,11}L^2) Y_{1,t} + (0 - a_{1,12}L - a_{2,12}L^2) Y_{2,t} &= \epsilon_{1t} \\ (0 - a_{1,21}L - a_{2,21}L^2) Y_{1,t} + (1 - a_{1,22}L - a_{2,22}L^2) Y_{2,t} &= \epsilon_{2t}. \end{aligned}$$

Auf jede Einzelreihe wird also ein Lagpolynom angewandt und zwar in beiden Gleichungen. Nun kann man diese Lagpolynome in einer Matrix zusammenfassen

$$\begin{pmatrix} (1 - a_{1,11}^1L - a_{2,11}^2L^2) & (0 - a_{1,12}^1L - a_{2,12}^2L^2) \\ (0 - a_{1,21}^1L - a_{2,21}^2L^2) & (1 - a_{1,22}^1L - a_{2,22}^2L^2) \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

So kommt man zu einer Darstellung, die der univariaten wieder sehr ähnlich ist,

$$A(L)Y_t = \epsilon_t. \quad (1.15)$$

Nur dieses Mal wird $A(L)$ als Lagpolynommatrix verstanden und Y_t und ϵ_t als vektorielle Prozesse.

Man kann $A(L)$ weiterhin als Polynom betrachten, dass nun im multivariaten Fall wie in Gleichung (1.14) Matrizen als Parameter hat; $A(L)$ ist hier ein *Matrixpolynom*. Andererseits kann man auch die Koeffizienten aus den einzelnen Matrizen in jeder Komponente zusammenfassen, so erhält man für $A(L)$ die Bedeutung einer *Polynommatrix* wie in Gleichung (1.15). Wir werden beide Sichtweisen verwenden. Schon im bivariaten Fall wird deutlich, warum man die vektorielle Schreibweise der mit allen Einzelreihen vorzieht.

Wir hatten im univariaten Fall ein Stationaritätskriterium für autoregressive Prozess über die charakteristische Gleichung formuliert, die vom AR-Lagpolynom abgeleitet wurde (Theorem 1.7.1, S.32). Da $A(L)$ im multivariaten Fall nun kein einfaches Polynom mehr ist, kann man das Kriterium $A(z) = 0$ nicht mehr verwenden. Die z -Transformierte wird wie im univariaten Fall durch Ersetzen von L durch z in der Lagpolynommatrix erzeugt. Als Stationaritätskriterium (Maddala & Kim, 2007) verwendet man jedoch

Theorem 1.14.2 (Stationaritätskriterium für VAR-Prozesse) *Ein vektorieller autoregressi-*

ver Prozess (VAR-Prozess) ist genau dann stationär, wenn die Determinante seiner Lagpolynommatrix, die selbst wieder ein Polynom ist, nur Nullstellen außerhalb des Einheitskreises besitzt, wenn also gilt

$$|A(z)| = 0 \Rightarrow |z| > 1.$$

Beispiel 1.14.3 Wir betrachten folgenden VAR(1)-Prozess

$$\begin{aligned} Y_{1t} &= -0.5Y_{1,t-1} - 1.0Y_{2,t-1} + \epsilon_{1t} \\ Y_{2t} &= 0.1Y_{1,t-1} + 0.4Y_{2,t-1} + \epsilon_{2t}. \end{aligned}$$

Die z -Transformierte seines Lagpolynoms ist

$$\begin{aligned} A(z) &= \mathbb{I}_2 - A_1 z \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} -0.5 & -1.0 \\ 0.1 & 0.4 \end{pmatrix} z \\ &= \begin{pmatrix} 1 + 0.5z & z \\ -0.1z & 1 - 0.4z \end{pmatrix}. \end{aligned}$$

Die Determinante dieser Polynommatrix ist

$$\begin{aligned} |A(z)| &= (1 + 0.5z)(1 - 0.4z) + 0.1z^2 \\ &= z^2 - z - 10. \end{aligned}$$

Die Lösungen dieser quadratischen Gleichung sind $z = -2.7$ und $z = 3.7$. Sie sind vom Betrag beide größer als Eins. Der Prozess ist also stationär.

1.15 Impulsantwortfunktionen: langfristige Wirkung externer Schocks

Im univariaten Fall hatten wir gesehen, dass man mit Hilfe der Autokorrelationsfunktion und der partiellen Autokorrelationsfunktion untersuchen kann, wie lange der Einfluss einer Innovation im Prozess bleibt. Bei einem vektoriiellen autoregressiven Prozess können die Innovationen einer Einzelreihe auch alle anderen Zeitreihen im System beeinflussen. Die Auswirkungen eines einzelnen Impulses auf das ganze System kann man anhand der sogenannten Impulsantwortfunktion analysieren.

Um dies zu sehen, betrachten wir zunächst einen stationären VAR(p)-Prozess mit *kontemporär unkorrelierten* Störvariablen ϵ_{it} (die Kovarianzmatrix Σ hat also Diagonalgestalt) in

seiner Moving-Average-Darstellung zum Zeitpunkt $t + s$

$$Y_{t+s} = \epsilon_{t+s} + B_1\epsilon_{t+s-1} + B_2\epsilon_{t+s-2} + \dots + B_s\epsilon_t + \dots$$

Dabei beschreiben ϵ_t die Vektoren der zufälligen Einflüsse (Innovationen) und B_i die Koeffizientenmatrizen, die den Einfluss dieser Vektoren gewichten. Formal betrachtet ist die Matrix B_s die partielle Ableitung des vektoriellen Prozesses Y_{t+s} nach dem Vektor ϵ_t

$$B_s = \frac{\partial Y_{t+s}}{\partial \epsilon_t}.$$

Die Einträge der Matrix B_s sind wiederum univariate partielle Ableitungen (siehe Anhang B.4, S.335) der Teilprozesse $Y_{i,t+s}$ zum Zeitpunkt $t + s$ in ihrer MA-Darstellung abgeleitet nach den jeweiligen Teilstörvariablen ϵ_{jt} zum Zeitpunkt t .

$$B_s = \begin{pmatrix} \frac{\partial Y_{1,s+t}}{\partial \epsilon_{1,t}} & \frac{\partial Y_{2,s+t}}{\partial \epsilon_{1,t}} & \frac{\partial Y_{3,s+t}}{\partial \epsilon_{1,t}} & \dots \\ \frac{\partial Y_{1,s+t}}{\partial \epsilon_{2,t}} & \frac{\partial Y_{2,s+t}}{\partial \epsilon_{2,t}} & \frac{\partial Y_{3,s+t}}{\partial \epsilon_{2,t}} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Der Eintrag $\frac{\partial Y_{i,s+t}}{\partial \epsilon_{j,t}}$ beschreibt die Reaktion des i -ten Teilprozesses zum Zeitpunkt $t + s$ auf eine Veränderung im Teilprozess Y_j zum Zeitpunkt t , also wie sich der Einfluss von einem sogenannten Impuls in einer Einzelreihe nach einer gewissen Zeitspanne in einer anderen Einzelreihe auswirkt.

Unter einem *Impuls* versteht man die einmalige Veränderung der Innovation ϵ_j im Zeitpunkt t um eine Einheit bei Konstanz aller anderen Innovationen (Kirchgässner & Wolters, 2007, S.138). Dies ist ein sehr isoliertes Szenario. Diese Sichtweise ist jedoch manchmal für die Analyse des Gesamtsystems sehr hilfreich. Die *Impulsantwort* ist die Reaktion der Variablen Y_i zu einem späteren Zeitpunkt $t + s$. Betrachtet man nun den Verlauf dieser Reaktion auf den einen Impuls, spricht man von einer *Impulsantwortfunktion*. Impulsantwortfunktionen sind theoretisch unendlich lang, konvergieren aber bei Stationarität gegen Null (Stier, 2001).

Bei dieser isolierten Betrachtung hat man jedoch vorausgesetzt, dass sich die Innovationen untereinander nicht beeinflussen, sie kontemporär unkorreliert sind. Dies ist jedoch eine sehr starke Forderung, die in der Praxis nur selten aufrechterhalten werden kann, da man z.B. nicht alle Störfaktoren ausschließen kann, die theoretisch auf mehrere Einzelprozesse gleichzeitig wirken können. In der Regel hat die Kovarianzmatrix Σ keine Diagonalgestalt, so dass man nicht ohne Weiteres Impulsantwortfunktionen verwenden kann.

In diesem Fall gibt es jedoch die Möglichkeit, die Innovationen durch trianguläre Faktorisierung oder die sogenannte Cholesky-Zerlegung zu orthogonalisieren (z.B. Alho & Spencer, 2005; Schröder & Buscher, 2002, S. 237), was jedoch den ursprünglichen VAR-Prozess verändert.

Wir werden Impulsantwortfunktionen jedoch nur für theoretische Erweiterungen verwenden, so dass wir auf diese Verfahren nicht weiter eingehen werden, sondern an dieser Stelle die Einführung in die allgemeine Analyse von Zeitreihen beenden und uns dem Hauptthema dieser Arbeit, der Kointegrationsanalyse, zuwenden.

2 Kointegration: gemeinsame Trends in Zeitreihensystemen

Bislang haben wir vor allem Analyseverfahren und Darstellungen für stationäre Prozesse betrachtet, die geeignet sind, stabile psychologische Phänomene zu beschreiben. Bei solchen Zeitreihen unterscheiden sich die Messwerte an verschiedenen Zeitpunkten nur durch eine zufällige Abweichung, die man der Fehlerkomponente der Zeitreihe zuordnen kann.

Es gibt jedoch auch viele andere Phänomene in der Psychologie, die nicht stabil sind, sondern einen zeitabhängigen Verlauf aufweisen. Viele Fragestellungen der Entwicklungspsychologie beschäftigen sich mit Entwicklungs- und Reifungsprozessen (Oerter, 2008; Berk, 2007). Insbesondere im Kindes- und Jugendalter beobachtet man einen ständigen Zuwachs an Wissen und Fertigkeiten. Ab dem zweiten Lebensjahr beispielsweise wächst der Wortschatz eines Kindes um ca. fünf Wörter pro Tag (Anglin, 1993). Man spricht von der sogenannten „Wortschatzexplosion“ (Karnath, 2006, S.636). Viele Langzeitstudien erfassen bestimmte psychologische Merkmale über eine sehr lange Zeit der Lebensspanne. So ist das Ziel der Seattle Longitudinal Study, an der im gesamten Verlauf mehr als 6000 Personen teilnahmen, verschiedene psychologische Entwicklungen über die gesamte Lebensspanne zu studieren (siehe Schaie (1996) für einen Überblick oder die offizielle Homepage Schaie und Willis (2011)). Bekannt ist auch die „Terman Studie“, in der Lewis Terman über Jahrzehnte eine Gruppe hochbegabter Schüler auf ihrem Lebensweg begleitete (Terman, 1926).

Durch die Intervention während eines Therapieverlaufs wird auch hier ein nicht stabiler Prozess über die Zeit erwartet, da die Therapie sonst wirkungslos wäre. Bereits 1937 zeichnete Lasswell (1937) verbale und physiologische Merkmale während einer psychoanalytischen Therapie auf, um Veränderungen bewusster und unbewusster Affekte aufzuzeigen. Allgemein hat die psychologische Forschung an dynamischen Systemen in den letzten Jahren einen enormen Zuwachs bekommen (Delignières et al., 2004; Wagenmakers, Farrell & Ratcliff, 2004; Gilden, 2001). Beobachtet man solche Phänomene über die Zeit, erwartet man keinen stationären Verlauf, der um einen konstanten Mittelwert schwankt. Vielmehr sollte das Zeitreihenmodell eine Trendkomponente aufweisen.

Mathematisch unterscheidet man zwei Arten von Trends: deterministische und stochastische. Bei deterministischen Trends folgt die Entwicklung einem klar vorgegebenen Verlauf, der über eine deterministische Komponente der Zeitreihe (siehe Definition 1.1.3 in Kapitel 1.1, S.18) modelliert werden kann. Im Beispiel der Wortschatzentwicklung eines Kindes ist es unwahrscheinlich, dass mehr als einige wenige Worte wieder vergessen werden. Insgesamt erwartet man einen klaren Aufwärtstrend über die Zeit.

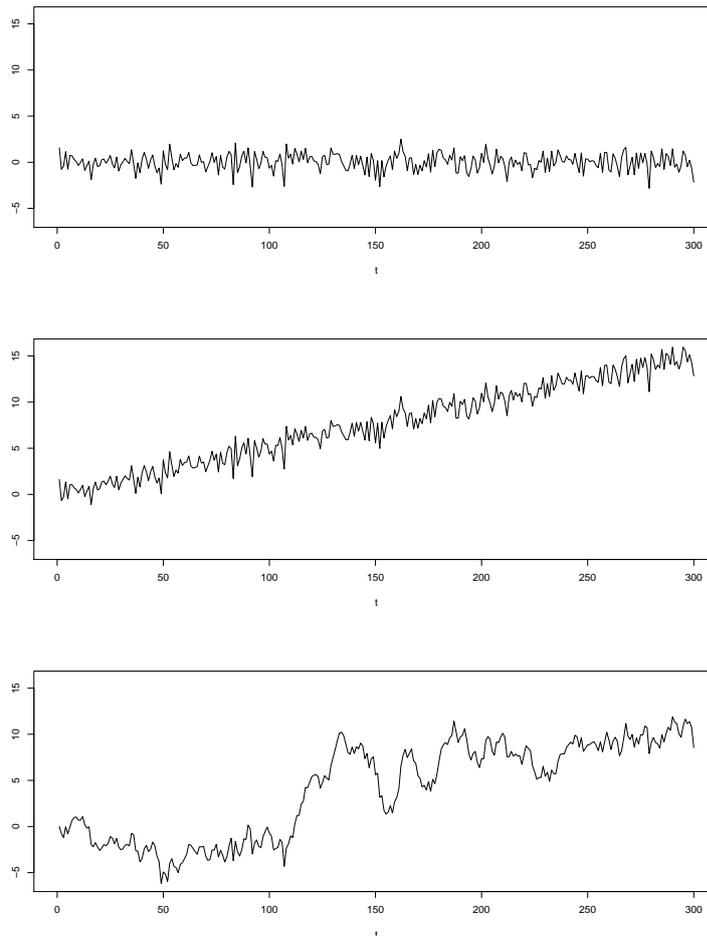


Abbildung 2.1: Vergleich stationärer, trendstationärer und differenzenstationärer Prozess

Meist ist der vorgegebene Trend ein linearer Verlauf. Die Zeitreihe bewegt sich dann an einer Geraden entlang. Durch Messfehler oder zufällige Abweichungen von dieser Entwicklung folgt der Verlauf nicht exakt dieser Geraden, sondern schwankt regelmäßig um sie herum. Im Idealfall erhält man, wenn man die Abweichungen berechnet, indem man den Trend zu jedem Zeitpunkt vom Messwert abzieht, einen stationären Prozess. Da meist die Abweichung vom Trend als Messfehler gewertet werden, erwartet man oft sogar Weißes Rauschen. Man sagt dann, dass die Reihe *trendstationär* sei, da die Fehler zu jedem Zeitpunkt dieselbe (konstante) Varianz haben. Der Mittelwert verändert sich regelmäßig über die Zeit. Er ist der „wahre“ Wert der Variablen auf der Geraden. Die Zeitreihe verläuft also stationär um den (idealen) Trend (z.B. Neusser, 2006, S.99).

Zur Analyse einer solchen Zeitreihe können dann nach *Trendbereinigung* die beschriebenen Verfahren und die üblichen Parameterschätzer eingesetzt werden, um die stationäre Reihe zu analysieren (z.B. Stier, 2001, S.282). In Abbildung 2.1 zeigt die mittlere Grafik einen trendstationären Prozess $Y_t = 0.05t + \epsilon_t$. Die obere Grafik zeigt das stationäre Weiße Rauschen ϵ_t , in das der trendstationäre Prozess nach Trendbereinigung überführt werden kann, $Y_t - 0.05t = \epsilon_t$.

Bei vielen psychologischen Prozessen beobachtet man jedoch keinesfalls einen deterministischen Verlauf, sondern ein dynamisches Verhalten, das in kurzen zeitlichen Abständen sehr starken Schwankungen unterworfen ist, und dem man bei der ersten grafischen Analyse keine Systematik zuschreiben würde. Viele dieser Zeitreihen besitzen einen stochastischen Trend. Der zunächst völlig unregelmäßig erscheinende Verlauf wird stationär, wenn man die Zeitreihe differenziert (z.B. Stier, 2001, S.282), weshalb man solche Prozesse als *differenzenstationär* bezeichnet. Der in der unteren Grafik der Abbildung 2.1 dargestellte Prozess ist ein differenzenstationärer Prozess, der durch Differenzierung in den im ersten Teil der Abbildung gezeigten White-Noise-Prozess überführbar ist. In einigen Fällen ist eine wiederholte Differenzierung notwendig, um einen stationären Prozess zu erhalten. Alle Prozesse, die durch (mehrfache) Differenzierung in einen stationären Prozess überführbar sind, nennt man *integrierte* Prozesse (z.B. Schröder & Buscher, 2002, S.19, S.265).

Glass et al. (1975) konnten zeigen, dass viele psychologische Prozesse integriert sind, und Nelson und Plosser (1982) weisen diese Eigenschaft für die meisten makroökonomischen Zeitreihen nach.

Nach der Differenzierung kann man den erhaltenen stationären Prozess mit den üblichen Verfahren analysieren (Box & Jenkins, 1976). Doch durch die Differenzierung verliert man auch wichtige Informationen zum Beispiel über die Lage der Zeitreihe. Denn eigentlich betrachtet man nicht länger die Zeitreihe Y_t selbst, sondern ihren *Differenzenprozess* ΔY_t . Deshalb braucht man statistische Verfahren, die man auf integrierte Prozesse anwenden kann. Eine Regressionsanalyse mit integrierten Zeitreihen kann zum Phänomen der Scheinregression führen, da die bekannten Schätzer und Statistiken ihre Standardverteilung verlieren und deshalb andere kritische Werte berücksichtigt werden müssen (C. W. Granger & Newbold, 1977, siehe auch Kapitel 2.2.4, S.79 dieser Arbeit). Nur wenn eine stationäre Linearkombination der Zeitreihen existiert, ist die Regressionsgleichung haltvoll. Ein solches System von Zeitreihen nennt man *kointegriert*. Man braucht also statistische Verfahren, um Systeme unabhängiger integrierter Prozesse von kointegrierten Systemen zu unterscheiden. Die Linearkombination von integrierten Prozessen kann nur dann stationär sein, wenn sich alle integrierten Anteile der Reihen durch die richtige Gewichtung in der Linearkombination kürzen lassen. Dies ist nur möglich, wenn die Zeitreihen gemeinsame stochastische Trends aufweisen.

Zeitreihen, die einen stochastischen Trend beinhalten, nennt man *integriert*. Systeme von Zeitreihen, die einen *gemeinsamen* stochastischen Trend besitzen, nennt man *ko-integriert*. In der formalen Definition eines kointegrierten Systems (Engle & Granger, 1987) werden sogenannte *common trends* zwar nicht erwähnt, sie sind jedoch der Schlüssel zum Verständnis der Kointegration. Die Kointegrationsbeziehung stellt ein langfristiges Gleichgewicht des Systems dar, wodurch die Regressionsgleichung aussagekräftig wird.

Ziel der Kointegrationsanalyse ist es, gemeinsame Trends in Zeitreihensystemen aufzudecken, um so zum einen den Zusammenhang aufzuklären und den Systemcharakter zu unterstreichen

und zum anderen in der langfristigen Gleichgewichtsbeziehung (Kointegrationsbeziehung) eine Erklärung für die Dynamik des Systems zu finden. In Kapitel 2.8 werden wir sehen, dass in einer Untersuchung (Kupfer et al., 2005) die Depressivitätswerte eines Ehepaares einen solchen Gleichgewichtszustand hatten, den man ähnlich wie in Abbildung 2.2 darstellen könnte.

Die Idee eines langfristigen Gleichgewichts lässt sich anschaulich sehr schön am Beispiel des Wechselkurses zweier Währungen oder am Preis einer Ware in zwei Landesteilen beschreiben. Betrachten wir also wie Granger im einleitenden Kapitel zu „Long-run economic relationships - readings in cointegration“ (Engle, 1991) z.B. den Kartoffelpreis im Süden PS_t und im Norden PN_t eines Landes²⁴.

Eine einfache Möglichkeit den Verlauf eines Systems aus zwei Zeitreihen grafisch zu analysieren ist es, die einzelnen Paare (PS_t, PN_t) in einem Koordinatensystem mit den Achsen PS_t und PN_t aufzutragen. Dadurch entsteht ein Plot, der etwa wie in Abbildung 2.2 aussehen kann.

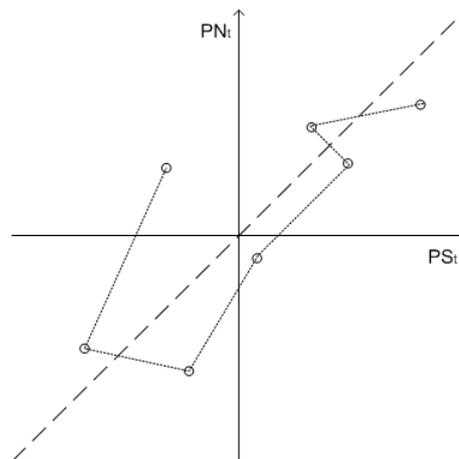


Abbildung 2.2: Zwei Zeitreihen gegeneinander aufgetragen, z.B. der Kartoffelpreis im Süden PS_t und im Norden PN_t eines Landes. Als Attraktor dient hier die erste Winkelhalbierende, auf der beide Preise identisch sind.

Es ist anzunehmen, dass sich die Preise in den beiden Landesteilen mit der Zeit annähern. Bezeichnen wir die Preise im Norden mit PN_t und die im südlichen Landesteil mit PS_t , dann werden Werte mit $PN_t \approx PS_t$ häufig angenommen. Ist dies nicht der Fall, ist der Preis beispielsweise im Süden des Landes viel günstiger als im Norden, dann ist es für den Handel ein lukratives Geschäft, Kartoffeln im Süden zu kaufen, in den Norden zu transportieren und dort wieder teurer zu verkaufen. Dadurch wird der Preis im Norden etwas gesenkt. Im Süden sinkt das Angebot, wodurch sich der Preis verteuert, was auf Dauer zu einer Angleichung des Kartoffelpreises im ganzen Land führt.

Die Gerade $PN_t = PS_t$, die 45°-Linie bzw. erste Winkelhalbierende, bildet in diesem Fall

²⁴Da wir zwei Regionen eines Landes und nicht zwei Länder betrachten, kann es freien Handel ohne Zölle geben.

einen sogenannten Attraktor (Engle, 1991, Introduction). Zwar können Schocks, wie Schwankungen auf dem Markt oder politische Eingriffe, zu Abweichungen von diesem Gleichgewicht führen. Ohne exogenes Einwirken strebt das System jedoch einen Punkt auf der Attraktorgeraden an. Die absolute Höhe des Preises ist dabei nicht entscheidend, solange beide Preise ähnlich hoch sind. Betrachtet man den Preis eines Gutes in zwei verschiedenen Währungen, wird der Wechselkurs in der Steigung der Attraktorgeraden deutlich.

Die Zustände, in denen die langfristige Gleichgewichtsbeziehung erfüllt ist, bilden einen Vektorraum, den sogenannten *Kointegrationsraum* (z.B. Neusser, 2006, Def.16.3). In unserem Beispiel mit dem Kartoffelpreis ist der Kointegrationsraum eine (eindimensionale) Gerade. Ein System aus mehr als zwei integrierten Zeitreihen kann jedoch auch einen höherdimensionalen Kointegrationsraum und mehrere Kointegrationsbeziehungen besitzen.

In diesem Kapitel werden wir Methoden kennenlernen, um kointegrierte Systeme zu identifizieren und darzustellen. Aus den verschiedenen Darstellungsmöglichkeiten werden wir Informationen über die Kointegrationsbeziehungen und den Kointegrationsraum gewinnen. Außerdem werden wir sehen, wie zusätzliche deterministische Trends im System berücksichtigt werden können und wie Testverfahren auf Kointegration arbeiten. Schließlich wird an einem psychologischen empirischen Beispiel zur Couple Collusion (Willi, 1984) gezeigt, wie man die Ergebnisse des Kointegrationstests interpretieren kann.

2.1 Integrierte Prozesse: Zeitreihen mit stochastischem Trend

Da integrierte Prozesse die Grundlage für kointegrierte Systeme darstellen, betrachten wir zunächst den Unterschied zwischen trendstationären und differenzenstationären Prozessen sowie die allgemeine Definition eines integrierten Prozesses. Über Kombinationsregeln für stationäre und integrierte Prozesse nähern wir uns dem Phänomen der Kointegration.

2.1.1 Stationäre Prozesse

Um die Unterschiede von stationären, trendstationären und differenzenstationären Prozessen herauszuarbeiten, wiederholen wir an dieser Stelle die wichtigsten Erkenntnisse über stationäre Prozesse aus Kapitel 1.1.

Unter einem *stationären* Prozess versteht man einen Prozess, dessen Erwartungswert und Varianz unabhängig von der Zeit, also konstant sind. Die Verteilung ist somit zu jedem (Mess-)Zeitpunkt gleich. Man könnte sagen, dass der Prozess sein Verhalten über die Zeit nicht verändert, sondern zufällig um den „wahren“ Wert des Merkmals schwankt.

Der einfachste stationäre Prozess ist normalverteiltes Weißes Rauschen. Hier ist die zufällige

Abweichung von Mittelwert μ normalverteilt mit Varianz σ^2 . Meist nimmt man sogar eine Standardnormalverteilung mit Erwartungswert Null und Varianz Eins an und schreibt dafür $\epsilon_t \sim WN(0, \sigma^2)$ (Fan & Yao, 2003).

Jeder stationäre Prozess kann als Summe aus einem deterministischen Prozess und einem MA(∞)-Prozess dargestellt werden (Theorem 1.6.4, S.30). Diese Darstellungsform nennt man *Wolddarstellung*. Im einfachsten Fall ist der deterministische Anteil ein von Null verschiedener Erwartungswert $\delta_t = \mu$

$$Y_t = \mu + \Psi(L)\epsilon_t.$$

Die Wolddarstellung eines stationären Prozesses ohne deterministischen Anteil $\delta_t = 0$ ist somit ein reiner MA(∞)-Prozess. MA-Prozesse sind immer stationär, während AR-Prozesse nur dann stationär sind, wenn alle Nullstellen des AR-(Lag-)Polynoms außerhalb des Einheitskreises liegen.

2.1.2 Trendstationäre Prozesse

Prozesse, die eine zeitliche Entwicklung beschreiben, können nicht stationär sein. Der von der Zeit unabhängige Erwartungswert stationärer Reihen macht eine langfristige Dynamik des Prozesses unmöglich. Zeitliche Veränderungen werden also durch nicht-stationäre Prozesse ausgedrückt.

Im einfachsten Fall ist nur der Erwartungswert des Prozesses zeitabhängig, bei der Wortschatzentwicklung steigt er etwa um fünf Worte pro Tag (Anglin, 1993). Statt durch eine Konstante μ wird er nun durch eine Funktion der Zeit $\mu(t)$ beschrieben, wodurch der Prozess folgende Wold-Darstellung bekommt (Neusser, 2006, Kap.7.1)

$$Y_t = \mu(t) + \Psi(L)\epsilon_t.$$

Durch den zeitabhängigen Erwartungswert erhält der Prozess einen Trend. Da dieser Trend einer Funktion $\mu(t)$ folgt, also durch eine (Funktions-)Vorschrift gesetzmäßig vorgegeben wird, heißt er *deterministisch*. Der stationäre Anteil lässt sich weiterhin durch einen MA(∞)-Prozess darstellen. Die Veränderung des Prozesses lässt sich also vor allem auf den Trend $\mu(t)$ zurückführen. Die Abweichungen ϵ_t wirken nur kurzfristig.

In der Abbildung 2.3 sind zwei solche Prozesse und ihr jeweiliger deterministischer Anteil (in rot) dargestellt. Es ist deutlich zu erkennen, dass der Prozess dem Trend folgt und nur durch zufällige Schwankungen von ihm abweicht. Da die Varianz der Schwankungen konstant bleibt, kann man sagen, dass der Prozess *stationär um einen deterministischen Trend schwankt*. Deshalb nennt man solche Prozesse *trendstationär* (Neusser, 2006, S.99, Schlittgen & Streitberg, 1999).

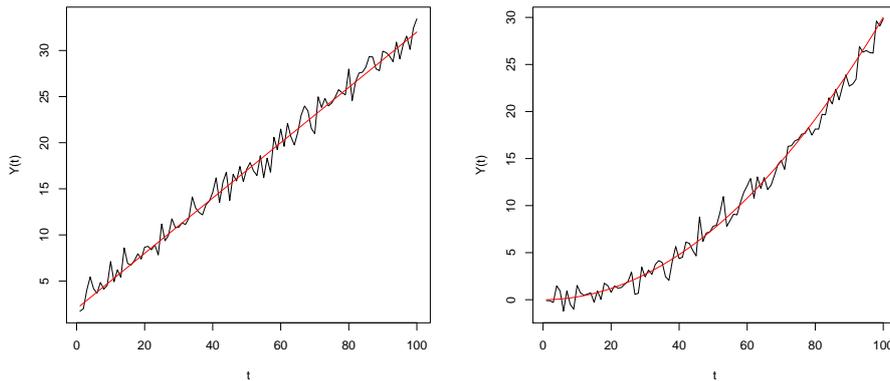


Abbildung 2.3: Linearer $Y_t = 0.3t + 2$ und quadratischer Trend $Y_t = 0.003t^2$ mit dem jeweiligen deterministischen Trend in rot

Definition 2.1.1 (trendstationärer Prozess) *Ein stochastischer Prozess, der wie folgt dargestellt werden kann*

$$Y_t = \mu(t) + \Psi(L)\epsilon_t, \quad (2.1)$$

heißt ein trendstationärer Prozess.

Am häufigsten entspricht der deterministische Trend einer linearen Funktion $\mu(t) = at + b$, wie er im linken Teil der Abbildung 2.3 dargestellt ist. Der Erwartungswert entwickelt sich linear in der Zeit, so dass der Prozess einer Geraden folgt. Natürlich kann der Trend $\mu(t)$ auch einer quadratischen oder polynomialen Funktion entsprechen, der lineare Fall ist in der Anwendung jedoch der übliche. Exponentielles Wachstum (vgl. Abbildung 2.3 rechts), wie es in der Ökonomie häufig vorzufinden ist, kann durch Logarithmieren in einen linearen Zusammenhang gebracht werden²⁵ (Stier, 2001, S.282).

Man kann einen trendstationären Prozess durch Subtraktion des deterministischen Trends $\mu(t)$ in einen stationären Prozess überführen (z.B. Kirchgässner & Wolters, 2007). Die Transformation in einen stationären Prozess ist nicht für alle nicht-stationären Prozesse so trivial.

2.1.3 Differenzenstationäre Prozesse

Wir haben bereits gesehen, dass Zeitreihen mit einer Einheitswurzel nicht stationär sind. Da sie sich auch nicht in der Form (2.1) darstellen lassen, sind sie auch nicht trendstationär. Vielmehr folgen sie einem *stochastischem* Trend (Stier, 2001, S.282). Bildet man jedoch die erste Differenz $\Delta Y_t := Y_t - Y_{t-1}$ eines Random Walks (z.B. Diggle, 1990, Kap.2.3), so verschwindet die Einheitswurzel und man erhält einen stationären Prozess.

Beispiel 2.1.2 (Einfacher Random Walk) *Betrachtet man die erste Differenz eines Random*

²⁵siehe auch Anhang C.1.3, S.341

Walks

$$\begin{aligned} Y_t &= Y_{t-1} + \epsilon_t \\ \Leftrightarrow Y_t - Y_{t-1} &= \epsilon_t, \end{aligned}$$

so erkennt man, dass sie einem stationärem Prozess, hier Weißem Rauschen (siehe Definition 1.2.3, S.20), folgt. Durch die Bildung der Differenz ist folglich ein stationärer Prozess entstanden.

Dieses Ergebnis führt uns direkt zu einer allgemeinen Definition (Hatanaka, 1996).

Definition 2.1.3 (Differenzenstationärer Prozess) *Ein Prozess Y_t , dessen erste Differenz ein stationärer Prozess ist, nennt man einen differenzenstationären Prozess.*

Der Prozess ΔY_t enthält keine Einheitswurzel mehr, sie wurde durch die Differenzierung aus dem Prozess gekürzt (z.B. Schlittgen & Streitberg, 1999, S.40; Schröder & Buscher, 2002, S. 20). Die Gleichung

$$\Delta Y_t = \epsilon_t$$

beschreibt jedoch nicht den Prozess selbst, sondern seine erste Differenz. Um daraus eine Darstellung von Y_t zu gewinnen, muss man den Differenzenoperator Δ invertieren. Dies geschieht über die Darstellung als Lagpolynom $\Delta Y_t = (1 - L)Y_t$. Die Darstellung des inversen Operators $(1 - L)^{-1}$ wird im Anhang C.1.2 ab Seite 338 hergeleitet.

In der Analysis ist die Umkehrung der ersten Ableitung, die über einen Differenzenquotienten definiert ist, die Integration (z.B. Wolter & Dahn, 2000; Meyberg, 1999). Analog zur ersten Ableitung beschreibt der erste Differenzenprozess ΔY_t die Steigungsunterschiede zwischen zwei Zeitpunkten. So ließe sich die Bezeichnung *Integrationsoperator* für den inversen Operator $(1 - L)^{-1}$ rechtfertigen, die in der Literatur jedoch nicht geläufig ist. Prozesse, die folgende Darstellung besitzen

$$Y_t = (1 - L)^{-1} \epsilon_t$$

werden jedoch üblicher Weise als *integrierter White Noise* bezeichnet (z.B. Deng, 2010; Avedyan, Tsytkin & Bittanti, 1998), da sie durch „Integration“ aus Weißem Rauschen entstehen. Der einfache Random Walk ist so ein Prozess.

Da stationäre Prozesse nach dem Satz von Wold (Theorem 1.6.4) eine $MA(\infty)$ -Darstellung haben, gilt dies auch für die erste Differenz eines differenzenstationären Prozesses (Neusser, 2006, S.100).

Bemerkung 2.1.4 (Differenzenstationär und trendstationär) *Für einen differenzenstationären*

nären gibt es eine Wolddarstellung

$$(1 - L)Y_t = \Psi(L)\epsilon_t$$

mit $\Psi(1) \neq 0$, $\epsilon_t \sim WN(0, \sigma^2)$ und $\sum_{j=0}^{\infty} j |\psi_j| < \infty$.

Die Forderung $\sum_{j=0}^{\infty} j |\psi_j| < \infty$ sorgt für die Konvergenz der durch $\Psi(L)\epsilon_t$ definierten Reihe.

Wichtiger ist die Bedingung $\Psi(1) \neq 0$. Sie schließt aus, dass ein trendstationärer Prozess in Form eines differenzenstationären Prozesses gebracht werden kann.

Für den Differenzenprozess jedes trendstationären Prozesses mit linearem Trend $X_t = \alpha + \delta t + \tilde{\Psi}(L)\epsilon_t$ gilt zwar $(1 - L)X_t = \delta + \Psi(L)\epsilon_t$ mit $\Psi(L) = (1 - L)\tilde{\Psi}(L)$. Doch in diesem Fall ist $\Psi(1) \neq 0$, wie man leicht sieht²⁶, verletzt. Ein trendstationärer Prozess ist also kein differenzenstationärer Prozess.

Nicht-stationäre Prozesse mit einer oder mehreren Einheitswurzeln (vgl. Ahtola & Tiao, 1987) nennt man auch *homogen instationär* (Stier, 2001, S.57).

2.1.4 Integrierte Prozesse

Durch die Bildung der ersten Differenz wurde eine Einheitswurzel aus dem Prozess gekürzt. Bei Prozessen mit mehreren Unit Roots reicht dies jedoch nicht aus, um sie auf einen stationären Prozess zu transformieren. Hier muss der Differenzenoperator erneut angewendet werden (z.B. Diggle, 1990). Durch jede Anwendung wird eine Einheitswurzel aus dem Prozess eliminiert. Dies führt uns zur allgemeinen Definition von integrierten Prozessen höherer Ordnung (z.B. Alho & Spencer, 2005; Schröder & Buscher, 2002, S.19 Def.1).

Definition 2.1.5 (Integrierter Prozess der Ordnung d) *Unter einem integrierten Prozess der Integrationsordnung d versteht man einen stochastischen Prozess, der nach d -fachem Differenzieren stationär wird. Man schreibt $Y_t \sim I(d)$, falls*

$$(1 - L)^d Y_t = X_t, \text{ wobei } X_t \text{ stationär.}$$

Für $(1 - L)^d$ kann man mittels des Binomischen Lehrsatzes (Theorem A.3.5, S.326) eine Darstellung als einfaches Lagpolynom herleiten; dies gilt auch für den inversen Operator $(1 - L)^{-d}$ (siehe Anhang C.1, S.337). Bislang war der stationäre Prozesse, über den der differenzierte Prozess dargestellt wurde, stets Weißes Rauschen. In der Definition ist X_t als allgemeiner stationärer Prozess gewählt, der wiederum eine Darstellung als Moving Average oder autoregressiver Prozess haben kann.

Häufig ist die gegebene stationäre Zeitreihe zu kurz, um eine reine Prozessdarstellung zu schätzen, da man für eine große Anzahl an zu schätzenden Parametern viele Datenpunkte

²⁶ $(1-L)$ ist für $L = 1$ Null, woraus $\Psi(1) = (1 - L)\tilde{\Psi}(1) = 0$ folgt.

benötigt (u.a. Stier, 2001, V.4.4 ARMA-Prozesse; Rinne & Specht, 2002). Deshalb ist es üblich eine Mischform der Darstellung zu wählen und einige (p) Parameter der AR- und einige (q) der MA-Darstellung zu schätzen. Einen stationären Prozess kann man also als sogenannten *ARMA*-Prozess darstellen. Deshalb ist es auch üblich, integrierte Prozesse, die nach d -facher Differenzierung eine ARMA-Darstellung haben als *ARIMA-Prozesse* zu bezeichnen (z.B. Stier, 2001, 5.4.5 ARIMA-Prozesse; Fan & Yao, 2003, 1.3.5 ARIMA Models).

Definition 2.1.6 (*ARIMA*(p,d,q)) *Ist die d -fache Differenz eines Prozesses $(1 - L)^d Y_t$ ein *ARMA*(p,q)-Prozess, dann heißt der Prozess selbst *ARIMA*(p,d,q). Er lässt sich wie folgt mit den Lagpolynomen $A(L)$ für den autoregressiven Anteil und $B(L)$ für den Moving Average Anteil darstellen*

$$A(L)(1 - L)^d Y_t = B(L)\epsilon_t,$$

wobei ϵ_t Weißes Rauschen bezeichnet.

Natürlich ist jeder *ARMA*(p,q)-Prozess auch ein *ARIMA*($p,0,q$)-Prozess. Insbesondere ist jeder AR-Prozess ein *ARIMA*($1,0,0$)-Prozess und jeder MA-Prozess ein *ARIMA*($0,0,1$)-Prozess. Der einfache Random Walk ist ein Beispiel für einen *ARIMA*($0,1,0$)-Prozess und allgemeines integriertes Weißes Rauschen ist stets ein *ARIMA*($0,d,0$)-Prozess.

Im Anhang C.1.1 wird die allgemeine Darstellung eines Differenzenoperators $(1 - L)^d$ als Lagpolynom hergeleitet. Diese Darstellung schließt auch solche Operatoren mit negativem Exponenten $d < 0$ ein. Dies erklärt, warum die Bezeichnung *Integrationsoperator* nicht gebräuchlich ist, da sie eine Teilmenge der Differenzenoperatoren sind. „Negative“ Differenzen erzeugen einen integrierten Prozess. Um diese Integration rückgängig zu machen, muss ein Differenzenoperator derselben Ordnung angewendet werden. Wird der Differenzenoperator zu oft auf einen stationären Prozess angewendet, erhält man einen *überdifferenzierten* Prozess. Durch Überdifferenzierung verlieren Schätz- und Testverfahren für stationäre Prozesse an Effizienz. Dennoch ist es im Allgemeinen besser, einen überdifferenzierten Prozess als einen (noch) integrierten Prozess mit statistischen Methoden für stationäre Prozesse auszuwerten (Schröder & Buscher, 2002, S. 142).

Konsequenterweise nennt man stationäre Prozesse, für die $d = 0$ gilt, in diesem Zusammenhang auch *integriert der Ordnung Null*, so dass man abkürzend $Y_t \sim I(0)$ schreiben kann (Johansen, 1995, S.35, Def. 3.2).

2.1.5 Stationäre und differenzenstationäre Prozesse

Da wir uns im Folgenden vor allem mit stationären ($Y_t \sim I(0)$) und differenzenstationären Prozessen ($Y_t \sim I(1)$) beschäftigen wollen, vergleichen wir zunächst einige Eigenschaften dieser Prozesstypen (Engle & Granger, 1987; Feller, 1970).

Bemerkung 2.1.7 (Vergleich stationärer und differenzenstationärer Prozesse) *Ein stationärer Prozess $Y_t \sim I(0)$ (mit Erwartungswert $=0$)*

- *hat zu jedem Zeitpunkt eine endliche Varianz,*
- *externe Schocks wirken nur kurzfristig,*
- *die Autokorrelation nimmt mit dem zeitlichen Abstand ab,*
- *die erwartete Zeit zwischen zwei 0-Zuständen ist endlich, d.h. der Prozess kehrt zu seinem Erwartungswert zurück.*

Ein differenzenstationärer (integriert vom Grad 1) Prozess $Y_t \sim I(1)$

- *hat eine mit der Zeit ins Unermessliche wachsende Varianz,*
- *ein Schock hat eine langfristige Wirkung (bleibt im Prozess enthalten),*
- *die Autokorrelation geht gegen Unendlich,*
- *der Prozess verläuft jedoch glatter als ein stationärer Prozess.*

Im ersten Kapitel haben wir gesehen, dass der Varianzprozess von stationären Prozessen konstant (und dann natürlich auch endlich, siehe Kapitel 1.2) und der eines Random Walks zeitabhängig ist (siehe Kapitel 1.8). Das letzte Ergebnis lässt sich auf allgemeine differenzenstationäre Prozesse übertragen, ihr Varianzprozess folgt einem linearen Trend, wodurch er ins Unendliche wächst. In der MA-Darstellung stationärer Prozess wurde deutlich, dass die Autokorrelationsfunktion endlich ist, weshalb externe Schocks nur eine endliche Zeit im Prozess wirken (kurzes Gedächtnis), während in einem Random Walk keine Innovation verloren geht, so dass die Autokorrelationsfunktion theoretisch unendlich lang ist (z.B. Stier, 2001, S.283).

Zusätzlich soll hier erwähnt werden, dass ein stationärer Prozess stets zu seinem Erwartungswert zurückzukehren versucht (z.B. Bhattacharya & Waymire, 1990, Kap.4), während ein differenzenstationärer Prozess einem stochastischem Trend folgt, wodurch sein Verlauf glatter wirkt (z.B. Alho & Spencer, 2005, S.200).

Da für Kointegration Linearkombinationen dieser Prozesse eine wichtige Rolle spielen, halten wir nun Regeln fest, wie sich stationäre und differenzenstationäre Prozesse in Linearkombinationen verhalten (z.B. Schröder & Buscher, 2002, S.20).

Bemerkung 2.1.8 (Kombinationsregeln für $I(0)$ - und $I(1)$ -Prozesse) *Es seien $a, b \in \mathbb{R}$.*

1. *Eine lineare Transformation ändert den Integrationsgrad nicht.*

$$X_t \sim I(0) \Rightarrow a + bX_t \sim I(0) \tag{2.2}$$

$$X_t \sim I(1) \Rightarrow a + bX_t \sim I(1) \tag{2.3}$$

2. Die Linearkombination zweier stationärer Prozesse ist wieder stationär.

$$X_t \sim I(0), Y_t \sim I(0) \Rightarrow aX_t + bY_t \sim I(0)$$

3. Die Linearkombination eines stationären und eines integrierten Prozesses ist integriert.

$$X_t \sim I(1), Y_t \sim I(0) \Rightarrow aX_t + bY_t \sim I(1)$$

4. Die Linearkombination zweier integrierter Prozesse ist (meist) wieder integriert.

$$X_t \sim I(1), Y_t \sim I(1) \Rightarrow aX_t + bY_t \sim I(1), \text{ fast immer.}$$

In der letzten Kombinationsregel ist bereits angedeutet, dass es Linearkombinationen von integrierten Prozessen gibt, die nicht integriert sind. Im Folgenden wollen wir diesen Spezialfall genauer untersuchen. Sie kann auch für integrierte Zeitreihen der Ordnung d verallgemeinert werden (Engle & Granger, 1987, S.283).

2.2 Kointegrierte Prozesse: Systeme mit gemeinsamen Trends

Im Allgemeinen ist die Linearkombination von integrierten Prozessen wieder integriert (derselben Ordnung). In einem Spezialfall existieren Linearkombinationen, die nicht vom selbem Grad integriert, manchmal sogar stationär sind. Dies ist der Fall, wenn die Reihen gemeinsame Trends haben. Solche integrierten Reihen nennt man dann *ko-integriert*. Ihre stationäre Linearkombination heißt in diesem Zusammenhang dann *Kointegrationsbeziehung*.

2.2.1 Der Begriff der Kointegration

In den Kombinationsregeln (Bemerkung 2.1.8, S.65) haben wir gesehen, dass sich der höchste Integrationsgrad in der Linearkombination integrierter Reihen auf die Linearkombination überträgt. Ist dies nicht der Fall, heißt das System der beiden Zeitreihen *kointegriert* (Engle & Granger, 1987). Diese Definition wollen wir auf ein System mit K Zeitreihen verallgemeinern (C. W. J. Granger, 1981; Engle & Granger, 1987).

Definition 2.2.1 (Kointegration) K univariate Prozesse Y_{1t}, \dots, Y_{Kt} nennt man *kointegriert der Ordnung (d, b)* , kurz $CI(d, b)$, falls

- sie alle integriert vom selben Integrationsgrad d sind, $Y_{it} \sim I(d) \forall i \leq K$

- und eine Linearkombination $Z_t = \beta_1 Y_{1t} + \dots + \beta_K Y_{Kt}$ existiert, die von kleinerem Integrationsgrad ist, $Z_t \sim I(d - b)$.

Ein K -dimensionaler Prozess Y_t heißt kointegriert, falls seine Komponentenprozesse Y_{1t}, \dots, Y_{Kt} kointegriert sind. Man schreibt kurz $Y_t \sim CI(d, b)$. Die Existenz einer Linearkombination ist also gleichbedeutend mit der Existenz eines K -dimensionalen Vektors $\beta = (\beta_1, \dots, \beta_K)$, so dass das Produkt von β und dem vektoriellen Prozess Y_t einen univariaten Prozess von kleinerer Integrationsordnung ergibt

$$Z_t = \beta' Y_t = \beta_1 Y_{1t} + \dots + \beta_K Y_{Kt}.$$

Jeden Vektor β , dessen Produkt mit Y_t einen solchen univariaten Prozess Z_t ergibt, nennt man einen Kointegrationsvektor. Die Linearkombination $\beta' Y_t$ nennt man analog eine Kointegrationsbeziehung.

In einem kointegrierten System existiert also eine Linearkombination der Zeitreihen, die einen kleineren Integrationsgrad hat. Für die Anwendung ist insbesondere der Spezialfall von Interesse, indem die Linearkombination den Integrationsgrad Null hat, also stationär ist (z.B. Schröder & Buscher, 2002, S.277).

Bemerkung 2.2.2 (Spezialfall $d = b = 1$) *Im Fall $d = 1$ sind alle Teilprozesse Y_{it} differenzstationär (d.h. ihre erste Differenz ΔY_{it} ist stationär). $b = 1$ bedeutet dann, dass es eine stationäre Linearkombination aus diesen Prozessen gibt*

$$Y_{1t}, \dots, Y_{Kt} \sim I(1)$$

$$Z_t = \beta' Y_t = \beta_1 Y_{1t} + \dots + \beta_K Y_{Kt} \sim I(0).$$

Der Kointegrationsvektor β erzeugt sozusagen aus K nicht-stationären Prozessen einen stationären Prozess (Stier, S.315).

Integrierte Prozesse zeichnen sich durch einen stochastischen Trend aus. Im Fall der Kointegration wird der Integrationsgrad verringert. Dieses Phänomen kann man nun darüber erklären, dass kointegrierte Einzelprozesse gemeinsame Trends enthalten, die sich bei geeigneter Kombination kürzen, sie sind also *gemeinsam (ko-)integriert*. Wir wollen zunächst ein kointegriertes bivariates System betrachten.

Beispiel 2.2.3 (Bivariater Fall) *Betrachten wir zur Veranschaulichung einen bivariaten kointegrierten Prozess Y_t , also zwei integrierte Einzelreihen Y_{1t} und Y_{2t} . Wie man in der Abbildung 2.4 erkennen kann, sind die zwei dargestellten Prozesse zwar nicht identisch, haben aber einen ähnlich (fast „parallelen“) Verlauf. Genauer hat der blaue Prozesse stets den doppelten Wert des grün dargestellten Prozesses (abgesehen von zufälligen Schwankungen um den stochastischen Trend)*

$$Y_{1t} \approx 2 \cdot Y_{2t}.$$

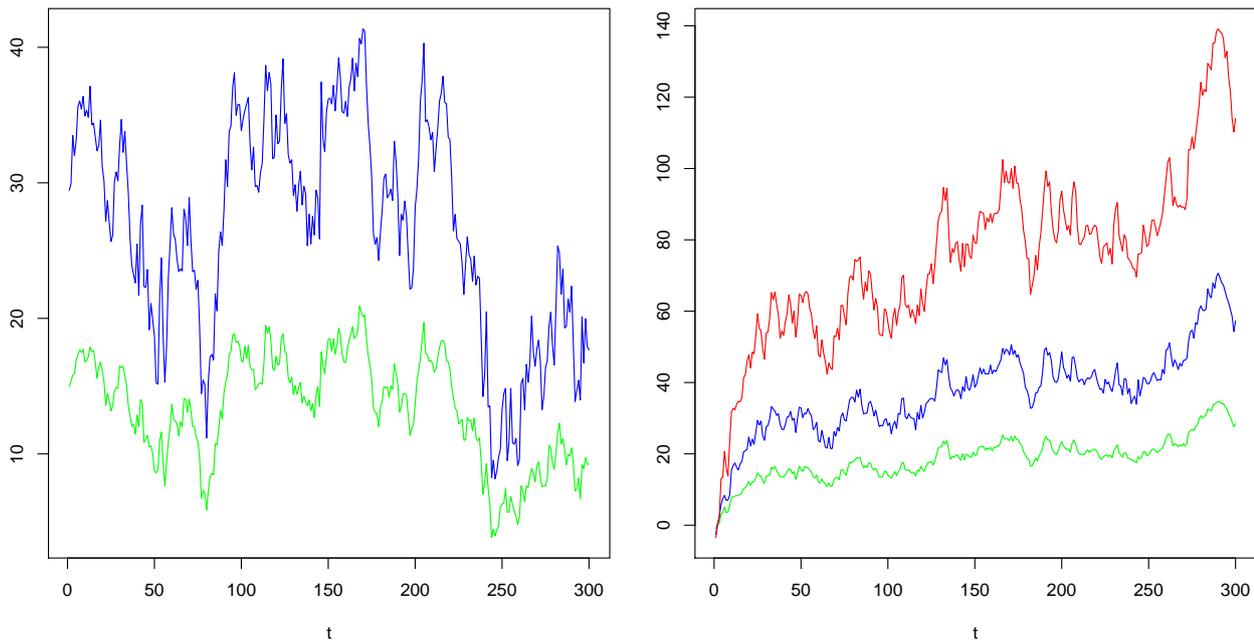


Abbildung 2.4: Zwei kointegrierte Systeme: links bivariat, rechts dreidimensional. Man erkennt deutlich den gemeinsamen Trend der Zeitreihen.

Durch Skalierung werden die beiden Prozesse auf dasselbe Niveau gebracht, so dass man den gemeinsamen Trend vollständig herauskürzen kann und die Linearkombination stationär ist

$$Y_{1t} - 2 \cdot Y_{2t} \sim I(0).$$

Nur bei richtiger Gewichtung der Teilprozesse gibt es eine Linearkombination, durch die der stochastische (gemeinsame) Trend eliminiert wird.

2.2.2 Kointegrationsbeziehungen

Bereits die Definition von Kointegration (Def. 2.2.1, S.66) legt nahe, dass es mehrere Kointegrationsvektoren geben kann²⁷. Tatsächlich gibt es stets unendlich viele solche Vektoren, obwohl wir gesehen haben, dass die meisten Linearkombinationen integrierter Prozesse wieder integriert sind. Die Kointegrationsvektoren sind aber auch nicht rein zufällig verbunden. Denn alle Vielfachen einer gefundenen Kointegrationsbeziehung sind ebenfalls stationär. Es gibt jedoch auch Systeme, in denen weitere, linear unabhängige Kointegrationsbeziehungen existieren. In den folgenden Abschnitten werden wir die Menge der Kointegrationsbeziehungen für verschiedene Systeme betrachten.

²⁷ „Jeder Vektor“, der die Bedingungen erfüllt, wird als Kointegrationsvektor bezeichnet.

Linear abhängige Kointegrationsbeziehungen

Die Kointegrationsbeziehung zweier Prozesse ist nicht eindeutig. Mehr noch: wenn es eine Kointegrationsbeziehung gibt, gibt es gleich unendlich viele. Doch diese Beziehungen sind nicht unabhängig voneinander, sondern Vielfache. Jedes Vielfache einer stationären Linearkombination ist wieder stationär

$$\begin{aligned} \beta_1 Y_{1t} + \beta_2 Y_{2t} &\sim I(0) \\ \Rightarrow \lambda (\beta_1 Y_{1t} + \beta_2 Y_{2t}) &\sim I(0) \quad \forall \lambda \in \mathbb{R} \\ &= \lambda \beta_1 Y_{1t} + \lambda \beta_2 Y_{2t} \sim I(0) \quad \forall \lambda \in \mathbb{R}. \end{aligned}$$

Dies ist ein Spezialfall der ersten Kombinationsregel aus Bemerkung 2.1.8, S.65. So ist in unserem obigen Beispiel 2.2.3 (S.67) auch $2Y_{1t} - 4Y_{2t}$ eine stationäre Linearkombination der beiden Prozesse, sowie jedes weitere Vielfache $\lambda Y_{1t} - 2\lambda Y_{2t}$ mit $\lambda \in \mathbb{R}$.

Eindeutigkeit kann durch die Normierung von $\beta = (\beta_1, \beta_2)$ erreicht werden. Dabei ist es üblich, die erste Komponente auf 1 festzulegen, $\beta^* = (1, \beta_2^*)$. (Dies wird erreicht für $\lambda = \frac{1}{\beta_1}$, also $\beta_2^* = \frac{\beta_2}{\beta_1}$). Unser Beispiel 2.2.3 ist bereits normiert.

Was wir hier für den bivariaten Fall dargestellt haben, gilt auch ganz allgemein für die Linearkombinationen von K Prozessen. Die Vielfachen einer stationären Linearkombination sind wieder stationär, so dass auch sie Kointegrationsbeziehungen bilden

$$\beta' Y_t \sim I(0) \Rightarrow \lambda \cdot \beta' Y_t \sim I(0) \quad \forall \lambda \in \mathbb{R}.$$

Da dies wegen $\lambda \in \mathbb{R}$ unendlich viele sind, braucht man eine Normierung, um sinnvoll über Kointegrationsbeziehungen sprechen zu können. Wie im bivariaten Fall setzt man $\beta_1^* = 1$ und legt damit die anderen $K - 1$ Koeffizienten durch $\frac{\beta_i}{\beta_1}$ fest.

Linear unabhängige Kointegrationsbeziehungen

Bei höher dimensionalen Prozessen mit $K > 2$ kann es weitere Kointegrationbeziehungen geben, die linear unabhängig zur ersten sind.

$$\begin{aligned} \beta^{(1)'} Y_t &\sim I(0) \\ \beta^{(2)'} Y_t &\sim I(0) \end{aligned}$$

mit $\beta^{(1)} \neq \lambda \beta^{(2)}$. Dabei muss man zwei verschiedene Fälle unterscheiden. Zum einen können linear unabhängige Kointegrationen dadurch entstehen, dass alle Reihen im System denselben Trend enthalten und man so für je zwei Reihen eine Linearkombination findet (Bsp. 2.2.4). Dies wäre eine Verallgemeinerung des bereits dargestellten bivariaten Systems mit *einem* gemeinsamen Trend und ist immer dann der Fall, wenn das System den Kointegrationsrang $K - 1$ hat. Interessanter und vielfältiger ist noch der Fall, in dem verschiedene

gemeinsame Trends das System treiben. Wir betrachten zunächst den ersten einfacheren Fall in einem dreidimensionalen System.

Beispiel 2.2.4 (Dreidimensionales System mit einem gemeinsamen Trend) *Die drei Reihen in der rechten Abbildung 2.4 auf Seite 68 haben alle denselben stochastischen Trend. Allerdings wirkt er in jeder Reihe unterschiedlich stark.*

So bilden je zwei der drei Zeitreihen ein bivariates System, in dem es eine Kointegrationsbeziehung gibt. Für jedes bivariate System gibt es eine Linearkombination von jeweils zwei Zeitreihen. Es gilt

$$Y_{1t} = 2 \cdot Y_{2t} \tag{2.4}$$

$$Y_{1t} = 4 \cdot Y_{3t} \tag{2.5}$$

$$Y_{2t} = 2 \cdot Y_{3t}. \tag{2.6}$$

Doch nur jeweils zwei der drei Linearkombinationen sind linear unabhängig voneinander, da man jede der Kointegrationsbeziehungen über die anderen beiden darstellen kann. So entsteht (2.6) durch Einsetzen von (2.4) in (2.5) und Normierung.

Wenn es mehrere linear unabhängige Kointegrationsbeziehungen gibt, die durch die Kointegrationvektoren $\beta^{(i)}$ beschrieben werden, dann ist es üblich, sie zu einer *Kointegrationsmatrix* β zusammenzufassen. Wie im Fall einer Kointegrationsbeziehung gilt dann, dass

$$Z_t = \beta' Y_t$$

ein stationärer Prozess ist. Der Prozess Z_t ist nun nicht mehr univariat, sondern besteht aus den r verschiedenen Kointegrationsbeziehungen der Komponentenprozesse von Y_t

$$\beta' Y_t = \begin{pmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1K} \\ \vdots & \ddots & & \dots \\ \beta_{r1} & \beta_{r2} & \dots & \beta_{rK} \end{pmatrix} \begin{pmatrix} Y_{1t} \\ \vdots \\ Y_{Kt} \end{pmatrix} = \begin{pmatrix} \beta_{11}Y_{1t} + \beta_{12}Y_{2t} + \dots + \beta_{1K}Y_{Kt} \\ \dots \\ \beta_{r1}Y_{1t} + \beta_{r2}Y_{2t} + \dots + \beta_{rK}Y_{Kt} \end{pmatrix}$$

Beispiel 2.2.5 (Fortsetzung) *Da es in obigem Beispiel zwei linear unabhängige Kointegrationsbeziehungen gibt ($r = 2$), hat die Kointegrationsmatrix β also z.B. folgende Gestalt*

$$\beta = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 4 \end{pmatrix}, \text{ da gilt}$$

$$\begin{aligned} \beta' Y_t &= \begin{pmatrix} 1 & 2 & 0 \\ 1 & 0 & 4 \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot Y_{1t} + 2 \cdot Y_{2t} + 0 \cdot Y_{3t} \\ 1 \cdot Y_{1t} + 0 \cdot Y_{2t} + 4 \cdot Y_{3t} \end{pmatrix}, \end{aligned}$$

also in Schreibweise als System von zwei stationären Linearkombinationen

$$\begin{aligned} 1 \cdot Y_{1t} + 2 \cdot Y_{2t} &\sim I(0) \\ 1 \cdot Y_{1t} + 4 \cdot Y_{3t} &\sim I(0). \end{aligned}$$

Die Matrix β ist jedoch nicht eindeutig, da auch die Kointegrationsbeziehungen von Y_{1t}, Y_{2t} und Y_{2t}, Y_{3t} linear unabhängig sind.

Die Anzahl der linear unabhängigen Kointegrationsbeziehungen bezeichnet man als *Kointegrationsrang* (Neusser, 2006, Definition 16.3) des Prozesses Y_t . Natürlich ist auch weiterhin jedes Vielfache einer der Linearkombinationen ebenfalls eine (linear abhängige) Kointegrationsbeziehung.

Definition 2.2.6 (Kointegrationsrang) *Der Kointegrationsrang r ist die maximale Anzahl von linear unabhängigen kointegrierenden Vektoren $\beta^{(1)}, \dots, \beta^{(r)}$. Diese spannen einen Vektorraum auf, den Kointegrationsraum.*

Dieser Kointegrationsraum ist die Menge an Wertetupeln (z.B. Paaren), die als Attraktor für das Zeitreihensystem dienen. Die normierten Kointegrationsvektoren dienen als Basisvektoren dieses Vektorraumes, da jede Linearkombination aus zwei Kointegrationsbeziehungen wieder eine Kointegrationsbeziehung ergibt (Johansen, 1995, Def.3.4). Im bivariaten Fall bildet die Menge der kointegrierenden Vektoren eine Gerade, wie wir im Beispiel mit dem Kartoffelpreis in der Einleitung zu Kapitel 2 gesehen haben. Das System versucht langfristig einen Wert in diesem Gleichgewichtsbereich zu erreichen und zu halten.

Der Kointegrationsraum ist ein r -dimensionaler Untervektorraum des K -dimensionalen Raumes \mathbb{R}^K . Allgemein kann es bei einem K -dimensionalen Prozess nur maximal $r \leq K - 1$ linear unabhängige Kointegrationsbeziehungen bzw. -vektoren geben, da $K - 1$ die größte Dimension eines Untervektorraumes ist. Dann ist $\beta'Y_t$ ein r -dimensionaler stationärer Prozess, wobei $\beta' = (\beta^{(1)}, \dots, \beta^{(r)})'$ die $(r \times K)$ Kointegrationsmatrix ist.

Gibt es K unabhängige stationäre Linearkombinationen in einem System, ist das nur möglich, wenn die K Einzelprozesse selbst alle stationär sind. Denn K linear unabhängige Kointegrationsvektoren β_i spannen den gesamten K -dimensionalen Vektorraum auf (z.B. Bosch, 2009; Beutelspacher, 2003; Fischer, 2002) und sind dann äquivalent zu jeder anderen Basis des Raumes, so also auch zu der Standardbasis der Einheitsvektoren. Das Produkt des i -ten Einheitsvektors mit einem K -dimensionalen Prozess ergibt den i -ten Komponentenprozess, so dass man über die K -Einheitsvektoren als Kointegrationsvektoren alle K Einzelprozesse erhält.

Die Basis des Kointegrationsraumes ist nicht eindeutig bestimmt. Wie in jedem Vektorraum kann sie durch jede andere Basis ersetzt werden. Übertragen bedeutet das, dass auch das System von r linear unabhängigen Linearkombinationen bzw. Kointegrationsvektoren nicht eindeutig ist. Vielmehr kann man jede Zeile oder jede Spalte der Matrix durch ein belie-

biges Vielfaches ersetzen, ohne die Stationarität zu verlieren. Denn jedes Vielfaches des vektoriellen stationären Prozesses ist wieder stationär:

$$\beta'Y_t \sim I(0) \Rightarrow \lambda'\beta'Y_t \sim I(0) \quad \forall \lambda \in \mathbb{R}^r$$

wobei

$$\lambda'\beta'Y_t = \begin{pmatrix} \lambda_1 (\beta_{11}Y_{1t} + \beta_{12}Y_{2t} + \dots + \beta_{1K}Y_{Kt}) \\ \dots \\ \lambda_r (\beta_{r1}Y_{1t} + \beta_{r2}Y_{2t} + \dots + \beta_{rK}Y_{Kt}) \end{pmatrix} = \begin{pmatrix} \lambda_1\beta_{11}Y_{1t} + \lambda_1\beta_{12}Y_{2t} + \dots + \lambda_1\beta_{1K}Y_{Kt} \\ \dots \\ \lambda_r\beta_{r1}Y_{1t} + \lambda_r\beta_{r2}Y_{2t} + \dots + \lambda_r\beta_{rK}Y_{Kt} \end{pmatrix}$$

Eine eindeutige Darstellung des Kointegrationsraumes bzw. seiner Basis und damit der Kointegrationsmatrix erhält man nur, wenn man die Kointegrationsbeziehungen so normiert, dass die ersten r Spalten der Matrix Einheitsvektoren sind

$$\beta^{*'} = \begin{pmatrix} \mathbb{I} & \beta_{(K-r)} \end{pmatrix}$$

Der r -dimensionale Gleichgewichtsprozess hat also folgende Gestalt

$$\beta^{*'}Y_t = \begin{pmatrix} 1 & 0 & \dots & 0 & \beta_{1(K-r)}^* & \dots & \beta_{1K}^* \\ 0 & 1 & \dots & 0 & \beta_{2(K-r)}^* & \dots & \beta_{2K}^* \\ \vdots & & \ddots & & \dots & & \\ 0 & 0 & \dots & 1 & \beta_{r(K-r)}^* & \dots & \beta_{rK}^* \end{pmatrix} \begin{pmatrix} Y_{1t} \\ \vdots \\ Y_{Kt} \end{pmatrix}.$$

Die r normierten Kointegrationsbeziehungen werden also jeweils als Linearkombinationen der Prozesse $Y_{(K-r)t}, Y_{(K-r+1)t}, Y_{(K-r+2)t}, \dots, Y_{Kt}$ und einem der Teilprozesse $Y_{1t}, Y_{2t}, \dots, Y_{(K-r-1)t}$ dargestellt.

2.2.3 Gemeinsamer Trend - Anschauung

Im Folgenden betrachten wir weiterhin vor allem den Spezialfall $d = b = 1$, in dem die Kointegrationsbeziehungen integrierter Zeitreihen der Ordnung Eins stationär sind. Der gemeinsame Trend in kointegrierten Systemen kann selbst als integrierter Prozess $W_t \sim I(1)$ dargestellt werden.

Wenn wir einen integrierten Prozess beobachten, dann können wir nicht sagen, wie viele integrierte Anteile er tatsächlich hat. Denn in den Kombinationsregeln hatten wir ja gesehen, dass die Linearkombination mehrerer integrierter Prozesse in der Regel wieder integriert ist. Die Kombination eines integrierten und eines stationären Prozesses ist ebenfalls wieder integriert. Es ist also nicht möglich aus einem beobachteten integrierten Prozess die einzelnen integrierten Anteile zu separieren, da die Summe aus mehreren integrierten (und stationären)

Anteilen wieder integriert ist

$$\underbrace{I(1) + I(1)}_{I(1)} + I(0) = I(1).$$

In einem bivariaten kointegrierten System haben beide Reihen alle integrierten Anteile gemeinsam. Es gibt nur einen gemeinsamen stochastischen Trend. Dieser ist die „Ursache“ für die Bewegung, die Wirkung für die Dynamik des Systems.

Beispiel 2.2.7 (Bivariates System mit gemeinsamem Trend (nach Stier, 2001, S.316)) *Wir betrachten zwei integrierte Reihen, die sich wie folgt darstellen lassen*

$$\begin{aligned} Y_{1t} &= aY_{2t} + \epsilon_{1t}, a \neq 0 \\ Y_{2t} &= Y_{2,t-1} + \epsilon_{2t}, \end{aligned}$$

wobei $\epsilon_{1t}, \epsilon_{2t} \sim WN$ unkorreliert sind. Y_{2t} ist ein Random Walk, also $I(1)$, Y_{1t} ist eine Linearkombination von $I(1)$ und $I(0)$ also auch $I(1)$. Es gibt jedoch eine stationäre Linearkombination

$$\begin{aligned} Y_{1t} - aY_{2t} &= aY_{2t} + \epsilon_{1t} - aY_{2t} \\ &= \epsilon_{1t} \\ &\sim I(0), \end{aligned}$$

so dass Y_{2t} herausfällt. In diesem Beispiel ist der gemeinsame $I(1)$ -Anteil und damit der gemeinsame Trend der beiden Reihen der Prozess Y_{2t} .

Das Beispiel zeigt einen Spezialfall, indem der eine Komponentenprozess über den anderen dargestellt werden kann. Im Allgemeinen gibt es einen integrierten Anteil $W_t \sim I(1)$, der in beiden Prozessen des Systems enthalten ist.

In einem bivariaten System kann man nur einen gemeinsamen Trend erkennen. Hat eine der beiden Reihen zwei $I(1)$ -Anteile, von denen die zweite Reihen *einen* teilt, reicht dies nicht aus, um alle $I(1)$ -Anteile aus dem System zu kürzen. In einem höherdimensionalen Prozess können verschiedene Konstellationen von gemeinsamen Trends in den Einzelprozessen vorkommen. Für die Anwendung ist es also interessant zu wissen, wie viele mögliche Common Trends man in einem System identifizieren kann (Stock & Watson, 1988).

Theorem 2.2.8 (Anzahl der Common Trends) *In einem System von K kointegrierten Reihen mit Kointegrationsrang r gibt es genau $K - r$ gemeinsame Trends.*

Die Anzahl der Common Trends ist also über den Kointegrationsrang bereits festgelegt. Wie dies im Einzelnen aussieht, werden wir uns nun ansehen.

Der Fall $K = 2$

Den bivariaten Fall haben wir bisher schon intensiv untersucht. Da es maximal $r = K - 1$ Kointegrationsbeziehungen in einem System geben kann, hat ein bivariates System entweder eine oder keine Kointegrationsbeziehung. Im Fall $r = 0$ handelt es sich um ein System von zwei unabhängigen integrierten Prozessen.

Bei Kointegration haben beide Reihen *einen* gemeinsamen Trend, der hier mit W_t bezeichnet wird

$$\begin{aligned} Y_{1t} &= \lambda_1 W_t + I(0) \\ Y_{2t} &= \lambda_2 W_t + I(0). \end{aligned}$$

Die Kointegrationsbeziehung wird dann beschrieben durch

$$\lambda_2 Y_{1t} + \lambda_1 Y_{2t} \sim I(0)$$

bzw. alle Vielfachen davon. Normiert also durch

$$Y_{1t} + \underbrace{\frac{\lambda_1}{\lambda_2}}_{\beta} Y_{2t} \sim I(0).$$

Der Fall $r = K - 1$

Bei multivariaten Prozessen kann es mehrere linear unabhängige Kointegrationsbeziehungen geben, ihre Anzahl wird über den Kointegrationsrang r beschrieben. Im Fall $r = K - 1$ haben alle Teilprozesse *einen* gemeinsamen Trend. Dieser treibt das Kointegrationssystem, ist mehr oder weniger stark in jeder Teilreihe enthalten. Das Verhältnis der Anteile in den verschiedenen Reihen wird durch (β_{ij}) ausgedrückt.

Beispiel 2.2.9 ($K = 3$) *In einem dreidimensionalen kointegrierten Prozess Y_t vom Rang 2 gibt es zwei linear unabhängige Kointegrationsbeziehungen zwischen den Komponentenprozessen*

$$\begin{aligned} Y_{1t} &= \beta_{13} Y_{3t} \\ Y_{2t} &= \beta_{23} Y_{3t}. \end{aligned}$$

Man kann die Reihen über ihren gemeinsamen Trend W_t darstellen

$$\begin{aligned} Y_{1t} &= \lambda_1 W_t + I(0) \\ Y_{2t} &= \lambda_2 W_t + I(0) \\ Y_{3t} &= \lambda_3 W_t + I(0). \end{aligned}$$

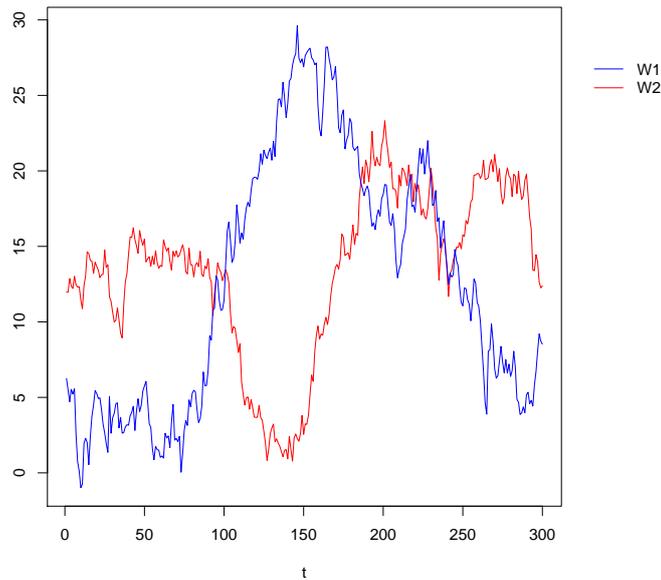


Abbildung 2.5: Zwei verschiedene Random Walk als Grundlage der folgenden Beispiele und Abbildungen

Ist der Kointegrationsrang r kleiner als $K - 1$, gibt es mehrere gemeinsame Trends im System.

Der Fall $0 < r < K - 1$

Bei Kointegrationsrang r gibt es $K - r$ gemeinsame stochastische Trends im Zeitreihensystem. Im Fall $r < K - 1$ sind es mindestens zwei Trends, die das System treiben. Dabei sind verschiedene Konstellationen möglich. In der Regel wird nicht jeder Trend in jeder Einzelreihe wirken. Da es *gemeinsame* Trends sind, taucht jedoch jeder Trend in mindestens zwei Einzelprozessen auf.

Betrachten wir ein Beispiel für $K = 3$. Für den Fall $r < K - 1$ bleibt nur $r = 1$. Es gibt also nur eine (normierte) Kointegrationsbeziehung

$$Y_{1t} + \beta_1 Y_{2t} + \beta_2 Y_{3t} \sim I(0).$$

Im System wirken $K - r = 2$ gemeinsame Trends. Dabei gibt es verschiedene Möglichkeiten, wie diese Trends in den drei Einzelreihen wirken. Die drei folgenden Beispiele sollen dies verdeutlichen. In den Abbildungen zu den Beispielen werden die beiden Random Walks aus Abbildung 2.5 verwendet.

Beispiel 2.2.10 ($K = 3, r=1$) *Eine Möglichkeit wäre, dass nicht in jeder Reihe beide Trends enthalten sind, sondern dass Y_{1t} zwei $I(1)$ -Anteile hat, von denen jeweils einer in den anderen*

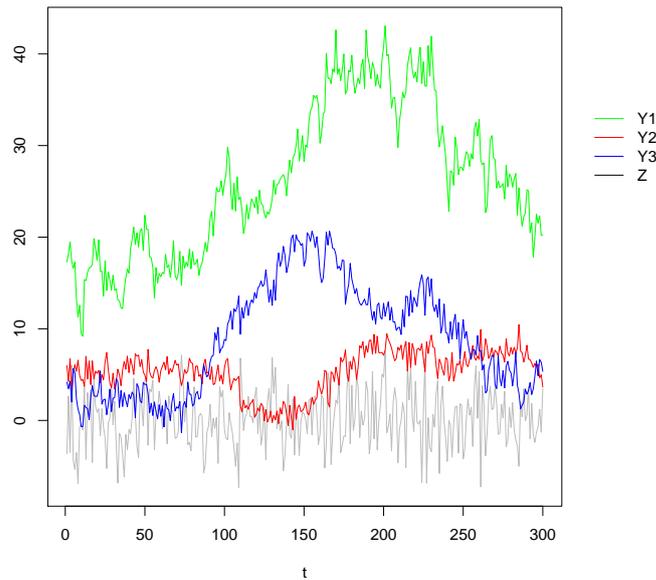


Abbildung 2.6: Beispiel 2.2.10 ($K = 3, r = 1$): Nur ein Prozess Y_{1t} (grün) enthält beide stochastischen Trends, die anderen beiden Prozesse Y_{2t} (rot) und Y_{3t} (blau) enthalten jeweils nur einen integrierten Anteil (vgl. Abbildung 2.5)

beiden Reihen enthalten ist

$$\begin{aligned} Y_{1t} &= W_{1t} + W_{2t} + I(0) \\ Y_{2t} &= \lambda_1 W_{1t} + I(0) \\ Y_{3t} &= \lambda_2 W_{2t} + I(0). \end{aligned}$$

Die Kointegrationsbeziehung ist dann

$$Y_{1t} + \underbrace{\frac{1}{\lambda_1}}_{\beta_1} Y_{2t} + \underbrace{\frac{1}{\lambda_2}}_{\beta_2} Y_{3t} \sim I(0).$$

Nur in dieser Kombination werden beide Trends herausgekürzt. In der Abbildung 2.6 ist $\lambda_1 = 0,4$ und $\lambda_2 = 0,7$. Man sieht deutlich, dass beide Reihen nur einen der beiden Trends aus Abbildung 2.5 enthalten²⁸.

Beispiel 2.2.11 ($K = 3, r=1$) In dem folgenden System besitzen zwei Reihen zwei gemeinsame Trends, allerdings zu unterschiedlichen Gewichten

$$\begin{aligned} Y_{1t} &= W_{1t} + W_{2t} + I(0) \\ Y_{2t} &= W_{1t} + \frac{1}{2}W_{2t} + I(0). \end{aligned}$$

²⁸Zur Verdeutlichung haben der Einzelprozess und der enthaltene stochastische Trend dieselbe Farbe.

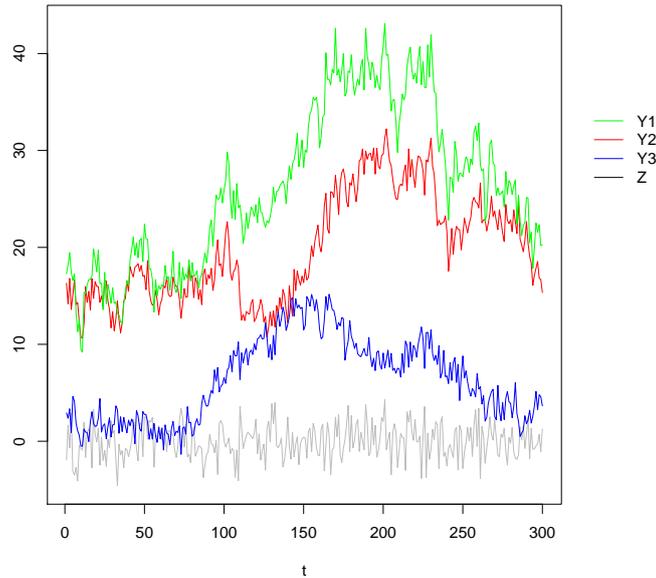


Abbildung 2.7: Beispiel 2.2.11 ($K = 3, r = 1$): Zwei Prozesse Y_{1t} (grün) und Y_{2t} (rot) besitzen zwei gemeinsame Trends, allerdings in unterschiedlicher Gewichtung. Nur durch Hinzunahme des dritten Prozesses Y_{3t} (blau) wird das System kointegriert.

Es ist nicht möglich, eine Linearkombination der beiden zu finden, in der beide Trends eliminiert werden. Dies ist erst durch Hinzunahme eines dritten Prozesses ins System möglich

$$Y_{3t} = \frac{1}{2}W_{2t} + I(0).$$

Die Kointegrationsbeziehung ist dann

$$Y_{1t} - Y_{2t} - Y_{3t} \sim I(0).$$

Erst in dem System dieser drei Reihen findet man eine Kointegrationsbeziehung. Ein gemeinsamer Trend W_{2t} ist hier in allen drei Prozessen enthalten, während W_{1t} nur zwei Prozesse treibt. Dieses Beispiel ist in Abbildung 2.7 mit den beiden stochastischen Trends aus Abbildung 2.5 dargestellt.

Das Beispiel ist natürlich zahlenmäßig so konstruiert, dass man leicht sieht, wie die Gewichte gewählt werden müssen. Allgemeiner können auch alle drei Prozesse die beiden Trendanteile aufweisen, jedoch so gewichtet, dass man mit nur zwei von ihnen nicht beide Trends eliminieren kann.

Beispiel 2.2.12 ($K = 3, r=1$) *Weisen alle drei Reihen beide Common Trends auf, wie*

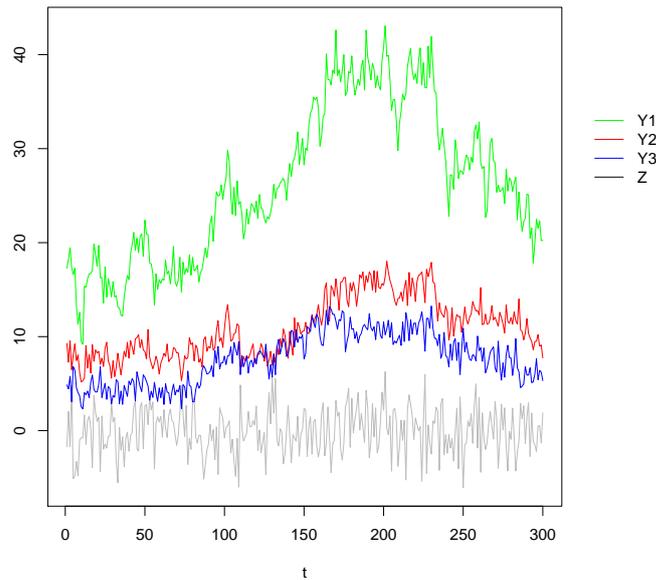


Abbildung 2.8: Beispiel 2.2.12 ($K = 3, r = 1$): Alle drei Prozesse enthalten die beiden gemeinsamen Trends, allerdings in unterschiedlicher Gewichtung. Nur in dem dreidimensionalen System gibt es eine Linearkombination, in der beide Trends gemeinsam eliminiert werden können.

beispielsweise

$$\begin{aligned} Y_{1t} &= W_{1t} + W_{2t} + I(0) \\ Y_{2t} &= \frac{1}{2}W_{1t} + \frac{1}{3}W_{2t} + I(0) \\ Y_{3t} &= \frac{1}{4}W_{1t} + \frac{1}{3}W_{2t} + I(0), \end{aligned}$$

ist es bei geeigneter Wahl der Gewichte β_{ij} eine stationäre Linearkombination zu finden

$$Y_{1t} - Y_{2t} - 2Y_{3t} \sim I(0).$$

Dieses Beispiel ist in Abbildung 2.8 zu sehen.

In allen drei Beispielen hat die Linearkombination von zwei Einzelprozessen mit der dritten Einzelreihe stets einen gemeinsamen integrierten Anteil. So werden die beiden Trends nur in richtiger Gewichtung aller drei Reihen aus dem System eliminiert. Es ist kein Umstellen möglich, die Beziehung ist eindeutig, falls normiert.

Ähnliche Beispiele lassen sich auch für höherdimensionale Prozesse finden. Wir werden in Kapitel 2.3.2 eine Darstellung kennenlernen, in der man die gemeinsamen Trends und ihre Gewichtung in einem System ablesen kann.

2.2.4 Spurious Regression

Wenn zwei integrierte Zeitreihen einen gemeinsamen stochastischen Trend aufweisen, dann besteht eine bedeutungsvolle Relation zwischen ihnen, die Kointegrationsbeziehung $Y_{1t} = \beta Y_{2t}$, die man auch in der Regressionanalyse erkennen sollte.

Dabei stößt man jedoch auf ein Problem, das als *Scheinregression* bzw. *spurious regression* bekannt ist. Yule (1926) machte das erste Mal darauf aufmerksam, dass man bei Regressionen mit Zeitreihendaten häufig hohe Korrelationen beobachtete, obwohl kein Zusammenhang zwischen den Prozessen besteht. Er nannte diesen Effekt *nonsense correlation*.

Der Begriff ist klar angelehnt an das Phänomen der Scheinkorrelation (spurious correlation), das von Pearson (1896) eingeführt wurde. Darunter versteht man, dass eine Korrelation zwischen zwei Variablen nicht automatisch einen Kausalschluss zulässt, sondern auch aus anderen Gründen auftreten kann (Haig, 2003). Die Korrelation besteht jedoch tatsächlich in den Daten.

Bei der Scheinregression liefert die Regressionsanalyse große t-Werte und hohe Bestimmtheitsmaße, obwohl die beiden Zeitreihen unabhängig voneinander sind. Das erste Mal machten C. W. Granger und Newbold (1974) auf dieses Phänomen aufmerksam, während die erklärenden Ansätze vor allem von P. Phillips (1986) stammen.

C. W. Granger und Newbold (1974) simulierten zwei voneinander unabhängige Random Walks (I(1)-Prozesse) X_t, Y_t und schätzten in der Regressionsgleichung

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t, \quad (2.7)$$

die Gewichte mittels des bekannten OLS-Schätzers (z.B. Werner, 1997; Seber & Lee, 2003), wobei ϵ_t hier als normalverteiltes Weißes Rauschen (also I(0)-Prozess) verstanden wird. Da die beiden Prozesse unabhängig sind, sollte die Nullhypothese $\beta_1 = 0$ in maximal 5% der Simulationen fälschlicherweise abgelehnt werden. Die tatsächliche Ablehnungsquote lag jedoch bei ca. 75%. Dieses Ergebnis konnte in zahlreichen Studien (u.a. Stroe-Kunold & Werner, 2009; Banerjee, 2000; Nelson & Kang, 1984) in verschiedenen Varianten reproduziert werden.

Zusätzlich beobachteten sie, dass sich dieser Effekt noch verstärkt, wenn mehr Regressoren in die Gleichung aufgenommen werden (Ablehnungsquote bis 96%). Mit dem Testwert gingen hohe Werte für das Bestimmtheitsmaß R^2 und sehr niedrige Werte für die Durbin-Watson-Statistik einher. Die Durbin-Watson-Statistik (z.B. Schlittgen & Streitberg, 1999, S. 19) weist darauf hin, dass es sich um ein missspezifiziertes Modell handelt. C. W. Granger und Newbold (1974) schließen deshalb, dass die normalen Signifikanztest für die Regressionskoeffizienten ungültig für Zeitreihendaten ungeeignet ist. Denn in Zeitreihen bestehen Abhängigkeiten, wie z.B. Autokorrelationen, so dass Verfahren, die die Unabhängigkeit der Daten voraussetzen, nicht eingesetzt werden können.

Eine zuverlässigere Aussage kann man erhalten, indem man statt der Niveaus die Regres-

sion der Differenzen betrachtet (C. W. Granger & Newbold, 1977). Dabei gehen jedoch Informationen verloren, nämlich der differenzenstationäre Anteil in den Reihen, so dass eine langfristige Beziehung, wie sie durch den gemeinsamen Trend gegeben ist, eliminiert wird (Stier, 2001, S.346f).

Bei genauer Betrachtung sieht man jedoch, dass das obige Regressionsmodell in den Niveaus (2.7) unter der Nullhypothese fehlspezifiziert ist, was durch die Durbin-Watson-Statistik bereits angedeutet wird. Denn tatsächlich soll unter der Nullhypothese $\beta_1 = 0$ folgende Gleichung gelten

$$\underbrace{Y_t}_{\sim I(1)} = \beta_0 + \underbrace{\epsilon_t}_{\sim I(0)}.$$

Der integrierte Prozess Y_t soll sich also über den stationären Prozess ϵ_t darstellen lassen.

Bereits C. W. Granger und Newbold (1974) räumen ein, dass die Regressionsgleichung (2.7) nur gültig sein kann, wenn $Y_t - \beta_1 X_t$ stationär ist. Dies ist jedoch bei Kointegration der Fall.

Da die Regressionsgleichung nur bei Kointegration richtig ist, kann eine Regression auch nur in diesem Fall bedeutungsvoll sein (Maddala & Kim, 2007). Bei integrierten Reihen ohne gemeinsamen Anteil handelt es sich um eine Scheinregression.

P. Phillips (1986) untersuchte die Verteilung des OLS-Schätzers für integrierte Systeme, die nicht kointegriert sind, und konnte so die Beobachtungen von C. W. Granger und Newbold (1974) erklären. Die Schätzung von β_1 konvergiert unter gültiger Nullhypothese nämlich nicht wie im Standardfall gegen Null, sondern hat eine komplexe Grenzverteilung, die unter anderem von zwei Wiener Prozessen abhängig ist (siehe auch Stier, 2001). Unter dieser Verteilung ist das Überschreiten der üblichen kritischen Werte immer wahrscheinlicher, je länger das System beobachtet wird. Dies lässt sich auch auf die Regression mit mehreren Regressoren übertragen.

Voraussetzung für diese Verteilung ist jedoch, dass es keine stationäre Linearkombination der abhängigen und unabhängigen Variablen gibt, also keine Kointegration vorliegt. Außerdem darf kein integrierter Prozess einen zusätzlichen Drift haben. Haldrup (1994) weitet Ergebnisse auf I(2)-Prozesse aus.

Nach DeJong (2003) ist eine Scheinregression eine Regression von integrierten Reihen, ohne das Vorliegen einer Kointegrationsbeziehung. Deshalb empfiehlt C. W. J. Granger (1980) Kointegrationstests, um einer Scheinregression mit integrierten Zeitreihendaten vorzubeugen.

Wir werden im Kapitel 2.7 geeignete Tests auf Kointegration kennenlernen.

2.3 Darstellungsformen kointegrierter Systeme

In diesem Kapitel wollen wir verschiedene Darstellungen eines kointegrierten Prozesses betrachten und insbesondere erarbeiten, wie sich die Kointegration darin ausdrückt und welchen Erkenntnisgewinn man aus welcher Darstellung ziehen kann. Einen wichtigen Schlüssel zu den verschiedenen Darstellungsformen bietet das sogenannte *Granger Repräsentationstheorem*, das erstmals in Engle und Granger (1987) veröffentlicht wurde. Die wichtigsten Aussagen sind z.B. bei Kirchgässner und Wolters (2007, S.204ff) zusammengefasst.

2.3.1 Granger Repräsentationstheorem

Das folgende Vorgehen hält sich an das Originalkapitel „3. Properties of co-integrated variables and their representations“ von Engle und Granger (1987). Die Darstellung wird durch eigene ausführliche Ergänzungen und Anschauungen sowie Beispiele abgerundet.

Im Folgenden sei Y_t ein K -dimensionaler integrierter Prozess der Ordnung $I(1)$ ohne deterministischen Anteil. Sein Differenzenprozess ΔY_t ist dann stationär und rein-stochastisch und hat den Erwartungswert $\mathbb{E}[\Delta Y_t] = 0$, weshalb er die Wolddarstellung

$$\Delta Y_t = \Psi(L)\epsilon_t$$

besitzt. Dabei ist $\Psi(L)$ eine (Lag-)Polynom-Matrix, deren (Polynom-) Einträge alle endliche Ordnung haben. Die Determinante dieser Matrix (ist ein Polynom und) habe nur Nullstellen auf oder außerhalb des Einheitskreises. $\Psi(0)$ sei wie üblich gleich der Einheitsmatrix \mathbb{I}_K . Dadurch wird bewirkt, dass die konstanten Anteile der Polynome auf der Diagonalen gleich Eins und die der übrigen gleich Null sind. Dadurch werden gleichzeitige (simultane) Einflüsse der Einzelreihen untereinander ausgeschlossen.

Zerlegung der Matrix $\Psi(L)$

Die Wolddarstellung eines allgemeinen Prozesses ohne deterministischen Anteil entspricht einer $MA(\infty)$ -Darstellung (siehe Korollar 1.6.6, S.31). Da die Lagpolynommatrix des Differenzenprozesses ΔY_t jedoch nach Voraussetzung endliche Ordnung hat, hat auch der Differenzenprozess eines integrierten Prozesses der Ordnung Eins eine *endliche* MA-Darstellung. Um diese Darstellung auf Kointegration untersuchen zu können, müssen wir die Matrix $\Psi(L)$ zerlegen. Dass dies möglich ist, besagt folgendes Lemma (Engle & Granger, 1987, S.255).

Lemma 2.3.1 (Zerlegung einer Polynom-Matrix endlicher Ordnung) *Eine (Lag-)Polynommatrix endlicher Ordnung mit $\Psi(0) = \mathbb{I}_K$, deren Determinante nur Nullstellen auf oder außerhalb des Einheitskreises hat, kann man durch einfache Umformung wie folgt zerlegen*

$$\Psi(L) = \Psi(1) + (1 - L)\Psi^*(L).$$

Da $\Psi(L)$ nur Einträge endlicher Ordnung hatte, gilt dies auch für die Lagpolynommatrix $\Psi^*(L)$.

Um im Anwendungsfall diese Zerlegung durchzuführen, muss man zunächst die Voraussetzungen prüfen. Dann kann man anschließend $\Psi(1)$ und $\Psi^*(L)$ berechnen und erreicht somit die Zerlegung.

Bemerkung 2.3.2 (Vorgehen zur Zerlegung von Ψ) *Die Zerlegung der (Lag-)Polynommatrix $\Psi(L)$ erfolgt in folgenden Schritten:*

- Prüfen, ob die konstanten Anteile die Bedingung $\Psi(0) = \mathbb{I}_K$ erfüllen, also auf der Diagonalen gleich Eins, außerhalb gleich Null sind.
- Determinante ausrechnen und prüfen, ob für alle Nullstellen gilt, dass sie auf oder außerhalb des Einheitskreises liegen.
- Für $z = 1$ einsetzen und die Werte aller Polynome ausrechnen.
- Diese Werte von den Polynomen abziehen $\Psi(z) - \Psi(1)$.
- Durch Polynomdivision²⁹ aus jedem verbleibenden Polynom den Faktor $(1 - z)$ ausklammern.

Dieses Vorgehen wird deutlicher, wenn man es auf ein einfaches Beispiel (Stier, 2001, S.317) anwendet.

Beispiel 2.3.3 (bivariater Fall) *Wir gehen von folgender 2×2 -Lagpolynommatrix $\Psi(L)$ und ihrer z -Transformierte $\Psi(z)$ aus*

$$\Psi(L) = \begin{pmatrix} 1 - L & 3L \\ 0 & 1 \end{pmatrix}; \quad \Psi(z) = \begin{pmatrix} 1 - z & 3z \\ 0 & 1 \end{pmatrix}.$$

Die Komponenten dieser Matrix sind also folgende Polynome (endlicher Ordnung):

$$\begin{aligned} \psi_{11}(z) &= 1 - z & \psi_{12}(z) &= (0+)3z \\ \psi_{21}(z) &= 0 & \psi_{22}(z) &= 1 \end{aligned}$$

Die Bedingung $\Psi(0) = \mathbb{I}_2$ ist erfüllt, da bei beiden Polynomen auf der Diagonalen ψ_{11} und ψ_{22} als konstantes Glied jeweils 1 und bei den beiden anderen ψ_{12} und ψ_{21} jeweils 0 steht. Die Determinante dieser Matrix ist (vgl. Bemerkung A.1.23, S.311 im Anhang)

$$\begin{aligned} \det(\Psi(z)) &= \psi_{11}(z)\psi_{22}(z) - \psi_{12}(z)\psi_{21}(z) \\ &= ((1 - z) \cdot 1) - (3z \cdot 0) = 1 - z. \end{aligned}$$

Die Determinante hat also eine Nullstelle $1 - z = 0 \Leftrightarrow z = 1$ vom Betrag Eins und liegt

²⁹siehe z.B. Bosch (2009); Meyberg (1999)

somit auf dem Einheitskreis. Den Wert für $\Psi(1)$ berechnet man durch Einsetzen

$$\begin{array}{ll} \psi_{11}(1) = 1 - 1 = 0 & \psi_{12}(1) = 3 \cdot 1 = 3 \\ \psi_{21}(1) = 0 & \psi_{22}(1) = 1 \end{array} \Rightarrow \Psi(1) = \begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix}.$$

Wenn man diese Werte nun von der ursprünglichen Polynommatrix abzieht, dann erhält man

$$\begin{array}{ll} \psi_{11}(z) - \psi_{11}(1) = 1 - z & \psi_{12}(z) - \psi_{12}(1) = 3z - 3 \\ \psi_{21}(z) - \psi_{21}(1) = 0 & \psi_{22}(z) - \psi_{22}(1) = 0 \end{array} \Rightarrow \Psi(z) - \Psi(1) = \begin{pmatrix} 1 - z & 3z - 3 \\ 0 & 0 \end{pmatrix}.$$

Klammert man nun aus jedem Polynom einen Faktor $(1 - z)$ aus, erhält man die Zerlegung von $\Psi(L)$

$$\begin{array}{ll} \psi_{11}^*(z) = (1 - z) \cdot 1 & \psi_{12}^*(z) = (1 - z) \cdot (-3) \\ \psi_{21}^*(z) = (1 - z) \cdot 0 & \psi_{22}^*(z) = (1 - z) \cdot 0 \end{array} \Rightarrow \Psi^*(z) = \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix}.$$

Durch Zusammensetzen der Ergebnisse ergibt sich die Zerlegung der Lagpolynommatrix

$$\begin{aligned} \Psi(L) &= \Psi(1) + (1 - L)\Psi^*(L) \\ \begin{pmatrix} 1 - L & 3L \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix} + (1 - L) \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Die Bestandteile dieser Zerlegung sind zentral für das Granger Repräsentationstheorem.

Granger Repräsentationstheorem

Wir haben bereits gesehen, dass die Differenz eines kointegrierten Prozesses eine endlich Moving-Average-Darstellung hat. Mit den wenigen obigen Annahmen (Y_t ist K -dimensional und kointegriert der Ordnung $CI(1,1)$ mit Rang r) kann man jedoch noch weitere Darstellungen des Prozesses ableiten. Diese sind im Granger Repräsentationstheorem zusammengefasst. Es wird hier zunächst nach Engle und Granger (1987) in der Form ohne deterministischen Anteil mit Erwartungswert Null wiedergegeben. Anschließend betrachten wir die einzelnen Darstellungen genauer und gehen dabei auf den etwas allgemeineren Fall eines von Null verschiedenen Erwartungswertes ein. Wir werden genau sehen, wie sich die Kointegration in den einzelnen Darstellungen ausdrückt und welchen Informationsgewinn der Anwender dadurch hat.

Theorem 2.3.4 (Granger Repräsentationstheorem, (Engle & Granger, 1987)) *Sei Y_t ein K -dimensionaler integrierter Prozess der Ordnung $I(1)$ ohne deterministischen Anteil mit einem stationären, rein-stochastischen Differenzenprozess mit Erwartungswert $\mathbb{E}[\Delta Y_t] = 0$, der die Wolddarstellung $\Delta Y_t = \Psi(L)\epsilon_t$ hat. Dieser Prozess Y_t sei außerdem kointegriert der Ordnung $CI(1,1)$ mit Kointegrationsrang r .*

Dann gelten folgende Aussagen.

1. $\Psi(1)$ hat den Rang $K - r$.
2. Es gibt eine VAR-Darstellung des Prozesses.

$$\begin{aligned} A(L)Y_t &= \epsilon_t \\ Y_t &= A_1Y_{t-1} + \dots + A_pY_{t-p} + \epsilon_t \end{aligned}$$

- $A(L) = \mathbb{I}_K - A_1L - \dots - A_pL^p$
 - $A(0) = \mathbb{I}_K$
 - $A(1)$ hat Rang r
3. Es existieren zwei $(K \times r)$ -Matrizen β und α , die beide den Rang r haben und die außerdem folgende Bedingungen erfüllen.
 - $A(1) = -\alpha\beta'$
 - $\beta'\Psi(1) = 0$
 - $\Psi(1)\alpha = 0$
 4. Es gibt eine Fehlerkorrektur-Darstellung des Prozesses mit dem r -dimensionalen Prozess $Z_t = \beta'Y_t$.

$$A^*(L)\Delta Y_t = \alpha Z_{t-1} + \epsilon_t$$

Dabei ist $A^*(0) = \mathbb{I}_K$.

5. Der Prozess Z_t ist gegeben durch

$$Z_t = \beta'\Psi^*(L)\epsilon_t \tag{2.8}$$

Der Beweis dieses zentralen Theorems findet sich auch bei Engle und Granger (1987, S.257f). Wir werden uns hier auf die Aussagen konzentrieren und einige Ergebnisse nur heuristisch herleiten.

Ganz allgemein besagt das Granger Repräsentationstheorem, dass ein K -dimensionaler Prozess Y_t eine bestimmte VAR-Darstellung besitzt, wenn seine Differenz eine bestimmte MA-Darstellung hat. Diese „Äquivalenz der beiden Darstellungen“ wird häufig zitiert (z.B. Hatanaka, 1996). Zusätzlich wird eine dritte Darstellungsform, die Fehlerkorrektur-Darstellung, erwähnt. Doch wir werden sehen, dass sie sehr eng mit der VAR-Darstellung verbunden ist, so dass wir uns zunächst auf die zwei *Haupt*darstellungen konzentrieren.

Da auch stationäre Prozesse in VAR-Form dargestellt werden können und ihr Differenzen

ebenfalls eine MA-Darstellung hat, für Y_t hier jedoch ausdrücklich ein *kointegrierter* Prozess angenommen wird, muss die Kointegration in den Darstellungen zu sehen sein.

Da die MA-Darstellung durch die (Lag-)Polynommatrix $\Psi(L)$ und die VAR-Darstellung durch die (Lag-)Polynommatrix $A(L)$ gekennzeichnet ist, verwundert es nicht, dass das Repräsentationstheorem vor allem Bedingungen an diese beiden Matrizen stellt. So entsprechen $\Psi(0)$ und $A(0)$ wie üblich der K -dimensionalen Einheitsmatrix, um simultane Einflüsse auszuschließen. Während $\Psi(1)$ den Rang $K - r$ hat, ist der Rang von $A(1)$ gleich r . Dies ist ein erster Hinweis auf die Kointegration. Denn wir wissen bereits, dass r die Anzahl der Kointegrationsbeziehungen beschreibt und dass das System von $K - r$ gemeinsamen stochastischen Trends getrieben wird. Wir wollen im Folgenden sehen, wie man dies in den Bedingungen an die beiden Polynommatrizen in den verschiedenen Darstellungen sieht und welche Rolle die beiden Matrizen α und β aus der Zerlegung von $A(1)$ spielen. Im nächsten Kapitel werden wir dann das Fehlerkorrekturmodell genau betrachten.

In den nächsten beiden Unterkapiteln wird gezeigt, wie man von den im Repräsentationstheorem genannten Voraussetzungen ausgehend zu den fünf Folgerungen gelangt. Dabei soll kein formaler mathematischer Beweis erfolgen, wie er bei Engle und Granger (1987) selbst zu finden ist. Wir werden uns vielmehr auf die Zusammenhänge der verschiedenen Darstellungen und die Anzeichen für Kointegration konzentrieren.

2.3.2 Moving-Average- und Common-Trend-Darstellung

Als Ausgangspunkt des Granger Repräsentationstheorems dient die Moving-Average-Darstellung der Differenz des kointegrierten Prozesses. Sie ist deshalb möglich, weil angenommen wird, dass der Prozess Y_t integriert der Ordnung Eins ist, so dass bereits seine erste Differenz stationär ist. Für stationäre Prozesse endlicher Ordnung existiert stets eine endliche MA-Darstellung, wie wir in Kapitel 1.3 (ab S.22) gesehen haben.

Durch die Zerlegung der MA-Matrix $\Psi(L)$ gelangen wir nicht nur zu Bedingungen an die Moving-Average-Darstellung eines kointegrierten Prozesses, sondern können auch die sogenannte *Common-Trend-Darstellung* ableiten, die es uns ermöglicht gemeinsame Trends im System direkt zu identifizieren.

Moving Average Darstellung

Da der Differenzenprozess ΔY_t stationär ist, hat er eine MA-Darstellung, die sich mit einer Lagpolynommatrix $\Psi(L)$ darstellen lässt. Wir betrachten hier den allgemeineren Fall, dass ΔY_t außerdem einen konstanten Mittelwert hat (vgl. Lütkepohl, 2005; Neusser, 2006). Der Prozess schwankt also nicht stationär um Null, sondern um $\mu \in \mathbb{R}^K$, lässt sich also als

$$\Delta Y_t = \mu + \Psi(L)\epsilon_t$$

darstellen. Dabei gilt weiterhin die Zerlegung aus Lemma 2.3.1 (S.81)

$$\Psi(L) = \Psi(1) + (1 - L)\Psi^*(L).$$

Bei Kointegration hat $\Psi(1)$ den Rang $K - r$, wie in Aussage 1 des Granger Repräsentationstheorems (Theorem 2.3.4, S.83) gefolgert wird. Wir werden in der Common-Trend-Darstellung, die sich aus der MA-Darstellung ableitet, sehen, dass $\Psi(1)$ nicht nur die Anzahl der gemeinsamen stochastischen Trends, sondern auch deren Gewichtung im System bestimmt. Aus dem Theorem 2.2.8 (S.73) wissen wir bereits, dass es in einem K -dimensionalen kointegrierten System vom Rang r genau $K - r$ gemeinsame stochastische Trends gibt.

Nun lassen sich für die Moving-Average-Darstellung Bedingungen für Kointegration formulieren (Engle & Yoo, 1987, siehe auch Stier, 2001, S.321f). Nach der Zerlegung der Lagmatrix $\Psi(L)$ (s.o.) gilt

$$\begin{aligned}\Delta Y_t &= \mu + \Psi(L)\epsilon_t \\ \Delta Y_t &= \mu + [\Psi(1) + (1 - L)\Psi^*(L)]\epsilon_t.\end{aligned}$$

Ersetzt man nun den Differenzenoperator durch seine Darstellung als Lagpolynom $\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$ (siehe Anhang C.1.1, S.337) und wendet die Bestandteile der Zerlegung einzeln auf ϵ_t an, ergibt sich

$$(1 - L)Y_t = \mu + \Psi(1)\epsilon_t + (1 - L)\Psi^*(L)\epsilon_t.$$

Die Matrix β besteht, wie wir aus Kapitel 2.2 wissen, aus den Kointegrationsvektoren des Prozesses Y_t . Durch Multiplikation mit dieser Kointegrationsmatrix taucht auf der linken Seite der Gleichung der Abweichungsprozess $Z_t = \beta'Y_t$ auf³⁰

$$(1 - L)\underbrace{\beta'Y_t}_{Z_t} = \beta'\mu + \beta'\Psi(1)\epsilon_t + (1 - L)\beta'\Psi^*(L)\epsilon_t. \quad (2.9)$$

Nun weiß man, dass im Fall der Kointegration dieser Abweichungsprozess Z_t als Gleichgewichtsbeziehung stationär ist. Dies ist in Gleichung (2.9) jedoch nur der Fall, wenn gilt

$$\beta'\mu = 0 \quad (2.10)$$

$$\text{und } \beta'\Psi(1) = 0. \quad (2.11)$$

³⁰Hinweis: Der Operator $(1 - L)$ kann in der Matrizenmultiplikation ausgeklammert, also nach vorne geschrieben werden.

Dann kann der Faktor $(1 - L)$ aus der Darstellung gekürzt werden und man erhält

$$\begin{aligned}(1 - L)Z_t &= \underbrace{\beta'\mu}_{=0} + \underbrace{\beta'\Psi(1)}_{=0}\epsilon_t + (1 - L)\beta'\Psi^*(L)\epsilon_t \\(1 - L)Z_t &= (1 - L)\beta'\Psi^*(L)\epsilon_t \\Z_t &= \beta'\Psi^*(L)\epsilon_t\end{aligned}$$

die Darstellung des Abweichungsprozesses Z_t , die im Granger Repräsentationstheorem (Theorem 2.3.4, S.83) in der Aussage 5 postuliert wird. Auch die Bedingung (2.11) für Kointegration findet sich in Aussage 3, die zweite Kointegrationsbedingung (2.10) wird durch die Hinzunahme eines konstanten Erwartungswertes μ notwendig (Lütkepohl, 2005).

Beispiel 2.3.5 (bivariater Fall) *Einen allgemeinen bivariaten Prozess kann man in folgender MA-Darstellung schreiben*

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}.$$

Für die Differenzen der Einzelprozesse gilt also, wie man leicht sieht, wenn man die Systemdarstellung betrachtet

$$\begin{aligned}\Delta Y_{1,t} &= \mu_1 + \psi_{11}(L)\epsilon_{1,t} + \psi_{12}(L)\epsilon_{2,t} \\ \Delta Y_{2,t} &= \mu_2 + \psi_{21}(L)\epsilon_{1,t} + \psi_{22}(L)\epsilon_{2,t},\end{aligned}$$

dass die Differenzen beider Einzelprozesse von beiden Fehlerprozessen ϵ_{1t} und ϵ_{2t} beeinflusst werden.

Mit der Matrix $\Psi(L)$ aus dem Beispiel 2.3.3 (S.82) gilt mit $\mu_1 = 2$ und $\mu_2 = 7$ beispielsweise

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix} + \begin{pmatrix} 1 - L & 3L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix},$$

was nach Matrizenmultiplikation in System Schreibweise wie folgt darstellt wird

$$\begin{aligned}\Delta Y_{1,t} &= 2 + (1 - L)\epsilon_{1,t} + (3L) \cdot \epsilon_{2,t} \\ \Delta Y_{2,t} &= 7 + 0 \cdot \epsilon_{1,t} + 1 \cdot \epsilon_{2,t}.\end{aligned}$$

Wenn man nun die Lagpolynome auf die Fehlerprozesse anwendet, erhält man die ausführliche Darstellung

$$\begin{aligned}\Delta Y_{1,t} &= 2 + \epsilon_{1,t} - \epsilon_{1,t-1} + 3\epsilon_{2,t-1} \\ \Delta Y_{2,t} &= 7 + \epsilon_{2,t}.\end{aligned}$$

Verwendet man nun, dass man wegen $\Delta Y_t = Y_t - Y_{t-1}$ den Prozess zum Zeitpunkt t über seinen Werte zum vorigen Zeitpunkt $t - 1$ darstellen kann,

$$\begin{aligned} Y_{1,t} &= Y_{1,t-1} + 2 + \epsilon_{1,t} - \epsilon_{1,t-1} + 3\epsilon_{2,t-1} \\ Y_{2,t} &= Y_{2,t-1} + 7 + \epsilon_{2,t} \end{aligned}$$

und setzt dies rekursiv fort, erhält man eine Darstellung von Y_t in Abhängigkeit von seinem Startwert Y_0 zum Zeitpunkt $t = 0$. Die ausführliche Herleitung dieser Darstellung ist im Anhang C.2.1 ab Seite 341 zu finden.

Common-Trend-Darstellung

Aus der Moving-Average-Darstellung des Differenzenprozesses ΔY_t kann man (vgl. Anhang C.2.1, S.341) eine Darstellung herleiten, in der man die gemeinsamen Trends des kointegrierten Systems identifizieren kann. In dieser Common-Trend-Darstellung (z.B. Lütkepohl, 2005)

$$Y_t = Y_0 + t\mu + \Psi(1) \sum_{k=1}^t \epsilon_t + \underbrace{\Psi^*(L)\epsilon_t - \Psi^*(L)\epsilon_0}_{V_t}$$

wird der Prozess Y_t in folgende Bestandteile zerlegt:

- Startwert Y_0
- deterministische lineare Trends (enthalten im Vektor $t\mu$)
- $K - r$ gemeinsame stochastische Trends $\sum_{k=1}^t \epsilon_{i,t}$ (enthalten in $\Psi(1) \sum_{k=1}^t$)
- eine stationäre Komponente V_t

Außer dem vektoriellen Startwert Y_0 und einer stationären Komponente V_t ist der Prozess von verschiedenen Trends abhängig. Es ergibt sich ein vektorieller deterministischer Trend durch die Hinzunahme eines konstanten Mittelwertvektors μ in der MA-Darstellung. Dieser Vektor bewirkt, dass jede Einzelreihe unabhängig von den anderen einem eigenen linearem Trend folgt, der den Mittelwert dieser Reihe zeitlich bedingt verändert.

Von besonderem Interesse ist der Teil, der die gemeinsamen stochastischen Trends beschreibt. So bietet diese Darstellung eine Möglichkeit die Common Trends des kointegrierten Systems mathematisch zu identifizieren. Alle K möglichen stochastischen Trends sind im Vektor $\sum_{k=1}^t \epsilon_t$ zusammengefasst³¹. Da die Matrix $\Psi(1)$ in der Darstellung diesen Vektor gewichtet, wird durch ihren Rang nicht nur die Anzahl der tatsächlich wirkenden Random Walks, sondern durch ihre Einträge ψ_{ij} auch die Stärke des Einflusses auf jede Einzelkomponente.

³¹Wir haben in Abschnitt 1.8 (S.33) gesehen, dass die Summen von Weißem Rauschen integriert ist, also einem stochastischen Trend folgt.

Beispiel 2.3.6 (Bivariater Fall: Common-Trend-Darstellung) *Im bivariaten Fall gibt es zwei integrierte Einzelreihe, also zwei mögliche stochastische Trends, $\sum_{k=1}^t \epsilon_{1,t}$ und $\sum_{k=1}^t \epsilon_{2,t}$. In der Common-Trend-Darstellung drückt sich das wie folgt aus*

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} Y_{1,0} \\ Y_{2,0} \end{pmatrix} + t \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \Psi(1) \begin{pmatrix} \sum_{k=1}^t \epsilon_{1,t} \\ \sum_{k=1}^t \epsilon_{2,t} \end{pmatrix} + \Psi^*(L) \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \Psi^*(L) \begin{pmatrix} \epsilon_{1,0} \\ \epsilon_{2,0} \end{pmatrix}.$$

Betrachtet man das obige Beispiel 2.3.5 (S.87)

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} 2 \\ 7 \end{pmatrix} + \begin{pmatrix} 1-L & 3L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$$

und nimmt an, dass der Prozess im Punkt $Y_0 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$ begonnen hat, dann ergibt sich mit

$$\Psi(1) = \begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix} \text{ und } \Psi^*(L) = \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix} \text{ (vgl. Beispiel 2.3.3, S.82) die Darstellung}$$

$$\begin{aligned} \begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} &= \begin{pmatrix} 3 \\ 5 \end{pmatrix} + t \cdot \begin{pmatrix} 2 \\ 7 \end{pmatrix} + \begin{pmatrix} 0 & 3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{k=1}^t \epsilon_{1,t} \\ \sum_{k=1}^t \epsilon_{2,t} \end{pmatrix} \\ &\quad + \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix} - \begin{pmatrix} 1 & -3 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{1,0} \\ \epsilon_{2,0} \end{pmatrix} \end{aligned}$$

Führt man alle Multiplikationen aus, erhält man für die beiden Einzelreihen die System-schreibweise

$$\begin{aligned} Y_{1,t} &= 3 + t \cdot 2 + 0 \cdot \sum_{k=1}^t \epsilon_{1,k} + 3 \sum_{k=1}^t \epsilon_{2,k} + \epsilon_{1,t} - 3\epsilon_{2,t} - \epsilon_{1,0} + 3\epsilon_{2,0} \\ Y_{2,t} &= 5 + t \cdot 7 + 0 \cdot \sum_{k=1}^t \epsilon_{1,k} + 1 \sum_{k=1}^t \epsilon_{2,k} + 0. \end{aligned}$$

Da der erste Random Walk in beiden Gleichungen mit Null gewichtet wird, sieht man in der Common-Trend-Darstellung

$$\begin{aligned} Y_{1,t} &= 3 + t \cdot 2 + 3 \sum_{k=1}^t \epsilon_{2,k} + \epsilon_{1,t} - 3\epsilon_{2,t} - \epsilon_{1,0} + 3\epsilon_{2,0} \\ Y_{2,t} &= 5 + t \cdot 7 + 1 \sum_{k=1}^t \epsilon_{2,k} + 0, \end{aligned}$$

dass das System nur von einem Random Walk $\sum \epsilon_{2,k}$ getrieben wird und der zweite $\sum \epsilon_{1,k}$ ohne Bedeutung bleibt.

Folglich ist die erste Aussage in Grangers Repräsentationstheorem entscheidend. Sie ist die theoretische Grundlage für die Behauptung, dass es bei Kointegrationsrang r genau $K - r$ gemeinsame Trends im kointegrierten System gibt. Wir haben eben gesehen, dass dies auf den Rang der Matrix $\Psi(1)$ zurückzuführen ist. In der zweiten Aussage des Theorems werden nun Bedingungen an die vektorielle autoregressive Darstellung des kointegrierten Systemes gestellt.

2.3.3 Vektorautoregressive und Fehlerkorrekturdarstellung

Die vektorautoregressive Darstellung ist in der Anwendung die häufigste, da sie sehr einfach geschätzt werden kann (z.B. Lütkepohl, 2005, Kap. 3; Rinne & Specht, 2002, Kap. 14). Deshalb ist es für kointegrierte Prozesse besonders wichtig, welche Informationen man an der VAR-Darstellung ablesen kann. Zumal hier auch der Prozess Y_t selbst dargestellt wird und nicht wie in der MA-Darstellung seine Differenz ΔY_t .

Außerdem ist die VAR-Darstellung eng verbunden mit der Fehlerkorrekturdarstellung für kointegrierte Prozesse, die sowohl lang- als auch kurzfristige Dynamiken des Systems abbildet und somit viele Informationen auf einen Blick liefert.

Vektorautoregressive Darstellung

Für einen kointegrierten Prozess Y_t postuliert das Granger Repräsentationstheorem auch eine vektorautoregressive Darstellung der Form

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t.$$

Auch hier nehmen wir zusätzlich zu Engle und Granger (1987) bereits eine Konstante $c \in \mathbb{R}^K$, die einen von Null verschiedenen Mittelwert beschreibt, in die Darstellung auf. Dadurch werden die Aussagen des Repräsentationstheorems jedoch nur leicht erweitert (Neusser, 2006; Lütkepohl, 2005).

Von zentraler Bedeutung für die VAR-Darstellung ist die Lagpolynommatrix $A(L)$, deren z -Transformierte (siehe Definition 1.5.4, S.28) sich wie folgt darstellen lässt

$$A(z) = \mathbb{I}_K - A_1 z - A_2 z^2 - \dots - A_p z^p. \quad (2.12)$$

Da die Einzelreihen eines kointegrierten Prozesses integriert sein müssen, damit sie kointegriert sein können, hat das charakteristische Polynom dieser Matrix $\det(A(z))$ nicht nur Nullstellen außerhalb, sondern auch einige *auf* dem Einheitskreis $|z| = 1$ (Neusser, 2006, Kap.16.2.2), was dazu führt, dass die Matrix $A(L)$ nicht invertierbar ist. Der Prozess Y_t ist also instationär, was eine triviale Voraussetzung für Integration ist.

Setzt man für $z = 0$ in die Gleichung (2.12) ein, fallen die Summanden $A_i z^i$ aus der Gleichung heraus, so dass das Ergebnis $A(0) = \mathbb{I}_K$ ist. Dies bewirkt, dass simultane Einflüsse der Einzelreihen untereinander ausgeschlossen werden. Die Einzelreihen sind nur von sich selbst oder vergangenen Zuständen der anderen Einzelreihen abhängig.

Das Phänomen der Kointegration lässt sich auch in der VAR(p)-Darstellung ablesen, und zwar an der Matrix $A(1)$. Im Fall einer Kointegration der Ordnung r ist diese Matrix nicht invertierbar, sondern hat den Rang r . Wie wir bei der Betrachtung des Fehlerkorrekturmodells sehen werden, wird ihrer additiv Inversen $\Pi = -A(1)$ besondere Bedeutung zuteil (z.B. Lütkepohl, 2005, S.248)

$$\Pi = -\mathbb{I}_K + A_1 + \dots + A_p.$$

In der VAR(p)-Darstellung gibt also der Rang der Matrix Π bzw. $A(1)$ den Kointegrationsrang des Prozesses an. Im Fall der Kointegration ist $r < K$, so dass man die Matrix Π in das Produkt zweier $(K \times r)$ -Matrizen zerlegen kann, wie dies in Aussage 3 des Granger Repräsentationstheorems (Theorem 2.3.4, S.83) für $A(1)$ postuliert wird

$$\underbrace{\Pi}_{K \times K} = \underbrace{\alpha}_{K \times r} \underbrace{\beta'}_{r \times K}.$$

Diese Zerlegung ist natürlich nicht eindeutig (z.B. Neusser, 2006, S.212). Vielmehr kann man mit jeder beliebigen invertierbaren $r \times r$ -Matrix M durch $\alpha^* = \alpha M'$ und $\beta^* = \beta M^{-1}$ neue $K \times r$ -Matrizen definieren, in die man Π zerlegen kann. Denn es gilt mit den Regeln aus Anhang A.1.3 (ab S.310)

$$\alpha^* \beta^{*'} = \alpha M' \cdot (\beta M^{-1})' = \alpha M' (M')^{-1} \beta' = \alpha \beta' = \Pi.$$

Dies zeigt noch einmal, dass die Kointegrationsbeziehungen ohne Normierung nicht eindeutig sind (siehe Unterkapitel 2.2.2, ab S.68). Wählt man für β eine feste Darstellung, ist auch α festgelegt.

Engle und Granger (1987) fordern in den Annahmen des Theorems, dass der Prozess Y_t integriert vom Grad Eins ist ($Y_t \sim CI(1, 1) \Rightarrow Y_t \sim I(1)$). Dies drückt sich in der Matrix Π darin aus, dass auch ihr Quadrat den Rang r hat (Neusser, 2006)

$$\text{Rang}(\Pi) = \text{Rang}(\Pi \cdot \Pi) = \text{Rang}(\Pi^2).$$

In diesem Fall gibt es genau $K - r$ Einheitswurzeln (Neusser, 2006) und es ist möglich die Polynommatrix $A(z)$ in eine spezielle Form der Smith-McMillan-Faktorisierung zu bringen (Yoo, 1987, S.28).

Bemerkung 2.3.7 (Smith-McMillan-Faktorisierung) *Für eine beliebige Polynommatrix $A(L)$*

gilt folgende Zerlegung:

$$A(z) = Q(z)M(z)R(z)$$

wobei $Q(z), R(z)$ Polynommatrizen sind, deren Nullstellen alle außerhalb des Einheitskreises liegen und für $M(z)$ gilt

$$M(z) = \begin{pmatrix} (1-z)\mathbb{I}_{K-r} & 0 \\ 0 & \mathbb{I}_r \end{pmatrix}.$$

Diese Zerlegung ist eine Spezialform der SmithMcMillan-Faktorisierung von Polynommatrizen.

Diese Zerlegung der AR-Matrix $A(z)$ wird benötigt, um aus der vektorautoregressiven Darstellung zur Moving Average Darstellung des Systems zu gelangen. Dies ist jedoch mathematisch sehr aufwendig, weshalb die Ausführungen in den Anhang C.2.3 (S.346f) verschoben wurden.

Fehlerkorrekturdarstellung

Sehr eng mit der vektorautoregressiven Darstellung ist die Fehlerkorrekturdarstellung eines kointegrierten Prozesses verbunden. Denn man gelangt von der VAR(p)-Darstellung

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t$$

durch die einfache Operation $-Y_{t-1}$ und Umstrukturierungen, die im Anhang C.2.2 ab Seite 345 ausführlich dargestellt sind, auf die Fehlerkorrekturdarstellung des Prozesses (z.B. Lütkepohl, 2005, S.248; Kirchgässner & Wolters, 2007, S.130)

$$\Delta Y_t = c + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t,$$

die also wiederum eine Darstellung des Differenzenprozesses ist. Hier werden langfristige Gleichgewichtsbeziehungen und kurzfristige Fehlerkorrekturen sichtbar. Eine wichtige Rolle spielt vor allem die Matrix Π , die wir aus den obigen Darstellungen bereits kennen.

Zusammengefasst lässt sich die Fehlerkorrekturdarstellung aus den Modellparametern der VAR-Darstellung wie folgt berechnen. Zur Herleitung sei noch einmal auf den Anhang C.2.2 (S.345) verwiesen.

$$\begin{aligned} \Pi &= -\mathbb{I}_K + A_1 + \dots + A_p \\ \Gamma_i &= -A_{i+1} - \dots - A_p \\ \Gamma_{p-1} &= -A_p \end{aligned}$$

Umgekehrt kann man natürlich aus einer geschätzten Fehlerkorrekturdarstellung auf die VAR-Darstellung schließen. Dazu sind folgende Umrechnungen notwendig:

$$\begin{aligned} A_1 &= \Pi + \mathbb{I}_K + \Gamma_1 \\ A_i &= \Gamma_i - \Gamma_{i-1} \\ A_p &= -\Gamma_{p-1}. \end{aligned}$$

Die enge Verbundenheit dieser beiden Darstellungsformen drückt sich in der Aussage 4 des Granger Repräsentationstheorems dadurch aus, dass die Lagpolynommatrix als $A^*(L)$ bezeichnet wird. Wir werden sie jedoch, wie in gängiger Literatur üblich, mit $\Gamma(L)$ (Neusser, 2006; Lütkepohl, 2005) bezeichnen. Die Darstellung des Fehlerkorrekturmodells, wie es Engle und Granger (1987) in Aussage 3 erwähnen, leitet sich wie folgt aus der obigen Darstellung ab

$$\begin{aligned} \Delta Y_t &= c + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t \\ \Leftrightarrow \Gamma(L) \Delta Y_t &= c + \Pi Y_{t-1} + \epsilon_t \\ &= c + \alpha \beta' Y_{t-1} + \epsilon_t \\ &= c + \alpha Z_{t-1} + \epsilon_t. \end{aligned}$$

Im nächsten Kapitel werden wir das Fehlerkorrekturmodell genau studieren. Hier seien zunächst nur verschiedene Darstellungsformen erwähnt.

$$\begin{aligned} \Delta Y_t &= c + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t \\ \Delta Y_t &= c + \alpha \beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t \\ \Delta Y_t &= c + \alpha Z_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t \end{aligned}$$

Durch Umstrukturierung der Parameter ist es möglich, den Fehlerkorrekturterm an das p -te Lag zu verschieben (z.B. Hylleberg & Mizon, 1989; Lütkepohl, 2005, S.249)

$$\Delta Y_t = c + D_1 \Delta Y_{t-1} + \dots + D_{p-1} \Delta Y_{t-p+1} + \Pi Y_{t-p} + \epsilon_t,$$

wobei die Parametermatrizen D_i wie folgt definiert sind

$$D_i = -(-\mathbb{I}_K - A_1 - \dots - A_i), \quad i = 1, \dots, p-1.$$

Dadurch wird deutlich, wie lange die Fehlerkorrektur einen Einfluss auf den Prozess hat. Mit Hilfe dieser Darstellung ist es Johansen (1988) gelungen einen Maximum-Likelihood-Schätzer für den Kointegrationsraum und damit insbesondere für den Kointegrationsrang r zu finden. Dieses Schätz- und Testverfahren wird im Kapitel 2.7 (ab S.124) behandelt.

Tabelle 2.1: Übersicht: Darstellungen eines kointegrierten Systems

MA-Darstellung

$$\Delta Y_t = \mu + \Psi(L)\epsilon_t$$

Common-Trend-Darstellung

$$Y_t = Y_0 + \Psi(1)\mu t + \Psi(1) \sum_{j=1}^t \epsilon_t + \Psi^*(L)\epsilon_0 - \Psi^*(L)\epsilon_t$$

VAR-Darstellung

$$A(L)Y_t = c + \epsilon_t$$

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t$$

Fehlerkorrekturdarstellungen

$$\Delta Y_t = c + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

$$\Delta Y_t = c + \alpha \beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

$$\Delta Y_t = c + \alpha Z_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + \epsilon_t$$

$$\Delta Y_t = c + D_1 \Delta Y_{t-1} + \dots + D_{p-1} \Delta Y_{t-p+1} + \Pi Y_{t-p} + \epsilon_t$$

2.3.4 Weitere Darstellungen eines kointegrierten Systems

Bislang haben wir nur die Darstellungen kointegrierter Prozesse behandelt, die sich im Repräsentationstheorem wiederfinden. Sie sind in Tabelle 2.1 in einer Übersicht zusammengestellt. Darüber hinaus gibt es noch einige weitere Darstellungen, die vor allem von mathematischem Interesse sind, hier jedoch der Vollständigkeit wegen kurz erwähnt werden sollen.

Zum einen wird unsere Common-Trend-Darstellung in mancher Literatur (z.B. bei Neusser, 2006) als Beveridge-Nelson-Zerlegung bezeichnet, aus der sich die Common-Trend-Darstellung dadurch ableitet, dass die Matrizenmultiplikationen ausgeführt und damit das Systems auf die entscheidenden Trends gekürzt wird. Aus dieser gekürzten Common-Trend-Darstellung lässt sich wiederum eine trianguläre Darstellung (z.B. Neusser, 2006; Lütkepohl, 2005; Maddala & Kim, 2007)

$$\begin{aligned} Y_t^{(1)} &= -\beta'_{(K-r)} Y_t^{(2)} + u_t^{(1)} \\ \Delta Y_t^{(2)} &= u_t^{(2)} \end{aligned}$$

herleiten, in der der Prozess Y_t in zwei Subprozesse $Y_t^{(1)}$ und $Y_t^{(2)}$ geteilt wird, so dass Kointegration nur noch im ersten Teilsystem stattfindet. Diese Darstellung ist eine weitere Fehlerkorrekturdarstellung, die sich von der autoregressiven Fehlerkorrekturdarstellung, die wir im folgenden Kapitel ausführlicher behandeln werden, dadurch unterscheidet, dass die kurzfristige Dynamik als Abweichung vom langfristigen Gleichgewicht $Y_t^{(1)} = -\beta'_{(K-r)} Y_t^{(2)}$ im Fehlerprozess u_t zu finden ist (P. Phillips, 1991). Dies erleichtert einige Schätzprozeduren. Hylleberg und Mizon (1989) erwähnen außerdem die Bewley-Darstellung, in der die Kointegrationsmatrix β direkt geschätzt werden kann.

Wir werden uns im Folgenden jedoch nur der Fehlerkorrekturdarstellung genauer zuwenden,

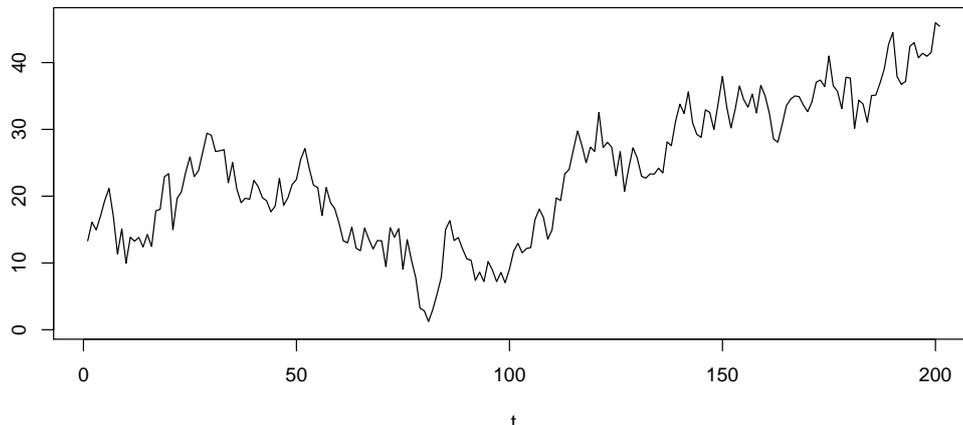


Abbildung 2.9: Random Walk oder Drunkard's Walk

da in ihr sowohl lang- als auch kurzfristige Dynamiken sichtbar werden und ihre Schätzung in attraktiver Software (Lütkepohl & Krätzig, 2004; R Development Core Team, 2008) implementiert ist, so dass sie für ein kointegriertes System die Darstellung der Wahl ist.

2.4 A Drunk and her Dog: Fehlerkorrekturmechanismen am konstruierten Beispiel

Bevor wir uns im nächsten Kapitel der Fehlerkorrekturdarstellung zuwenden, wird hier zunächst ein sehr anschauliches Beispiel eingeführt, das nicht nur das Konzept der Kointegration verständlicher macht, sondern auch zeigt, warum die Fehlerkorrekturdarstellung so nützlich ist. Murray (1994) führte das Beispiel einer Betrunkenen und ihres Hundes ein, das von A. Smith und Harrison (1995) um einen ebenfalls betrunkenen Verehrer erweitert wurde.

2.4.1 Random Walk or Drunkard's Walk

Ein Random Walk ohne Drift ist der einfachste integrierte Prozess. Eine allgemein bekannte Veranschaulichung ist der Gang eines Betrunkenen („Drunkard's Walk“), bei dem jeder Schritt zwar nach vorne, aber auch zufällig nach rechts oder links verläuft (z.B. Gottman, 1981, S.118). Mathematisch wird dieser Gang beschrieben durch den univariaten Prozess

$$Y_t = Y_{t-1} + \epsilon_t.$$

Anschaulich kann man sich vorstellen, dass die Abbildung 2.9 den Gang eines Betrunkenen aus der Vogelperspektive zeigt. Um uns an die Beispiele zu halten (Murray, 1994; A. Smith & Harrison, 1995) werden wir im Folgenden ebenfalls von einer betrunkenen Frau sprechen, die

wir Laura taufen wollen. Die Position der Betrunkenen zum Zeitpunkt t , durch L_t beschrieben, entsteht dadurch, dass man zu der Position am vorigen Zeitpunkt L_{t-1} eine zufällige Größe ϵ_t addiert. Im einfachsten Fall nimmt ϵ_t mit gleicher Wahrscheinlichkeit entweder den Wert $+1$ bzw. -1 , die betrunkene Laura geht also entweder einen Schritt nach rechts oder einen Schritt nach links. Dabei würde jedoch vorausgesetzt, dass alle ihre Schritte gleich groß sind. Davon kann man im Allgemeinen bei einer Betrunkenen jedoch nicht ausgehen. Die Schrittweite variiert folglich, dabei bleibt jedoch die Wahrscheinlichkeit gleich, ob der nächste Schritt nach rechts oder links geht. Da Laura jedoch eine maximale Schrittweite hat, ist die Varianz der Schrittgröße eine endliche Zahl. Man kann also annehmen, dass ihre Schrittgröße symmetrisch um Null mit konstanter Varianz verteilt ist, z.B. normalverteilt. Als Zeitreihe betrachtet ist die Schrittgröße also stationär. Folglich gilt, dass die Differenz der Position von einem Zeitpunkt zum nächsten einen stationären Prozess bildet

$$\Delta L_t = L_t - L_{t-1} = \epsilon_t \sim I(0).$$

Deshalb ist die beste Vorhersage für die Position einer Betrunkenen zu einem späteren Zeitpunkt die letzte beobachtete Position. Außerdem ist die Abweichung von der zuletzt beobachteten Position und der jetzigen umso größer, je länger die letzte Betrachtung zurückliegt. Murray (1994) macht dies anschaulich mit der Situation, dass er die Betrunkene zuletzt gesehen hat, als er mit seinem Freund ein Lokal betrat. Beim Verlassen fragt ihn sein Freund „Wo ist die Betrunkene, die wir vorhin gesehen haben?“. Die beste Auskunft, die er geben kann, ist „Vorhin war sie noch da drüben.“. Da der Gang der Betrunkenen jedoch instationär ist, muss er hinzufügen „Aber weiß der Himmel, wo sie jetzt ist.“

Alles, was hier über einen Betrunkenen gesagt wurde, gilt in ähnlicher Weise für einen Welpen, der ohne Leine frei herumläuft (Murray, 1994). Getrieben von seinen Instinkten bestimmt er jeden Moment aufs Neue, wohin er jetzt geht. Das leiseste Geräusch oder der feinste Geruch kann ihn auf eine neue Fährte bringen. So folgt also auch der Welpen einem Random Walk ohne Drift. Da er kürzere Beine hat, folgt die stationäre Zeitreihe seiner Schritte einer Verteilung mit Erwartungswert Null aber anderer Varianz. Diesen Prozess bezeichnen wir hier mit η_t , den Gang des Welpen mit O_t . Folglich soll gelten

$$\Delta O_t = O_t - O_{t-1} = \eta_t.$$

2.4.2 A Drunk and her Dog

Was ist aber, wenn der Welpen einer Betrunkenen gehört? Jeder für sich folgt einem integrierten Prozess der Ordnung $I(1)$. Da die Betrunkene ihren Hund jedoch nicht verlieren will, ruft sie in regelmäßig: „Oliver (Murrays Name für den Hund), wo bist du?“. Oliver hat gelernt auf diesen Ruf hin zu bellen. So hört er Laura und sie hört auch ihn. Beide wollen einander nicht verlieren, wünschen sich also an möglichst derselben Stelle zu sein. Das lang-

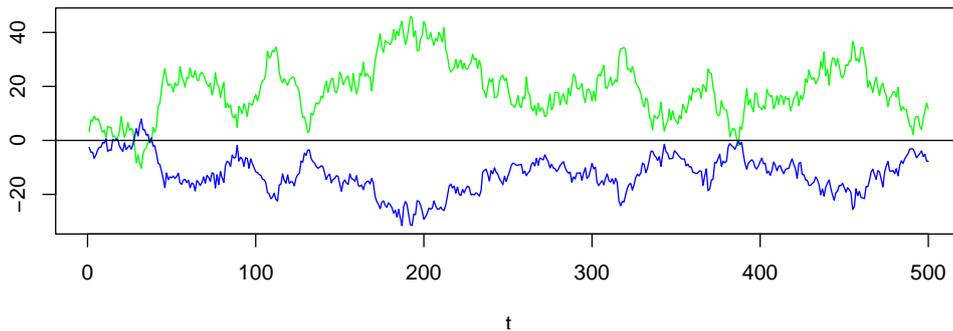


Abbildung 2.10: Bivariates kointegriertes System mit positivem β : Zirkusartisten auf einer Wippe

fristige (erwünschte) Gleichgewicht lautet also $L_t = O_t$ bzw. $L_t - O_t = 0$. Deshalb versuchen sie mit ihren nächsten Schritten den Abstand zueinander zu verkleinern. Somit folgen beide nicht mehr einem Random Walk, sondern haben ihrem Gang einen Fehlerkorrekturprozess hinzugefügt. Beide irren immer noch ziellos durch die Gegend, aber diesmal *gemeinsam*. Der Abstand zwischen den beiden (ist eine Zufallsvariable und) verändert sich mit der Zeit, bleibt aber immer ähnlich groß. Im besten Fall ist er stationär, dann sind die beiden Prozesse L_t und O_t kointegriert der Ordnung $CI(1, 1)$. Die beiden folgende nun

$$\begin{aligned}\Delta L_t &= \epsilon_t + c(L_{t-1} - O_{t-1}) \\ \Delta O_t &= \eta_t + d(L_{t-1} - O_{t-1}).\end{aligned}$$

Man sieht, dass beide Reihen vom Abstand zwischen der betrunkenen Laura und ihrem Hund abhängen. Je nach der Größe des Abstandes $L_{t-1} - O_{t-1}$ wird der Schritt zum nächsten Zeitpunkt t angepasst. Wie stark diese Anpassung ist, beschreiben die Parameter c und d . Da wir annehmen, dass ΔL_t und ΔO_t stationär sind, muss der Prozess, der den Abstand $Z_t = L_t - O_t$ beschreibt, stationär sein. Dies gilt natürlich nur, wenn die beiden zusammen gehören. Ein fremder Hund geht seines Weges, auch wenn er die Betrunkene rufen hört, und entfernt sich mit der Zeit immer weiter von ihr. Erst die Fehlerkorrektur der beiden bewirkt ihr gemeinsames Wandern.

In diesem Beispiel haben die Betrunkene und ihr Hund denselben Trend mit derselben Gewichtung. In der Kointegrationsbeziehung $L_t - O_t$ ist $\beta = 1$. Um den Fall ungleicher Gewichte zu veranschaulichen, konstruiert Murray (1994) ein Beispiel, in dem die Frau und der Hund Zirkusartisten sind und eine Wippe in Balance halten müssen, was sie dazu zwingt, trotz beider Ziellosigkeit sich in entgegengesetzte Richtungen zu bewegen. Da der Hund viel leichter ist als seine Besitzerin, muss er viel weiter von der Mittellinie weglaufen, um die Balance zu halten (vgl. Abbildung 2.10).

Während diese Erweiterung sehr konstruiert wirkt, ist die Fortsetzung der Geschichte, die

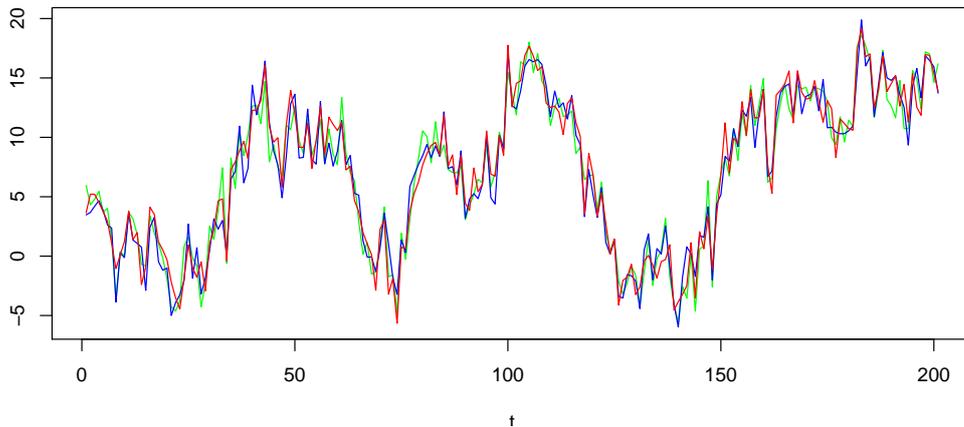


Abbildung 2.11: Ein kointegriertes System: Laura (rot) mit Welpen Oliver (grün) und Boyfriend Kinley (blau)

A. Smith und Harrison (1995) ergänzt haben, wieder sehr anschaulich.

2.4.3 A Drunk, her Dog and a Boyfriend

A. Smith und Harrison (1995) führen in das System der Betrunknen und ihres Hundes noch einen betrunkenen Verehrer (*boyfriend*) ein, der wie der Hund bei seiner Freundin bleiben will. Dadurch wird ein komplexeres Modell veranschaulicht, in dem verschiedene Korrekturprozesse auftreten können.

Bereits bei der Betrachtung der Betrunknen und ihres Hundes machen A. Smith und Harrison (1995) eine etwas andere Annahme als Murray (1994), die zu einer leichten Vereinfachung führt. Wir werden beide Modelle im nächsten Kapitel genauer betrachten und konzentrieren uns hier auf die Anschauung.

Ausgangspunkt für A. Smith und Harrison (1995) ist der Gang der Betrunknen

$$\Delta L_t = \epsilon_t,$$

der einem einfachen Random Walk ohne Fehlerkorrektur folgt. Dies ist der Unterschied zu Murray (1994). Hier passt sich nur der Hund Oliver dem Weg seines Frauchens an und nicht beide

$$\Delta O_t = \eta_t + d(L_{t-1} - O_{t-1}).$$

Der gemeinsame Trend wird hier also ganz klar von Laura vorgegeben, die Anpassung geschieht nur durch den Hund. Bei Murray ergibt sich der gemeinsame Trend durch die gegenseitige Anpassung, ohne dass einer der beiden ihn eindeutig vorgibt. In diesem Fall misst der

Parameter d also, wie gut Oliver erzogen ist und wie schnell er wieder zu seinem Frauchen zurückkommt. Je größer d ist, desto stärker ist der Fehlerausgleich, also der Gehorsam des Hundes.

Nun kommt noch der Verehrer oder Freund der Betrunkenen, Kinley, zu den beiden hinzu. Kinley ist selbst betrunken, aber sehr daran interessiert in der Nähe der betrunkenen Laura zu bleiben. Dabei ist ihm der Hund zunächst völlig egal. Da Kinley nicht die Hand der Betrunkenen hält, kann sich sein Abstand zu ihr theoretisch beliebig vergrößern. Gleiches gilt ja auch für den nicht angeleiteten Oliver. Kinleys Gang kann also durch folgenden Prozess beschrieben werden

$$\Delta K_t = K_t - K_{t-1} = \alpha_K(K_{t-1} - L_{t-1}) + \delta_t.$$

Kinleys nächster Schritt ist nicht länger nur zufällig, sondern wird von der Fehlerkorrektur des Abstandes zu Laura beeinflusst. In diesem Beispiel kann man Kinley als (vielleicht gar lästigen) Verehrer sehen, denn die betrunkene Laura setzt ihren Weg unbeirrt fort.

Wenn es sich bei Kinley jedoch um ihren Freund handelt, hat Laura auch ein Interesse in seiner Nähe zu bleiben. Ihr Gang wird also ebenfalls davon beeinflusst, wie weit sie sich von Kinley entfernt hat. Sie folgt also dem Prozess

$$\Delta L_t = L_t - L_{t-1} = \epsilon_t + \alpha_L(K_{t-1} - L_{t-1}).$$

Wenn sie nun gleichzeitig auch daran interessiert ist ihren Hund Oliver nicht zu verlieren, wird sie von zwei (linear) unabhängigen Kointegrationsbeziehungen beeinflusst. Sie wünscht sich gleichzeitig dort zu sein, wo Kinley ist, will aber auch Oliver nicht verlieren.

$$\Delta L_t = L_t - L_{t-1} = \epsilon_t + \alpha_{L1}(K_{t-1} - L_{t-1}) + \alpha_{L2}(L_{t-1} - O_{t-1}).$$

Je nach Gewichtung der Abstandsprozesse a und c , will sie lieber bei ihrem Freund Kinley $a > c$ oder lieber bei ihrem Hund $a < c$ bleiben.

Weiterhin kann man zusätzlich die Beziehung von Kinley zu Oliver betrachten. Hilft er seiner Freundin dabei, auf den Hund aufzupassen, bleibt er möglichst nahe bei Oliver

$$\Delta K_t = K_t - K_{t-1} = \alpha_{K1}(K_{t-1} - L_{t-1}) + \alpha_{K2}(K_{t-1} - O_{t-1})\delta_t.$$

Eine Anziehung drückt sich hier stets in negativem Vorzeichen aus. Kann Kinley den Hund nicht leiden und ist bemüht den Abstand zu ihm eher zu vergrößern, so zeigt sich dies in positivem α .

Ist der Koeffizient α vom Betrag her größer als Eins, kann dies ein Abdriften zur Folge haben. Kointegration kann es dann nur geben, wenn die Anziehungskraft der anderen Reihe stark genug ist (A. Smith & Harrison, 1995), dies auszugleichen.

Nun bleibt jedoch zu beachten, dass das Verhältnis von Kinley und Oliver $K_t - O_t$ nicht (linear) unabhängig von dem Verhältnis der beiden zur Betrunkenen ist. Mathematisch kann man $K_t - O_t$ als Linearkombination von $L_t - O_t$ und $K_t - L_t$ darstellen. In einem kointegrierten System von drei Reihen, kann es maximal zwei unabhängige Kointegrationsbeziehungen geben (siehe Abschnitt 2.2.3, S.75).

Wir haben an diesem anekdotischem Beispiel gesehen, wie Kointegration entstehen kann, und haben ein System erstellt, dass durch Fehlerkorrekturprozesse bestimmt war. Im folgenden Kapitel werden wir nun sehen, was ein allgemeines Fehlerkorrekturmodell leisten kann. Dabei werden wir uns an geeigneter Stelle an die eben beschriebene Anschauung erinnern.

2.5 Fehlerkorrekturmodell: langfristiges Gleichgewicht und kurzfristige Dynamik

Das Fehlerkorrekturmodell vereint die autoregressive Struktur mit der Abhängigkeit der Einzelprozesse von den Kointegrationsbeziehungen. Die Darstellung gilt allgemein für kointegrierte Prozesse beliebiger Ordnung $I(d)$. Wir wollen uns jedoch wie Lütkepohl (2005) auf den für die Anwendung zentralen Fall eines differenzenstationären K -dimensionalen Prozesses $Y_t \sim I(1)$ konzentrieren, so dass bei Kointegration der r -dimensionale Abweichungsprozess Z_t stationär ist.

Definition 2.5.1 (Fehlerkorrekturmodell) *Sei Y_t ein K -dimensionaler Prozess, der integriert vom Grad Eins ist, $Y_t \sim I(1)$. Für die erste Differenz gibt es ein Fehlerkorrekturmodell mit folgenden Parametern:*

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-(p-1)} + \epsilon_t \quad (2.13)$$

- β : $(K \times r)$ -Matrix, Kointegrationsmatrix
- α : $(K \times r)$ -Matrix, Ladungsmatrix
- $\Gamma_1, \dots, \Gamma_{p-1}$: $(K \times K)$ -Matrizen, Gewichte der Lags
- p : Anzahl der Lags
- $\Pi := \alpha \beta'$: $(K \times r)$ -Matrix vom Rang r
- r : Kointegrationsrang

Außerdem können deterministische Terme in das Fehlerkorrekturmodell aufgenommen werden, was wir in Kapitel 2.6 untersuchen werden. Dazu gehört auch das Hinzunehmen eines konstanten Vektors c , der Mittelwertsunterschiede in den Einzelreihen repräsentiert.

- c : konstanter K -dimensionaler Vektor

Wir wollen noch einmal erwähnen, wie man den Kointegrationsrang des Systems aus der Fehlerkorrekturdarstellung ablesen kann, da diese Größe eine entscheidende Rolle spielt (Lütkepohl, 2005, S.248).

Definition 2.5.2 (Kointegriert vom Rang r) *Ein K -dimensionaler VAR(p)-Prozess*

$$Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t$$

heißt kointegriert vom Rang r , falls die Matrix

$$\Pi := -(\mathbb{I}_K - A_1 - \dots - A_p)$$

den Rang r hat.

Die Matrizen Γ_i in der Darstellung (2.13) gewichten die Lags des Differenzenprozesses ΔY_t . Dadurch wird die autoregressive Struktur des Systems aufrecht erhalten. Zusätzlich wird die Veränderung ΔY_{it} der einzelnen Prozesse auch durch die Kointegrationsbeziehung(en) beeinflusst. Dabei beschreibt die (normierte) Matrix β die langfristigen Kointegrationsbeziehungen, während die Matrix α den kurzfristigen Einfluss einer Abweichung von diesem Gleichgewicht gewichtet. Die Fehlerkorrektur findet also durch den Einfluss von α statt.

2.5.1 Der allgemeine bivariate Fall

Wir widmen uns zunächst dem bivariaten Modell, in dem zwei Zeitreihen Y_{1t} und Y_{2t} gemeinsam betrachtet werden und das im allgemeinsten Fall wie folgt dargestellt wird

$$\begin{pmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & \beta \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \gamma_{1,11} & \gamma_{1,12} \\ \gamma_{1,21} & \gamma_{1,22} \end{pmatrix} \begin{pmatrix} \Delta Y_{1,t-1} \\ \Delta Y_{2,t-1} \end{pmatrix} \\ + \dots + \begin{pmatrix} \gamma_{(p-1),11} & \gamma_{(p-1),12} \\ \gamma_{(p-1),21} & \gamma_{(p-1),22} \end{pmatrix} \begin{pmatrix} \Delta Y_{1,t-(p-1)} \\ \Delta Y_{2,t-(p-1)} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Wir konzentrieren uns jedoch auf den Fall ohne konstanten Vektor $c = 0$ und lassen auch die autoregressiven Einflüsse der Lags zunächst außer Acht. Das Modell vereinfacht sich also zu

$$\begin{pmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 & \beta \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

oder ausformuliert in der System-Schreibweise mit Einzelprozessen

$$\begin{aligned} \Delta Y_{1t} &= \alpha_1 (Y_{1,t-1} + \beta Y_{2,t-1}) + \epsilon_{1t} \\ \Delta Y_{2t} &= \alpha_2 (Y_{1,t-1} + \beta Y_{2,t-1}) + \epsilon_{2t}. \end{aligned}$$

In dieser kurzen Form wird die Abhängigkeit der Prozesse von der Kointegrationsbeziehung

$Z_{t-1} = Y_{1,t-1} + \beta Y_{2,t-1}$ besonders deutlich.

Die Veränderung beider Variablen $\Delta Y_{it}, i = 1, 2$ hängt von der Abweichung von der Gleichgewichtsbeziehung ab (Lütkepohl, 2005, S.248). Der Prozess Y_t befindet sich im Gleichgewicht, wenn $Z_t = Y_{1t} + \beta Y_{2t} = 0$ gilt (Engle & Granger, 1987). Meistens befindet sich der Prozess jedoch nicht exakt im Gleichgewicht, so dass man $Z_t = Y_{1t} + \beta Y_{2t}$ als *Gleichgewichtsfehler* bezeichnen kann (Engle & Granger, 1987), der die Abweichung von diesem Gleichgewicht beschreibt. Man kann die Gleichgewichtsbeziehung auch als Gleichung zwischen den Einzelprozessen auffassen

$$\begin{aligned} Y_{1t} + \beta Y_{2t} &= 0 \\ \Leftrightarrow Y_{1t} &= -\beta Y_{2t}. \end{aligned}$$

Der Parameter β beschreibt hier eine Skalierung, durch die die beiden Prozesse ins Gleichgewicht gesetzt werden. Wenn die beiden Prozesse gleichmäßig gemeinsam fallen oder steigen, bleibt diese Beziehung erhalten. Muss nicht nur die Skala, auf der die Prozesse gemessen wurden, sondern auch ihr Mittelwert berücksichtigt werden, kann dies durch die Aufnahme einer ausgleichenden Konstante in die Kointegrationsbeziehung zum Ausdruck gebracht werden. Die Kointegrationsbeziehung lautet dann $Y_{1t} + \beta Y_{2t} + \mu_0$, wobei μ_0 den Mittelwertsunterschied zwischen Y_{1t} und Y_{2t} beschreibt. Dieser Fall wird im Kapitel 2.6 ausführlich dargestellt.

Wenn sich die beiden Prozesse im nächsten zeitlichen Schritt (leicht) unterschiedlich entwickeln, wird die Gleichgewichtsbeziehung verletzt. Diese Abweichung kann positiv oder negativ sein. Im Falle einer negativen Abweichung gilt

$$Z_t < 0 \Leftrightarrow Y_{1t} + \beta Y_{2t} < 0 \Leftrightarrow Y_{1t} < -\beta Y_{2t}.$$

Im einfachsten Fall kommt dies zustande, da sich nur ein Prozess verändert hat: Y_{1t} ist gefallen oder $-\beta Y_{2t}$ ist gestiegen. Analoges gilt mit umgekehrtem Vorzeichen bei positiver Abweichung

$$Z_t > 0 \Leftrightarrow Y_{1t} > -\beta Y_{2t}.$$

Durch den Parameter α_i wird der Einfluss des Gleichgewichtsfehlers Z_t auf die Veränderung in der Variable ΔY_{it} gewichtet. Dabei beschreibt der Betrag von α_i die Stärke dieses Einflusses. Ist $\alpha_i = 0$, dann ist die Veränderung der Variable nur vom Zufall bedingt

$$\Delta Y_{it} = \epsilon_{it}.$$

Die integrierte Reihe Y_{it} folgt also einem Random Walk $Y_{it} = Y_{i,t-1} + \epsilon_t$. Man denke an die betrunkene Laura, die unbeeindruckt von Kinleys Annäherungsversuchen oder dem Herumstreunen ihres Hundes ihren Weg geht und sich von keinem der beiden beeinflussen lässt (siehe Kapitel 2.4).

Tabelle 2.2: Auswirkung des Ladungsparameters α_i und des Gleichgewichtsprozesses Z_t auf das Verhalten des Prozesses Y_{it} zum nächsten Zeitpunkt

	$\alpha_i > 0$	$\alpha_i < 0$
$Z_{t-1} = 0$	$\Delta Y_{it} = 0$ keine Veränderung	$\Delta Y_{it} = 0$ keine Veränderung
$Z_{t-1} < 0$	$\Delta Y_{it} < 0$ Y_{it} fällt	$\Delta Y_{it} > 0$ Y_{it} steigt
$Z_{t-1} > 0$	$\Delta Y_{it} > 0$ Y_{it} steigt	$\Delta Y_{it} < 0$ Y_{it} fällt

2.5.2 Fehlerkorrektur im bivariaten System

Es ist interessant zu betrachten, welche Auswirkung das Vorzeichen von α_i auf den Prozessverlauf hat. Dabei konzentrieren wir uns nur auf die systematischen Einflüsse und blenden die zufälligen Veränderungen ϵ_{it} für die Untersuchung vorerst aus

$$\Delta Y_{it} = \alpha_i Z_{t-1}.$$

Wenn die Gleichgewichtsbeziehung $Z_{t-1} = 0$ erfüllt ist, gibt es keine systematische Veränderung $\Delta Y_{it} = 0$. Bei positivem α_i verhält sich die Veränderung (Steigung) des Prozesses Y_{it} in dieselbe Richtung wie die Abweichung vom Gleichgewicht. Bei negativem Parameter läuft sie entgegengesetzt.

Wir werden nun verschiedene Kombinationen von α_1 und α_2 betrachten, um zu sehen, in welchen Fällen in einem bivariaten System Fehlerkorrektur möglich ist. Dazu benutzen wir neben der Anekdote der Betrunknen und ihres Hundes die Ergebnisse einer empirischen Zeitschätzungsstudie (Stroe-Kunold et al., 2010) zur Veranschaulichung.

Zeitschätzungsstudie

In einem Zeitschätzungsexperiment (Stroe-Kunold et al., 2010) erhielten sechs Studenten die Aufgabe, nach Auftauchen eines Stimulus auf einem Bildschirm exakt eine Sekunde (= 1000ms) lang zu warten und nach dieser geschätzten Zeitperiode eine Taste zu betätigen. Dies sollten sie nach einer kurzen Übungsphase 512 mal wiederholen, so dass eine Zeitreihe entstand, die für viele Auswertungen eine geeignete Länge besitzt³².

Nach jeder Zeitschätzung erhielt der Proband ein Feedback über seinen Schätzwert in der Angabe von Millisekunden. Das Experiment bestand aus zwei within-subject Bedingungen,

³²Um eine Spektralanalyse zu ermöglichen, wurde für die Anzahl der Wiederholungen eine Potenz von 2 gewählt, $512 = 2^9$.

so dass für jede Versuchsperson zwei Zeitreihen mit 512 Messpunkten entstanden. In der Kontrollbedingung erhielt die Person ein korrektes Feedback, in der Experimentalbedingung wurde die echte Reaktionszeit $(RT)_t$ mit einer simulierten integrierten Reihe mit Erwartungswert Null $(SIM)_t$ verrechnet, so dass als Feedback $(F)_t = (RT)_t - (SIM)_t$ ausgegeben wurde.

Ziel des Experiments war es zu prüfen, ob eine solche Zeitschätzung von einem falschen Feedback beeinflusst wird oder ob der Mensch über eine „innere Uhr“ verfügt, die unabhängig vom Feedback funktioniert.

Statistische Auswertungen mit Unit Root Tests ergaben³³, dass in der Kontrollbedingung eine stationäre Zeitreihe um den Mittelwert $1000ms$ entstand. Dies war zu erwarten, weil durch das korrekte Feedback den Probanden die Möglichkeit gegeben wird, ihr Reaktionsverhalten korrekt anzupassen. Natürlich ist es fast nicht möglich $1000ms$ genau zu treffen, so dass die Werte um diesen Wert zufällig streuen.

Die Zeitschätzung $(RT)_t$ in der Versuchsbedingung mit manipuliertem Feedback zeigt jedoch einen nicht-stationären Verlauf. Der Johansen Trace Test, der in Kapitel 2.7 dieser Arbeit behandelt wird, ergab, dass eine Kointegrationsbeziehung zwischen der Reaktionszeit $(RT)_t$ und der für das falsche Feedback verrechneten Reihe $(SIM)_t$ vorliegt

$$\begin{aligned} (RT)_t &= 1.003(SIM)_t + 958.861 \\ \Leftrightarrow Z_t &= (RT)_t \underbrace{-1.003}_{\beta}(SIM)_t \underbrace{-958.861}_{\mu_0}. \end{aligned}$$

Die beiden Reihen bewegen sich also auf derselben Skala $\beta \approx -1$, sind allerdings um den Mittelwert $\mu_0 \approx 1000ms$ versetzt. Dieser Mittelwertsunterschied erklärt sich aus der Aufgabenstellung von selbst und ist für unsere weiteren Betrachtungen nicht von entscheidender Bedeutung.

Das bivariate Fehlerkorrekturmodell wurde mittels des Johansen tests auf

$$\begin{aligned} \begin{pmatrix} \Delta(RT)_t \\ \Delta(SIM)_t \end{pmatrix} &= \begin{pmatrix} -0.769^{***} \\ 0.032 \end{pmatrix} \left(\begin{pmatrix} 1 & -1.003^{***} \end{pmatrix} \begin{pmatrix} RT_{t-1} \\ SIM_{t-1} \end{pmatrix} - 958.861^{***} \right) \\ &+ \begin{pmatrix} -0.143^{***} & -0.151 \\ -0.024 & 0.156^{***} \end{pmatrix} \begin{pmatrix} \Delta(RT)_{t-1} \\ \Delta(SIM)_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \end{aligned}$$

geschätzt, wobei $***$ ein Signifikanzniveau von $p < 0.001$ bedeutet. Berücksichtigt man nur die signifikanten Werte erhält man in Systemschreibweise folgende Darstellung

$$\begin{aligned} \Delta(RT)_t &= -0.769 \left((RT)_{t-1} - 1.003(SIM)_{t-1} - 958.861 \right) - 0.143 (RT)_{t-1} + \epsilon_{1t} \\ \Delta(SIM)_t &= 0 \left((RT)_{t-1} - 1.003(SIM)_{t-1} - 958.861 \right) + 0.156 (SIM)_{t-1} + \epsilon_{2t}. \end{aligned}$$

³³Hier werden die Ergebnisse einer zufällig ausgewählten Versuchsperson wiedergegeben, da sich die Ergebnisse der sechs Probanden kaum unterscheiden.

In diesem Fall ist also $\alpha_1 < 0$ und $\alpha_2 = 0$. Dies ist ein Spezialfall der ersten Parameterkonstellation, die wir systematisch untersuchen wollen.

$$\alpha_1 \neq 0, \alpha_2 = 0$$

Wir untersuchen zunächst nur den Einfluss von α_1 , betrachten also das vereinfachte Modell

$$\begin{aligned}\Delta Y_{1t} &= \alpha_1 (Y_{1t} + \beta Y_{2t}) + \epsilon_{1t} \\ \Delta Y_{2t} &= \epsilon_{2t}.\end{aligned}$$

Hier ist die Veränderung von Y_{2t} rein zufällig, während der Verlauf von Y_{1t} durch die Abweichung vom Gleichgewicht $Z_t = Y_{1t} + \beta Y_{2t}$ beeinflusst wird. In diesem Fall kann eine Fehlerkorrektur nur durch den Prozess Y_{1t} geschehen, da Y_{2t} keinen systematischen Veränderungen unterliegt. Dies ist z.B. in der Zeitschätzungsstudie der Fall. Die simulierte Reihe (hier Y_{2t}) ist bereits vor der Durchführung des Experiments festgelegt und kann von der Zeitreihe der Zeitschätzung nicht beeinflusst werden, während ein Einfluss der simulierten Feedbackreihe auf der Reaktion der Versuchsperson möglich ist und wie die Studie zeigt auch stattfindet.

Bemerkung 2.5.3 (Auswirkung $\alpha_1 > 0$) Bei **positivem** α_1 kommt es zu einer Verstärkung der Abweichung vom Gleichgewicht. Der Prozess Y_{1t} bewegt sich immer weiter von Y_{2t} weg. Es kommt zur „Explosion“. Dadurch wird das Gleichgewicht nie wieder erreicht. Der Gleichgewichtsfehler Z_t ist ein instationärer Prozess, so dass keine Kointegrationsbeziehung besteht und der Fehler nicht korrigiert wird.

Beweis. Um dies zu sehen, betrachten wir zunächst den Fall, dass es im Zeitpunkt $t - 1$ zu einer (zufälligen) negativem Abweichung vom Gleichgewicht $Z_{t-1} < 0$ kam. Wie oben beschrieben gilt in diesem Fall

$$Y_{1,t-1} < -\beta Y_{2,t-1}. \quad (2.14)$$

Bei positivem α_1 verhält sich die Veränderung des Prozesses Y_{1t} genauso wie die Abweichung vom Gleichgewicht (siehe Tabelle 2.2, S.103). In diesem Fall fällt der Prozess, so dass der Wert des Prozesses zum Zeitpunkt t kleiner ist als zum vorigen Zeitpunkt $t - 1$

$$Y_{1,t} < Y_{1,t-1}. \quad (2.15)$$

Da die Abweichung vom Gleichgewicht sich nicht auf $Y_{2,t}$ auswirkt, seine Veränderung rein zufällig ist, kann man annehmen, dass sich sein Wert in diesem Zeitschritt kaum verändert hat

$$Y_{2,t} \approx Y_{2,t-1}. \quad (2.16)$$

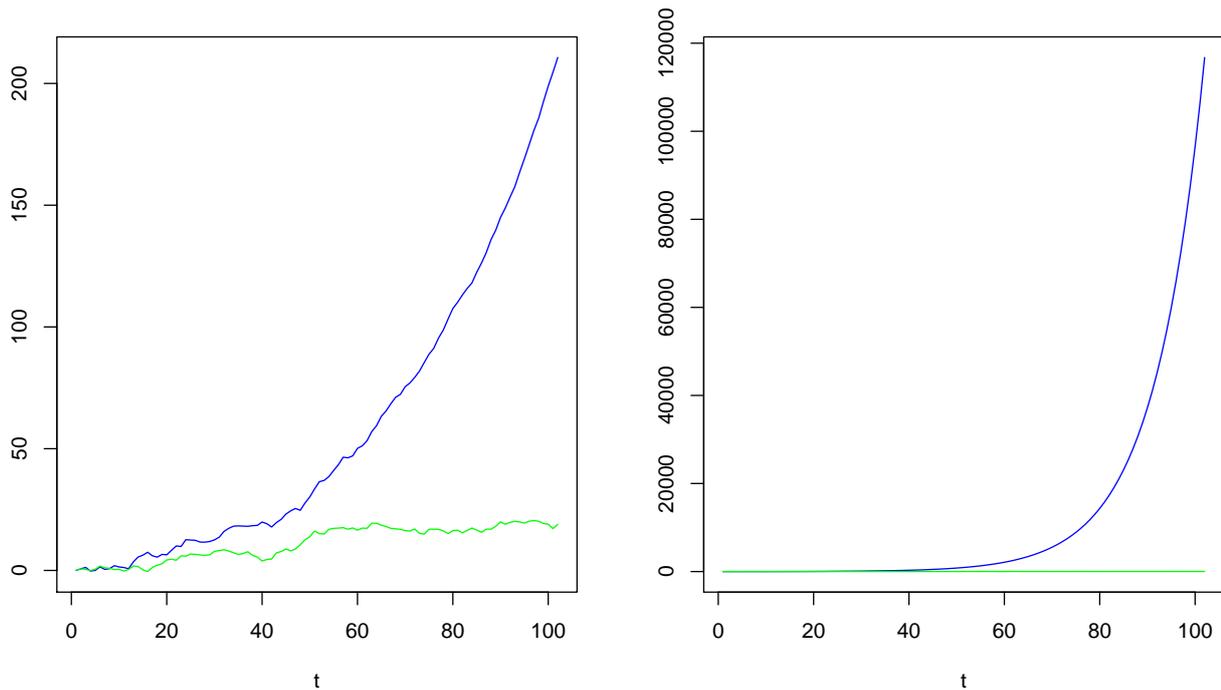


Abbildung 2.12: Auseinanderdriften bei positivem α_1 ; links $\alpha_1 = 0.03$, rechts $\alpha_1 = 0.1$

Setzt man diese Überlegungen zusammen, erhält man folgende Ungleichungskette

$$Y_{1,t} \stackrel{(2.15)}{<} Y_{1,t-1} \stackrel{(2.14)}{<} -\beta Y_{2,t-1} \stackrel{(2.16)}{\approx} -\beta Y_{2,t}.$$

Zum Zeitpunkt t hat sich die negative Abweichung vom Gleichgewicht also verstärkt

$$\begin{aligned} Y_{1,t} &< -\beta Y_{2,t} \\ \Leftrightarrow Z_t = Y_{1t} + \beta Y_{2t} &< 0. \end{aligned}$$

Zum nächsten Zeitpunkt t liegt also dieselbe Situation vor wie am Ausgangspunkt $t - 1$. Genau genommen hat sich die negative Abweichung sogar noch verstärkt $Z_t < Z_{t-1}$. Die beiden Prozesse sind noch weiter von ihrem Gleichgewicht entfernt. So beginnt ein Kreislauf, der dazu führt, dass Y_{1t} ins negativ Unendliche abdriftet.

$$Y_{1,t-1} \stackrel{(2.15)}{>} Y_{1,t} > Y_{1,t+1} > Y_{1,t+2} > Y_{1,t+3} > \dots \rightarrow -\infty$$

Wie schnell der Prozess „explodiert“, ist gut in Abbildung 2.12 zusehen. Im linken Teil ist α_1 als 0.03 gewählt und der Prozess erreicht bereits nach 100 Beobachtungspunkten einen Wert über 200. In der rechten Grafik ist $\alpha_1 = 0.1$. Hier liegen die Werte von Y_t schon nach 100 Zeitpunkten im sechsstelligen Bereich (> 150.000).

Analoges Verhalten zeigt sich bei positiver Abweichung vom Gleichgewicht. In diesem Fall

explodiert Y_{1t} ins positiv Unendliche.

$$\begin{aligned} Z_{t-1} &> 0 \\ \Leftrightarrow Y_{1,t-1} &> -\beta Y_{2,t-1} \end{aligned}$$

$$\begin{aligned} (\alpha_1 > 0, \alpha_2 = 0) \Rightarrow Y_{1,t} > Y_{1,t-1} &> -\beta Y_{2,t-1} \approx -\beta Y_{2,t} \\ \Leftrightarrow Z_t &> 0 \end{aligned}$$

$$\Rightarrow Y_{1,t-1} < Y_{1,t} < Y_{1,t+1} < Y_{1,t+2} < Y_{1,t+3} < \dots \rightarrow \infty$$

□

Während Y_{2t} nur zufälligen Einflüssen unterliegt, also differenzenstationär ist, hat Y_{1t} einen extremen Trend. Da dieser Trend nicht gemeinsamer Bestandteil beider Reihen ist, liegt keine Kointegration vor.

Bemerkung 2.5.4 (Auswirkung bei $\alpha_1 < 0$) Bei **negativem** α_1 wird die Abweichung vom Gleichgewicht Z_t wieder ausgeglichen, es kommt zur (Gleichgewichts-)Fehlerkorrektur. Nur in diesem Fall befinden sich die beiden Prozesse in einer stationären Beziehung, nur so kann es langfristig Fehlerkorrektur geben.

Beweis. Betrachten wir wieder den Fall einer (zufälligen) negativen Abweichung vom Gleichgewicht zum Zeitpunkt $t - 1$

$$Z_{t-1} < 0 \Leftrightarrow Y_{1,t-1} < -\beta Y_{2,t-1}.$$

Der Prozess Y_{2t} unterliegt weiterhin nur zufälligen Veränderungen, während Y_{1t} sich systematisch umgekehrt zu Z_t verhält, also zum nächsten Zeitpunkt einen größeren Wert annimmt. Für die beiden Prozesse gilt also

$$\begin{aligned} Y_{1,t} &> Y_{1,t-1} \\ Y_{2,t} &\approx Y_{2,t-1}. \end{aligned}$$

Nimmt man diese Überlegungen wieder zusammen, erhält man

$$Y_{1,t} \stackrel{(2.17)}{>} Y_{1,t-1} \stackrel{(2.17)}{<} -\beta Y_{2,t-1} \stackrel{(2.17)}{\approx} -\beta Y_{2,t}.$$

Es gibt keine eindeutige Richtung in dieser (Un)gleichungskette, so dass es im Idealfall sofort zu einem Ausgleich kommen kann

$$Y_{1,t} \approx -\beta Y_{2,t} \Leftrightarrow Z_t \approx 0.$$

Die Gleichgewichtsbeziehung ist im Idealfall wiederhergestellt, zumindest haben sich die Prozesse ihrem Gleichgewicht angenähert, der Gleichgewichtsfehler ist kleiner geworden $Z_t < Z_{t-1}$.

Im Fall einer positiven Abweichung vom Gleichgewicht kommt es ebenso zu einer Anpassung des Prozesses Y_{1t} an das Gleichgewicht. Man prüfe dies, indem man obige Überlegungen mit umgekehrten Vorzeichen erneut verwendet. \square

Bei negativem α_1 enthält der Prozess Y_t einen Fehlerkorrekturmechanismus. Dieser kann nur über die Anpassung von Y_{1t} passieren, da Y_{2t} nur zufälligen Einflüssen unterliegt. Um den Prozess der Fehlerkorrektur zu verstehen, gehen wir zunächst davon aus, dass zum Zeitpunkt t ein Ungleichgewicht $Z_t > 0 \Leftrightarrow Y_{1t} > -\beta Y_{2t}$ vorliegt. Der Fehlerkorrekturmechanismus bewirkt nun, dass Y_{1t} zum nächsten Zeitpunkt fällt, wodurch er dem Gleichgewicht wieder näher kommt, während sich Y_{2t} rein zufällig verhält. Nun gibt es im nächsten Zeitpunkt $t + 1$ drei Möglichkeiten

- $Z_{t+1} = 0$: Das Gleichgewicht ist wieder hergestellt. Y_{1t} verändert sich zum nächsten Zeitpunkt nicht systematisch, sondern nur zufällig: $\Delta Y_{1,t+2} = \underbrace{\alpha_1 Z_{t+1}}_{=0} + \epsilon_{t+2} = \epsilon_{t+2}$.
- $Z_{t+1} > 0$: Die Ausgleichsbewegung war noch nicht ausreichend groß genug. Y_{1t} fällt weiter.
- $Z_{t+1} < 0$: Die Ausgleichsbewegung war zu groß. Y_{1t} steigt nun, um das Gleichgewicht wieder zu erreichen.

In jedem Fall ist die Dynamik des Prozesses um Wiederherstellung des Gleichgewichtszustands bemüht. Im Fall $\alpha_2 = 0$ kann ein Ausgleich ja nur durch die Dynamik von Y_{1t} bewirkt werden.

$\alpha_1 = 0, \alpha_2 \neq 0$

Im Fall $\alpha_1 = 0$ verändert sich Y_{1t} rein zufällig, eine ausgleichende Dynamik muss also in Y_{2t} liegen. Das betrachtete Modell

$$\begin{aligned}\Delta Y_{1t} &= \epsilon_{1t} \\ \Delta Y_{2t} &= \alpha_2 (Y_{1t} + \beta Y_{2t}) + \epsilon_{2t}\end{aligned}$$

scheint analog zum oben betrachteten Modell in Gleichung (2.14) auf S.105. Durch die Normierung der Gleichgewichtsbeziehung bzw. des Kointegrationsvektors auf $(1, \beta)$ muss man für die Interpretation von α_2 auch das Vorzeichen von β berücksichtigen. Hier kommt es zur Explosion, wenn $\alpha_2 \cdot \beta > 0$, die Vorzeichen also identisch sind. Im Fall unterschiedlicher Vorzeichen $\alpha_2 \cdot \beta < 0$ findet eine Fehlerkorrektur, also Kointegration statt. Man beachte,

dass man ohne normierte Kointegrationsbeziehung auch in obigem Fall statt α_1 stets $\alpha_1\beta_1$ betrachten muss.

Beweis. Betrachten wir dazu den Fall einer (zufälligen) negativen Abweichung vom Gleichgewicht im Zeitpunkt $t - 1$

$$Z_{t-1} < 0 \Leftrightarrow (Y_{1t} + \beta Y_{2t}) < 0 \Leftrightarrow Y_{1,t-1} < -\beta Y_{2,t-1}$$

Bei **positivem** α_2 fällt der Prozess Y_{2t} zum nächsten Zeitpunkt:

$$\begin{aligned} \Delta Y_{2t} = Y_{2,t} - Y_{2,t-1} &= \underbrace{\alpha_2}_{>0} \underbrace{(Y_{1t} + \beta Y_{2t})}_{<0} + \underbrace{\epsilon_{2t}}_{\approx 0} \\ &< 0 \\ \Rightarrow Y_{2,t} &< Y_{2,t-1}. \end{aligned}$$

Ein Fallen von Y_{2t} wirkt sich je nach Vorzeichen von β auf das Gleichgewicht aus. Wir nehmen weiterhin an, dass sich Y_{1t} nur zufällig verhält

$$Y_{1,t} \approx Y_{1,t-1}.$$

Bei positivem β gilt

$$-\beta Y_{2,t-1} < \underbrace{-\beta}_{<0} Y_{2,t}.$$

Daraus ergibt sich für das Gleichgewicht im nächsten Zeitpunkt

$$\begin{aligned} Y_{1,t} \approx Y_{1,t-1} &< -\beta Y_{2,t-1} < -\beta Y_{2,t} \\ \Rightarrow Z_t &= Y_{1,t} + \beta Y_{2,t} < 0. \end{aligned}$$

Für negatives β gilt

$$-\beta Y_{2,t-1} > \underbrace{-\beta}_{>0} Y_{2,t},$$

so dass die Ungleichungskette keine eindeutige Richtung besitzt

$$Y_{1,t} \approx Y_{1,t-1} < -\beta Y_{2,t-1} > -\beta Y_{2,t}$$

und deshalb ein Ausgleich möglich ist

$$\Rightarrow Z_t = Y_{1,t} + \beta Y_{2,t} \approx 0.$$

Tabelle 2.3: Fallunterscheidung im Fall $\alpha_1 = 0, \alpha_2 \neq 0$

$$\begin{array}{l} \left. \begin{array}{l} \alpha_2 > 0, \beta > 0 \\ \alpha_2 < 0, \beta < 0 \end{array} \right\} \alpha_2 \cdot \beta > 0 \quad \text{Explosion} \\ \left. \begin{array}{l} \alpha_2 > 0, \beta < 0 \\ \alpha_2 < 0, \beta > 0 \end{array} \right\} \alpha_2 \cdot \beta < 0 \quad \text{Ausgleich} \end{array}$$

□

Ein Ausgleich

$$Z_t \approx 0 \Leftrightarrow Y_{1,t} \approx -\beta Y_{2,t}$$

ist nur im Fall $\beta < 0$ möglich, also wenn α_2 und β verschiedene Vorzeichen haben. Im Fall $\beta > 0$ setzt sich die Ungleichungskette immer weiter fort, es kommt zur Explosion von Y_{2t} . Bei einem Abweichen vom Gleichgewicht in positive Richtung, verkehrt sich nur die Richtung der Dynamik. Die Fallunterscheidung liefert auch hier nur im Fall $\beta < 0$ eine Fehlerkorrektur.

Analoges Verhalten zeigt sich bei **negativem** α_2 . Nur wenn β ein entgegengesetztes Vorzeichen (hier also positiv) hat, kann man von Fehlerkorrektur und somit Kointegration sprechen. Sonst kommt es zur Verstärkung der vom Gleichgewicht abweichenden Dynamik, was zur Explosion und somit Instationarität des Gleichgewichtsfehlers Z_t führt.

Die Ergebnisse dieses Abschnitts sind in Tabelle 2.3 zusammengefasst.

$\alpha_1 \neq 0, \alpha_2 \neq 0$

Betrachtet man nun das „volle“ Modell, in dem beide Prozesse von der gemeinsamen Kointegrationsbeziehung abhängig sind

$$\begin{aligned} \Delta Y_{1t} &= \alpha_1 (Y_{1t} + \beta Y_{2t}) + \epsilon_{1t} \\ \Delta Y_{2t} &= \alpha_2 (Y_{1t} + \beta Y_{2t}) + \epsilon_{2t}, \end{aligned}$$

muss man vier Kombinationen von Parameterkonstellationen unterscheiden.

- Fall 1: $\alpha_1 < 0, \alpha_2 \cdot \beta < 0$
- Fall 2: $\alpha_1 < 0, \alpha_2 \cdot \beta > 0$
- Fall 3: $\alpha_1 > 0, \alpha_2 \cdot \beta < 0$
- Fall 4: $\alpha_1 > 0, \alpha_2 \cdot \beta > 0$

Nur im ersten Fall findet eine Fehlerkorrektur statt, da beide Prozesse um einen Ausgleich bemüht sind. Während beide Prozesse im Fall 4 abdriften, gibt es in den Fällen 2 und 3 jeweils

Tabelle 2.4: Fallunterscheidung der Ausprägung von α_1 und $\alpha_2\beta$ und Auswirkung auf die Dynamik des bivariaten Systems

Fall 1 $\alpha_1 < 0, \alpha_2\beta < 0$	Fall 2 $\alpha_1 < 0, \alpha_2\beta > 0$	Fall 3 $\alpha_1 > 0, \alpha_2\beta < 0$	Fall 4 $\alpha_1 > 0, \alpha_2\beta > 0$
$Y_{1,t} > Y_{1,t-1}$ $-\beta Y_{2,t} < -\beta Y_{2,t-1}$	$Y_{1,t} > Y_{1,t-1}$ $-\beta Y_{2,t} > -\beta Y_{2,t-1}$	$Y_{1,t} < Y_{1,t-1}$ $-\beta Y_{2,t} < -\beta Y_{2,t-1}$	$Y_{1,t} < Y_{1,t-1}$ $-\beta Y_{2,t} > -\beta Y_{2,t-1}$
Anpassung Y_{1t} Anpassung Y_{2t}	Anpassung Y_{1t} Abdriften Y_{2t}	Abdriften Y_{1t} Anpassung Y_{2t}	Abdriften Y_{1t} Abdriften Y_{2t}
Kointegration	Explosion	Explosion	Explosion

einen Prozess, der abdriftet, und einen, der um Ausgleich bemüht ist. Analoge Überlegungen wie in den obigen Kapiteln führen zu den Ergebnissen, die in Tabelle 2.4 festgehalten sind.

Für alle kointegrierten Systeme, die im Kapitel 3 für Simulationsstudien verwendet werden, liegt für die Schätzung der Komponenten der Matrix α Werte wie im Fall 1 vor. Nur wenn beide Prozesse an der Fehlerkorrektur beteiligt sind, kann es eine Kointegrationsbeziehung geben.

Beweis. Wir betrachten im Folgenden das Modell ohne zufällige Veränderungen der Variablen, also mit $\epsilon_{1t} = \epsilon_{2t} = 0$

$$\begin{aligned}\Delta Y_{1t} &= \alpha_1 Z_{t-1} \\ \Delta Y_{2t} &= \alpha_2 Z_{t-1}\end{aligned}$$

und bei negativer Abweichung vom Gleichgewicht im Zeitpunkt $t - 1$

$$\begin{aligned}Z_{t-1} &< 0 \\ Y_{1,t-1} &< -\beta Y_{2,t-1}.\end{aligned}$$

Zum Zeitpunkt $t - 1$ liegt also ein Ungleichgewicht vor. Der Wert von $Y_{1,t-1}$ ist (etwas) geringer als der Wert des zweiten (gewichteten) Prozesses $-\beta Y_{2,t-1}$.

Wir führen die Überlegungen hier nur für den Fall $\alpha_1 < 0, \alpha_2 \cdot \beta < 0$ durch. Die anderen Fälle erhält man durch analoge Überlegungen. Bei negativem α_1 und negativer Abweichung vom Gleichgewicht wächst der erste Prozess $Y_{1,t}$ zum nächsten Zeitpunkt

$$Z_{t-1} < 0 \stackrel{\alpha_1 < 0}{\Rightarrow} \Delta Y_{1,t} = \alpha_1 Z_{t-1} > 0 \Leftrightarrow Y_{1,t} > Y_{1,t-1}.$$

Für den zweiten Prozess $Y_{2,t}$ kommt es in beiden Parameterkonstellationen des Fall 1 $\alpha_2 <$

$0, \beta > 0$ und $\alpha_2 > 0, \beta < 0$ zu demselben Verlauf.

$$\begin{aligned} Z_{t-1} < 0 \quad \alpha_2 \leq 0 &\Rightarrow \Delta Y_{2,t} = \alpha_2 Z_{t-1} > 0 \\ &\Leftrightarrow Y_{2,t} > Y_{2,t-1} \\ &\stackrel{\beta > 0}{\Leftrightarrow} -\beta Y_{2,t} < -\beta Y_{2,t-1} \end{aligned}$$

$$\begin{aligned} Z_{t-1} < 0 \quad \alpha_2 \geq 0 &\Rightarrow \Delta Y_{2,t} = \alpha_2 Z_{t-1} < 0 \\ &\Leftrightarrow Y_{2,t} < Y_{2,t-1} \\ &\stackrel{\beta \leq 0}{\Leftrightarrow} -\beta Y_{2,t} < -\beta Y_{2,t-1} \end{aligned}$$

Der Prozess fällt im nächsten Zeitpunkt. Beide Prozesse verändern ihre Dynamik in Richtung eines Ausgleichs. So führt eine negative Abweichung vom Gleichgewicht

$$Z_{t-1} < 0 \Rightarrow \begin{cases} Y_{1,t} > Y_{1,t-1} \\ -\beta Y_{2,t} < -\beta Y_{2,t-1} \end{cases}$$

zu einer Fehlerkorrektur im nächsten Zeitpunkt

$$Y_{1,t} \approx -\beta Y_{2,t}.$$

□

2.5.3 Der multivariate Fall am dreidimensionalen Beispiel

Im allgemeinen multivariaten Fall kann es mehr als eine Kointegrationsbeziehung geben. Wir betrachten hier exemplarisch den dreidimensionalen Fall, um die Erweiterungsmöglichkeiten gegenüber dem bivariaten Fall zu veranschaulichen.

Im dreidimensionalen Fall ($K = 3$) kann es eine oder zwei Kointegrationsbeziehungen zwischen den drei Teilprozessen geben. Diese Anzahl wird über den Kointegrationsrang r beschrieben. α und β sind $(K \times r)$ -Matrizen, also nur im Fall *einer* Kointegrationsbeziehung $r = 1$ Vektoren wie im bivariaten Fall. Der Gleichgewichtsfehler Z_t ist ein r -dimensionaler Prozess, der bei vorliegender Kointegration stationär ist. Wir betrachten hier nur das Modell ohne autoregressive Einflüsse, um uns auf die Kointegration und Fehlerkorrektur zu konzentrieren.

Kointegrationsrang $r=1$ Im Falle *einer* Kointegrationsbeziehung stellt sich das Fehlerkorrekturmodell wie folgt dar

$$\begin{pmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \begin{pmatrix} 1 & \beta_2 & \beta_3 \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}.$$

$\beta = (1, \beta_2, \beta_3)$ ist also ein Vektor, so dass das Produkt mit dem Vektor Y_{t-1} einen eindimensionalen Gleichgewichtsfehler-Prozess ergibt, der alle Teilprozesse beeinflusst

$$\begin{aligned} \Delta Y_{1t} &= \alpha_1 (Y_{1,t-1} + \beta_2 Y_{2,t-1} + \beta_3 Y_{3,t-1}) + \epsilon_{1t} \\ \Delta Y_{2t} &= \alpha_2 (Y_{1,t-1} + \beta_2 Y_{2,t-1} + \beta_3 Y_{3,t-1}) + \epsilon_{2t} \\ \Delta Y_{3t} &= \alpha_3 (Y_{1,t-1} + \beta_2 Y_{2,t-1} + \beta_3 Y_{3,t-1}) + \epsilon_{3t}. \end{aligned}$$

Alle Teilprozesse sind von demselben Gleichgewichtsfehlerprozess $Z_t = Y_{1,t} + \beta_2 Y_{2,t} + \beta_3 Y_{3,t}$ abhängig. Durch die unterschiedliche Gewichtung α_i ist der Einfluss unterschiedlich stark.

Durch die Normierung ist die Kointegrationsbeziehung eindeutig. Alle drei Teilprozesse sind über sie zueinander ins Verhältnis gesetzt. β_1 ist normiert auf 1. Im System wirken zwei *common trends*, die nur in richtiger Gewichtung in der Linearkombination gekürzt werden können.

Dieser Fall unterscheidet sich also kaum vom bivariaten Modell. Im dreidimensionalen Modell können im Fall $p > 1$ die Lags aller Teilprozesse die einzelnen Prozesse beeinflussen. Die Interpretation der Kointegrationsbeziehung verändert dadurch sich jedoch nicht. Man beachte jedoch, dass Kointegration nur in diesem dreidimensionalen Modell vorliegt. Zwei Einzelprozesse allein sind nicht kointegriert. Dies ändert sich im dreidimensionalen Modell mit *zwei* Kointegrationsbeziehungen.

Kointegrationsrang $r=2$ Auch den Fall mehrerer Kointegrationsbeziehungen veranschaulichen wir nur am dreidimensionalen Beispiel. Hier kann es durch Hinzunahme eines weiteren (dritten) Prozesses zu einem (bivariaten) kointegrierten System eine neue Kointegrationsbeziehung geben. In diesem Fall ist die normierte Fehlerkorrekturdarstellung des Prozesses:

$$\begin{pmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} 1 & 0 & \beta_{31} \\ 0 & 1 & \beta_{32} \end{pmatrix} \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{pmatrix}.$$

α und β sind nun Matrizen vom Rang $r = 2$, so dass im ausgeschriebenen Modell zwei (normierte) Kointegrationsbeziehungen auftreten

$$\begin{aligned}\Delta Y_{1t} &= \alpha_{11} (Y_{1,t-1} + \beta_{31} Y_{3,t-1}) + \alpha_{12} (Y_{2,t-1} + \beta_{32} Y_{3,t-1}) + \epsilon_{1t} \\ \Delta Y_{2t} &= \alpha_{21} (Y_{1,t-1} + \beta_{31} Y_{3,t-1}) + \alpha_{22} (Y_{2,t-1} + \beta_{32} Y_{3,t-1}) + \epsilon_{2t} \\ \Delta Y_{3t} &= \alpha_{31} (Y_{1,t-1} + \beta_{31} Y_{3,t-1}) + \alpha_{32} (Y_{2,t-1} + \beta_{32} Y_{3,t-1}) + \epsilon_{3t}.\end{aligned}$$

Im vollen Modell können also alle Teilprozesse von beiden Kointegrationsbeziehungen beeinflusst werden. Alle drei Teilprozesse enthalten ein und denselben integrierten Anteil. Der Kointegrationsraum ist *zweidimensional*. Zwischen jedem Paar von Teilprozessen gibt es eine Kointegrationsbeziehung. Je zwei linear unabhängige Kointegrationsbeziehungen spannen den Kointegrationsraum auf. Das heißt, dass alle möglichen Kointegrationsbeziehung der drei Teilprozesse $\lambda_1 Y_{1t} + \lambda_2 Y_{2t} + \lambda_3 Y_{3t}$ als Linearkombinationen dieser beiden Kointegrationsbeziehungen dargestellt werden können.

Im obigen Modell bestimmt wegen der Normierung die Reihenfolge der drei Teilprozess, welche Kointegrationsbeziehungen im System dargestellt werden. Zu den beiden Kointegrationsbeziehungen gibt es eine dritte Kointegrationsbeziehung. In obigem Modell ist der Prozess Y_{3t} durch die beiden Kointegrationsbeziehungen zu Y_{1t} und Y_{2t} in Beziehung gesetzt. Durch den gemeinsamen Trend existiert jedoch eine dritte Kointegrationsbeziehung, nämlich zwischen Y_{1t} und Y_{2t}

$$\beta_{32} Y_1 = \beta_{31} Y_2 \Leftrightarrow Y_1 - \frac{\beta_{31}}{\beta_{32}} Y_2.$$

Wie man an ihren Koeffizienten sehen kann, ist sie jedoch nicht linear unabhängig von den beiden oberen. Welche zwei der drei Kointegrationsbeziehungen man in das Modell aufnimmt, ist mathematisch willkürlich und kann nur theoriegeleitet sinnvoll entschieden werden. In der Praxis kann man durch Ändern der Reihenfolge alle möglichen Darstellungen des dreidimensionalen Systems schätzen lassen, um so das Modell zu finden, welches die Theorie am besten wiedergibt.

Zusätzlich ist es meist notwendig, deterministische Anteile in die Kointegrationsbeziehung mitaufzunehmen, um Mittelwertsunterschiede der einzelnen Zeitreihen auszugleichen.

2.6 Deterministische Trends in kointegrierten Prozessen

Bislang haben wir im Fehlerkorrekturmodell stets einen Prozess Y_t betrachtet, dessen Erwartungswert $\mathbb{E}[Y_t] = 0$ war, so dass auch in seiner Fehlerkorrekturdarstellung keine Konstante auftrat. In den vorangegangenen Abschnitten hatten wir manchmal eine Konstante $c \in \mathbb{R}^K$ mit in das System mit aufgenommen, um einen von Null verschiedenen Mittelwert zu berücksichtigen. Da es sich beim Fehlerkorrekturmodell jedoch um eine Darstellung des Dif-

Tabelle 2.5: Übersicht über deterministische Terme im VECM: Fallunterscheidung nach Ausprägung der Parameter c_0 und c_1 im Modell (2.17)

	c_0	c_1	Modell	Z_t
1. Fall	$c_0 = 0$	$c_1 = 0$	„Zero Mean“	kein deterministischer Anteil
2. Fall	$c_0 = \alpha\tau'$	$c_1 = 0$	„Nonzero Mean“	Mittelwert
3. Fall	$c_0 \neq 0$	$c_1 = 0$	orthogonaler Trend	Mittelwert
4. Fall	$c_0 \neq 0$	$c_1 = \alpha\eta'$	linearer Trend	linearer Trend
5. Fall	$c_0 \neq 0$	$c_1 \neq 0$	allgemeiner Trend	linearer Trend

ferenzenprozesses handelt, kann eine Konstante in der Darstellung außer einem konstanten Mittelwert auch einen linearen Trend in der Reihe Y_t bedeuten.

2.6.1 Übersicht der Modelle

Um für den Prozess Y_t einen konstanten Erwartungswert oder sogar einen linearen Trend, bei dem der Erwartungswert von der Zeit abhängig ist, zu berücksichtigen, muss man einen (linearen) Term $c_0 + c_1t$ in das Fehlerkorrekturmodell aufnehmen

$$\Delta Y_t = c_0 + c_1t + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t. \tag{2.17}$$

Je nach Ausprägung der beiden konstanten Vektoren $c_0, c_1 \in \mathbb{R}^K$ kann man bis zu fünf Fälle unterscheiden, die hier in einer kurzen Übersicht nach Neusser (2006) in Tabelle 2.5 dargestellt werden. Ein konstanter Vektor c_0 fügt der Differenz jeder Einzelreihe einen konstanten Mittelwert hinzu, während der vektorielle Term c_1t in jeder Differenzenreihe einen linearen Trend bewirkt.

Für die Darstellung von Y_t bedeutet die Hinzunahme eines linearen Terms $Y_t = c_0 + c_1t + \dots$ mit $c_0 \neq 0$ einen nicht-konstanten Mittelwert und $c_1 \neq 0$ (zusätzlich) einen linearen deterministischen Zeittrend (siehe Abschnitt 2.1.2, S.60). Da im Fehlerkorrekturmodell jedoch der Differenzenprozess ΔY_t dargestellt wird, kann durch die Aufnahme eines konstanten Vektors $c_0 \neq 0$ ein linearer Trend ins System gelangen. Schon im univariaten Fall sieht man leicht, dass $\Delta Y_t = c_0$ auch für $Y_t = c_0t$ gilt.

$$\begin{aligned} Y_t &= c_0 \cdot t \\ \Rightarrow Y_t - Y_{t-1} &= c_0 \cdot t - c_0 \cdot (t - 1) \\ \Delta Y_t &= c_0 \cdot \underbrace{(t - (t - 1))}_{=1} \\ \Delta Y_t &= c_0 \end{aligned}$$

Ebenso kann der Term $c_1 t$ im Fehlerkorrekturmodell einen quadratischen Trend in das System bringen. Bei Kointegration stellt sich die Frage, ob die zusätzlichen Terme eine Auswirkung auf die Kointegrationsbeziehung haben. Wir hatten in Kapitel 2.5.1 (S.101) schon erwähnt, dass ein Mittelwertunterschied zwischen zwei Einzelreihen in einem bivariaten System in die Kointegrationsbeziehung mit aufgenommen werden kann.

Wir werden zunächst alle Fälle in einer kurzen Übersicht darstellen und aufzeigen, welche Auswirkung die Ausprägung der Vektoren c_0 und c_1 zum einem auf das K -dimensionale System Y_t , als auch auf den r -dimensionalen Gleichgewichtsprozess Z_t hat. Im nächsten Abschnitt werden die Darstellungen anschaulich hergeleitet und damit die folgenden Aussagen bewiesen.

1.Fall: „Zero Mean“ Bei $c_0 = c_1 = 0$ handelt es sich um das bisher in Kapitel 2.5 betrachtete „einfache“ Kointegrationsmodell ohne deterministischen Term, also mit Erwartungswert $\mathbb{E}[Y_t] = 0$, das wir in der Darstellung

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t$$

kennengelernt hatten. Hier gibt es keinen deterministischen Anteil im System Y_t , so dass die Kointegrationsbeziehung weiterhin als

$$Z_t = \beta' Y_t$$

dargestellt werden kann.

2.Fall: „Nonzero Mean“ Wenn im Fehlerkorrekturmodell kein zeitabhängiger Anteil ($c_1 = 0$) auftritt und man den konstanten Vektor c_0 als Linearkombination der Spalten der Matrix α aus der Zerlegung $\Pi = \alpha' \beta$ darstellen kann ($c_0 = \alpha \tau'$ mit $\tau \in \mathbb{R}^r$), dann führt das Einführen von c_0 nicht zu einem Trend im Prozess, sondern fügt einen von Null verschiedenen („nonzero“) Mittelwert μ_0 hinzu (Lütkepohl, 2005, S.257)

$$\Delta Y_t = \underbrace{\alpha \tau'}_{\mu} + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t.$$

Dieser Mittelwertsvektor kann in die Kointegrationsbeziehung gezogen werden. Hier wird er benötigt, um die (konstanten) Mittelwertsunterschiede der Einzelreihen $Y_{i,t}$ auszugleichen. In diesem Fall gibt es sowohl im K -dimensionalen Prozess Y_t als auch im r -dimensionalen Gleichgewichtsprozess Z_t einen konstanten Mittelwert, jedoch keinen deterministischen zeitabhängigen Trend

$$Z_t = \beta'(Y_t - \mu_0).$$

Erfüllt der konstante Vektor die Bedingung $c_0 = \alpha\tau'$ jedoch nicht, so gelangt durch seine Hinzunahme ein linearer Trend ins System Y_t .

3.Fall: Orthogonaler Trend Nimmt man eine allgemeine Konstante c_0 , aber keine zeitabhängige Komponente ($c_1 = 0$) in das Modell auf (Lütkepohl, 2005, S.258)

$$\Delta Y_t = c_0 + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t,$$

dann „verursacht“ diese Konstante einen linearen Trend im Prozess Y_t , der sich jedoch nicht in die Kointegrationsbeziehung niederschlägt, da er *orthogonal* (hier: entgegengesetzt, siehe Kapitel 2.6.2) zu ihr ist. Wir werden im nächsten Abschnitt diese Aussage beweisen und auch den Begriff der Orthogonalität besser verstehen. In dieser Übersicht ist vorerst nur entscheidend, dass der lineare Trend, der durch einen konstanten Vektor in das System Y_t eingeführt wird, sich nicht in die Kointegrationsbeziehung aufnehmen lässt

$$Z_t = \beta'(Y_t - \mu_0).$$

Dies ist erst möglich, wenn der Differenzenprozess einen linearen Anteil hat.

4.Fall: linearer Trend Enthält das vektorielle Fehlerkorrekturmodell außer dem konstanten Vektor c_0 noch eine zeitabhängige Komponente $c_1 t$, tritt im System Y_t (wie im Fall 3) ein linearer Trend auf. Lässt sich dieser zeitabhängige Term als Linearkombination der Spalten von α darstellen $c_1 = \alpha\eta'$, kann er in die Kointegrationsbeziehung miteinbezogen werden. Im Modell

$$\Delta Y_t = c_0 + \alpha\eta' t + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t,$$

ist also nicht nur die Reihe selbst, sondern auch die langfristige Gleichgewichtsbeziehung zeitabhängig

$$Z_t = \beta'(Y_t - \mu_1 t).$$

Erfüllt c_1 diese Bedingungen jedoch nicht, dann können weitere deterministische Trends ins System gelangen.

5.Fall: Allgemeiner Trend So wie wir gesehen haben, dass eine allgemeine Konstante c_0 einen linearen Trend in der Reihe bedeuten kann, der sich jedoch nicht in der Kointegrationsbeziehung niederschlägt, so kann ein allgemeiner linearer deterministischer Term einen *quadratischen* Trend (Grad 2) in der Reihe Y_t bewirken, während in der Kointegrationsbe-

ziehung lediglich ein linearer Trend (Grad 1) auftritt. Im allgemeinen Modell

$$\Delta Y_t = c_0 + c_1 t + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t$$

hat der Gleichgewichtsprozess also ebenfalls die Gestalt

$$Z_t = \beta'(Y_t - \mu_1 t).$$

Ausblick: weitere Fälle Theoretisch könnte man dieses Muster erweitern, indem man auch quadratische Anteile $c_2 t^2$ etc. ins Modell aufnimmt. Lässt sich die Konstante c_i als Linearkombination der Spalten von α darstellen, kann ein Trend vom Grad i in die Kointegrationsbeziehung aufgenommen werden. Ist dies nicht der Fall, verursacht $c_i t^i$ einen Trend vom Grad $i + 1$ in der Reihe und einen Trend vom Grad i in der Kointegrationsbeziehung.

Bereits die Schätzprozedur für allgemeines c_1 , das einen quadratischen Trend bedeutet, ist schon nicht mehr in gängigen Software-Anwendungen (z.B. JMulti Lütkepohl & Krätzig, 2004 oder R Pfaff, 2008) implementiert. Neben der zusätzlichen Schwierigkeit, ein so komplexes Modell richtig zu schätzen, tritt auch das Problem einer plausiblen Interpretation auf. Was bedeutet es, wenn die Kointegrationsbeziehungen einem deterministischen quadratischen (kubischen, etc.) Trend folgen? Theoretisch ist die Schätzung eines solchen Modells möglich. Praktisch sollte man den Informationsgewinn hinterfragen, bevor man ein solch kompliziertes Modell wählt.

2.6.2 Berechnungen

Im vorigen Abschnitt haben wir das allgemeine Fehlerkorrekturmodell mit allgemeinem linearen Trend

$$\Delta Y_t = c_0 + c_1 t + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t \quad (2.18)$$

betrachtet und die Bedeutung der Konstanten c_0 und c_1 für den Prozess und die Kointegrationsbeziehung beschrieben. Dabei wurde stets erwähnt, dass man einen konstanten Mittelwert oder einen linearen Trend unter bestimmten Voraussetzungen in die Kointegrationsbeziehung aufnehmen kann.

Im Folgenden wird gezeigt, wie man mathematisch zu den Voraussetzungen gelangt und wie man das Modell so umschreiben kann, dass Konstanten und Trends in die Kointegrationsbeziehung aufgenommen werden. Dabei halten wir uns an das Vorgehen Lütkepohl (2005, Kap.6.4) und erweitern es durch eigene Überlegungen.

Mittelwert

Wir haben in Kapitel 2.5 (S.100) gesehen, dass ein rein stochastischer VAR(p)-Prozess X_t mit Erwartungswert $\mathbb{E}[X_t] = 0$ folgende Fehlerkorrekturdarstellung hat

$$\Delta X_t = \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \dots + \Gamma_{p-1} \Delta X_{t-(p-1)} + \epsilon_t. \quad (2.19)$$

Nun wollen wir einen Prozess Y_t betrachten, der außer dieser stochastischen Komponente noch einen deterministischen Teil μ_t enthält

$$Y_t = \mu_t + X_t.$$

Dabei ist der deterministische Term μ_t eine Funktion der Zeit. Wir betrachten nun verschiedene solche Funktionen. Sei zunächst $\mu_t = \mu_0$ ein konstanter Mittelwert(svektor)

$$\begin{aligned} Y_t &= \mu_0 + X_t \\ \Leftrightarrow X_t &= Y_t - \mu_0. \end{aligned}$$

Aus dieser Darstellung folgt für die Differenzen der beiden Prozesse direkt deren Gleichheit

$$\Delta Y_t = Y_t - Y_{t-1} = X_t - \mu_0 - X_{t-1} + \mu_0 = X_t - X_{t-1} = \Delta X_t. \quad (2.20)$$

Ersetzt man nun ΔX_t und ΔX_{t-j} gemäß (2.20) mit dem entsprechenden Term für Y_t und X_{t-1} nach (2.20), erhält man die Fehlerkorrekturdarstellung von Y_t

$$\Delta Y_t = \alpha\beta' \underbrace{(Y_{t-1} - \mu_0)}_{X_{t-1}} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t. \quad (2.21)$$

Der Mittelwertsvektor von Y_t kann also in die Kointegrationsbeziehung aufgenommen werden.

Im Folgenden wird mathematisch dargestellt, wie man von dieser Darstellung, in der der konstante Erwartungswert $\mathbb{E}[Y_t] = \mu_0$ in die Kointegrationsbeziehung aufgenommen ist, zu der üblichen Darstellung (2.18) und der Bedingung an die Konstante $c_0 = \alpha\tau'$ gelangt.

Wir betrachten dazu die Kointegrationsbeziehung aus (2.21), in der der konstante Mittelwertsvektor $\mu_0 = (\mu_{01}, \dots, \mu_{0K})$ zum Ausgleich der Mittelwertsunterschiede der Einzelreihen einbezogen ist, und stellen sie so um, dass man die ursprüngliche Darstellung der Kointegra-

tionsbeziehung im rein stochastischen Fall $\beta'Y_{t-1}$ erkennt.

$$\begin{aligned} \beta'(Y_{t-1} - \mu_0) &= \begin{pmatrix} \beta_{11}(Y_{1,t-1} - \mu_{01}) + \dots + \beta_{1K}(Y_{K,t-1} - \mu_{0K}) \\ \vdots \\ \beta_{r1}(Y_{1,t-1} - \mu_{01}) + \dots + \beta_{rK}(Y_{K,t-1} - \mu_{0K}) \end{pmatrix} \\ \Leftrightarrow \beta'Y_{t-1} - \beta'\mu_0 &= \begin{pmatrix} \beta_{11}Y_{1,t-1} + \dots + \beta_{1K}Y_{K,t-1} \\ \vdots \\ \beta_{r1}Y_{1,t-1} + \dots + \beta_{rK}Y_{K,t-1} \end{pmatrix} - \begin{pmatrix} \beta_{11}\mu_{01} + \dots + \beta_{1K}\mu_{0K} \\ \vdots \\ \beta_{r1}\mu_{01} + \dots + \beta_{rK}\mu_{0K} \end{pmatrix} \\ \Leftrightarrow \beta'Y_{t-1} + \tau' &= \begin{pmatrix} \beta_{11}Y_{1,t-1} + \dots + \beta_{1K}Y_{K,t-1} \\ \vdots \\ \beta_{r1}Y_{1,t-1} + \dots + \beta_{rK}Y_{K,t-1} \end{pmatrix} + \begin{pmatrix} \tau_1 \\ \vdots \\ \tau_r \end{pmatrix} \end{aligned}$$

Der hier definierte Vektor $\tau = -\beta\mu_0$ hat die Dimension r und besteht aus dem Produkt der Kointegrationsmatrix β und dem Mittelwertsvektor μ_0 .

Dieser konstante Vektor τ ist im Anwendungsfall wie die Matrix β und der Mittelwert μ_0 unbekannt. Deshalb kann er als zusätzliche Zeile in die Matrix β aufgenommen werden, die ohnehin geschätzt werden muss. Jeder Kointegrationsvektor (Spalte der Matrix) enthält nun neben den Koeffizienten der Kointegrationsbeziehung (Linearkombination) einen weiteren Koeffizienten. Da dieser konstant in das Modell eingeht, muss auch der Vektor Y_{t-1} verändert werden, damit τ richtig ins Modell integriert wird. Dies ist der Fall, wenn dem Vektor Y_{t-1} eine 1 als zusätzliche Zeile angehängt wird. Mit diesen Überlegungen definieren wir die „neue“ Kointegrationsmatrix $\beta^{0'}$ und den „neuen“ Prozess Y_t^0 durch

$$\begin{aligned} \beta^{0'} &:= [\beta'; \tau'] = \begin{pmatrix} \beta_{11} & \beta_{21} & \dots & \beta_{K1} & \tau_1 \\ \vdots & & & & \\ \beta_{1r} & \beta_{2r} & \dots & \beta_{Kr} & \tau_r \end{pmatrix} \\ Y_{t-1}^0 &:= \begin{pmatrix} Y_{t-1} \\ 1 \end{pmatrix} = \begin{pmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ \vdots \\ Y_{K,t-1} \\ 1 \end{pmatrix}. \end{aligned}$$

Die Matrix $\beta^{0'}$ und Y_{t-1}^0 sind nun so gewählt, dass für die „alte“ und „neue“ Kointegrationsbeziehung Gleichheit gilt $\beta^{0'}Y_{t-1}^0 = \beta'(Y_{t-1} - \mu_0)$. Mit dieser neuen Matrix $\beta^{0'}$ kann man nun eine neue Matrix $\Pi^0 = \alpha\beta^{0'}$ berechnen

$$\Pi^0 := \alpha\beta^{0'} = \alpha[\beta'; \tau'] = [\alpha\beta'; \alpha\tau'] = [\Pi; c_0].$$

Dabei wird der konstante Vektor $c_0 = \alpha\tau'$ definiert³⁴.

³⁴Die Matrizenmultiplikation in dieser Schreibweise ist im Anhang A.1.3 (S.313) beschrieben.

Eine erste Möglichkeit mit einem konstanten deterministischen Term umzugehen, ist also die Modellmatrix β bzw. Π und den Vektor Y_{t-1} abzuändern und das Fehlerkorrekturmodell wie folgt zu schreiben

$$\begin{aligned}\Delta Y_t &= \alpha\beta^{0'}Y_{t-1}^0 + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t \\ &= \Pi^0 Y_{t-1}^0 + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t.\end{aligned}\tag{2.22}$$

Andererseits kann man die Fehlerkorrekturdarstellung mit konstantem Mittelwert so umschreiben, dass man die Darstellung der Kointegrationsbeziehung nicht verändern muss, sondern lediglich eine Konstante c_0 in das Fehlerkorrekturmodell aufnimmt. Denn es gilt

$$\begin{aligned}\alpha\beta'(Y_{t-1} - \mu_0) &= \alpha\beta'Y_{t-1} - \alpha\beta'\mu_0 \\ &= \alpha\beta'Y_{t-1} + \alpha\tau' \\ &= \Pi Y_{t-1} - \Pi\mu_0 \\ &= \Pi Y_{t-1} + c_0.\end{aligned}$$

So erhält man den Fall 2 für das Fehlerkorrekturmodell

$$\Delta Y_t = c_0 + \alpha\beta'Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t,$$

in dem die Konstante folgender Restriktion

$$c_0 = \alpha\tau' = -\Pi\mu_0$$

genügen muss.

Umgekehrt bedeutet ein konstanter Vektor c_0 im Fehlerkorrekturmodell also nur dann einen allgemeinen Mittelwertsvektor (der zusätzlich sogar in die Kointegrationsbeziehung aufgenommen werden kann), wenn er obige Bedingung erfüllt, also als Linearkombination der Spalten von α geschrieben werden kann.

Allgemeiner linearer Trend

Nun betrachten wir den Fall, dass der Erwartungswert von Y_t zusätzlich einen linearen Trend hat, $\mu_t = \mu_0 + \mu_1 t$. Der Prozess Y_t lässt sich also über einen „zeromean“- Prozess X_t wie in (2.19) darstellen als

$$Y_t = \mu_0 + \mu_1 \cdot t + X_t.$$

Aus dieser Beziehung kann man ähnlich wie oben zwei Hilfsgleichungen ableiten

$$X_t = Y_t - \mu_0 - \mu_1 t \quad (2.23)$$

$$\begin{aligned} \Delta X_t &= Y_t - \mu_0 - \mu_1 t - (Y_{t-1} - \mu_0 - \mu_1(t-1)) \\ &= \Delta Y_t - \mu_1 \end{aligned} \quad (2.24)$$

Hier fällt μ_0 aus der Darstellung heraus. Bei entsprechender Ersetzung in der Darstellung von ΔX_t aus (2.19) ergibt sich die Fehlerkorrekturdarstellung für ΔY_t :

$$\begin{aligned} \Delta X_t &= \Pi X_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + \epsilon_t \\ \Delta Y_t - \mu_1 &= \underbrace{\Pi (Y_{t-1} - \mu_0 - \mu_1(t-1))}_{X_{t-1}, vgl.(2.23)} + \sum_{j=1}^{p-1} \Gamma_j \underbrace{(\Delta Y_{t-j} - \mu_1)}_{\Delta X_{t-j}, vgl.(2.24)} + \epsilon_t \end{aligned}$$

Der lineare Trend wurde so in die Kointegrationsbeziehung aufgenommen. Zusätzlich taucht μ_1 jedoch auch in den Lags auf, was die Darstellung unnötig kompliziert werden lässt. Zur Vereinfachung multiplizieren wir zunächst alle Terme aus, um sie anschließend geschickt wieder zusammenzufassen.

$$\begin{aligned} \Delta Y_t &= \mu_1 + \Pi Y_{t-1} - \Pi \mu_0 - \Pi \mu_1(t-1) + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} - \sum_{j=1}^{p-1} \Gamma_j \mu_1 + \epsilon_t \\ &= \underbrace{\mu_1 - \Pi \mu_0 - \sum_{j=1}^{p-1} \Gamma_j \mu_1}_{c_0^*} + \underbrace{\Pi (Y_{t-1} - \mu_1(t-1))}_{\Pi^+ Y_{t-1}^+} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t \end{aligned}$$

So werden alle konstanten Anteile im Vektor c_0^* und alle zeitabhängigen Anteile in $\Pi^+ Y_{t-1}^+$ zusammengefasst, so dass die Lags nicht von den deterministischen Termen verändert werden.

Man erhält so ähnlich wie unter (2.22) im Fall eines konstanten Erwartungswertes eine Darstellung von ΔY_t , in der der lineare Trend in die Kointegrationsbeziehung aufgenommen wurde und die Kointegrationsmatrix β bzw. Π entsprechend abgeändert wurde.

$$\Delta Y_t = c_0^* + \Pi^+ Y_{t-1}^+ + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t$$

Welche Veränderung dafür notwendig ist, zeigt folgende Rechnung

$$\begin{aligned}
 \Pi^+ Y_{t-1}^+ &= \Pi (Y_{t-1} - \mu_1(t-1)) \\
 &= \alpha \beta' (Y_{t-1} - \mu_1(t-1)) \\
 &= \alpha \left(\beta' Y_{t-1} - \underbrace{\beta' \mu_1(t-1)}_{\eta'} \right) \\
 &= \alpha [\beta'; \eta'] \begin{pmatrix} Y_{t-1} \\ t-1 \end{pmatrix}.
 \end{aligned}$$

Der lineare Trend $\mu_1 t$ wurde hier in die Kointegrationsbeziehung mit aufgenommen. Man muss also als neue Modellparameter $\beta^{+'} := [\beta'; \eta']$ bzw. $\Pi^+ := \alpha \beta^{+'}$ mit $\eta' = -\beta' \mu_1$ verwenden und Y_{t-1} eine Zeile anhängen, die die Zeitabhängigkeit ausdrückt, wozu hier das Anhängen einer Konstante (also gleich Eins) wie im Fall eines konstanten Terms nicht ausreicht, so dass $t-1$ als weitere Komponente aufgenommen wird: $Y_{t-1}^+ := \begin{pmatrix} Y_{t-1} \\ t-1 \end{pmatrix}$.

Möchte man andererseits Bedingungen für die beiden Konstanten c_0 und c_1 in der allgemeinen Darstellung ableiten, muss man folgende Umformungen betrachten

$$\begin{aligned}
 \Pi^+ Y_{t-1}^+ &= \alpha \beta' (Y_{t-1} - \mu_1(t-1)) \\
 &= \alpha \beta' Y_{t-1} - \alpha \beta' \mu_1(t-1) \\
 &= \alpha \beta' Y_{t-1} + \alpha \eta' (t-1) \\
 &= \alpha \beta' Y_{t-1} + \underbrace{\alpha \eta'}_{c_1} t - \underbrace{\alpha \eta'}_{\text{konstant}}.
 \end{aligned}$$

Der konstante Anteil aus $\Pi^+ Y_{t-1}^+$, nämlich $\alpha \eta' = -\alpha \beta' \mu_1 = -\Pi \mu_1$ wird zur Konstanten c_0^* addiert, so dass sich

$$c_0 := -\Pi \mu_0 + (\mathbb{I}_K - \Gamma_1 - \dots - \Gamma_{p-1} - \Pi) \mu_1 \in \mathbb{R}^K$$

ergibt. In diesem Fall unterliegt c_0 keinen Restriktionen.

Für den Parameter c_1 muss $c_1 = \alpha \eta'$ gelten, damit er in die Kointegrationsbeziehungen aufgenommen werden kann.

$$\Delta Y_t = \underbrace{c_0^* - \alpha \eta'}_{c_0} + \alpha \beta' Y_{t-1} + \underbrace{\alpha \eta'}_{c_1} t + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + \epsilon_t$$

Orthogonaler linearer Trend

Der lineare Trend μ_1 der Reihe Y_t kann als $\eta' = -\beta' \mu_1$ in die Kointegrationsbeziehung gezogen werden, wenn der Parameter $c_1 = \alpha \eta'$ erfüllt. Unter bestimmten Voraussetzungen

wirkt sich der Trend der Reihe nicht auf die Kointegrationsbeziehungen Z_t aus. Dies ist der Fall, wenn der Trend μ_1 orthogonal zur Kointegrationsmatrix ist, wenn also das Produkt der Kointegrationsmatrix mit dem Vektor des Trends μ_1 den Nullvektor ergibt

$$\eta' = -\beta' \mu_1 = 0$$

für $\mu_1 \neq 0$.

Im Fehlerkorrekturmodell entfällt dann der Term $c_1 t = \alpha \eta' t = 0$, so dass der Trend allein in der Konstante c_0 enthalten ist.

$$\Delta Y_t = c_0 + \Pi Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta_{t-j} + \epsilon_t$$

Bemerkung 2.6.1 (Zum Begriff der Orthogonalität) *Die Spalten der Kointegrationsmatrix β spannen einen r -dimensionalen Unterraum des \mathbb{R}^K auf. Algebraisch gesprochen bedeutet $-\beta' \mu_1 = 0$, dass der Trend μ_1 im orthogonalen Komplement dieses Kointegrationsraumes liegt, da die (lineare) Abbildung von μ_1 in den Kointegrationsraum gleich Null ist.*

Diese Überlegungen sind nur in einem kointegrierten System sinnvoll. In einem stationären System mit $r = K$ gibt es keinen zum „Kointegrationsraum“ orthogonalen Unterraum, in dem der lineare Trend als Vektor liegen könnte.

Bemerkung 2.6.2 (Stationäres System) *Für $r = K$ ist β vom Rang K , also invertierbar, so dass Produkt mit einem Vektor im Allgemeinen ungleich dem Nullvektor ist, $\beta' \mu_1 \neq 0$ für $\mu_1 \neq 0$. Denn der einzige Vektor, der bei Multiplikation mit einer invertierbaren Matrix auf den Nullvektor abgebildet wird, ist der Nullvektor selbst. Im Fall $\mu_1 = 0$ ist jedoch jede Einzelreihe selbst stationär.*

2.7 Kointegrationstests

Um in der Praxis kointegrierte Prozesse und Systeme entdecken zu können, braucht man statistische Möglichkeiten, um sie auf Kointegration zu prüfen. In diesem Kapitel sollen deshalb gängige Kointegrationstests vorgestellt werden.

2.7.1 Tests bei Kointegrationsrang Eins

In einem bivariaten System kann es maximal eine Kointegrationsbeziehung geben. Aber auch in höherdimensionalen Prozessen ist es möglich, dass der Kointegrationsrang gleich Eins ist. Man betrachtet also den univariaten (eindimensionalen) Abweichungsprozess $Z_t = \beta' Y_t = \beta_1 Y_{1,t} + \dots + \beta_K Y_{K,t}$. Ist der Prozess im Gleichgewicht, so ist $Z_t = 0$ und nach Normierung

$\beta_1^* = 1$ ergibt sich folgende Gleichgewichtsbeziehung

$$Y_{1,t} = \beta_2^* Y_{2,t} + \dots + \beta_K^* Y_{K,t}. \quad (2.25)$$

Je nachdem, ob der Kointegrationsvektor β bekannt ist oder nicht, ergeben sich unterschiedliche Ansätze für einen Kointegrationstest (Stier, 2001, 19.1.2).

Bekannter Kointegrationsvektor

Im einfachsten Fall ist der (nicht unbedingt normierte) Vektor β bekannt. Dieser Fall kann vorliegen, wenn eine inhaltliche Hypothese die stationäre Linearkombination der Variablen vorgibt. Bei der erwähnten Zeitschätzungsstudie aus Kapitel 2.5.2 (S.103) könnte man beispielsweise prüfen, ob für die anderen Probanden dieselbe Kointegrationsbeziehung zwischen der Zeitreihe der Schätzungen und dem vorgegebenen falschen Feedback gültig ist. Denkbar wäre auch der Fall, dass aus anderen Untersuchungen ein kointegrierter Zusammenhang für eine bestimmte Person oder eine bestimmte Gruppe bekannt ist und man prüfen will, ob dieser Zusammenhang übertragbar ist. Also ob er auch für eine andere Gruppe oder für dieselbe Person nach einer Intervention gültig ist.

Bei bekanntem Kointegrationsvektor β ist der Test auf Kointegration denkbar einfach. Denn wenn β tatsächlich ein Kointegrationsvektor für Y_t ist, bedeutet dies, dass die Linearkombination $\beta'Y_t$ stationär sein muss, obwohl die Einzelprozesse von Y_t alle integriert sind. Diese Stationaritätsprüfung kann man mit einem Unit-Root-Test durchführen. Wird die Nullhypothese, dass es eine Einheitswurzel gibt, nicht verworfen, ist β kein Kointegrationsvektor.

Man testet zuerst, ob alle Einzelprozesse $Y_{i,t}$ integriert sind, was mit einem Unit-Root-Test (z.B. ADF-Test oder PP-Test aus Abschnitt 1.12) möglich ist. Anschließend prüft man, ob die vorgegebene Linearkombination $\beta'Y_t$ stationär, also eine Kointegrationsbeziehung, ist. Wir werden im Kapitel 3 unter anderem diesen Kointegrationstest bei bekanntem Kointegrationsvektor für verschiedene stationäre, integrierte und kointegrierte Systeme genauer untersuchen.

Unbekannter Kointegrationsvektor

Schwieriger ist es, wenn man zwar eine Kointegrationsbeziehung vermutet, aber nicht weiß, wie diese aussehen soll. In diesem Fall muss der Kointegrationsvektor β zuerst geschätzt werden. Dazu betrachtet man die mögliche Kointegrationsbeziehung als Regression und schätzt den Zusammenhang β wie gewohnt mittels Kleinsten-Quadrate-Schätzung (OLS) (z.B. Werner, 1997; Seber & Lee, 2003).

Zu einem idealen Zeitpunkt ist der Prozess im Gleichgewicht, dann gilt die Gleichung (2.25) genau. Zu den meisten Zeitpunkten befindet sich der Prozess jedoch nicht im Gleichgewicht,

so dass diese Gleichung mit einem Fehler u_t belastet ist

$$Y_{1,t} = \beta_2^* Y_{2,t} + \dots + \beta_K^* Y_{K,t} + u_t.$$

Diese Gleichung kann man nun als Regressionsgleichung von $Y_{1,t}$ auf $Y_t^* = (Y_{2,t}, \dots, Y_{K,t})$ auffassen und den Regressionsparameter $\beta = (\beta_2^*, \dots, \beta_K^*)$ schätzen. Für die Residuen $\hat{u}_t = Y_{1,t} - \hat{\beta} Y_t^*$ muss im Fall der Kointegration Stationarität gelten (optimal $u_t \sim \mathcal{N}(0, 1) \forall t$, also Weißes Rauschen) (z.B. Stier, 2001, S.325).

In Begriffen der Kointegration ist \hat{u}_t genau der Abweichungsprozess Z_t , der bei Kointegration eben stationär ist. Liegt keine Kointegration vor, so ist jede Linearkombination von $Y_{1,t}$ und $Y_t^* = (Y_{2,t}, \dots, Y_{K,t})$ integriert der Ordnung Eins, so auch \hat{u}_t . Zum Test auf Kointegration wird also nach OLS-Schätzung von β ein Unit-Root-Test für $\hat{u}_t = p\hat{u}_{t-1} + \epsilon_t$ durchgeführt. Wird dabei p nahe 1 geschätzt, ist \hat{u}_t also integriert, sind die beiden Prozesse nicht kointegriert. Hier wird $Y_{1,t}$ als abhängige Variable von den anderen Einzelprozessen $Y_{k,t}$ betrachtet. Dies führt zu einer Normierung von β auf $\beta^* = (1, \beta_2, \dots, \beta_K)$.

Bei der OLS-Schätzung werden die Koeffizienten so bestimmt, dass die (quadrierten) Abweichungen vom Gleichgewicht $Y_{1,t} = \beta' Y_t^*$ bzw. $\beta' Y_t = 0$ über den beobachteten Zeitraum $[0, T]$ minimiert werden. Dass diese Schätzung tatsächlich konsistent ist, also mit längerer Betrachtung $T \rightarrow \infty$ immer genauer wird, zeigen P. Phillips und Durlauf (1986). Die Verteilung des Schätzers ist keine der üblichen Verteilungen (t-, χ^2 -Verteilung), sie ist jedoch bekannt und kann deshalb simuliert und somit tabellarisiert werden (z.B. Hamilton, 1994).

Ein Problem bei diesem Vorgehen ist jedoch, dass eine Zeitreihe aus dem System von Einzelreihen als abhängige Variable ausgezeichnet wird. Diese Normierung kann Einflüsse auf die Ergebnisse der Regressions-schätzungen haben. Deshalb verwendet man meist eine Schätzprozedur, die ohne eine solche Normierung auskommt. Der Johansen-test in Abschnitt 2.7.2 erfüllt diese Eigenschaft.

Schätzung im Fehlerkorrekturmodell

Wenn man schon einmal den Kointegrationsvektor β wie oben geschätzt hat, kann man die übrigen Parameter im Fehlerkorrekturmodell (α, Γ_i) durch eine weitere Regressionsbetrachtung schätzen. Dazu muss man zunächst die Anzahl der Lags p festlegen (vgl. Kapitel 1.11, S.41). Dann kann man das Fehlerkorrekturmodell

$$\Delta Y_t = \alpha \beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \dots + \Gamma_{p-1} \Delta Y_{t-(p-1)} + \epsilon_t \quad (2.26)$$

als Regression von ΔY_t betrachten. Die Kointegrationsbeziehung $\beta' Y_{t-1}$ dient neben den Differenzen ΔY_{t-k} als Prädiktor.

Die vorgestellten Test- und Schätzmechanismen funktionieren jedoch nur, wenn β ein Vektor und keine Matrix ist, also nur im Fall einer Kointegrationsbeziehung. Wir brauchen

also einen Test, der nicht nur die Anzahl der Kointegrationsbeziehungen, sondern auch die zugehörigen Kointegrationsvektoren schätzen kann und somit einen Test auf den Kointegrationsrang darstellt.

2.7.2 Johansen-Test

Für den sogenannten Johansen-Test auf Kointegration gelten die Einschränkungen im obigen Abschnitt nicht. Er ist bei jeder beliebigen Anzahl von Kointegrationsbeziehungen einsetzbar und schätzt die entsprechenden Kointegrationsbeziehungen, ohne einen Einzelprozess besonders auszuzeichnen.

Dieser Abschnitt hält sich an die Originalarbeit von Johansen (1988). Eine ähnliche Darstellung ist auch bei Neusser (2006, Kap. 16.3) oder Kirchgässner und Wolters (2007, Kap. 6.3.2) zu finden. Es fließen eigene Erklärungen und Ergänzungen ein, die die mathematische Herleitung anschaulich dokumentieren sollen.

Der Johansen-Test vereint einen Test, ob ein System kointegriert ist, mit der (Maximum-Likelihood-)Schätzung des Kointegrationsraumes, für den sowohl die Dimension als auch die Basisvektoren geschätzt werden. Außerdem können auf Grundlage seiner Ideen die übrigen Parameter(matrizen) im Fehlerkorrekturmodell geschätzt werden. Er geht von der VAR(p)-Darstellung eines Prozesses aus und vollzieht sich in folgenden Schritten.

1. Bereinigung des Einflusses der Lags (durch Regressionsschätzungen von zwei Hilfsgleichungen)
2. Schätzung der Matrix $\beta^{(r)}$ für alle möglichen Werte von r (durch die Kanonische Korrelation der entsprechenden Residuen)
3. Schätzung des Kointegrationsranges, und damit Festlegung von β (durch Johansen-Trace- oder Johansen-Max-Test)
4. Schätzung der übrigen Parameter im Fehlerkorrekturmodell

Im ersten Schritt werden die Einflüsse der Lags herausgekürzt, so dass nur noch betrachtet wird, welche Informationen für ΔY_t in Linearkombinationen von Y_{t-1} stecken. Anschließend werden *stationäre* Linearkombinationen der Einzelprozesse von Y_{t-1} geschätzt, die einen Zusammenhang mit ΔY_t haben. Dabei werden sie nach Stärke ihrer Korrelation sortiert. Bei einem Kointegrationsrang von r bilden die ersten r stationären Linearkombinationen (bzw. die Koeffizienten als Vektor β_i geschrieben) eine Basis des Kointegrationsraumes, also die Spalten der Matrix $\beta^{(r)}$.

Nun wird geprüft, für wie viele der K möglichen Linearkombinationen die Korrelation groß genug (statistisch von Null verschieden) ist, um tatsächlich *Kointegrationsbeziehung* zu sein. So bestimmt man die Anzahl der signifikanten Zusammenhänge zwischen Linearkombinationen von Y_{t-1} und ΔY_t . Ist der Rang der Matrix β und damit sie selbst bekannt, kann man

die fehlenden Parameter α und Γ_i im Fehlerkorrekturmodell über ein Regressionsmodell schätzen.

Ausgangspunkt: Fehlerkorrekturmodell

Ausgangspunkt ist ein K -dimensionaler Prozess Y_t in seiner VAR(p)-Darstellung, der auf Kointegration getestet werden soll. An dieser Stelle kann Y_t auch noch stationär oder integriert ohne Kointegrationsbeziehung sein. Im stationären Fall ist der Kointegrationsrang $r = K$. Bei einem integrierten, jedoch nicht kointegrierten Prozess ist $r = 0$. Bei Kointegration nimmt r Werte zwischen Null und K an. Für die Durchführung des Johansen tests gibt es keine Einschränkungen für den Kointegrationsrang. So erkennt der Johansen test auch stationäre oder integrierte Systeme, was wir in Kapitel 3 noch genauer untersuchen werden.

Wichtig ist jedoch, dass p Anfangswerte Y_0, \dots, Y_{-t-p+1} bekannt sind oder man annehmen kann, dass der Prozess im Zeitpunkt $t = 0$ startet, also zuvor konstant Null (Lütkepohl, 2005, Kap. 8.2.1) war; auch dann sind die Startwerte bekannt, nämlich gleich Null. Außerdem muss die Verteilung des Fehlerprozesses ϵ_t bekannt (z.B. multivariat normalverteilt) sein. Dies erreicht man über eine geeignete Wahl von p in der VAR(p)-Darstellung (Stier, 2001, S.333), was durch die Minimierung eines der Informationskriterien aus Kapitel 1.11 (S.41) möglich ist.

Wir betrachten Y_t in seiner Fehlerkorrekturdarstellung (2.26), die leicht aus der VAR-Darstellung gewonnen werden kann (siehe Abschnitt 2.3.3, S. 90) und in der die Kointegrationsmatrix β direkt auftritt. Ziel ist es, alle Parameter des Fehlerkorrekturmodells, insbesondere β , aus dem beobachteten Prozess Y_t (und den dann einfach berechneten Differenzen ΔY_{t-i}) zu schätzen. Dies soll durch einen Maximum-Likelihood-Ansatz (siehe Anhang C.3.3, S.352 dieser Arbeit oder Krengel, 2003) geschehen.

Da man durch Umstellen der Gleichung (2.26) eine Darstellung des Fehlerprozesses erhält, in der alle gesuchten Parameter enthalten sind

$$\epsilon_t = \Delta Y_t - \alpha \beta' Y_{t-1} - \Gamma_1 \Delta Y_{t-1} - \dots - \Gamma_{p-1} \Delta Y_{t-(p-1)} \quad (2.27)$$

und man die Verteilung dieser Größe (einziger zufälliger Parameter des Modells) $\epsilon_t \sim \mathcal{N}_K(0, \Sigma)$ kennt, kann man die Likelihood-Funktion (bedingt auf die Startwerte) berechnen. Sie ist abhängig von allen anderen (unbekannten!) Parametern des Modells und lässt sich wie folgt schreiben (siehe auch Anhang C.3.2, S.351)

$$L(\alpha, \beta, \Sigma, \Gamma_1, \dots, \Gamma_{p-1}) = (2\pi)^{-\frac{TK}{2}} (\det \Sigma)^{-\frac{T}{2}} \exp \left(-\frac{1}{2} \sum_{t=1}^T \epsilon_t' \Sigma^{-1} \epsilon_t \right). \quad (2.28)$$

Wegen der Darstellung (2.27) sind alle Parameter α, β, Γ_i in ϵ_t und damit in der Likelihood-funktion (2.28) enthalten. Hier wird nur zur Übersicht die kürzere Schreibweise gewählt. Die

ausführliche Darstellung erhält man, indem man (2.27) in (2.28) einsetzt.

1. Schritt: Bereinigung des Einflusses der Lags

Bei Kointegration ist jeder Einzelprozess $\Delta Y_{k,t}$ von mindestens einer Linearkombination (maximal r) der verzögerten Einzelprozesse $Y_{i,t-1}$, die dann Kointegrationsbeziehungen bilden, abhängig. Außerdem kann es durch $\Delta Y_{i,t-i}$ autoregressive Einflüsse geben. Im allgemeinen Fall gilt zum Beispiel für den ersten Einzelprozess in der Fehlerkorrekturdarstellung

$$\begin{aligned} \Delta Y_{1,t} = & \alpha_{11} \underbrace{(\beta_{11}Y_{1,t-1} + \dots + \beta_{1K}Y_{K,t-1})}_{\beta'_1 Y_{t-1}} + \dots + \alpha_{1r} \underbrace{(\beta_{r1}Y_{1,t-1} + \dots + \beta_{rK}Y_{K,t-1})}_{\beta'_r Y_{t-1}} \\ & + \left(\sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} \right)_{(1)} + \epsilon_t, \end{aligned}$$

wobei die Schreibweise $A_{(1)}$ hier die erste Zeile des Matrixpolynoms A bezeichnet, da wir zur Anschauung nur den ersten Einzelprozess dargestellt haben. Die Differenzen der anderen Einzelprozesse können auf ähnliche Weise dargestellt werden.

Ziel ist es, einen Zusammenhang zwischen der ersten Differenz des Gesamtprozesses ΔY_t und Linearkombinationen der verzögerten Einzelprozesse $Y_{1,t-1}, \dots, Y_{K,t-1}$ festzustellen. Also die Frage zu klären, ob Kointegrationsbeziehungen $\beta'_1 Y_{t-1}, \dots, \beta'_r Y_{t-1}$ bestehen und wie sie das System beeinflussen, also die Koeffizienten aus α zu bestimmen.

Dazu werden in einem ersten Schritt die Einflüsse der Lags $\Gamma_i \Delta Y_{t-i}$ eliminiert. Dies geschieht, indem man sowohl ΔY_t als auch Y_{t-1} auf die ersten Differenzen $\Delta Y_{t-1}, \dots, \Delta Y_{t-p+1}$ regressiert und die Parameter beider Hilfsgleichungen schätzt

$$\Delta Y_t = \hat{\zeta}_1 \Delta Y_{t-1} + \dots + \hat{\zeta}_{t-p+1} \Delta Y_{p-1} + R_{\Delta,t} \quad (2.29)$$

$$LY_t = Y_{t-1} = \hat{\xi}_1 \Delta Y_{t-1} + \dots + \hat{\xi}_{t-p+1} \Delta Y_{p-1} + R_{L,t}. \quad (2.30)$$

In den Residuen der beide Modelle $R_{\Delta,t}, R_{L,t}$ ist die Information enthalten, die nicht vom autoregressiven Anteil bzw. den Lags erklärt werden kann. Wenn Y_t kointegriert ist, reicht die Information der Lags nicht aus, um ΔY_t zu erklären. Dazu sind noch die Kointegrationsbeziehungen aus $\beta' Y_{t-1}$ wichtig. In diesem Fall sind die Residuen in den Gleichungen (2.29) und (2.30) kein Weißes Rauschen, da in ihnen die zusätzliche Information der Kointegration enthalten sein muss.

Bemerkung 2.7.1 (Deterministische Terme im Johansentest) *Enthält das Fehlerkorrekturmodell deterministische Anteile (z.B. eine Konstante), dann müssen diese bei der Regression berücksichtigt werden (Stier, 2001, S.333) und können später ebenfalls geschätzt werden, wie es in Kapitel 2.6 dargestellt wurde (Lütkepohl, 2005, Kap. 8.2).*

Bemerkung 2.7.2 (Überlagerte Lageinflüsse beachten) *Eigentlich würde eine Regression*

auf $\Delta Y_t - \alpha\beta'Y_{t-1}$ ausreichen, um die Γ_i zu bestimmen

$$\Delta Y_t - \alpha\beta'Y_{t-1} = \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + \epsilon_t. \quad (2.31)$$

Doch in den Γ_i stecken gemischte Einflüsse für ΔY_t und Y_{t-1} . Da man den (bereinigten) Zusammenhang zwischen diesen beiden Größen erkennen will, muss man die Lageinflüsse getrennt herausrechnen, um Überlagerungen zu vermeiden.

Das Modell (2.31) ergibt sich ja aus den Hilfsgleichungen durch Subtraktion (2.29) - Π (2.30), was sich in den Schätzungen in Abschnitt 2.7.2 widerspiegelt. Aus den geschätzten Matrizen in obigen Modellen kann man dann die Γ_i in (2.31) durch $\hat{\Gamma}_i = \hat{\zeta}_i - \hat{\Pi}\hat{\xi}_i$ bestimmen. Man will jedoch mit den Residuen weiterarbeiten, die sich in ähnlicher Weise aus dem Fehlerprozess ergeben $\epsilon_t = R_{\Delta,t} - \Pi R_{L,t}$.

Durch das Kürzen des Modells um die Lageinflüsse erhält man folgendes VAR(1)-Modell,

$$R_{\Delta,t} = \Pi R_{L,t} + \epsilon_t = \alpha\beta' R_{L,t} + \epsilon_t, \quad (2.32)$$

das nun weiterhin betrachtet wird. Dabei ist $\epsilon_t \stackrel{iid}{\sim} \mathcal{N}_K(0, \Sigma) \forall t$ und $\Pi = \alpha\beta'$. Die Lageinflüsse sind durch die Residuen weiterhin in der Likelihoodfunktion enthalten, was man durch Umstellen der Gleichung (2.32) sehen kann (vgl. die Darstellung von ϵ_t (2.27)).

Bemerkung 2.7.3 (Spezialfall) *Geht man im Abschnitt 2.7.2 von einem VAR(1)-Modell aus, ist der hier beschriebene erste Schritt nicht notwendig, da es ohnehin keine Lageinflüsse gibt. Neusser (2006, Kap. 16.3) betrachtet deshalb ab dieser Stelle ohne Einschränkung der Allgemeinheit das VAR(1)-Modell*

$$\Delta Y_t = \Pi Y_{t-1} + \epsilon_t.$$

Wir halten uns jedoch an Johansen und arbeiten mit den Residuen weiter. Es ist jedoch für die Anschauung nützlich, diesen Spezialfall bei der weiteren Betrachtung in Erinnerung zu behalten.

2. Schritt: Schätzung der Matrix β

Durch die Schätzung der Lag-Einflüsse wird die Likelihood-Funktion über die Γ_i bereits maximiert, so dass sie nun nur noch von den unbekannt Parametermatrizen α, β und Σ abhängt. Die (Schätzwerte der) Lageinflüsse sind in den Residuen $R_{\Delta,t}$ und $R_{L,t}$ enthalten, durch die Schätzung sind sie jedoch festgelegt. Die um die Lageinflüsse bereinigte Darstellung führt so durch Umstellen von Gleichung (2.32) und Einsetzen in die ursprüngliche Darstellung der Likelihoodfunktion (2.28) zu der konzentrierte Likelihoodfunktion (Johansen, 1988,

S.234)

$$L(\alpha, \beta, \Sigma) = (2\pi)^{-\frac{TK}{2}} (\det \Sigma)^{-\frac{T}{2}} \cdot \exp \left(-\frac{1}{2} \sum_{t=1}^T \underbrace{(R_{\Delta,t} + \alpha\beta' R_{L,t})}_{\epsilon_t}' \Sigma^{-1} \underbrace{(R_{\Delta,t} + \alpha\beta' R_{L,t})}_{\epsilon_t} \right).$$

Sie ist nur von der Ladungsmatrix α , der Kointegrationsmatrix β und der Kovarianzmatrix der Fehler Σ abhängig. Die Residuen $R_{\Delta,t}, R_{L,t}$ sind durch den beobachteten Prozess vorgegeben. Ziel ist es nun, die (unbekannten) Werte für α, β und Σ so zu wählen, dass der Wert $L(\alpha, \beta, \Sigma)$ maximiert wird (vgl. Anhang C.3.3, S.352).

An dieser Stelle betrachten viele Autoren (Stier, 2001, Neusser, 2006) bereits die logarithmierte Likelihoodfunktion

$$\mathcal{L}(\alpha, \beta, \Sigma) = -\frac{TK}{2} \ln(2\pi) - \frac{T}{2} \ln(\det \Sigma^{-1}) - \frac{1}{2} \sum_{t=1}^T (R_{\Delta,t} + \alpha\beta' R_{L,t})' \Sigma^{-1} (R_{\Delta,t} + \alpha\beta' R_{L,t}).$$

Wegen der Monotonie des Logarithmus wird \mathcal{L} genau dann maximal, wenn L maximal wird (Krengel, 2003, Kap. 4.2). Die logarithmierte Darstellung hat hier den Vorteil, dass die Exponentialfunktion aus der Darstellung verschwindet. Wir werden später die nützliche Eigenschaft verwenden, dass durch Logarithmieren aus Produkten Summen werden³⁵. Da wir an dieser Stelle jedoch mit der Determinante weiterarbeiten wollen, verschieben wir das Logarithmieren auf später, um die Betrachtung von $\det \Sigma$ nicht unnötig kompliziert zu gestalten.

Falls β bekannt wäre, sind Schätzer für α und Σ über die Produktmoment-Matrizen der Residuen durch

$$\begin{aligned} \hat{\alpha}(\beta) &= -S_{01}\beta(\beta'S_{11}\beta)^{-1} \\ \hat{\Sigma}(\beta) &= S_{00} - S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta'S_{10} \end{aligned} \quad (2.33)$$

gegeben. Dabei können die vier Kovarianzmatrizen S_{ij} aus dem beobachteten Daten (Residuen der Hilfsgleichungen) geschätzt werden (Johansen, 1988, Gleichung (8), Neusser, 2006, S.225, Stier, 2001, S.331)

$$\begin{aligned} S_{00} &= \frac{1}{T} \sum_{t=1}^T R_{\Delta,t} R'_{\Delta,t} & S_{11} &= \frac{1}{T} \sum_{t=1}^T R_{L,t} R'_{L,t} \\ S_{01} &= \frac{1}{T} \sum_{t=1}^T R_{\Delta,t} R'_{L,t} & S_{10} &= \frac{1}{T} \sum_{t=1}^T R_{L,t} R'_{\Delta,t} \end{aligned}$$

Diese Schätzer für α und Σ sind die bekannten OLS-Schätzer (z.B. Werner, 1997; Seber & Lee, 2003), die wegen der Normalverteilung der Fehler auch den Maximum-Likelihood-Schätzern entsprechen. Setzt man diese Schätzer in die Likelihoodfunktion (2.33) (vgl. Neusser, 2006,

³⁵ $\log(a \cdot b) = \log(a) + \log(b)$

S.225) ein

$$\begin{aligned} L(\beta) &= L(\hat{\alpha}(\beta), \beta, \hat{\Sigma}(\beta)) \\ &= \underbrace{(2\pi)^{-\frac{TK}{2}}}_{\text{konstant}} \left(\det \hat{\Sigma}(\beta) \right)^{-\frac{T}{2}} \cdot \underbrace{\exp\left(\frac{TK}{2}\right)}_{\text{konstant}}, \end{aligned}$$

ist sie nur noch von der Kointegrationsmatrix β abhängig. Durch das Einsetzen von $\hat{\alpha}(\beta)$ wird das Argument der Exponentialfunktion stark vereinfacht

$$-\frac{1}{2} \sum_{t=1}^T (R_{\Delta,t} + \alpha\beta' R_{L,t})' \Sigma^{-1} (R_{\Delta,t} + \alpha\beta' R_{L,t}) = \frac{TK}{2}. \quad (2.34)$$

Diese Umformung wird im Anhang C.3.4 ähnlich wie bei Neusser (2006, S.225) gezeigt.

Entscheidend ist, dass die Likelihoodfunktion $L(\beta)$ umgekehrt proportional zur Determinante von $\hat{\Sigma}(\beta)$ ist.

$$L(\beta) \sim \frac{1}{\det(\hat{\Sigma}(\beta)^{T/2})} \sim \frac{1}{\det(\hat{\Sigma}(\beta))}$$

Wenn man die Likelihoodfunktion maximieren will, ist dies also gleichbedeutend damit, dass man die Determinante durch Wahl von β minimiert. Dadurch wird auch der Term mit $\frac{T}{2}$ potenziert minimal. Anschaulich versucht man also ein Modell zu wählen, in dem die Fehlervarianz möglichst klein ist.

Bemerkung 2.7.4 (Minimum der Determinante) *Das Minimum der Determinante in Abhängigkeit von β (vgl. die Schätzung von Σ (2.33)) ist gegeben durch*

$$\det \hat{\Sigma}_{\min}(\beta) = \det(S_{00}) \prod_{i=1}^K (1 - \lambda_i)$$

dabei sind λ_i die kanonischen Korrelationen zwischen $R_{\Delta,t}$ und $R_{L,t}$.

Der mathematisch sehr anspruchsvolle Beweis ist der Vollständigkeit wegen in Anhang C.3.5 (ab S.353) ausgeführt.

Kanonische Korrelation Die kanonische Korrelation beschreibt den Zusammenhang zwischen den Linearkombinationen der Komponenten eines p -dimensionalen Zufallsvektors $a = (a_1, \dots, a_p)$ und den Linearkombinationen eines q -dimensionalen Zufallsvektors $b = (b_1, \dots, b_q)$ (z.B. Thompson, 1998; Bortz, 2005, Kap.19). In unserem Fall sind a und b die Residuen $R_{\Delta,t}, R_{L,t}$, in denen die Information steckt, die nur durch die Kointegrationsbeziehungen erklärt werden kann.

Um die kanonischen Korrelationen zu berechnen, geht man zunächst davon aus, dass der

Erwartungswert jedes Vektors (und damit jeder Komponente) Null ist ($\mathbb{E}[a] = \mathbb{E}[b] = 0$). Ist dies nicht der Fall, werden die um ihren Erwartungswert korrigierten Variablen betrachtet (Stier, 2001, 19.2.2). Die Variablen der beiden Vektoren können sowohl innerhalb eines Vektors, als auch zwischen den beiden Vektoren korreliert sein. So erstellt man in einem ersten Schritt eine sogenannte *Supermatrix* (Bortz, 2005, S.628) von bivariaten Kovarianzen

$$S = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}, \quad (2.35)$$

wobei Σ_{aa} und Σ_{bb} die $p \times p$ - bzw. $q \times q$ -Kovarianzmatrizen der Vektoren a und b und $\Sigma_{ab} = \Sigma'_{ba}$ die Kreuzkovarianzmatrizen sind.

Man versucht nun Linearkombinationen von a_1, \dots, a_p und Linearkombinationen von b_1, \dots, b_q zu finden, die maximal miteinander korrelieren. Die Linearkombinationen von a fasst man im Vektor $\eta = C'a$ zusammen, die von b im Vektor $\gamma = D'b$. Dabei sollen die Matrizen C und D so gewählt werden, dass alle Linearkombinationen in η und γ eine Varianz von Eins haben, die Korrelation der Linearkombinationen untereinander jedoch maximal wird. Mathematisch ist dies über folgende Bedingungen gegeben (Stier, 2001, S.330):

$$\begin{aligned} \mathbb{E}[\eta\eta'] &= C'\Sigma_{aa}C = \mathbb{I}_n \\ \mathbb{E}[\gamma\gamma'] &= D'\Sigma_{bb}D = \mathbb{I}_n \\ \mathbb{E}[\gamma\eta'] &= D'\Sigma_{ba}C = R, \end{aligned}$$

wobei

$$R = \begin{pmatrix} r_1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & r_n \end{pmatrix}$$

eine Diagonalmatrix ist, deren Einträge man *kanonische Korrelationskoeffizienten* von a und b nennt. Die Anzahl n der kanonischen Korrelationen entspricht der Anzahl von Komponenten im kürzeren Vektor $n = \min(p, q)$ (Bortz, 2005, S.629), da es maximal so viele Zusammenhänge geben kann, wie der kürzere Vektor Komponenten hat.

Die kanonischen Korrelationskoeffizienten r_i und die Matrizen C und D , deren Spaltenvektoren *kanonische Variablen* heißen, kann man mit Hilfe der Teilmatrizen von S aus (2.35) berechnen. Und zwar sind sie die Eigenwerte und (normierten) Eigenvektoren folgender Eigenwertprobleme:

$$\begin{aligned} (\Sigma_{aa}^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}) c_i &= \lambda_i c_i, \quad i = 1, 2, \dots, p \\ (\Sigma_{bb}^{-1}\Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}) d_i &= \mu_i d_i, \quad i = 1, 2, \dots, q \end{aligned}$$

mit der Normierung $c_i' \Sigma_{aa} c_i = 1$ und $d_i' \Sigma_{bb} d_i = 1$. Hier bezeichnet c_i eine Spalte der Matrix C und d_i analog eine Spalte der Matrix D . Dabei gilt für die n größten Eigenwerte Gleichheit $\lambda_i = \mu_i$, so dass es ausreicht eines der Eigenwertprobleme zu lösen.

Außerdem sind die Eigenwerte λ_i die quadrierten kanonischen Korrelationskoeffizienten $\lambda_i = r_i^2$. Es ist üblich, die quadrierten kanonischen Korrelationen absteigend zu sortieren

$$1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0.$$

Wenn man die (quadrierten) kanonischen Korrelationen von $R_{\Delta,t}$ und $R_{L,t}$ berechnet, sucht man also vereinfacht den Zusammenhang zwischen Linearkombinationen von ΔY_t und Linearkombinationen von Y_{t-1} (nach Bereinigung). Da ΔY_t stationär ist, gilt dies auch für jede Linearkombination der Einzelprozesse. Die Suche nach einer Linearkombination der (integrierten!) Einzelprozesse Y_{t-1} , die mit einer Linearkombination von ΔY_t korreliert, ist also gleichzeitig eine Suche nach einer *stationäre* Linearkombination, also nach einem Kointegrationsvektor.

Die Linearkombination (die kanonische Variable, der Kointegrationsvektor β_1), die zum größten kanonischen Korrelationskoeffizienten $\sqrt{\lambda_1}$ gehört, ist gleichzeitig die Linearkombination, die einer stationären Zeitreihe am nächsten kommt (Neusser, 2006, Anmerkung 16.2). Da die Eigenvektoren zu weiteren kanonischen Korrelationskoeffizienten orthogonal zu den bereits bekannten stehen, ist die Suche nach einer weiteren (stationären) Linearkombination von Y_{t-1} , die mit einer von ΔY_t korreliert, auch die Suche nach einer von β_1 (linear) unabhängigen Kointegrationsbeziehung.

Da ΔY_t und Y_{t-1} beide die Dimension K besitzen, ist S aus (2.35) auf Seite 133 in diesem Fall eine $K \times K$ -Matrix, die dann auch K Eigenwerte besitzt. Man kann auf diese Weise also K mögliche Kointegrationsvektoren schätzen. Da der Kointegrationsrang gleich der Dimension des Kointegrationsraumes ist, bilden die ersten r Eigenvektoren, die zu den r größten Eigenwerten gehören, seine Basis (also zusammengesetzt auch die Matrix β). Für jedes r ist somit die Matrix $\beta^{(r)}$ bestimmt. Wenn wir r schätzen können, ist β als $\beta^{(r)}$ festgelegt.

3. Schritt: Schätzung des Kointegrationsranges

Die Schätzung des Kointegrationsranges $\text{Rang}(\Pi) = r$ ist gleichbedeutend mit der Schätzung der Dimension des Kointegrationsraumes, also der Anzahl der linear unabhängigen Vektoren β_i , die eine Kointegrationsbeziehung beschreiben.

Da die Kointegrationsvektoren als kanonische Variablen gleichzeitig Eigenvektoren aus einem Eigenwertproblem sind³⁶, ist die Problemstellung gleichbedeutend mit der Frage, wie viele der Eigenwerte (kanonischen Korrelationen) λ_i von Null verschieden sind.

³⁶Das Eigenwertproblem tritt im Beweis von Bemerkung 2.7.4 im Anhang C.3.5 auf.

Die maximale Likelihoodfunktion hat bei Kointegrationsrang r nach Verwendung von Bemerkung 2.7.4 (S.132) die Gestalt

$$L^{\max}(\beta) = \exp\left(\frac{TK}{2}\right) (2\pi)^{-\frac{TK}{2}} \left(\det(S_{00}) \prod_{i=1}^K (1 - \lambda_i)\right)^{-\frac{T}{2}},$$

in der man die Abhängigkeit von den kanonischen Korrelationskoeffizienten λ_i erkennen kann. Nun berechnen wir die logarithmierte maximale Likelihoodfunktion und verwenden dabei wie bereits bemerkt, dass beim Logarithmieren aus Produkten Summen werden (siehe Anhang C.1.3, S.341)

$$\mathcal{L}^{\max}(\beta) = \underbrace{\frac{TK}{2} - \frac{TK}{2} \ln(2\pi) - \frac{T}{2} \det S_{00}}_{\text{konstant}} - \frac{T}{2} \sum_{i=1}^K \ln(1 - \lambda_i). \quad (2.36)$$

Da der Logarithmus von Eins gleich Null ist, $\ln(1) = 0$, wird die Summe $\sum_{i=1}^K \ln(1 - \lambda_i)$ gleich Null, falls dies für $\sum_{i=1}^K \lambda_i$ zutrifft³⁷. Sind einige der kanonischen Korrelationen λ_i gleich Null, hat die Summe weniger als K Summanden. Da bei bekanntem Kointegrationsrang genau r dieser Werte von Null verschieden sind, hat die Summe in Gleichung (2.36) genau r Summanden. Deshalb kann man die Likelihoodfunktion auch in Abhängigkeit vom Kointegrationsrang r darstellen.

$$\mathcal{L}_r^{\max}(\beta) = \underbrace{\frac{TK}{2} - \frac{TK}{2} \ln(2\pi) - \frac{T}{2} \det S_{00}}_{\text{konstant}} - \frac{T}{2} \sum_{i=1}^r \ln(1 - \lambda_i). \quad (2.37)$$

Da alle K möglichen Eigenwerte (quadrierte kanonische Korrelationen) bekannt sind, ist die Likelihoodfunktion nur noch vom Kointegrationsrang r abhängig. Es gilt also zum Schluß noch, diesen zu bestimmen. Dies geschieht durch das sukzessive Testen folgender Hypothesen.

Hypothesen Man formuliert eine Folge von Hypothesen für $r = 0, 1, \dots, K$

$$H(r) : \text{Rang}(\Pi) \leq r.$$

Die Hypothese $H(r)$ besagt also, dass es höchstens r Kointegrationsbeziehungen gibt. Dies ist gleichbedeutend damit, dass nur die ersten r kanonischen Korrelationen von Bedeutung sind, also dass nur die ersten r der λ_i statistisch signifikant von Null verschieden sind. Dies bedeutet im Übrigen, dass bereits $\lambda_{r+1} = 0$ ist und dass auch die Summe der verbleibenden Eigenwerte $\lambda_{r+1} + \dots + \lambda_K = 0$ ist.

³⁷Eine Summe aus nicht-negativen Summanden wird nur genau dann Null, wenn alle Summanden gleich Null sind.

Teststrategie Es gibt nun zwei verschiedene Arten mit diesen Hypothesen auf den Kointegrationsrang von Y_t zu schließen. Beiden ist gemeinsam, dass man sukzessiv die Nullhypothesen $H(r)$ für $r = 0, 1, \dots$ testet. Sie unterscheiden sich jedoch in der Alternativhypothese. Im ersten Fall testet man $H(r)$ stets gegen $H(K)$, beim zweiten Verfahren verwendet man $H(r + 1)$ als Alternative.

In beiden Fällen betrachtet man als Teststatistik die Differenz der beiden (Maximalwerte der) Likelihoodfunktionen als Teststatistik. Im ersten Fall ergibt sich (Johansen, 1988, Theorem 1)

$$\mathcal{L}_K^{\max} - \mathcal{L}_r^{\max} = -\frac{T}{2} \sum_{i=r+1}^K \ln(1 - \lambda_i). \quad (2.38)$$

Die asymptotische Verteilung dieser Teststatistik ist gegeben durch die *Spur* der folgenden stochastischen Matrix (Stier, 2001, S.334), (Johansen, 1988, Theorem 3)

$$Q = \left(\int_0^1 B(r)dB(r)' \right) \left(\int_0^1 B(r)dB(r)' \right)^{-1} \left(\int_0^1 B(r)dB(r)' \right),$$

wobei $B(r)$ eine multivariate Brownsche Bewegung ist (siehe Stier, 2001, S. 334). Weshalb dieser Test auch *Johansen-trace-Test* genannt wird.

Im zweiten Fall folgt die Teststatistik

$$\mathcal{L}_{r+1}^{\max} - \mathcal{L}_r^{\max} = -\frac{T}{2} \ln(1 - \lambda_{r+1}) \quad (2.39)$$

einer asymptotischen Verteilung, die durch den maximalen Eigenwert von Q gegeben ist, wodurch der Test den Namen *Johansen-max-Test* erhalten hat.

Da alle kanonischen Korrelationen positiv $\lambda_i \geq 0$ sind, sind die Statistiken aus (2.38) bzw. (2.39) stets größer Null. Die Summe $\sum_{i=1}^r \ln(1 - \lambda_i)$ wird genau dann Null, wenn alle $\lambda_i = 0$ sind, also auch für ihre Summe gilt $\sum_{i=1}^K \lambda_i = 0$. Deshalb reicht es, sich im Folgenden auf die Summe der kanonischen Korrelationen zu konzentrieren.

Beim Johansen-trace-Test testet man stets gegen die Alternative $H(K)$. Dies ist gleichbedeutend damit, es K stationäre Zusammenhänge von ΔY_t und Y_{t-1} gibt und dass ΔY_t stationär ist. Dann sind alle Eigenwerte λ_i größer als Null. Auf jeden Fall ist es die Gesamtsumme $\lambda_1 + \dots + \lambda_K > 0$.

$H(r)$ wird solange abgelehnt, bis die verbleibenden Eigenwerte so klein sind, dass ihre Summe $\lambda_{r+1} + \dots + \lambda_K$ (statistisch) nahe Null ist. Wird $H(r)$ angenommen, werden natürlich auch alle weiteren Hypothesen $H(r + 1), \dots, H(K - 1)$ angenommen, da sie gleichbedeutend damit sind, dass die Summe der restlichen Eigenwerte gleich Null ist. Gilt dies schon für die $K - r$ letzten, ist dies trivialerweise auch für die Summe von weniger übrigen Eigenwerten richtig. Deshalb kann man das Testverfahren abbrechen, wenn eine Nullhypothese $H(r)$ angenommen

Tabelle 2.6: Teststrategie des Johantests: sukzessives Testen der Alternativhypothese $H(r)$ gegen die Nullhypothese $H(K)$ (Trace-Test) oder $H(r+1)$ (Max-Test)

1.Schritt: $H(0)$ gegen $H(K)$ bzw. $H(0)$ gegen $H(r+1)$			
$H(0):$	$\text{Rang}(\Pi) \leq 0$	$\lambda_1 + \lambda_2 + \dots + \lambda_K = 0$	$\Leftrightarrow \lambda_1 = 0$
$H(K):$	$\text{Rang}(\Pi) \leq K$	$\lambda_1 + \lambda_2 + \dots + \lambda_K > 0$	$\Rightarrow \lambda_1 > 0 = H(1)$
2.Schritt: $H(1)$ gegen $H(K)$			
$H(1):$	$\text{Rang}(\Pi) \leq 1$	$\lambda_2 + \dots + \lambda_K = 0$	$\Leftrightarrow \lambda_2 = 0$
$H(K):$	$\text{Rang}(\Pi) \leq K$	$\lambda_2 + \dots + \lambda_K > 0$	$\Rightarrow \lambda_2 > 0 = H(r)$
\vdots			
$r+1$.Schritt: $H(r)$ gegen $H(K)$			
$H(r):$	$\text{Rang}(\Pi) \leq r$	$\lambda_{r+1} + \dots + \lambda_K = 0$	$\Leftrightarrow \lambda_{r+1} = 0$
$H(K):$	$\text{Rang}(\Pi) \leq K$	$\lambda_{r+1} + \dots + \lambda_K > 0$	$\Rightarrow \lambda_{r+1} > 0 = H(r+1)$

wurde. Der Kointegrationsrang von Y_t ist dann gleich dem r der letzten Nullhypothese.

Die Kointegrationsmatrix β besteht dann aus den r ersten kanonischen Variablen und ist als $\beta^{(r)}$ festgelegt.

4. Schritt: Schätzung der Modellparameter

Durch die Normierung der kanonischen Variablen β_i durch $\beta' S_{11} \beta = \mathbb{I}$ (vgl. Anhang C.3.5, S.353) ergeben sich Vereinfachungen der Schätzer für einige Modellparameter. So sind die Maximum-Likelihood-Schätzer der Ladungsmatrix α , Π und der Kovarianzmatrix des Fehlerprozesses Σ (Johansen, 1988, S.235f) gegeben durch

$$\begin{aligned}\hat{\alpha}(\beta) &= -S_{01}\beta(\beta'S_{11}\beta)^{-1} &= -S_{01}\beta \\ \hat{\Pi}(\beta) &= -S_{01}\beta(\beta'S_{11}\beta)^{-1}\beta' &= -S_{01}\beta\beta' \\ \hat{\Sigma}(\beta) &= S_{00} - S_{01}\beta\beta'S_{10} &= S_{00} - \hat{\alpha}\hat{\alpha}'.\end{aligned}$$

Die Maximum-Likelihood-Schätzungen für die Γ_i ergeben sich aus den Hilfsgleichungen durch (2.29) - $\hat{\Pi}$ (2.30) = $\Delta Y_t - \hat{\Pi}Y_{t-1}$ (Stier, 2001, S.332):

$$\hat{\Gamma}_i = \hat{\zeta}_i - \hat{\Pi}\hat{\xi}_i, \quad i = 1, 2, \dots, p-1.$$

Damit wird durch die Anwendung des Johantests nicht nur der Kointegrationsrang ermittelt, sondern das gesamte Fehlerkorrekturmodell geschätzt.

In Kapitel 3 werden wir diesen Test verwenden, um zu prüfen, welche Qualität psychologische

Daten haben müssen, damit eine vorhandene Kointegrationsbeziehung erkannt werden kann. Zunächst wenden wir den Johansenstest jedoch bei der Analyse eines empirischen Datensatzes an.

2.8 Couples in Collusion

Kointegrationsanalyse in psychologischer Praxis

Wenn man die Einführung in die Kollusionstheorie von Jürg Willi (1984) liest, wird man unweigerlich an die Darstellung der Kointegrationsanalyse erinnert. Mit *Kollusion* wird eine unbewusste *geheime Absprache* gemeint, die in gemeinsamen Einverständnis dafür sorgt, dass durch ein „neurotisches Zusammenspiel“ die ungelösten Grundkonflikte beider Partner in einer Art „Gleichgewichtsbalance“ scheinbar gelöst werden. Der Erfolg dieser Beziehung begründet sich aus der ständigen „Anpassung“ beider Partner, um das Gleichgewicht nicht zu gefährden. Durch dieses Anpassungsverhalten entsteht ein gemeinsamer Entwicklungstrend.

2.8.1 Kollusionstheorie

Die Psychoanalyse geht davon aus, dass eine Ehe nur dann funktionieren kann, wenn Mann und Frau als gleichberechtigte Partner zusammenleben und ständig an einer reifen Beziehung arbeiten (Meissner, 1978). Tatsächlich beobachtet man jedoch auch stabile und sogar glückliche Beziehungen zwischen Partner mit ungelösten neurotischen Konflikten. Mit *Kollusion* wird ein unbewusstes Zusammenspiel bezeichnet, in dem sich die Verhaltensweisen so ergänzen, dass der gleichartige unbewältigte Grundkonflikt in je einer polarisierten Variante ausgelebt werden kann (Willi, 1984, S.1).

Willi (1984) beschreibt eine Partnerschaft als „sich entwickelnden fortschreitenden Prozess“ oder auch „Koevolution“. Da eine Beziehung das individuelle Verhalten beider Partner formt und beeinflusst, kann die Persönlichkeitsstruktur oder das Verhalten nur während der Interaktion beurteilt werden (Willi, 1984, S.2). Ohne den Partner lässt sich das Verhalten der Einzelpersonen nicht erklären. Die Beziehung beeinflusst, welches Verhaltenspotential in der Partnerschaft ausgelebt wird. Willi (1982) prägt dafür den Begriff *interaction personality*. In einer „geheimen“ Absprache passen sie ihr Verhalten so an, dass ein Partner stets „überfunktionierte“, während der andere Partner „unterfunktionierte“. Durch die komplementären Ergänzungsrollen wird das Gleichgewicht der Partnerschaft aufrecht erhalten. So nimmt ein Partner stets die „progressive“ Rolle ein, während der andere Partner die „regressive“ Variante auslebt.

Dabei werden vier Grundmuster einer Kollusion beschrieben, die aus Themen resultieren, mit denen sich auch gesunde Partner auseinander setzen müssen. Der *narzisstische Beziehungskonflikt* beschreibt den Gegensatz von Entwicklung eines Selbstwertfühls in Abgrenzung

der Verwirklichung durch den Partner. Im Zentrum des *oralen Beziehungsthemas* steht die Verteilung von Fürsorge und das Übernehmen von Verantwortung, während das *anale Beziehungsthema* den Konflikt zwischen Kontrolle und Abhängigkeit meint. Ob in der Beziehung „klassische“ Geschlechterrollen gelebt werden ist Kern des *phallischen Beziehungsthemas*.

2.8.2 Kollusion in der Psychosomatik

In einer klinischen Studie an der Universität Gießen wurde während der Dauer einer Paartherapie ein Datensatz bestehend aus verschiedenen Zeitreihen mit 144 täglichen Messungen erhoben und erstmals durch Kupfer, Brosig und Brähler (2005) zeitreihenanalytisch ausgewertet.

Seit Trennung einer früheren langjährigen Beziehung litt die 26-jährige Ehefrau seit 2,5 Jahren unter Bulimia nervosa und befand sich aus diesem Grund zusätzlich in einer Einzeltherapie, während der vor allem die schwierige Beziehung zu ihrer Mutter und der Umgang mit Gefühlen von Einsamkeit aufgearbeitet wurden. Der männliche 22-jährige Partner war etwa zeitgleich mit dem Beginn der gemeinsamen Beziehung an einer atopischen Dermatitis (auch bekannt als Neurodermitis) erkrankt. Weitere Details zum Therapieverlauf sind bei Kupfer et al. (2005) beschrieben.

Da Bulimie und Neurodermitis häufig mit bestimmten psychologischen Merkmalen wie Aggressivität und Depressivität in Verbindung gebracht werden, wurden neben den körperlichen Beschwerden auch der Verlauf dieser psychologischen Variablen erhoben. In Form eines strukturierten Tagebuches füllten beide Partner an allen Tagen im Erhebungszeitraum einen Stimmungsfragebogen (MSF) aus, der aus 58 Adjektiven besteht („Mehrdimensionaler Stimmungsfragebogen, Hecheltjen & Mertesdorf, 1973). Zusätzlich wurde zur Einschätzung der körperlichen Symptome der Gießener Beschwerdebogen (GGB) ausgefüllt (Brähler & Scheer, 1983), der aus 57 Items mit den Stufen 0 - 4 besteht. Für die körperlichen Symptome zeigte sich bei dem Ehemann einzig das Item „Itching“ als aussagekräftig. Für die Beschwerden der Frau wurde die Variable „Druck und Schweregefühl im Magen“ gewählt, da „Erbrechen“ von ihr als zu offensichtliches Symptom der Bulimie abgelehnt wurde. Aus dem Stimmungsfragebogen wurden anhand einer P-Faktor-Analyse (Cattell, 1966) jeweils drei individuelle Dimensionen bestimmt, die zu individuellen Stimmungsskalen führten.

So besteht der folgende Datensatz aus jeweils vier Zeitreihen für jeden Partner

- Aktivität
- Aggressivität
- Depressivität
- Körperliches Symptom

Die erste zeitreihenanalytische Auswertung fand mit Hilfe von Autokorrelationen und Kreuzkorrelationen (siehe Kapitel 1.9, S.35) statt (Kupfer et al., 2005). Die dazu verwendeten VAR-Modelle werden jedoch von den Autoren nicht angegeben. Die Verläufe der beiden körperlichen Symptome wurden klar als integriert identifiziert. Bei der Autoregression wurden jeweils die ersten sieben Lags berücksichtigt. Nach diesen Auswertungen wird die Intensität der Neurodermitis am besten durch die eigene depressive Stimmung am selben Tag, sowie die Aggressivität der Ehefrau am Vortag erklärt. Außerdem konnte ein Zusammenhang zur eigenen Aggressivität am folgenden Tag festgestellt werden. Ihre körperlichen Anzeichen konnten nur über eigene Variablen wie der Depressivität desselben und der Aggressivität am Vortag erklärt werden. Außerdem stand sie im Gegensatz zur Aktivität des Mannes am nächsten Tag. Zusätzlich konnte eine hohe Korrelation für die Lags 1 und 7 festgestellt werden.

2.8.3 Interpretation durch Fehlerkorrekturdarstellung

Durch die Anwendung von Kointegrationsanalyse ist es möglich, unbewusste gemeinsame Trends im Sinne von Willis Konzept der *couple collusion* in den Systemen von psychologischen und psychosomatischen Zeitreihen zu erfassen und das langfristige Gleichgewicht, die „Balance“, die aus dem unbewussten Zusammenspiel beider Partner entsteht, ebenso wie kurzfristige Anpassungsdynamiken zu beschreiben.

Für die folgenden Variablen wurde in der univariaten Analyse eindeutig ein integrierter Verlauf festgestellt

- Derm - Dermatitis
- Bul - Bulimia
- AH - seine Aggressivität
- AW - ihre Aggressivität
- DH - seine Depressivität
- DW - ihre Depressivität,

so dass sie zur weiteren Auswertung mittels Johansenstest und Fehlerkorrekturdarstellung berücksichtigt wurden.

Für die Auswertung wurde das benutzerfreundliche Analysetool *JMulti* verwendet (Lütkepohl & Krätzig, 2004), mit dem es möglich ist eine Übersicht über das gesamte Fehlerkorrekturmodell zu erhalten. Für alle Kointegrationen wurde eine Konstante in die Kointegrationsbeziehung aufgenommen, um den Mittelwertsunterschieden gerecht zu werden (siehe Abbildung 2.14, S.144). In *JMulti* werden vier verschiedene Informationskriterien berechnet. Aus dieser Angabe lässt sich entnehmen, wie viele Lags in das Fehlerkorrektur aufgenommen werden sollten (vgl. Kapitel 1.11, S.41). Da sich dieser Wert zwischen den Informationskriterien

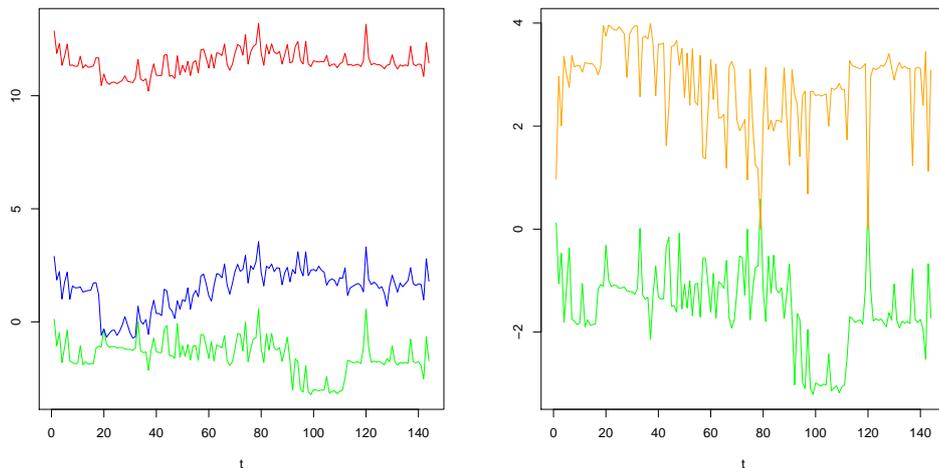


Abbildung 2.13: Links: Kointegriertes System: Bulimie (rot), seine Aggressivität (blau), seine Depressivität (grün); rechts: Kointegriertes System: Dermatitis (orange) und seine Depressivität (grün).

unterscheiden kann, wurde für einige Systeme die Schätzung für verschiedene Laganzahlen durchgeführt.

Der zentrale Aspekt der Analyse lag auf dem Zusammenhängen der körperlichen Symptome mit den anderen Variablen.

Bulimia nervosa Von den 15 möglichen bivariaten Systemen³⁸ wurden zunächst jene untersucht, die die Zeitreihe Bu_1 beinhalten, um Zusammenhänge mit dem Verlauf der Bulimia nervosa aufzudecken. Für die Systeme mit DW, AH und DH wurde ein Kointegrationsrang von $r = 1$ festgestellt.

Für die Schätzung des Fehlerkorrekturmodells des Systems BUL - DW wird die Hinzunahme von 1 bzw. 2 Lags empfohlen. In beiden Fällen wird die Kointegrationsbeziehung signifikant auf $Z_t = BUL_t - 0.6 DW_t - 11$ geschätzt. Nur für Bulimie wurden der Einfluss der Lags signifikant geschätzt. Die Zeitreihe der Bulimie und der Depressivität der Ehefrau folgen also einem gemeinsamen stochastischem Trend. Das negative Vorzeichen von β_0 bedeutet einen synchronen Verlauf. Die Anzeichen der Bulimie ist stärker, wenn die Depressivität größer ist und verbessern sich bei weniger niedergeschlagener Stimmung. Der Mittelwertsunterschied -11 ist durch die unterschiedlichen Skalen begründet, die zur Erhebung der Daten verwendet wurden, siehe Abbildung 2.13. Der Zusammenhang von Depressivität und ihren körperlichen Symptomen verwundert nicht, da anfangs bereits erwähnt wurde, dass die Klientin sich insbesondere mit Einsamkeit und Verlassensängsten plagt. Kupfer et al. (2005) kommen zu einem ähnlichen Resultat.

Auch für die Schätzung des Systems BUL - AH sollen 1 bis 2 Lags aufgenommen werden. Die

³⁸Es gibt 15 Möglichkeiten Paare aus sechs vorhandenen Variablen auszuwählen, $\binom{6}{2} = 15$.

Kointegrationsbeziehung wird auf $Z_t = \text{AH}_t - 2.45 \text{ BUL} + 26$ geschätzt. Bei Hinzunahme von Lags sind nur die eigenen Lags im Differenzenprozess für AH signifikant, was darauf hindeutet, dass seine aggressive Stimmung zusätzlich von seiner Stimmung am Vortag abhängt. Der Wert $\beta_0 = -2,45$ zeigt ein Zusammenhang zwischen seiner Aggressivität und ihrer Bulimie in gleicher Bewegungsrichtung an. Bei verstärkten bulemischen Symptomen zeigt er sich aggressiver.

Für das Fehlerkorrekturmodell zwischen BUL und DH wird der Einfluss der Lags als nicht signifikant gewertet, so dass hier das Fehlerkorrekturmodell ohne Lags dargestellt wird

$$\begin{pmatrix} \Delta \text{DH} \\ \Delta \text{BUL} \end{pmatrix} = \begin{pmatrix} -0.267 \\ -0.206 \end{pmatrix} \cdot \left[\begin{pmatrix} 1 & 1.926 \end{pmatrix} \begin{pmatrix} \text{DH}_{t-1} \\ \text{BUL}_{t-1} \end{pmatrix} - 20.553 \right] + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Die Kointegrationsbeziehung lautet folglich $Z_t = \text{DH}_t + 1.926 \text{BUL}_t - 20.553$. Sie beeinflusst beide Einzelreihen etwa gleich stark, $|\alpha_1| \approx |\alpha_2| \approx 0.2$. Sowohl α_1 als auch $\alpha_2\beta$ sind negativ, so dass Kointegration vorliegt (vgl. Tabelle 2.4, S.111). Je geringer ihre Symptome desto stärker seine Depressivität

$$\begin{aligned} \Delta \text{DH} &= -0.267 \cdot (\text{DH}_{t-1} + 1.926 \text{ BUL}_{t-1} - 20.553) + \epsilon_{1t} \\ \Delta \text{BUL} &= -0.206 \cdot (\text{DH}_{t-1} + 1.926 \text{ BUL}_{t-1} - 20.553) + \epsilon_{2t}. \end{aligned}$$

Ihre Bulimie hat also einen Einfluss auf seine Stimmung, sowohl auf die Depressivität also auch die Aggressivität. Betrachtet man nun den Zusammenhang dieser beiden Größen, so ergibt sich eine Kointegrationsbeziehung

$$\begin{pmatrix} \Delta \text{AH} \\ \Delta \text{DH} \end{pmatrix} = \left[\begin{pmatrix} -1.133 \\ -0.193 \end{pmatrix} \begin{pmatrix} 1 & 1.772 \end{pmatrix} \begin{pmatrix} \text{AH}_{t-1} \\ \text{DH}_{t-1} \end{pmatrix} + 1.230 \right] + \begin{pmatrix} -0.226 & 0 \\ 0 & -0.224 \end{pmatrix} \begin{pmatrix} \Delta \text{AH}_{t-1} \\ \Delta \text{DH}_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Seine Aggressivität und Depressivität haben einen gemeinsamen Trend jedoch umgekehrte Bewegungsrichtungen. Je aggressivier er ist, desto weniger depressiv. Zusätzlich hängen beide Variablen von der Intensität der jeweiligen Stimmung am Vortag ab.

Beide Variablen wiesen einen gemeinsamen Trend mit Bul auf. Nun ist es interessant das dreidimensionale System BUL - AH - DH zu betrachten. Die Analyse mit dem Johansenstest ergibt einen Kointegrationsrang von $r = 2$. Das Fehlerkorrekturmodell ohne Lags wird wie folgt geschätzt

$$\begin{pmatrix} \Delta \text{AH} \\ \Delta \text{DH} \\ \Delta \text{BUL} \end{pmatrix} = \left[\begin{pmatrix} -0.042 & -0.209 \\ 0.602 & 0.042 \\ 0.509 & 0.055 \end{pmatrix} \begin{pmatrix} 1 & 0 & -2.790 \\ 0 & 1 & 1.966 \end{pmatrix} \begin{pmatrix} \text{AH}_{t-1} \\ \text{DH}_{t-1} \\ \text{BUL}_{t-1} \end{pmatrix} + \begin{pmatrix} 30.506 \\ -21.014 \end{pmatrix} \right] + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}.$$

Man kann die Kointegrationsbeziehungen der obigen beiden bivariaten Systeme erkennen. Die drei Zeitreihen haben alle denselben integrierten Anteil, wie auch im linken Teil der Abbildung 2.13 zu sehen ist. Wenn man nur die signifikant von Null verschiedenen Schätzungen berücksichtigt, ergibt sich folgende Systemschreibweise

$$\begin{aligned}\Delta AH &= 0 \cdot (AH_{t-1} - 2.79BUL_{t-1}) - 0.2 \cdot (DH_{t-1} + 1.97BUL_{t-1}) \\ \Delta DH &= 0.6 \cdot (AH_{t-1} - 2.79BUL_{t-1}) + 0 \cdot (DH_{t-1} + 1.97BUL_{t-1}) \\ \Delta BUL &= 0.5 \cdot (AH_{t-1} - 2.79BUL_{t-1}) + 0 \cdot (DH_{t-1} + 1.97BUL_{t-1}).\end{aligned}$$

In dieser Darstellung ist jeder Differenzenprozess der Einzelreihen nur von einer Kointegrationsbeziehung abhängig. Da es sich jedoch um ein gemeinsames System handelt mit einem eindimensionalen Kointegrationsraum, bewegen sich die drei Prozesse entsprechend des gemeinsamen Trends, der ihre Fehlerkorrektur bestimmt. Im vierdimensionalen VECM mit zusätzlich DW gibt es zwei gemeinsame Trends. Ebenso in jedem dreidimensionalen System der vier Variablen, in dem DW enthalten ist. Ihre Depressivität enthält also noch einen weiteren Aspekt, der den anderen drei Zeitreihen nicht gemeinsam ist. Bei grafischer Betrachtung der Verläufe fällt jedoch eine starke Abweichung dieser Zeitreihe im Bereich um $t = 100$ auf. Bei einer Analyse des Systems bis zum Zeitpunkt $t = 90$ hat das vierdimensionale System eindeutig Kointegrationsrang 3. Alle vier Zeitreihen folgen also demselben stochastischen Trend. Auch die Kointegrationsbeziehungen mit seiner Depressivität $\beta \approx (1, -1)$ und seiner Aggressivität $\beta \approx (1, 0.3)$ sind signifikant vorhanden.

Mit einer ähnlichen Analyse kann man erkennen, dass seine Dermatitis im Zusammenhang mit seiner Aggressivität und seiner Depressivität steht, was sowohl im dreidimensionalen als auch in den jeweiligen zweidimensionalen Fehlerkorrekturmodellen gezeigt werden kann. So wird die Kointegrationsbeziehung zwischen Dermatitis und der Depressivität auf $Derm_t - 0.686 DH_t - 3.872$ geschätzt (siehe rechter Teil in Abbildung 2.13). Zwischen Dermatitis und Aggressivität besteht die Beziehung $Derm_t + 0.535 AH_t - 3.593$. Auch hier wird die Gegensätzlichkeit von Depressivität und Aggressivität erneut deutlich.

Analysiert man das „volle“ sechsdimensionale System wird der Kointegrationsrang über den gesamten Erhebungszeitraum auf $r = 4$ geschätzt. Nimmt man jedoch nur die ersten 90 Beobachtungen (siehe Abbildung 2.14, S.144), zeigt ein Kointegrationsrang von $r = 5$ auf dem 0.1 und 0.05 Niveau an, dass das gesamte System demselben gemeinsamen stochastischen Trend folgt.

Die körperlichen Symptome sind also wie die Stimmungen der beiden Ehepartner Teil einer gemeinsamen Kollusion.

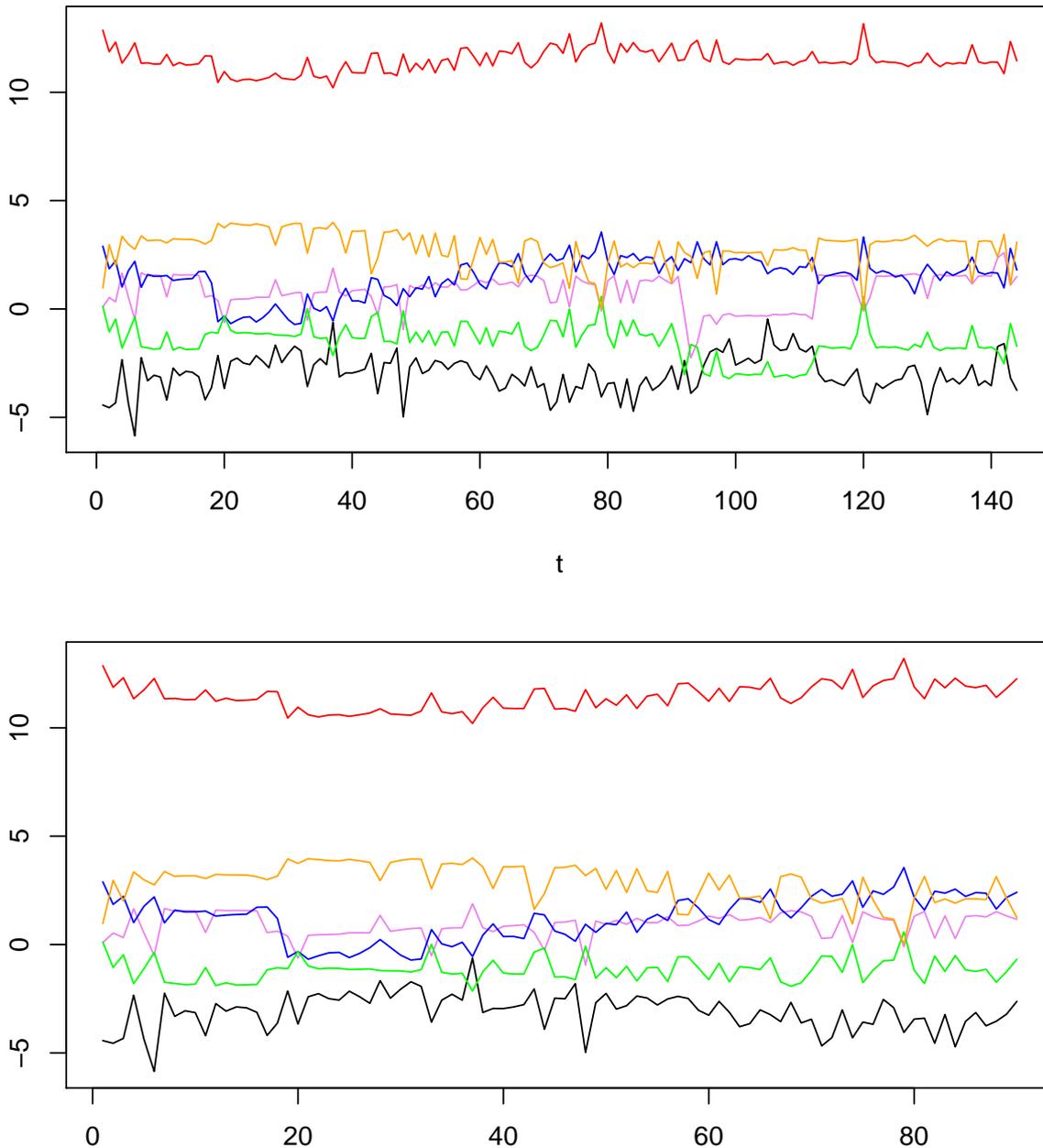


Abbildung 2.14: Oben: Das gesamte System mit allen 144 Messpunkten: Die beiden Depressivitätswerte bringen um $t = 100$ zusätzliche Information ins System; Unten: Das gesamte System bis zur Länge 90: Ein gemeinsamer Trend ist in allen Zeitreihen zu erkennen. Bulimie (rot), Dermatitis (orange), seine Aggressivität (blau), ihre Aggressivität (schwarz), seine Depressivität (grün) und ihre Depressivität (lila).

3 Simulationsstudien: Kointegration in psychologischen Daten

Für die Erkennung von kointegrierten Systemen hatten wir den Johansen-Test kennengelernt und ihn in Kapitel 2.8 bereits auf einen psychologischen Datensatz angewendet.

Bisher haben wir uns dabei über den Wertebereich E der Zeitreihen jedoch keine Gedanken gemacht. Allgemein findet man in der Literatur zur Zeitreihenanalyse kaum Aussagen dazu. Diese Frage ist jedoch bedeutend, um zu prüfen, ob die psychologische Forschung von den Erkenntnissen, die vor allem mit wirtschaftswissenschaftlichen Daten gewonnen wurden, überhaupt profitieren kann.

Bei der Operationalisierung von psychologischen Phänomenen ist es häufig notwendig ein geeignetes Messinstrument zu erstellen. Bei der rationalen Skalenkonstruktion geht man von einer Theorie aus, mit der sich Personen bzw. deren Verhalten oder andere Merkmale beschreiben und unterscheiden lassen (Amelang & Zielinski, 2004). Daraus werden verschiedene Kategorien von Verhaltensweisen, Eigenschaften oder Dispositionen gebildet. Für jede Kategorie werden weitere qualitative Abstufungen vorgenommen, die sich nach der Häufigkeit oder Intensität der beobachteten Unterschiede richten. „Im einfachsten Fall wird unterstellt, dass es möglich ist, jede Eigenschaft, für die es einen Namen gibt, zu messen.“ (Amelang & Zielinski, 2004, S.67).

Bei der Konstruktion spielt auch die Skalenqualität der Messung eine wichtige Rolle. So sind statistisch bedeutsame Mittelwertsunterschiede nur dann relevant, wenn das Merkmal auf einer Intervallskala gemessen wurde (Bortz & Döring, 1995). Viele ökonomische Zeitreihendaten werden in Geldeinheiten gemessen. Das führt zu einer sehr hohen Qualität der Daten, da man durch die Einteilung in Cent (und weitere Nachkommastellen) zum einen fast beliebig genaue Differenzierungsmöglichkeiten hat, zum anderen der Wertebereich uneingeschränkt groß ist. Ein Aktienkurs kann sehr stark wachsen und theoretisch auch negative Werte annehmen, was in der Praxis jedoch den Konkurs und damit das Ausscheiden der Aktie von der Börse bedeutet.

Für die Anwendung der Kointegrationsanalyse in der psychologischen Forschung ist es jedoch notwendig zu wissen, welche Anforderungen an den Wertebereich der Daten gestellt werden müssen. Hierbei müssen nicht zuletzt die Fragen nach der Skalenqualität und Stufenanzahl beantwortet werden, sondern es muss auch untersucht werden, wie man mit Phänomenen wie Deckeneffekten und Missing Values umgehen kann.

Für die meisten psychologischen Auswertungsverfahren werden intervallskalierte Daten ge-

fordert. Hier ist vor allem die Unterscheidung zu ordinalskalierten Daten entscheidend. Im Folgenden soll jedoch auch kurz auf weitere Skalenarten eingegangen werden.

In der *Einführung in die Theorie des Messens* definiert (Orth, 1974) den Vorgang des Messens durch: „Messen ist die Bestimmung der Ausprägung einer Eigenschaft eines Dinges. Messen erfolgt durch Zuordnung von numerischen Größen.“

Wenn durch die Zuordnung einer numerischen Größe nur die Unterscheidung der Merkmalsausprägungen bewirkt wird (wie z.B. Geschlecht: 1 - weiblich, 2- männlich), handelt es sich um eine Nominalskala³⁹ (Orth, 1974; Bortz, 2005). Dies ist bei Zuordnung zu verschiedenen Gruppen (z.B. Einteilung nach Geschlecht, Studienfach oder Wohnort) sinnvoll. Wenn durch die Nummerierung eine Rangreihe festgelegt wird, spricht man von einer Ordinalskala. Ein typisches Beispiel sind Schulnoten, die die Leistung in verschiedenen Aufgaben und Schulfächern bewerten. Im Gegensatz zur Intervallskala sind die Abstände zwischen zwei Skalenstufen jedoch inhaltlich nicht als gleich groß zu sehen. So ist der qualitative Unterschied in der Leistung zwischen den Schulnoten „mangelhaft“ und „ausreichend“ meist ein geringerer als zwischen den Skalenstufen „gut“ und „sehr gut“.

Anders ist dies bei intervallskalierten Daten. Wie bei Temperaturmessungen bedeutet die Differenz einer Skalenstufe jeweils den gleichen Zuwachs an Merkmalsausprägung. Im Vergleich zu einer Verhältnisskala sind Aussagen wie „es ist doppelt so heiß wie gestern“ jedoch nicht zulässig. Dies ist erst möglich, wenn die Skala einen absoluten Nullpunkt hat. Der Nullpunkt bei Temperaturmessungen ist willkürlich am Gefrierpunkt von Wasser festgelegt, während bei Längen- oder Zeitmessungen der Nullpunkt ganz natürlich vorgegeben ist, wodurch die Daten ein höheres Niveau erreichen, nämlich das der Verhältnisskala (z.B. Assenmacher, 2003, 2.2 Messen und Skalieren).

Auf ordinalskalem Niveau sind alle Transformationen der Daten, die die Rangreihe bzw. Ordnungsstatistik nicht verändern, zulässig. Bei intervallskalierten Daten müssen zusätzlich die Abstände beibehalten werden, wie es bei einer Verschiebung der Skalenwerte der Fall ist. Da man häufig vor allem am Vergleich mit einer Referenzgruppe interessiert ist, sind lineare Transformationen, wie sie bei der Normierung der Daten geschehen, in der Psychologie allgegenwärtig. Bei intervallskalierten Daten bleiben alle Informationen der Daten bei solchen Transformationen erhalten. Bei einer Verhältnisskala geht durch eine Verschiebung der Skala der Nullpunkt verloren, so dass auf diesem Niveau nur noch Streckungen erlaubt sind, durch die nur die Maßeinheit verändert wird.

Oft werden die Voraussetzungen an die Skalenqualität der Daten nicht überprüft. Stattdessen vertraut man in einer *Per-Fiat-Messung* darauf, dass die Daten intervallskaliert sind (Davison & Sharma, 1988; Bortz, 2005, S. 26).

In den Sozialwissenschaften werden häufig sogenannte *Ratingskalen* eingesetzt, durch die auf unkomplizierte Weise direkt intervallskalierte Daten gewonnen werden können (Bortz & Dö-

³⁹Nomen = lat. Name

ring, 1995, Kap. 4.2.4 Rating-Skalen). So finden sie Verwendung bei der Leistungsbeurteilung von Lehrenden, zur Einschätzung des Arbeitsplatzes, aber auch um das Verhalten psychisch Erkrankter zu beschreiben oder die emotionale Intelligenz zu messen (Schutte et al., 1998).

Bei Ratingskalen werden durch Zahlen oder Beschreibungen, manchmal auch Beispiele bestimmte Bereiche auf einem Merkmalskontinuum markiert. Dabei sollen die einzelnen Abschnitte als äquidistant verstanden werden, so dass eine Intervallskala erzeugt wird (Rohrman, 1978). Dabei gibt es sowohl unipolare als auch bipolare Skalen, bei denen zwei gegensätzliche Begriffe einander gegenübergestellt werden. Beispiele für als gleich groß empfundene Beschreibungen sind bei Bortz und Döring (1995) und im englischsprachigen bei Wyatt und Meyers (1987) zu finden. Als Beispiel ist die Abgrenzung *nie - selten - gelegentlich - oft - immer* als fünfstufige bipolare Intervallskala zu nennen. Ebenso bieten sich bei Häufigkeitsratings Prozentzahlen an *0% - 25% - 50% - 75% - 100%* (Newstead & Arnold, 1989).

Wie in diesen Beispielen bereits deutlich wird, muss auch die Anzahl der Skalenstufen festgelegt werden. Häufig wählt man dabei ungerade Werte, um bei bipolaren Skalen eine neutrale Mitte einzuschließen. Will man hingegen ein Urteil erzwingen, wählt man eine gerade Anzahl an Stufen (Bortz & Döring, 1995).

Zu beachten ist jedoch, dass die Differenzierungsmöglichkeit der Skala zwar mit steigender Stufenanzahl zunimmt, dass sie dennoch so gewählt sein muss, dass die Differenzierungskapazität des Urteilers nicht überfordert wird (Matell & Jacoby, 1972; Tränkle, 1987). Henss (1989) zeigte, dass auf einer Ratingskala mit sehr vielen Stufen (z.B. 100) solche Werte bevorzugt werden, die durch 10 oder 5 teilbar sind und erklärt dies mit der „Prominenzstruktur des Dezimalsystems“. Nach Rohrman (1978) werden fünfstufige Ratingskalen bevorzugt (vgl. auch Lissitz & Green, 1975).

Bei der Erstellung einer Skala ist auch auf die Verteilung der Werte auf dem Merkmalskontinuum zu achten (Bortz & Döring, 1995). Wenn Extremwerte nicht mehr adäquat eingestuft werden können, da die maximalen Skalenwerte bereits für schwächerer Ausprägungen verwendet wurden, spricht man von *Ceiling- bzw. Flooreffekten*. So ist die Mindestanforderung an Ratingskalen, dass ihr Ursprung und die Einheit der Stufen von den Urteilern gleich eingeschätzt werden.

Die Frage nach der Skalenqualität der Daten ist jedoch kein analytisches Problem und aus den mathematischen Voraussetzungen an die Daten nicht abzuleiten. Vielmehr handelt es sich um ein messtheoretisches Problem (Bortz & Döring, 1995). „The numbers do not know where they come from.“ (Lord, 1953). Denn erst bei der Interpretation der Daten wird die Skalenqualität entscheidend (Stine, 1989; Michell, 1986).

Gaito (1980) zeigt anhand von Literaturbeispielen, wie die fehlende Unterscheidung von statistischen Voraussetzungen und messtheoretischen Anforderungen zu Interpretationsproblemen führen kann. Baker, Hardyck und Petrinovich (1966) prüfen in einer aufwendigen Simulationsstudie, wie die Ergebnisse eines t-Tests durch verschiedene Verzerrungen einer

Intervallskala beeinflusst werden. Dazu werden die äquidistanten Skalenwerte systematisch auf unterschiedliche Ordinalskalen abgebildet.

Durch ein ähnliches Verfahren soll in diesem Kapitel gezeigt werden, welches Skalenniveau und wie viele Skalenstufen notwendig sind, um in einem bivariaten System eine Kointegrationsbeziehung richtig zu erkennen. Gleichzeitig sollen jedoch auch solche Systeme richtig klassifiziert werden, die nicht kointegriert sind. Zusätzlich werden wir Aussagen zu Deckeneffekten und dem Umgang mit fehlenden Messwerten anhand von weiteren Simulationsstudien treffen können.

Dazu wurden zunächst jeweils 1000 bivariate Systeme in neun verschiedenen Gruppen erzeugt (siehe Tabelle 3.1, S.149), für die anschließend jeweils ein Messvorgang auf einer Skala simuliert wurde. Zur Auswertung wurden verschiedene Kriterien angewandt, mit deren Hilfe eine Unterscheidung zwischen kointegrierten, stationären und integrierten Systemen möglich ist. Diese Vorgehensweise ist im nächsten Abschnitt ausführlich beschrieben.

3.1 Datenbasis und Auswertungskriterien

In den folgenden Simulationsstudien wollen wir die Frage klären, welche Bedingungen wir an die Messdaten stellen müssen, damit es überhaupt möglich ist, eine bestehende Kointegrationsbeziehung richtig zu erkennen. Ebenso wichtig ist es natürlich, dass die Daten eine ausreichend Güte haben, um festzustellen, dass ein System *nicht* kointegriert ist.

Mit dem Johansenstest kann in einem System von Zeitreihen die Anzahl der Kointegrationsbeziehungen bestimmt werden (siehe Kapitel 2.7.2, S.134). Der Kointegrationsrang nimmt in einem K -dimensionalen Prozess Werte zwischen Null und K an (siehe Kapitel 2.2.2, S.68). In einem bivariaten System ($K = 2$) kann es also 0 - 2 (linear unabhängige) stationäre Linearkombinationen der beiden Einzelprozesse geben. Im Fall $r = 2$ sind beide Einzelprozesse selbst stationär, bei $r = 1$ liegt Kointegration vor, während es im Fall $r = 0$ keine stationäre Linearkombination gibt, da beide Einzelprozesse unabhängig voneinander integriert sind.

In allen Simulationsstudien wird also zu prüfen sein, ob die drei Gruppen von bivariaten Systemen

1. Kointegrierte Systeme - CI(1)
2. Stationäre Systeme - I(0)
3. Integrierte Systeme - I(1)

richtig klassifiziert und unterschieden werden können.

Tabelle 3.1: Allgemeiner Versuchsplan mit allen Untergruppen

	Integriert - I(1)			Kointegriert - CI(1)			Stationär - I(0)		
T	(1,1)	(2,3)	(3,10)	(1,1)	(2,3)	(3,10)	(1,1)	(2,3)	(3,10)
10									
⋮									
500									

3.1.1 Datenbasis

Ziel ist es also, die Anzahl der Kointegrationsbeziehungen in bivariaten Systemen richtig zu schätzen und dadurch die drei Systemgruppen zuverlässig voneinander unterscheiden zu können. Die kointegrierten Systeme zeichnen sich dadurch aus, dass beide Einzelreihen integriert sind, es jedoch eine (bei Normierung sogar *genau eine*) stationäre Linearkombination gibt. Bei den stationären Systemen sind beide Einzelreihen stationär, ebenso jede beliebige Linearkombination. Die integrierten Systeme bestehen aus zwei integrierten Einzelreihen, für die man keine stationäre Linearkombination finden kann. Für die Praxis ist vor allem die Differenzierung zwischen integrierten und kointegrierten Systemen interessant, da dies die Aussage einer Regression als bedeutungsvoll oder als Scheinregression kennzeichnet.

Da kointegrierte Systeme nur im Idealfall $\beta = -1$ den gemeinsamen stochastischen Trend zu gleichen Anteilen enthalten, ist es notwendig Systeme mit verschiedenen Gewichtungen zu betrachten. Dazu besteht jede Gruppe aus drei Untergruppen, die sich bei kointegrierten Daten in der Gewichtung des Common Trends unterscheiden.

Die kointegrierten Systeme wurde so konstruiert, dass der gemeinsame integrierte Trend W_t in den beiden Einzelreihen mit unterschiedlicher Gewichtung a, b enthalten ist

$$\begin{aligned} Y_{1t} &= a \cdot W_t + \epsilon_{1t} \\ Y_{2t} &= b \cdot W_t + \epsilon_{2t}. \end{aligned}$$

Die Untergruppen unterscheiden sich in der Parameterausprägung (a, b) . Wir betrachten jeweils die Parametergruppen $(1, 1)$, $(2, 3)$ und $(3, 10)$ (siehe Abbildung 3.1, S. 150, mittlere Spalte). Die normierte Linearkombination $Z_t = Y_{1t} - \frac{a}{b}Y_{2t}$ ist stationär, jede andere (linear unabhängige) Linearkombination ist integriert. Alle Fehlerprozesse ϵ_{it} werden stets standardnormalverteilt erzeugt (siehe Anhang D.4, S.386).

Die Komponentenprozesse in einem stationären System enthalten im Allgemeinen keine gemeinsamen Anteile. Um die Vergleichbarkeit mit den Untergruppen der kointegrierten Systeme zu gewährleisten, werden sie durch die Gewichtung von individuellen stationären Anteile V_{it} erzeugt. Dadurch sind die Reihen zwar beide stationär, durch die Gewichtung in den unterschiedlichen Untergruppen, werden jedoch verschiedene Varianzunterschiede erzeugt

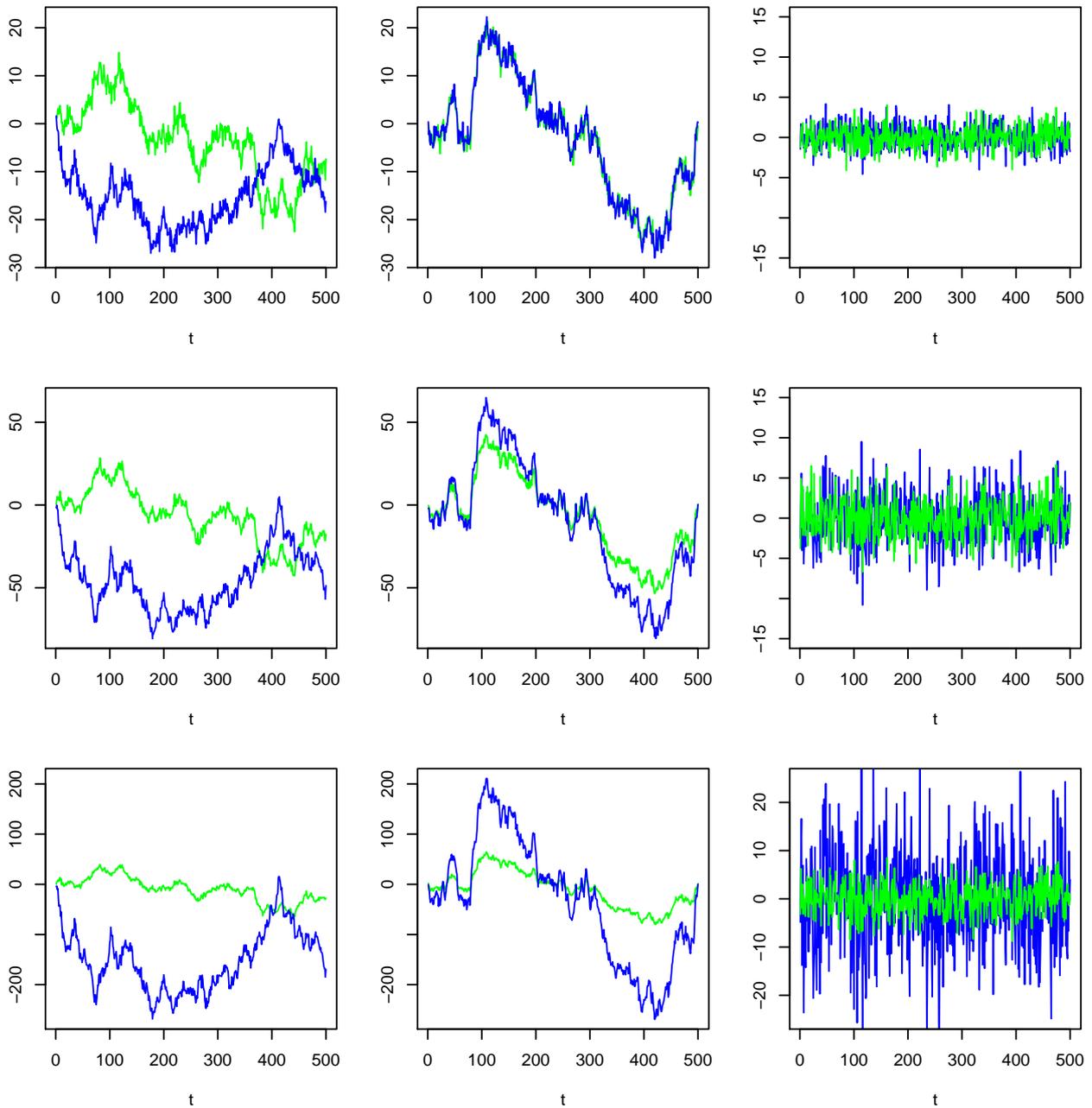


Abbildung 3.1: Die neun Untergruppen: von rechts nach links: integrierte, kointegrierte und stationäre Systeme; von oben nach unten: Gruppe (1,1), Gruppe (2,3) und Gruppe (3,10). In allen Parametergruppen werden jeweils dieselben integrierten bzw. stationären Anteile gewichtet. Durch die zunehmende Varianz zwischen den Gruppen verändert sich die Skala der Werteachse in den Abbildungen. Die kointegrierten Reihen folgen demselben Trend, während die Zeitreihen der integrierten System je einen eigenen Trend beinhalten. Auch bei den stationären Reihen sieht man deutlich, wie sich die unterschiedliche Gewichtung der stationären Anteile auf den Wertebereich auswirkt.

(siehe Abbildung 3.1, rechte Spalte)

$$\begin{aligned} Y_{1t} &= a \cdot V_{1t} + \epsilon_{1t} \\ Y_{2t} &= b \cdot V_{2t} + \epsilon_{2t}. \end{aligned}$$

Die Linearkombination $Z_t = Y_{1t} - \frac{a}{b}Y_{2t}$ ist wie jede andere Linearkombination der beiden Reihen stationär.

Ähnlich wie bei den kointegrierten Systemen werden in den drei Untergruppen die integrierten Anteile der Einzelreihen unterschiedlich gewichtet. Im Gegensatz zu den kointegrierten Systemen enthalten die Reihen jedoch verschiedene Random Walks W_{1t}, W_{2t}

$$\begin{aligned} Y_{1t} &= a \cdot W_{1t} + \epsilon_{1t} \\ Y_{2t} &= b \cdot W_{2t} + \epsilon_{2t}. \end{aligned}$$

Auch die normierte Linearkombination $Z_t = Y_{1t} - \frac{a}{b}Y_{2t}$ ist wie alle anderen möglichen Linearkombinationen integriert.

Für die folgenden Simulationsstudien wurden für jede Untergruppe jeweils 1000 bivariate Systeme der Länge 500 erzeugt. Der zu diesem Zweck entworfene Algorithmus ist im Anhang D.4 ab Seite 386 kommentiert. In jeder Gruppe wurden 1000 Common Trends W_t bzw. Paare (V_{1t}, V_{2t}) oder (W_{1t}, W_{2t}) erzeugt und in den Untergruppen entsprechend gewichtet, um eine Vergleichbarkeit zwischen den Untergruppen zu erzielen. Man kann in der Abbildung 3.1 deutlich sehen, wie sich die unterschiedliche Ausprägung der Parameter a, b auf den Verlauf der kointegrierten und integrierten Systeme auswirkt. Dadurch kann auch ein eventueller Effekt der Gewichtung bzw. eines Varianzunterschiedes erfasst werden. Da sich die Varianz in den Gruppen stark unterscheidet, muss in der Grafik die Werteachse jeweils angepasst werden. Dadurch entsteht der Eindruck, dass der Verlauf der blau dargestellten Zeitreihe gleich bleibt, während die grün dargestellte Reihe ihren Verlauf ändert. Tatsächlich ändern beide Zeitreihen je nach Gewichtung ihren Verlauf.

3.1.2 Auswertungskriterien

Die insgesamt neun Untergruppen mit jeweils 1000 Zeitreihen-Paare bilden die Datenbasis, für die eine Reihe von Auswertungsmerkmalen erhoben wird, in denen sich die Gruppen unterscheiden. Dadurch, dass der Daten erzeugende Prozess durch die Simulation bekannt ist, bieten sich zwei verschiedene Ansätze für einen Test auf Kointegration (vgl. Stroe-Kunold, Stadnytska, Werner & Braun, 2009).

Stationäre Linearkombination

In allen Gruppen sind die Parameter (a, b) einer möglichen Kointegrationsbeziehung bekannt. Dies können wir nutzen und für einen Test auf Kointegration den Test aus Abschnitt 2.7.1 (ab S.124) verwenden. Dazu ist einfach zu testen, ob die Kointegrationsbeziehung $Z_t = Y_{1t} - \frac{a}{b}Y_{2t}$ stationär ist. Bei den stationären und kointegrierten Systemen sollte dies der Fall sein, während für die integrierten Systeme keine stationäre Linearkombination existiert. Da die Einzelreihen der stationären Systeme selbst stationär sind, die der kointegrierten Systeme jedoch integriert, ist damit eine vollständige Klassifikation der drei Systemarten möglich. Wir konzentrieren und jedoch auf den Kointegrationstest, da die Eigenschaft der Einzelreihen auf Stationarität bereits bei der Konstruktion der Daten mit Unit-Root-Tests überprüft wurde (siehe Anhang D.4, S.386).

In Kapitel 1.12 (S.42ff) haben wir verschiedene Einheitswurzeltest kennengelernt, die man zu diesem Zweck verwenden kann. Wir werden im Folgenden den Augmented Dickey-Fuller-Test (siehe Gujarati (2003) als Einstieg oder Hamilton (1994) bzw. Maddala und Kim (2007)) und den Phillips-Perron-Test (Perron, 1988; P. C. B. Phillips & Perron, 1988) einsetzen. Die Nullhypothese lautet „Die Linearkombination enthält eine Einheitswurzel“. Sie gilt jedoch nur in der Gruppe der integrierten Systeme und sollte im Idealfall für alle stationären und kointegrierten Systeme verworfen werden.

Im Folgenden wird ein Klassifikationsergebnis dann als ausreichend gut bewertet, wenn in jeder Untergruppe mindestens 950 Systeme richtig klassifiziert wurden. Der Anteil der Fehlklassifikationen sollte also maximal 5% betragen.

Dazu betrachten wir, bei wie vielen System in der jeweiligen Untergruppe die p-Werte der entsprechenden Teststatistik die kritischen Werte des 90%, 95% und 99% Quantils überschreiten.

Bereits im Kapitel 1.12 wurde darauf hingewiesen, dass der PP-Test eine höhere Macht hat als der ADF-Test, also eine falsche Nullhypothese mit höherer Wahrscheinlichkeit ablehnt. Allerdings zeigt er auch eine Tendenz die Nullhypothese zu schnell zu verwerfen, was sich in den Simulationsstudien dieses Kapitels bestätigen wird.

Für die Anwendung der Tests in R verwenden wir die Funktionen `adf.test` für den Augmented Dickey-Fuller-Test und `pp.test` für den Phillips-Perron-Test aus dem Paket `tseries` (Trapletti & Hornik, 2009).

Wird der ADF-Test mit seiner Default-Einstellung ausgeführt, z.B. auf ein normalverteiltes Weißes Rauschen der Länge 100 (`WN100`)

```
> WN100 <- rnorm(100)
> adf.test(WN100),
```

dann wird ein Test mit der Alternativhypothese „Prozess ist stationär“ ausgeführt, der einen

eventuellen linearen Trend mit einschließt und für die Anzahl der zu berücksichtigenden Lags einen optimalen Wert im Sinne von Said und Dickey (1984) bestimmt. Die p-Werte werden aus der Tabelle 4.2 auf Seite 103 aus Banerjee (2000) interpoliert (siehe auch Fuller, 1996). Falls ein Wert außerhalb des dort angegebenen Wertebereiches liegt, wird eine Warnung ausgegeben. In der Ergebnisausgabe

Augmented Dickey-Fuller Test

```
data: WN100
Dickey-Fuller = -4.0922, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

wird neben der Dickey-Fuller-Statistik auch die Anzahl der verwendeten Lags in der Regressionsgleichung und der p-Wert (hier 0.01) ausgegeben. An Stelle der Alternativhypothese „stationary“ wie hier, kann man auch auf eine Nullstelle der charakteristischen Gleichung innerhalb des Einheitskreises (alternative = „explosive“) testen. Da der Wert der Statistik im Beispiel einen p-Wert von 0.01 hat, ist die Nullhypothese der Einheitswurzel klar abgelehnt. Zusätzlich wurde in diesem Beispiel folgende Warnung ausgegeben, da p-Werte, die kleiner als 0.01 sind, nicht unterschieden werden können.

```
In adf.test(WN100) : p-value smaller than printed p-value
```

Bei einem integrierten Prozess, sieht die Ausgabe wie folgt aus

```
> Y <- Fkt.RW(100)
> adf.test(Y)
```

Augmented Dickey-Fuller Test

```
data: Y
Dickey-Fuller = -1.9714, Lag order = 4, p-value = 0.5885
alternative hypothesis: stationary
```

Da die Reihen beide die gleiche Länge 100 haben, werden in beiden Beispielen vier Lags in die Regressionsgleichung eingeschlossen. Der Wert der Dickey-Fuller-Statistik liegt hier jedoch deutlich unter dem kritischen Wert, so dass der p-Wert klar die Annahme der Nullhypothese anzeigt.

Der PP-Test im Paket `tseries` ist ähnlich anzuwenden. Auch hier wird in der Defaulteinstellung ein linearer Trend im System berücksichtigt und der Wert der eingeschlossenen Lags optimiert. Die p-Werte werden aus den Tabellen 4.1 und 4.2 aus Banerjee (2000) verwendet. Als Alternativhypothese ist im Standardfall „stationary“ eingestellt, was man durch „explosive“ ersetzen kann. Für dieselben Datenreihen wie oben erhält man folgende Ausgaben

```
> pp.test(WN100)
```

Phillips-Perron Unit Root Test

```
data: WN100
Dickey-Fuller Z(alpha) = -101.5696, Truncation lag parameter = 3,
p-value = 0.01
alternative hypothesis: stationary
```

Der PP-Test berücksichtigt bei einer Länge von 100 drei Lags im Gegensatz zum ADF-Test, der vier Lags in die Regressionsgleichung einschließt. Doch auch er erkennt den Prozess als stationär und gibt für den p-Wert dieselbe Warnung wie der `adf.test` aus.

Phillips-Perron Unit Root Test

```
data: Y
Dickey-Fuller Z(alpha) = -6.7327, Truncation lag parameter = 3, p-value
= 0.7258
alternative hypothesis: stationary
```

Auch den integrierten Prozess erkennt der `pp.test` im Beispiel klar bei der Alternativhypothese „stationary“.

Bei beiden Test kann man mit `$p.value` auf den P-Wert (Quantil für den erzielten Wert der Statistik) zugreifen und diesen dann weiter verarbeiten. Genauere Informationen, wie die Ausgabe der beiden Tests für die Auswertung der Simulationsstudien verwendet wurde, sind im Anhang D.5 ab Seite 390 beschrieben.

Johansenstest

Als allgemeinen Test auf Kointegration haben wir in Kapitel 2.7.2 den Johansenstest kennen gelernt. Für ein bivariates System fallen die beiden Varianten dieses Tests zusammen. Für den Trace-Test wird die Statistik $\mathcal{L}_K^{\max} - \mathcal{L}_r^{\max}$ betrachtet, für den Max-Test $\mathcal{L}_{r+1}^{\max} - \mathcal{L}_r^{\max}$. Da im Fall *einer* Kointegrationsbeziehung $r + 1 = 2 = K$ gilt, haben beide Statistiken denselben Wert. Deshalb ist es ausreichend, wie in der Default-Einstellung der Testvariante im Statistikprogramm R die Max-Variante zu betrachten.

Allgemein bestimmt der Johansenstest, wie viele (unabhängige) stationäre Linearkombinationen zwischen den Einzelreihen existieren. In einem bivariaten Systemen können dies 0, 1 oder 2 sein. Deshalb können wir diesen Test für unsere Klassifikation verwenden. In einem integrierten System ohne gemeinsamen integrierten Anteil gibt es keine stationäre Linearkombination. Da die beiden Einzelreihen in einem stationären System selbst stationär sind, gibt es hier zwei linear unabhängige stationäre Linearkombinationen. In einem kointegrierten System gibt es genau eine normierte Möglichkeit, den gemeinsamen Trend auszuschließen, und somit genau eine (linear unabhängige) stationäre Linearkombination (siehe Tabelle 3.2).

Tabelle 3.2: Klassifikation der verschiedenen Systemgruppen nach Kointegrationsrang r

r	Klassifikation	λ_2	λ_1	$r = 0$	$r \leq 1$
$r = 0$	integriertes System	klein	klein	klein (angenommen)	klein (angenommen)
$r = 1$	kointegriertes System	groß	klein	groß (abgelehnt)	klein (angenommen)
$r = 2$	stationäres System	groß	groß	groß (abgelehnt)	groß (abgelehnt)

Da es in einem bivariaten System maximal zwei stationäre Linearkombinationen geben kann, betrachtet man zwei quadrierte kanonische Korrelationskoeffizienten λ_i . Diese sind stets positiv und liegen geordnet vor $\lambda_2 > \lambda_1 > 0$. Der Johantest wird also in (maximal) zwei Schritten ausgeführt. Im ersten Schritt prüft man, ob der größere Wert λ_2 signifikant von Null verschieden ist. Ist dies nicht der Fall, kann man den Johantest abbrechen. Es gibt keine stationäre Linearkombination im System, also muss es integriert sein. Bei einem von Null verschiedenen Eigenwert, gibt es mindestens eine stationäre Linearkombinationen. Man muss nun in einem zweiten Schritt prüfen, ob es sich um ein kointegriertes oder ein stationäres System handelt. Wie in Tabelle 3.2 zu sehen, liegt ein stationäres System dann vor, wenn auch der zweite Eigenwert als von Null verschieden getestet wird.

Tatsächlich testet man jedoch nicht wie gerade beschrieben die Werte der λ_i , sondern eine von ihnen abhängige Max-Statistik (siehe Gleichung (2.39), S.136). Die Verteilung dieser Statistik kann simuliert werden und wurde von Osterwald-Lenum (1992) tabellarisiert (siehe auch Johansen & Juselius, 1990, Johansen, 1988, Johansen, 1991). Für unsere Zwecke sind die Werte für den bivariaten Fall aus Tabelle 3.3 entscheidend (Osterwald-Lenum, 1992, Table 1.1).

Tabelle 3.3: Quantile der Max-Statistik im bivariaten Johantest

	90%	95%	99%
L.max.2	12,91	14,90	19,19
L.max.1	6,50	8,18	11,65

Nach der Logik des Johantest werden die einzelnen Schritte nacheinander ausgeführt und man entscheidet jeweils nur, ob eine weitere stationäre Linearkombination im System möglich ist. Durch die gemeinsame Betrachtung der Statistiken aus den beiden Schritten, die man bei einem bivariaten System durchführen kann, wird ein neuer Klassifikationstest erzeugt, dessen Niveau sich jedoch nicht aus dem Niveau, auf dem die beiden Einzeltest abgelehnt werden, erschließen lässt. Wir betrachten jedoch jeweils die Werte der beiden Statistiken auf demselben Niveau und werden die Klassifikationsergebnisse dementsprechend nach dem betrachteten Niveau unterscheiden, obwohl man dies nicht als das Niveau des entstandenen Tests ansehen kann. Die Werte dieser Klassifikationsrate sollen in den Simulationsstudien ja erst bestimmt werden.

Da die Simulationsstudien im Statistikprogramm R durchgeführt wurden, wird hier mit wenigen Worten erläutert, wie der Johantest in R angewendet werden kann und wie die Ausgabe zu interpretieren ist. Der Johantest ist Bestandteil des R-Paketes *urca* („Unit-Root-and Cointegration-Analysis“, Pfaff, 2008). Außer der Datenmatrix, die die Einzelprozesse als Spaltenvektoren enthält, und der Anzahl von Einzelprozessen sollte man noch unter `dimnames` die *Label* der einzelnen Prozesse angeben. Dann erhält man mit `summary` das Ergebnis des Johantest. Zusätzlich kann man angeben, ob man den Trace- oder Max-Test ausführen will, und spezifizieren wie viele Γ_i in welcher Darstellung des Fehlerkorrekturmodells geschätzt werden sollen. Man kann außerdem noch wählen, ob eine saisonale Komponente oder ein linearer Trend im System enthalten ist. Ohne nähere Angaben wird als Defaultfall der Max-Test mit der Berücksichtigung von zwei Lags und einem linearen Trend ausgeführt.

```
> cajo <- ca.jo(matrix(c(X, Y), ncol = 2, dimnames = list(NULL, c("X","Y"))))
> summary(cajo)
```

In der `summary` erhält man die folgenden Informationen. Hier ein Beispiel aus der Untergruppe der kointegrierten Systeme mit Parameterausprägung (1,1).

```
#####
# Johansen-Procedure #
#####
```

```
Test type: maximal eigenvalue statistic (lambda max) , with linear trend
```

```
Eigenvalues (lambda):
[1] 0.38449062 0.01811469
```

```
Values of teststatistic and critical values of test:
```

```
          test 10pct  5pct  1pct
r <= 1 |  1.79  6.50  8.18 11.65
r = 0  | 47.56 12.91 14.90 19.19
```

```
Eigenvectors, normalised to first column:
(These are the cointegration relations)
```

```
          X.12  Y.12
X.12  1.0000000 1.00000
Y.12 -0.9530153 0.67077
```

```
Weights W:
(This is the loading matrix)
```

```

          X.12      Y.12
X.d -0.4816403 -0.02396106
Y.d  0.6689990 -0.02765677

```

Unter `Eigenvalues (lambda)` sind die beiden Schätzungen für λ_2, λ_1 angegeben. Man sieht deutlich, dass der erste Wert 0.384 viel größer als der zweite Wert 0.018 ist. Doch ob beide signifikant von Null verschieden sind, klärt erst die Ausgabe des Johantests. In der Default-Einstellung wird die Max-Statistik berechnet.

Unter `Values of teststatistic and critical values of test` findet man die Tabelle 3.4. Nach der Logik des Johantests muss man diese Tabelle von unten nach oben lesen. In

Tabelle 3.4: Ausgabe der Funktion `ca.jo`, Ergebnis des Johantests

	test	10pct	5pct	1 pct
$r \leq 1$	L.max.1	6.50	8.18	11.65
$r = 0$	L.max.2	12.91	14.90	19.19

der Zeile $r = 0$ steht das Ergebnis des Test im ersten Schritt des Johantests, nämlich, ob es mindestens eine stationäre Linearkombination gibt. In der Spalte `test` steht das Ergebnis der Teststatistik bei einem Test zur Nullhypothese $r=0$. Die drei Spalten geben die Quantile der Verteilung dieser Teststatistik an. Dabei steht `10pct`⁴⁰ für das 90% Quantil, `5pct` für 95% und `1pct` für 99%. Übersteigt der Wert der Teststatistik einen kritischen Wert, so wird die Nullhypothese auf diesem Niveau abgelehnt. Dies ist im Beispiel mit 47.56 deutlich der Fall. In der Zeile $r \leq 1$ steht das Ergebnis des Tests gegen die Nullhypothese „Es gibt maximal eine stationäre Linearkombination“. Das Ergebnis ist bei Annahme der Nullhypothese $r=0$ nicht von Bedeutung. Andernfalls besagt ein Ablehnen dieser Hypothese, dass es mehr als eine, im bivariaten System also zwei stationäre Linearkombinationen gibt. Damit ist die Klassifikation des Systems eindeutig an dieser Tabelle abzulesen⁴¹.

Für jede Untergruppe wird in den Simulationsstudien festgehalten, wie viele der 1000 bivariaten Systeme Teststatistiken hatte, deren Werte über den kritischen Werten lagen. Aus diesen Werten kann man dann ableiten, wie viele Systeme als integriert, kointegriert oder stationär klassifiziert wurden.

Des Weiteren werden noch die Schätzungen der Matrizen β und α ausgegeben. Die Matrix α wird in der Ausgabe mit `Weights W` bezeichnet mit dem zusätzlichen Kommentar `This is the loading matrix`. Im Beispiel sollte sie wegen der Kointegrationsbeziehung Rang Eins haben. Die niedrigen Werte in der zweiten Spalte deuten dies an. Berechnet man das Produkt von α und β im Beispiel, dann erkennt man, dass der Fall $\alpha_1 < 0, \alpha_2\beta < 0$ vorliegt, für den

⁴⁰pct bedeutet „per centum“, also Prozent

⁴¹Die Werte der Max-Statistik für $r \leq 1$ sind mit denen der Trace-Statistik in diesem Fall wie erwähnt identisch.

wir gesehen hatten (siehe Abschnitt 2.5.2), dass dann eine Anpassung von beiden Prozessen zur Erhaltung der Gleichgewichtsbeziehung erfolgt.

Kointegrationsvektor β

Insbesondere bei den kointegrierten Systemen ist es interessant, wie die Kointegrationsbeziehung geschätzt wurde. Die Kointegrationsbeziehung wird zu $X_t - \frac{a}{b}Y_t$ normiert. Für den Kointegrationsvektor gilt also $\beta = (1, -\beta_0)$, wobei $\beta_0 = -\frac{a}{b}$. Im Johansenstest wird der Vektor β geschätzt, auf den man in R über das Attribut `V` Zugriff hat (siehe auch Anhang D.5, S.390). Dies ist theoretisch eine zweidimensionale Matrix und wird deshalb im Ergebnis des Johansenstests auch so dargestellt. Im Kointegrationsfall hat sie jedoch Rang Eins, so dass nur der erste normierte Spaltenvektor $\beta = (1, -\beta_0)$ für die weitere Analyse notwendig ist. Da es bei integrierten Systemen häufig zu einer Scheinregression kommt, ist ebenfalls interessant, wie die Werte von β_0 in diesem Fall verteilt sind. Für stationäre und integrierte Systeme liegt keine Kointegration vor, so dass die Schätzungen um Null liegen sollten.

```
beta[i,] <- attr(ca.jo.i, "V")[,1]
```

Für weitere Details zum Zugriff und der Auswertung der Boxplot-Statistiken sei hier auf den Anhang D.5 (ab S.390) verwiesen.

Um die jeweiligen 1000 Schätzungen für β bzw. β_0 zusammenzufassen bietet es sich an, die empirische Verteilung in einem sogenannten Box-Plot zu veranschaulichen (siehe Abbildung 3.5). Unter einem Box-Plot versteht man die Darstellung einer Verteilung eines mindestens ordinalskalierten Merkmals durch fünf Statistiken der Verteilung (Tukey, 1977). Die sogenannte *Box* liegt zwischen dem oberen und unteren Quartil der Verteilung (25%- bzw. 75%-Quantil). Der Median (50%-Quantil) wird durch einen breiten Strich innerhalb der Box gekennzeichnet. Zusätzlich werden der minimale und maximale Wert der Verteilung angegeben (Assenmacher, 2003, S. 89f).

In einer Abwandlung werden zwei weitere Werte kenntlich gemacht. Im sogenannten Box-Whisker-Plot bezeichnen die *Whisker* (engl. Schurr- bzw. Tasthaar einer Katze) jeweils den (im Standardfall) 1,5-fachen Quartilsabstand. Messungen, die innerhalb des Whiskers liegen, können als unauffällige Abweichungen angesehen werden (Sachs & Hedderich, 2009, Kap.3.2.4). Wir werden zur Analyse der Schätzungen des Kointegrationsparameters den Box-Whisker-Plot verwenden, für den es in R folgenden Befehl gibt.

```
boxplot()
```

Wie in Tabelle 3.5 auf Seite 159 werden jeweils neben dem `Median`, die beiden Statistiken `Box1` für den unteren Wert der Box und `Box2` für den oberen Wert, sowie die beiden Whiskerwerte `W1` und `W2` und der minimale `Min` und maximale Wert der Schätzungen `Max` angegeben. Die Funktion `boxplot` liefert zusätzlich noch die Anzahl der Schätzwerte `Out`, die außerhalb des Whiskers liegen.

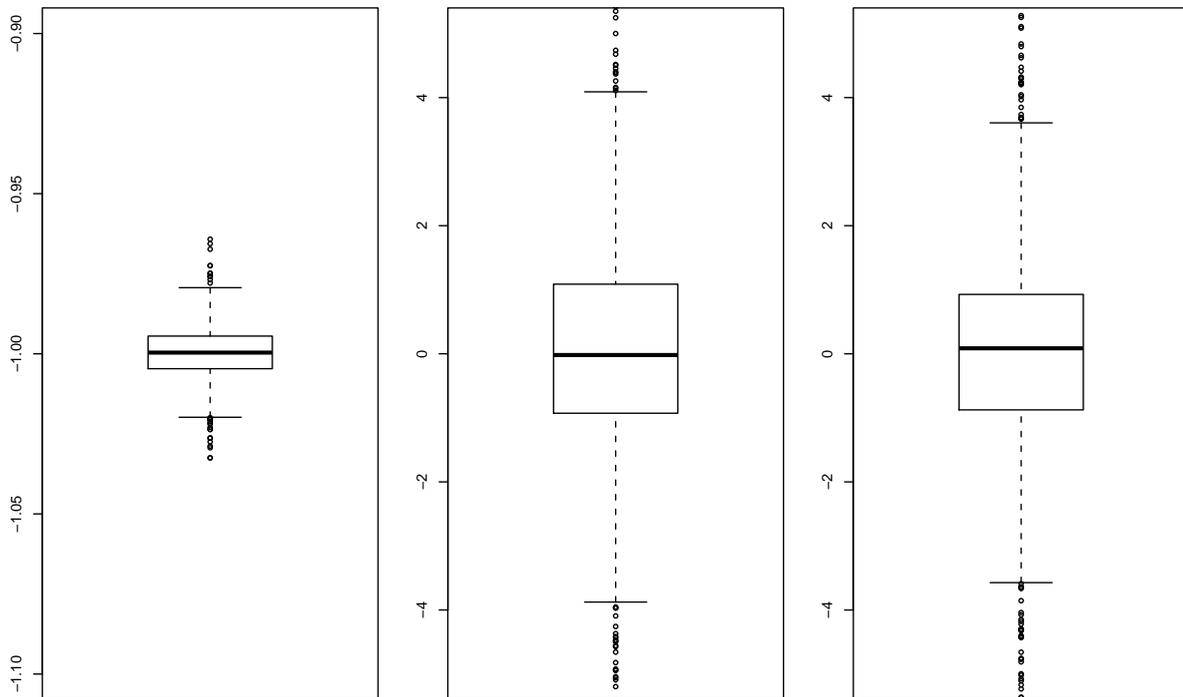


Abbildung 3.2: Boxplot der Schätzungen für β_0 bei Systemen der Länge 500 aus der Gruppe (1,1): Bei den kointegrierten Systemen (links) liegen alle Werte symmetrisch um -1, während die Schätzungen bei den stationären (Mitte) und integrierten Systemen (rechts) sehr stark um Null streuen. Die stärksten Ausreißer können im gewählten Wertebereich nicht angezeigt werden (siehe auch Tabelle 3.5). Man beachte die unterschiedliche Skala der Boxplots.

Tabelle 3.5: Boxplot-Statistiken für die Schätzung von β aus jeweils 1000 Systemen der Länge 500

	Min	W1	Box1	Med	Box2	W2	Max	Out
kointegriert	-1,032	-1,020	-1,005	-1,000	-0,994	-0,979	-0,964	28
stationär	-13730,836	-3,876	-0,928	-0,019	1,087	4,090	466,628	144
integriert	-5558,524	-3,573	-0,876	0,086	0,927	3,606	12921,672	158

3.2 Die Länge einer Zeitreihe

In der ersten Simulationstudie beschäftigen wir uns mit der Frage, wie lang ein bivariates System sein muss, damit der Johantest es richtig klassifizieren kann. Bei der Schätzung der Autokovarianzen $\gamma(h)$ (siehe Kapitel 1.9, S.35) wurde bereits erwähnt, dass die Schätzwerte erst ab einer Länge von 50 Datenpunkten sinnvoll verwendet werden können (Box & Jenkins, 1976). McCleary und Hay (1980, S.2) geben als Daumenregel ebenfalls eine Mindestlänge von $T = 50$ an. Für die Schätzung der Parameter eines ARMA-Modells konnte Bürge (1998) zeigen, dass erst bei einer Länge von 100 Messpunkten die Ergebnisse zufriedenstellend genau sind. Für Kointegration gibt es bislang keine solchen Orientierungswerte. Deshalb werden verschiedene kurze und lange Beobachtungsperioden simuliert, um eine möglichst große Parametermenge für T abzudecken.

Für jede Untergruppe wurden 1000 bivariate Systeme der Länge 500 erzeugt. Um bestimmen zu können, wie lang die Zeitreihen sein müssen, damit man Kointegration erkennt, aber nicht fälschlicherweise annimmt, betrachten wir 14 verschiedene Längen, nämlich jeweils ersten 10, 20, 30, ..., 100, 150, 200, 300, 500 Zeitpunkte der Systeme. Es werden in jeder Gruppe stets dieselben 1000 Systeme betrachtet, so dass man einen Effekt in der Klassifikation der Länge zuschreiben kann. Die Abbildung 3.3 zeigt je einen Vertreter aus der Untergruppe (1,1) und die jeweiligen Schnitte der Betrachtung. Sie sind so gewählt, dass die typischen Unterschiede in der Abbildung gut zu erkennen sind. Im Laufe der Zeit entfernen sich die Zeitreihen der kointegrierten und integrierten Systeme zunehmend vom Startwert Null, während beide Zeitreihen des stationären Systems um diesen Startwert konstant mit gleichbleibender Varianz schwanken. Dabei bewegen sich die kointegrierten Reihen gemeinsam, während die integrierten Reihen sich immer weiter voneinander entfernen. Im Abschnitt der Grafik ab ca. 200 Datenpunkten könnte man einen spiegelbildlichen Verlauf ($\beta = 1$) vermuten, der wiederum auf Kointegration hindeuten würde, obwohl die Reihen nach Konstruktion unabhängig voneinander sind. Aufgabe eines Kointegrationstests ist es genau, diese Scheinregressionen zu vermeiden.

In diesem Kapitel soll nun geklärt werden, wie lang ein bivariates System sein muss, damit diese Unterscheidung zuverlässig von den Testverfahren gemacht wird. Als Kriterium für eine ausreichend große Länge werden 95% richtig klassifizierter Systeme in den jeweiligen Parametergruppen angesetzt. Eine Länge reicht also dann aus, wenn mindestens 950 stationäre, 950 kointegrierte und 950 integrierte Systeme richtig klassifiziert wurden. Ist das Kriterium in einer Untergruppe nicht erfüllt, müssen weitere Beobachtungspunkte berücksichtigt werden.

3.2.1 Bekannter Kointegrationsvektor

In jeder Gruppe gibt es 1000 kointegrierte Systeme, 1000 stationäre Systeme und 1000 integrierte Systeme. Es gibt drei verschiedene Gruppen, die sich durch die Ausprägung von

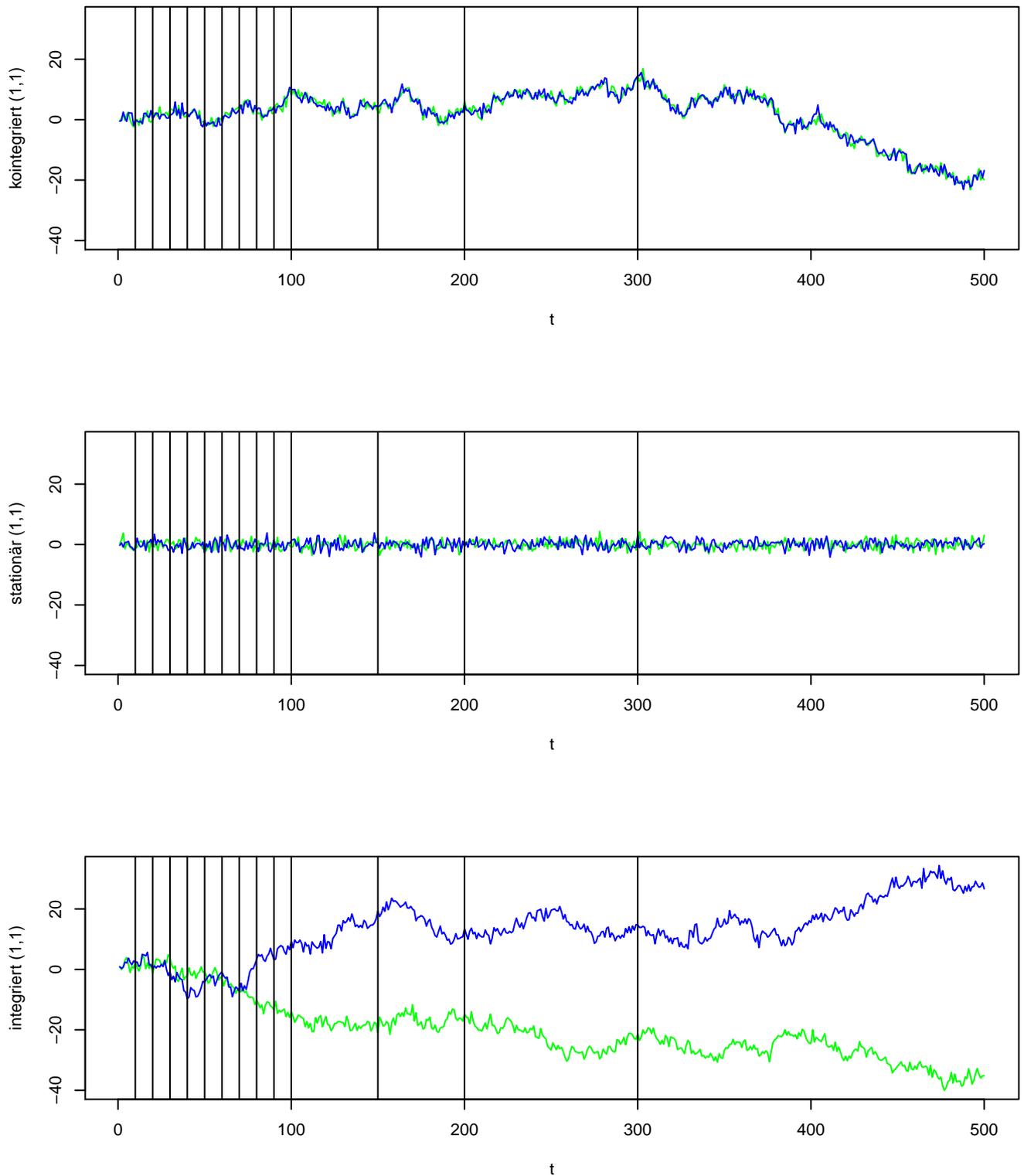


Abbildung 3.3: Schnittstellen für die Betrachtung des Längeneinflusses an jeweils einem System aus der Gruppe (1,1). Von oben: kointegriertes, stationäres und integriertes System.

(a,b) unterscheiden (vgl. allgemeiner Versuchsplan, Tabelle 3.1, S.149). Da der Kointegrationsvektor als bekannt angenommen wird, muss man nur testen, ob die eine bestimmte Linearkombination, die als Kointegrationsbeziehung angenommen wird, stationär ist. Da uns die theoretische Kointegrationsbeziehung in jeder Gruppe bekannt ist, testen wir die Linearkombination $Z_t = Y_{1t} - \frac{a}{b}Y_{2t}$ auf Stationarität. Dazu verwenden wir den Augmented- Dickey-Fuller-Test und den Phillips-Perron-Test (siehe Kapitel 1.12, S.42). Bei kointegrierten Systemen sollte die Nullhypothese „Linearkombination enthält Einheitswurzel“ abgelehnt werden, da sie gleich der vorgegebenen Kointegrationsbeziehung und deshalb nach Definition stationär ist. Auch die Linearkombination der stationären Systeme sollte als stationär erkannt werden, da dies für jede Linearkombination von stationären Prozessen der Fall ist. Keine Linearkombination zweier integrierter Reihen ohne gemeinsamen Trend ist stationär, deshalb sollte auch die getestete Linearkombination als nicht-stationär erkannt, die Nullhypothese „Linearkombination hat Einheitswurzel“ also angenommen werden.

Auswertung Gruppe (1,1)

Phillips-Perron-Test In der Gruppe (1,1) erkennt der Phillips-Perron-Test (siehe Tabelle 3.6, S.163) bei einer Länge von 10-20 Beobachtungspunkten viel zu wenig Linearkombinationen der **kointegrierten Systeme** als stationär. Doch bereits bei $T = 30$ werden auf allen drei Niveaus mehr als 90% der Kointegrationsbeziehungen, auf dem 0.05 und 0.1 Niveau sogar mehr als 99% richtig erkannt. Bei einer Länge von 50 Beobachtungspunkten werden alle 1000 Linearkombinationen auf allen Niveaus als stationär eingestuft.

Bei den **stationären Systemen** verhält sich der Test fast identisch wie bei kointegrierten Systemen. Das ist nicht überraschend, da der Prozess Z_t in beiden Fällen stationär ist, für den Test also dieselbe Situation vorliegt. Ab einer Länge von 30 sind die Ergebnisse gut, bereits ab 50 auf allen Niveaus perfekt. Auf dem 0.1 Niveau werden bei jeder Länge mehr Systeme richtig eingestuft als auf dem Signifikanzniveau 0.05, wobei zum 0.01 Niveau eine weitere Verbesserung vorliegt.

Bei **integrierten** Systemen allerdings sind die Ergebnisse des Tests viel schlechter. Auf dem 0.1 Niveau wird bis zu einer Länge von 200 über die Hälfte aller Linearkombinationen fälschlicherweise als stationär angenommen. Auf dem 0.01 Niveau sind dies immer noch mehr als 20%. Erst bei einer Länge von 500 Datenpunkten werden hier nur noch 61 Linearkombinationen falsch eingeschätzt. Das sind jedoch auch noch weniger als der geforderte Richtwert von 950.

In den Simulationsstudien wird also deutlich, dass der Phillips-Perron-Test dazu neigt, die Nullhypothese abzulehnen (vgl. Kapitel 1.12, S.42). Zwar wird durch die höhere Macht eine falsche Nullhypothese „hat Einheitswurzel“ zuverlässig abgelehnt. Jedoch bewirkt die Verzerrung, dass dies auch im Fall integrierter Prozesse mit gleicher Varianz viel zu häufig geschieht. Trotz der guten Ergebnisse bei stationären Linearkombinationen ist der Test zum Erkennen

Tabelle 3.6: Ergebnisse des PP-Tests in Gruppe (1,1): Anzahl der als stationär eingestuften Linearkombinationen

T	kointegriert			stationär			integriert		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
10	13	1	0	12	2	1	8	0	0
20	851	678	260	861	671	245	345	204	37
30	998	991	930	999	993	901	508	367	178
40	1000	1000	995	1000	1000	993	551	432	255
50	1000	1000	1000	1000	1000	1000	561	442	250
60	1000	1000	1000	1000	1000	1000	565	465	268
70	1000	1000	1000	1000	1000	1000	578	443	281
80	1000	1000	1000	1000	1000	1000	563	460	275
90	1000	1000	1000	1000	1000	1000	558	457	270
100	1000	1000	1000	1000	1000	1000	541	442	273
150	1000	1000	1000	1000	1000	1000	530	426	266
200	1000	1000	1000	1000	1000	1000	476	385	219
300	1000	1000	1000	1000	1000	1000	373	291	175
500	1000	1000	1000	1000	1000	1000	304	201	61

von integrierten Systemen nicht gut geeignet. Da in der Praxis jedoch insbesondere die Unterscheidung zwischen integrierten und kointegrierten Systemen von Interesse ist, kann der Test bei keiner der getesteten Länge in dieser Parametergruppe eingesetzt werden.

Augmented Dickey-Fuller-Test Bei **kointegrierten Systemen** zeigt der ADF-Test (siehe Tabelle 3.7, S.164) einen deutlichen Längeneffekt auf allen drei Niveaus. Je länger die Reihen beobachtet werden, desto mehr Systeme werden als kointegriert erkannt. Bei jeder Länge werden stets auf dem 0.1 Niveau mehr Systeme richtig eingestuft als auf dem 0.05 Niveau und dort wiederum mehr als auf dem 0.01 Niveau. Bei einer Länge von 90 Beobachtungspunkten werden auf dem 0.1 Niveau mehr als 950 Linearkombinationen als stationär erkannt und damit der kritische Wert das erste Mal erreicht. Dies ist bei 100 Punkten auch auf dem 0.05 Niveau der Fall. Bei 150 Punkten werden 943 Linearkombinationen auch auf dem 0.01 Niveau als stationär klassifiziert. Bei einer Länge von 200 Beobachtungspunkten werden nahezu alle Systeme richtig eingeschätzt, bei 300 ist die Klassifikation auf allen Niveaus perfekt.

Nahezu identische Ergebnisse erhält man für die **stationären Systeme**. Die 95% Marke wird auf dem 0.1 bei 90, auf dem 0.05 Niveau bei 100 und auf dem 0.01 bei 150 erstmals überschritten. Je länger die Beobachtungszeit, desto besser die Klassifikation, wobei es auf dem 0.1 Niveau häufiger gelingt als auf dem 0.05 und 0.01 Niveau. Auch hier ist bei 200-300 Datenpunkten die Klassifikation perfekt.

Bei den **integrierten Systemen** ist auf keinem Niveau ab einer Länge von 30 ein Zeiteffekt

zu sehen. Erst bei 500 Punkten ist das Ergebnis stark verändert. Man kann jedoch deutlich sehen, dass auf den drei Niveaus unterschiedlich viele Linearkombinationen als nicht-stationär erkannt werden. Auf dem 0.01 Niveau sind es weniger als 10 Linearkombinationen, die für stationär gehalten werden; auf dem 0.05 zwischen 30-50 und auf dem 0.1 Niveau zwischen 70-90 Systeme. Hier werden also auf dem 0.05 und 0.01 mehr als 950 als nicht-kointegriert erkannt. Wobei das 0.01 Niveau viel besser abschneidet als das 0.05 und 0.1 Niveau.

In dieser Gruppe schneidet der ADF-Test viel besser ab als der PP-Test. Bereits ab Länge 30 erkennt er auf dem 0.01 und 0.05 Niveau ausreichend viele integrierte Systeme und lehnt die Nullhypothese „Einheitswurzel in der Linearkombination“ nur im Einzelfall ab.

Tabelle 3.7: Ergebnisse des ADF-Tests in Gruppe (1,1): Anzahl der als stationär eingestuften Linearkombinationen

T	kointegriert			stationär			integriert		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
10	156	128	85	166	132	89	181	146	96
20	222	130	37	219	132	36	112	58	18
30	309	182	47	346	202	50	90	36	6
40	531	341	107	534	362	112	63	35	6
50	742	584	242	738	587	237	85	43	6
60	881	742	417	876	739	405	71	34	6
70	839	678	312	828	684	320	78	36	3
80	920	814	476	925	813	467	86	31	3
90	969	908	634	973	901	634	74	35	3
100	989	955	765	990	960	770	78	34	4
150	1000	996	943	1000	999	955	76	47	9
200	1000	1000	998	1000	1000	998	86	45	4
300	1000	1000	1000	1000	1000	1000	73	30	3
500	1000	1000	1000	1000	1000	1000	19	2	0

Ergebnis Gruppe (1,1) Der PP-Test ist für die Unterscheidung von stationären und nicht-stationären Linearkombinationen in dieser Gruppe nicht geeignet, da er eine Tendenz zum Ablehnen der Nullhypothese zeigt. Eine vom PP-Test als stationär eingestufte Linearkombination könnte auch nicht-stationär sein. Auf eine Klassifikation als nicht-stationäre Linearkombination kann man sich bei diesem Test dafür verlassen. Eine Klassifikation als stationär sollte man mit dem ADF-Test überprüfen.

Bei Reihen unter 30 Datenpunkte bringt keiner der Tests gute Ergebnisse. Ab einer Länge von 100 Beobachtungen ist der ADF-Test bei allen Systemen sehr zuverlässig. Mehr als 95% der Systeme werden richtig klassifiziert. Linearkombinationen mit weniger als 30 Datenpunkten können also nicht zuverlässig klassifiziert werden. Ab einer Länge von 100 sollte das Ergebnis beider Tests identisch sein. Ist dies nicht der Fall, sollte man bedenken, dass der PP-Test zur Stationarität neigt.

Auswertung Gruppe (2,3)

Tabelle 3.8: Ergebnisse des PP-Tests in Gruppe (2,3): Anzahl der als stationär eingestuften Linearkombinationen

T	kointegriert			stationär			integriert		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
10	16	1	0	15	3	1	2	1	0
20	847	644	231	861	659	249	125	53	7
30	998	992	904	997	995	910	197	125	28
40	1000	1000	994	1000	1000	992	209	129	42
50	1000	1000	1000	1000	1000	1000	206	128	33
60	1000	1000	1000	1000	1000	1000	198	119	39
70	1000	1000	1000	1000	1000	1000	214	124	36
80	1000	1000	1000	1000	1000	1000	215	129	36
90	1000	1000	1000	1000	1000	1000	200	133	42
100	1000	1000	1000	1000	1000	1000	209	130	36
150	1000	1000	1000	1000	1000	1000	180	123	45
200	1000	1000	1000	1000	1000	1000	163	89	27
300	1000	1000	1000	1000	1000	1000	131	66	22
500	1000	1000	1000	1000	1000	1000	44	0	0

Phillips-Perron-Test Bei **kointegrierten Systemen** ab 30 Beobachtungspunkten werden in der Gruppe (2,3) mehr als 90%, auf dem 0.05 und 0.1 Niveau sogar mehr als 99% der Systeme durch den PP-Test richtig klassifiziert (siehe Tabelle 3.8, S.165). Ab einer Länge von 50 Punkten ist die Klassifikation auf allen drei Niveaus perfekt. Ganz Ähnliches gilt für **stationäre Systeme**. Bei **integrierten Systemen** ist kaum ein Zeiteffekt zu erkennen. Auf dem 0.1 Niveau werden bei jeder Länge ca. 20% der Systeme fälschlicher Weise als kointegriert eingestuft, auf dem 0.05 Niveau sind es immer noch mehr als 10%. Nur auf dem 0.01 Niveau werden ausreichend viele Systeme (mehr als 950) als nicht-kointegriert erkannt. Dies jedoch unabhängig von der Länge des Systems. Erst bei sehr langen Zeitreihen ist ein Unterschied zu verzeichnen. Bei 500 Datenpunkten wird die Einheitswurzel auf dem 0.05 und 0.01 Niveau in allen Linearkombinationen, auf dem 0.1 Niveau immer noch bei mehr als 95% entdeckt.

Augmented Dickey-Fuller-Test Bei **kointegrierten und stationären Systemen** ist für den ADF-Test in dieser Gruppe ein deutlicher Zeiteffekt und auch ein Niveaufekt zu sehen (siehe Tabelle 3.9, S.166). Je länger die Reihen desto mehr Linearkombinationen werden als stationär erkannt und dies stets auf dem 0.1 Niveau besser als auf dem 0.05 Niveau und noch besser als auf dem 0.01 Niveau. Bei 90, 100 bzw. 150 wird die 95% Marke erstmals auf den drei Niveaus überschritten. Bereits bei 200 Punkten ist die Klassifikation auf allen Niveaus perfekt. Bei **integrierten Systemen** ist kein Zeittrend zu sehen. Aber die Klassifikation

Tabelle 3.9: Ergebnisse des ADF-Tests in Gruppe (2,3): Anzahl der als stationär eingestuften Linearkombinationen

T	kointegriert			stationär			integriert		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
10	162	132	90	164	128	81	183	151	102
20	253	142	45	257	146	30	111	67	13
30	305	185	56	329	189	51	85	33	6
40	530	341	106	535	370	117	73	32	3
50	731	560	247	732	559	232	91	42	9
60	876	747	412	892	752	380	83	38	7
70	833	670	346	840	691	322	94	43	7
80	923	826	487	920	810	472	82	41	8
90	963	903	634	962	904	640	86	46	8
100	989	960	768	984	955	774	86	54	9
150	1000	998	952	1000	1000	964	82	43	12
200	1000	1000	1000	1000	1000	999	89	36	4
300	1000	1000	1000	1000	1000	1000	74	31	3
500	1000	1000	1000	1000	1000	1000	20	3	0

auf den drei Niveaus ist unterschiedlich. Auf 0.01 werden maximal 12 Linearkombinationen als stationär angenommen. Bei 0.05 sind es durchschnittlich 40 und bei 0.1 schon 80.

Ergebnis Gruppe (2,3) Ähnlich wie in der Parametergruppe (1,1) gibt es unter 30 Datenpunkten keine verwertbaren Ergebnisse. Bei mehr als 30 oder gar 50 Punkten hat der PP-Test hier jedoch auf dem 0.01 Niveau auch bei integrierten Systemen gute Klassifikationsergebnisse. Auf den anderen Niveaus werden zu wenig integrierte Systeme richtig erkannt, wodurch die guten Werte in den anderen Gruppen kompensiert werden. Der ADF-Test kann ab 100 Datenpunkten auf dem 0.05 Niveau in allen Systemgruppen zu sehr guten Aussagen führen. Ein bivariates System in dieser Gruppe muss mindestens 30 Datenpunkte haben, damit die stationäre Linearkombination richtig erkannt wird. Ab einer Länge von 100 sind die Ergebnisse beider Tests gut zu verwenden. Je länger das System ist, desto zuverlässiger sind die Aussagen. In dieser Gruppe kann im Gegensatz zur Gruppe (1,1) der PP-Test eingesetzt werden, jedoch nur auf dem 0.01 Niveau.

Auswertung Gruppe (3,10)

Phillips-Perron-Test Bei **kointegrierten und stationären Systemen** verhält sich der PP-Test in der Gruppe (3,10) wie in den anderen Gruppen (siehe Tabelle 3.10, S.167). Ab 30 Datenpunkten werden über 90% der stationären Linearkombinationen auf dem 0.01 Niveau bzw. 99% auf den anderen Niveaus richtig erkannt. Ab einer Länge von 50 werden auf allen Niveaus alle 1000 Linearkombinationen richtig eingestuft. Bei **integrierten Reihen**

Tabelle 3.10: Ergebnisse des PP-Tests in Gruppe (3,10): Anzahl der als stationär eingestuften Linearkombinationen

T	kointegriert			stationär			integriert		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
10	13	0	0	23	5	1	5	2	0
20	835	653	246	859	653	270	80	28	3
30	998	995	916	1000	998	925	140	66	18
40	1000	1000	993	1000	1000	996	141	76	17
50	1000	1000	1000	1000	1000	1000	147	76	19
60	1000	1000	1000	1000	1000	1000	143	75	20
70	1000	1000	1000	1000	1000	1000	152	78	16
80	1000	1000	1000	1000	1000	1000	157	93	15
90	1000	1000	1000	1000	1000	1000	147	89	19
100	1000	1000	1000	1000	1000	1000	142	85	16
150	1000	1000	1000	1000	1000	1000	141	84	20
200	1000	1000	1000	1000	1000	1000	111	63	18
300	1000	1000	1000	1000	1000	1000	95	47	15
500	1000	1000	1000	1000	1000	1000	19	0	0

schneidet der Test in dieser Gruppe besser ab als in den anderen Gruppen. Auf dem 0.01 Niveau werden bei allen Längen nur 15-20 Linearkombinationen falsch eingeschätzt, bei 500 Datenpunkten werden alle als integriert erkannt. Auf den anderen Niveaus sind es ca. 70-90 bzw. 140 fälschlicherweise abgelehnte Nullhypothesen. Ein Zeiteffekt ist auch hier erst bei den längsten Reihen (ab 200 Datenpunkten) zu erkennen.

Augmented Dickey-Fuller-Test Der ADF-Test (siehe Tabelle 3.11, S.168) zeigt auch in dieser Gruppe ein ähnliches Verhalten wie in den anderen Gruppen und übertrifft den kritischen Wert von 950 Systemen bei 90, 100 bzw. 150-200 Beobachtungspunkten auf den verschiedenen Niveaus für die **stationären und kointegrierten Systeme**. Auch bei den **integrierten Systemen** gibt es kaum Unterschiede zu den anderen Gruppen. Auf dem 0.01 Niveau werden auf Länge 30 mehr ca. 99% richtig eingeschätzt; auf dem 0.05 Niveau mehr als 95% und auf dem 0.01 mehr als 90%.

Ergebnis Gruppe (3,10) Der ADF-Test schneidet in dieser Gruppe ähnlich ab wie in den anderen beiden Gruppen. Der PP-Test ist auch hier bei stationären Linearkombinationen bereits bei kurzen Längen perfekt, lehnt die Nullhypothese bei integrierten Systemen jedoch zu häufig ab, wie man hier im Vergleich zum ADF-Test sehr gut sehen kann.

Tabelle 3.11: Ergebnisse des ADF-Tests in Gruppe (3,10): Anzahl der als stationär eingestuften Linearkombinationen

T	kointegriert			stationär			integriert		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
10	178	153	100	147	112	78	180	144	106
20	228	152	36	249	146	37	108	63	21
30	327	185	46	346	196	48	78	34	6
40	525	358	109	532	361	121	67	33	5
50	741	541	221	727	555	249	88	38	4
60	881	741	384	871	725	396	86	41	1
70	805	676	328	832	668	322	87	48	11
80	901	800	477	924	809	465	78	44	8
90	959	895	615	967	903	637	82	46	8
100	986	949	751	986	951	774	85	42	9
150	1000	996	940	1000	999	965	80	42	7
200	1000	999	999	1000	1000	999	80	36	6
300	1000	1000	1000	1000	1000	1000	74	33	5
500	1000	1000	1000	1000	1000	1000	17	4	0

Ergebnis: bekannter Kointegrationsvektor

In keiner Gruppe gab es bei weniger als 30 Datenpunkten gute Ergebnisse. Dies sollte also als absolutes Mindestmaß für den Einsatz eines Unit-Root-Tests bei bekanntem Kointegrationsvektor verstanden werden. Wirklich zuverlässige Ergebnisse wurden jedoch erst ab einer mehr als drei Mal so großen Länge erzielt.

Ab 100 Beobachtungspunkten erbringt der ADF-Test sehr gute Ergebnisse in allen drei Gruppen. Der PP-Test hat hingegen eine Verzerrung, so dass er die Nullhypothese zu schnell ablehnt, wodurch die Niveaus auch in der Gruppe (3,10), in der er noch am besten abschneidet, nicht eingehalten werden. Seine Macht ist jedoch höher, so dass er bereits bei viel weniger Informationen die stationären Linearkombinationen der stationären und kointegrierten Systeme erkennt (Neusser, 2006; Maddala & Kim, 2007).

Der ADF-Test hält seine Niveaus perfekt ein und lehnt die Nullhypothese bei integrierten Systemen nur im Einzelfall ab. Man sieht hier deutlich, dass auf den unterschiedlichen Niveaus die Nullhypothese statistisch in einem bestimmten Prozentsatz von Einzelfällen abgelehnt wird. Auf dem 0.1 Niveau wird die Nullhypothese in weniger als 10% der Testfälle abgelehnt, auf dem 0.05 Niveau in weniger als 5% und auf dem 0.01 Niveau in ca. 1% der Simulationen. Die Ergebnisse entsprechen also genau den Erwartungen an das Testniveau. Eine perfekte Klassifikation bei den integrierten Systemen, bei denen die Nullhypothese in Wahrheit gültig ist, ist statistisch unwahrscheinlich. Jedoch wird in den anderen Gruppen die Nullhypothese bis 100 Datenpunkten noch zu häufig fälschlicherweise beibehalten, um den Test bei kürzeren Systemen einzusetzen.

Beim PP-Test ist ein Effekt zwischen den Gruppen zu sehen, die sich in der Ausprägung der Parameter a und b unterscheiden. Eine unterschiedliche Ausprägung von a und b , wie sie in den Gruppen (2,3) und (3,10) vorliegt, bedeutet einen Unterschied in der Varianz der beiden Reihen. Bei integrierten Reihen ist das Auseinanderdriften bei unterschiedlichen Trends dadurch deutlicher zu erkennen als bei gleicher Varianz.

Zu bedenken ist auch, dass alle Reihen den Ausgangspunkt Null zu unserem Startpunkt $t = 0$ haben. Dadurch bewegt sich ein integrierter Prozess gerade am Anfang noch nahe des Erwartungswertes Null. So ist der Verlauf zweier integrierter Prozesse ähnlicher als wenn man die Betrachtung beider Reihen zu einem späteren Zeitpunkt beginnt. So kann man die hier vorgestellten Ergebnisse auch als Extrembeispiel verstehen.

3.2.2 Johansen-Test: Klassifikation der Systeme verschiedener Länge

In der Gruppe der kointegrierten Systeme erwarten wir den Kointegrationsrang Eins und eine Schätzung von β nahe des jeweiligen Wertes für $-\frac{a}{b} = -1, -0.6667$ bzw. 0.3 . Stationäre Systeme haben den „Kointegrationsrang“ Zwei und keine Kointegrationsbeziehung, was zu einer Schätzung von β um Null führen sollte. Bei integrierten Systemen besteht die Gefahr der Scheinregression. Hier sollten sowohl der Kointegrationsrang als auch der Wert für β auf Null geschätzt werden. Es sind jedoch für die Schätzung des Kointegrationsvektors Abweichungen zu erwarten.

Kointegrierte Systeme

Gruppe (1,1) In der ersten Gruppe wird ab einer Länge von 70, 80 bzw. 100 Datenpunkten keines der Systeme mehr fälschlicherweise als integriert eingeschätzt (siehe Tabelle 3.12, S.170). Bei jeder Länge gibt es auf jedem Niveau jedoch einige, die als stationär eingestuft werden. Auf dem 0.1 Niveau sind es sogar 12-14%, so dass hier der kritische Wert von 950 als kointegriert klassifizierten Systemen nie erreicht wird. Auch auf dem 0.05 Niveau wird diese Marke ebenfalls nie überschritten, obwohl bereits ab einer Länge von 60 Punkten mehr als 900 als kointegriert erkannt werden. Auf dem 0.01 Niveau schließlich werden ab 70 mehr als 950 kointegrierte Systeme erkannt und ab 300 bereits mehr als 990.

Gruppe (2,3) In der zweiten Gruppe (siehe Tabelle 3.13, S.171) ist auf allen drei Niveaustufen bis zu einer Länge von 70 Beobachtungspunkten ein starker, danach noch ein leichter Längeneffekt zu erkennen. Ab einer Länge von 60 Punkten ist die Rate richtiger Klassifikation auf dem 0.01 Niveau stets besser als auf 0.05, wo immer noch mehr Systeme erkannt werden als auf dem 0.1 Niveau.

Auf dem 0.1 Niveau werden ab einer Länge von 50 Datenpunkten 88-92% der Systeme richtig als kointegriert eingestuft. Ab 70 Datenpunkten gibt es keine Fehleinschätzung als integriert

Tabelle 3.12: Klassifikation 1000 kointegrierter Systeme der Gruppe (1,1)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	414	495	91	529	416	55	714	273	13
20	546	392	62	692	287	21	890	109	1
30	268	634	98	437	523	40	766	230	4
40	92	794	114	204	745	51	489	501	10
50	18	857	125	46	895	59	230	755	15
60	3	872	125	8	920	72	66	917	17
70	0	863	137	1	929	70	16	964	20
80	0	865	135	0	936	64	2	979	19
90	0	864	136	0	921	79	1	976	23
100	0	861	139	0	925	75	0	981	19
150	0	872	128	0	935	65	0	976	24
200	0	868	132	0	942	58	0	982	18
300	0	874	126	0	931	69	0	992	8
500	0	879	121	0	949	51	0	994	6

mehr. Aber auch bei einer Länge von 500 gibt es noch ca. 9% als stationär deklarierte Systeme. Auf dem 0.05 Niveau werden ab einer Länge von 60 mehr als 900, ab 150 mehr als 950 kointegrierte Systeme erkannt. Ab 70 Beobachtungspunkten gibt es keine Fehleinschätzung als integriertes System mehr.

Auf dem 0.01 Niveau werden bei einer Länge von 60 mehr als 900, bereits ab 70 mehr als 980 und bei 300 und 500 Messpunkten mehr als 990 Systeme richtig eingestuft. Ab einer Länge von 90 Datenpunkten gibt es keine integrierte Einschätzung mehr.

Gruppe (3,10) In der Gruppe (3,10) ist ab einer Länge von 30 Datenpunkten auf allen Niveaus ein deutlicher Längeneffekt zu sehen (siehe Tabelle 3.14, S.171). Ab eine Länge von 70 ist die Klassifikation auf 0.01 Niveau stets besser als auf den anderen Niveaustufen. Auf dem 0.1 Niveau werden ab 50 Beobachtungspunkten mehr als 880 Systeme richtig klassifiziert, ab 70 gibt es auch keine Fehleinschätzungen als integriertes System mehr, ab 100 Datenpunkten werden mehr als 900 Systeme, jedoch nie mehr als 92% richtig erkannt. Bei der Klassifikation werden bei jeder Länge zu viele Systeme als stationär interpretiert. Auf dem 0.05 Niveau werden ab einer Länge von 50 Datenpunkten bereits über 900 Systeme, ab 150 sogar über 950, bei 500 schließlich dann 971 Systeme als kointegriert erkannt. Bereits ab 90 Beobachtungspunkten gibt es keine Fehleinschätzung als integriertes System mehr. Auf dem 0.01 Niveau wird die 900-Marke ab 60 Punkten überschritten, ab 70 auch der kritische Wert von 950 und ab 200 Beobachtungen sogar mehr als 990. Ab einer Länge von 90 gibt es nur noch Fehlklassifikationen als stationäre Systeme.

Tabelle 3.13: Klassifikation 1000 kointegrierter Systeme der Gruppe (2,3)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	399	481	120	494	442	64	695	296	9
20	512	431	57	656	326	18	855	145	0
30	252	654	94	405	554	41	712	284	4
40	76	828	96	161	793	46	448	543	9
50	14	881	105	52	900	48	209	779	12
60	1	897	102	6	938	56	67	917	16
70	0	890	110	0	943	57	8	980	12
80	0	887	113	0	948	52	1	988	11
90	0	888	112	0	939	61	0	984	16
100	0	898	102	0	944	56	0	985	15
150	0	912	88	0	954	46	0	984	16
200	0	909	91	0	961	39	0	989	11
300	0	913	87	0	957	43	0	995	5
500	0	913	87	0	967	33	0	997	3

Tabelle 3.14: Klassifikation 1000 kointegrierter Systeme der Gruppe (3,10)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	400	484	116	485	463	52	673	312	15
20	501	430	69	633	335	32	859	138	3
30	226	676	98	397	564	39	720	275	5
40	60	840	100	165	789	46	465	525	10
50	10	887	103	36	915	49	202	786	12
60	2	890	108	8	938	54	71	915	14
70	0	894	106	3	936	61	15	970	15
80	0	891	109	1	945	54	3	987	10
90	0	893	107	0	944	56	0	985	15
100	0	902	98	0	946	54	0	982	18
150	0	914	86	0	953	47	0	982	18
200	0	915	85	0	959	41	0	990	10
300	0	912	88	0	957	43	0	994	6
500	0	916	84	0	971	29	0	996	4

Schätzung von β Der Johansentest hat die meisten kointegrierten Systeme auch als solche klassifiziert. Nun bleibt zu prüfen, ob er nicht nur das Vorliegen der Kointegration richtig erkannt wird, sondern auch die Kointegrationsbeziehung richtig geschätzt wird. Dazu betrachten wir die zweite Komponente des auf $(1, \beta)$ normierten Kointegrationsvektors.

In der **Gruppe (1,1)** liegt der wahre Wert für $\beta = -\frac{a}{b}$ bei -1 . Wie man in Tabelle 3.15 (S.173) deutlich erkennen kann, liegt der Median der 1000 Schätzungen für β bei allen Längen auf eine Nachkommastelle (ab Länge 30 sogar zwei Nachkommastellen) genau bei diesem Wert.

Bei allen Statistiken des Boxplots ist ein deutlicher Längeneffekt zu erkennen. Nicht nur der Median wird mit zunehmender Beobachtungslänge immer genauer geschätzt, auch die Box und der Whisker werden immer schmaler. Dabei liegen sowohl die Werte der Box als auch des Whiskers symmetrisch um den Median. Außerdem gibt es immer weniger Schätzungen, die sich außerhalb des Whiskers befinden (*Out* in der Tabelle), je mehr Datenpunkte für die Schätzung berücksichtigt werden.

Bereits bei einer Länge von 60 Beobachtungen liegen alle Werte zwischen $-1,7$ und $-0,4$ und 50% sogar zwischen $-1,05$ und $-0,95$. Nur 57 der 1000 Schätzungen liegen außerhalb des Intervalls $[-1,19; -0,81]$. Die Schätzung von β ist also bereits bei kleinen Längen sehr gut und wird zunehmend besser. Bei 100 Beobachtungspunkten liegen nur 37 Schätzwerte mehr als eine Nachkommastelle vom wahren Wert entfernt.

In der **Gruppe (2,3)** ist $\beta = -\frac{a}{b} = -0,666667$. Der Median der 1000 Schätzungen liegt sehr genau bei dieser Größe (siehe Tabelle 3.15, S.173). Für die Statistiken des Boxplot ist ein klarer Längeneffekt zu erkennen, Box und Whisker werden immer schmaler. Für das Minimum ist eine steigende, für das Maximum eine fallende Tendenz auszumachen. Allerdings gibt es einen Ausreißer. Bei einer Länge von 70 wurde bei einem System β auf ca. 619 geschätzt. Im Anhang D.1.1, S.357 ist die Auswertung des Johansentest für dieses System zu finden. Daraus geht hervor, dass dieses System als stationär fehlklassifiziert wurde. Der zweithöchste Wert bei dieser Länge für eine Schätzung von β beträgt $-0,5787$.

Bereits bei einer Länge von 30 liegen 50% zwischen $-0,69$ und $-0,63$ und nur 82 der 1000 Schätzungen weiter als 0,12 vom wahren Wert $-0,66667$ entfernt.

Der untere Teil der Tabelle 3.15 zeigt die Ergebnisse der Schätzung für β in der **Gruppe (3,10)**, in der der wahre Wert für β bei $-0,3$ liegt. Dieser Wert wird vom Median fast exakt angenähert. Auch in dieser Gruppe kann man in allen Statistiken einen Längeneffekt ausmachen. Die unteren Werte von Box und Whisker steigen zunehmend, während die oberen Werte mit weiteren Beobachtungspunkten fallen. Auch für Minimum und Maximum der Schätzung kann man diese Tendenz erkennen. Dabei gibt es ab einer Länge von 40 auch keine Ausreißer mehr, so dass alle 1000 Schätzungen nahe des wahren Wertes angesiedelt sind. Bereits bei einer Länge von 30 liegt die Hälfte aller Werte zwischen $-0,31$ und $-0,29$ und nur 85 außerhalb von $[-0,335; -0,266]$. Bei 300 Datenpunkten sind sogar Minimum und

Tabelle 3.15: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systeme

	Min	W1	Box1	Median	Box2	W2	Max	Out
T	Gruppe (1,1)							
10	-893,7197	-2,7850	-1,3423	-0,8186	-0,3274	1,1912	38,1608	152
20	-230,8542	-1,8600	-1,2335	-1,0108	-0,8130	-0,1847	318,7888	115
30	-21,1005	-1,4890	-1,1332	-1,0095	-0,8942	-0,5410	7,8164	72
40	-3,4351	-1,3251	-1,0851	-1,0045	-0,9228	-0,6814	2,6952	61
50	-1,6797	-1,2637	-1,0724	-1,0050	-0,9422	-0,7507	-0,4077	51
60	-1,7154	-1,1990	-1,0522	-1,0038	-0,9533	-0,8173	-0,3934	57
70	-1,5207	-1,1640	-1,0458	-1,0060	-0,9635	-0,8404	-0,4141	56
80	-1,4821	-1,1446	-1,0397	-1,0047	-0,9697	-0,8653	-0,5653	54
90	-1,4766	-1,1206	-1,0323	-1,0041	-0,9731	-0,8855	-0,6340	44
100	-1,4591	-1,1169	-1,0314	-1,0023	-0,9738	-0,8910	-0,6658	37
150	-1,3970	-1,0757	-1,0199	-1,0010	-0,9826	-0,9268	-0,8269	34
200	-1,3355	-1,0546	-1,0145	-1,0011	-0,9872	-0,9471	-0,9059	39
300	-1,0724	-1,0327	-1,0086	-0,9990	-0,9911	-0,9649	-0,9423	32
500	-1,0325	-1,0198	-1,0047	-0,9996	-0,9945	-0,9794	-0,9642	28
T	Gruppe (2,3)							
10	-190,0188	-1,2976	-0,8227	-0,6484	-0,5014	-0,0278	10,4504	165
20	-2,5548	-0,8803	-0,7171	-0,6636	-0,6063	-0,4423	43,2753	106
30	-4,6254	-0,7889	-0,6961	-0,6654	-0,6341	-0,5479	2,1524	82
40	-0,9657	-0,7476	-0,6872	-0,6666	-0,6448	-0,5827	3,3077	72
50	-0,9531	-0,7331	-0,6829	-0,6653	-0,6489	-0,5981	1,5305	59
60	-0,8418	-0,7201	-0,6790	-0,6658	-0,6512	-0,6104	1,2305	51
70	-0,7882	-0,7117	-0,6774	-0,6654	-0,6541	-0,6208	619,7441	58
80	-0,7650	-0,7055	-0,6759	-0,6655	-0,6561	-0,6268	-0,5896	57
90	-0,7566	-0,7011	-0,6747	-0,6659	-0,6570	-0,6309	-0,5959	45
100	-0,8210	-0,6971	-0,6739	-0,6658	-0,6583	-0,6352	-0,5997	42
150	-0,7065	-0,6866	-0,6717	-0,6663	-0,6613	-0,6458	-0,6217	44
200	-0,7025	-0,6807	-0,6702	-0,6665	-0,6629	-0,6521	-0,6302	46
300	-0,7034	-0,6769	-0,6693	-0,6668	-0,6642	-0,6568	-0,6506	34
500	-0,6787	-0,6727	-0,6682	-0,6666	-0,6651	-0,6606	-0,6553	36
T	Gruppe (3,10)							
10	-29,7687	-0,4641	-0,3424	-0,3019	-0,2585	-0,1330	22,9765	150
20	-1,7967	-0,3635	-0,3174	-0,3014	-0,2866	-0,2410	2,6440	109
30	-35,4707	-0,3352	-0,3098	-0,3005	-0,2923	-0,2664	0,0980	85
40	-0,5794	-0,3236	-0,3058	-0,2996	-0,2937	-0,2762	0,0919	61
50	-0,3416	-0,3182	-0,3044	-0,2994	-0,2951	-0,2812	-0,2565	38
60	-0,3412	-0,3146	-0,3034	-0,2998	-0,2958	-0,2845	-0,2507	40
70	-0,3380	-0,3115	-0,3028	-0,2998	-0,2968	-0,2884	-0,2700	55
80	-0,3257	-0,3102	-0,3024	-0,2998	-0,2973	-0,2897	-0,2799	49
90	-0,3168	-0,3088	-0,3022	-0,3000	-0,2976	-0,2908	-0,2840	50
100	-0,3155	-0,3080	-0,3021	-0,2999	-0,2981	-0,2923	-0,2863	52
150	-0,3121	-0,3049	-0,3012	-0,2999	-0,2987	-0,2949	-0,2811	51
200	-0,3095	-0,3037	-0,3009	-0,3000	-0,2990	-0,2962	-0,2888	43
300	-0,3047	-0,3024	-0,3006	-0,2999	-0,2993	-0,2975	-0,2946	48
500	-0,3029	-0,3015	-0,3004	-0,3000	-0,2996	-0,2985	-0,2970	30

Maximum auf eine Nachkommastelle genau. Bei dieser Länge wurden auch alle Systeme als kointegriert erkannt.

Ergebnis - kointegrierte Systeme Der Johantest liefert in allen drei Gruppen der kointegrierten Systeme ähnliche Ergebnisse. Lediglich auf dem 0.1 Niveau ist ein leichter Unterschied zwischen der Gruppe (1,1) und den Parametergruppen mit Varianzunterschied zu erkennen. In allen Gruppen gibt es ab einer Länge von 70 Datenpunkten keine Fehlklassifikation als integriert mehr.

Den kritischen Wert auf dem 0.1 Niveau übersteigen zu viele Werte der zweiten Statistik des Johantests. So werden zu viele Systeme als stationär statt als kointegriert klassifiziert. Die Unterscheidung zwischen integrierten und kointegrierten Systemen funktioniert dafür ab einer Länge von 70 tadellos.

Da dem Johantest meist ein Test der Einzelreihen auf Integration vorausgeht, spricht eine Klassifikation als stationäres System durch den Johantest eher für Kointegration, wenn man beide Reihen zuvor klar als integriert identifizieren konnte.

Festzuhalten bleibt, dass der Johantest für kointegrierte Systeme auf dem 0.01 Niveau ab einer Länge von 70 Datenpunkten sehr gute Ergebnisse liefert. Die anschließende Verbesserung ist minimal. Interessant ist deshalb vor allem der Übergang zwischen 60 und 70 Datenpunkten.

Stationäre Systeme

Tabelle 3.16: Klassifikation 1000 stationärer Systeme der Gruppe (1,1)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	378	486	136	488	441	71	677	309	14
20	341	262	397	506	298	196	802	173	25
30	91	90	819	211	196	593	536	278	186
40	13	10	977	43	39	918	223	182	595
50	0	1	999	3	6	991	51	59	890
60	0	0	1000	0	1	999	4	7	989
70	0	0	1000	0	0	1000	0	1	999
80	0	0	1000	0	0	1000	0	0	1000
90	0	0	1000	0	0	1000	0	0	1000
100	0	0	1000	0	0	1000	0	0	1000
150	0	0	1000	0	0	1000	0	0	1000
200	0	0	1000	0	0	1000	0	0	1000
300	0	0	1000	0	0	1000	0	0	1000
500	0	0	1000	0	0	1000	0	0	1000

Tabelle 3.17: Klassifikation 1000 stationärer Systeme der Gruppe (2,3)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	369	492	139	488	437	75	670	315	15
20	345	280	375	529	304	167	805	170	25
30	86	97	817	203	205	592	546	262	192
40	5	13	982	30	53	917	216	204	580
50	0	2	998	2	5	993	58	55	887
60	0	0	1000	0	0	1000	4	14	982
70	0	0	1000	0	0	1000	1	4	995
80	0	0	1000	0	0	1000	0	1	999
90	0	0	1000	0	0	1000	0	0	1000
100	0	0	1000	0	0	1000	0	0	1000
150	0	0	1000	0	0	1000	0	0	1000
200	0	0	1000	0	0	1000	0	0	1000
300	0	0	1000	0	0	1000	0	0	1000
500	0	0	1000	0	0	1000	0	0	1000

Tabelle 3.18: Klassifikation 1000 stationärer Systeme der Gruppe (3,10)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	386	473	141	497	429	74	693	290	17
20	364	270	366	519	309	172	814	161	25
30	85	98	817	209	214	577	542	282	176
40	12	6	982	37	48	915	215	202	583
50	0	0	1000	3	6	991	60	60	880
60	0	0	1000	0	1	999	7	15	978
70	0	0	1000	0	0	1000	1	4	995
80	0	0	1000	0	0	1000	0	0	1000
90	0	0	1000	0	0	1000	0	0	1000
100	0	0	1000	0	0	1000	0	0	1000
150	0	0	1000	0	0	1000	0	0	1000
200	0	0	1000	0	0	1000	0	0	1000
300	0	0	1000	0	0	1000	0	0	1000
500	0	0	1000	0	0	1000	0	0	1000

Gruppe (1,1) In der Gruppe (1,1) ist bei den stationären Systemen auf allen drei Niveaus von Anfang an ein Effekt der Beobachtungslänge zu erkennen (siehe Tabelle 3.16, S.174). Die Klassifikationsrate ist auf dem 0.1 Niveau stets besser als auf dem 0.05 und noch besser als auf 0.01 Niveau. Denn hier werden bereits bei 40 Datenpunkten mehr als 980 Systeme, ab einer Länge von 60 schon alle 1000 richtig klassifiziert. Auf dem 0.05 Niveau sind dies bei 40 Datenpunkten immer noch über 950 und dann bei 70 auch alle 1000 Systeme. Ab einer Länge von 50 Punkten werden mehr als 900, ab 80 dann ebenfalls alle 1000 Systeme auch auf dem 0.01 Niveau richtig klassifiziert.

Gruppe (2,3) Auch in der zweiten Gruppe (siehe Tabelle 3.17, S.175) nimmt die Anzahl richtiger Klassifikationen mit steigender Beobachtungslänge zu. Dabei werden auf dem 0.1 Niveau stets die meisten, auf dem 0.01 die wenigsten Systeme richtig klassifiziert. Wie in der ersten Gruppe werden auf dem 0.1 Niveau bei einer Länge von 40 Datenpunkten schon mehr als 980 Systeme, bei 60 Punkten dann alle 1000 als stationär eingestuft. Auf dem 0.05 Niveau werden in dieser Gruppe schon bei 60 Datenpunkten alle Systeme richtig klassifiziert und bei einer Länge von 40 bereits 917. Um mehr als 900 Systeme auf dem 0.01 Niveau als stationär zu erkennen, muss man 60 Datenpunkte erreichen. Für eine perfekte Klassifikation braucht man mindestens 90 Beobachtungspunkte.

Gruppe (3,10) In der dritten Gruppe sind die Klassifikationsergebnisse bei den stationären Systemen noch besser als in den anderen Gruppen (siehe Tabelle 3.18, S.175). Hier werden auf dem 0.01 Niveau bereits bei 80 Datenpunkten alle Systeme richtig klassifiziert, bei 60 sind es schon mehr als 900. Auf dem 0.1 Niveau werden bereits bei 50 Datenpunkten alle stationären Systeme richtig erkannt, auf dem 0.05 Niveau ist dies bei 70 Beobachtungen der Fall.

Schätzung von β Da in stationären Systemen keine Kointegrationsbeziehung vorliegt, die beiden Zeitreihen vielmehr voneinander unabhängig sind, sollte die Schätzung für β hier nahe Null liegen.

So wird der Median in allen drei Parametergruppen auch auf eine Nachkommastelle genau auf Null geschätzt. In der Tabelle D.1 (Anhang D.1, S.358) sieht man jedoch, dass die Schätzwerte stark schwanken. Zwar sind Box und Whisker symmetrisch um den Median, was für zufällige Abweichungen in beide Richtungen spricht, jedoch gibt es bei allen Längen auch extreme Ausreißer bis in den fünfstelligen Bereich und viele Werte (ca. 150), die außerhalb des Whiskers liegen.

Man kann außerdem einen Gruppeneffekt betrachten. Während die Box in der ersten Gruppe die Ausmaße von ca. $[-1, 1]$ annimmt und der Whisker bei $[-4, 4]$ liegt, sind die Werte in der Gruppe (2,3) nur bei $[-0, 7; 0, 7]$ für die Box und $[-2, 8; 2, 8]$ für den Whisker. In der dritten

Gruppe liegen die Schätzwerte gesamt noch näher bei Null, sodass die Box ca. $[-0, 3; 0, 3]$ und der Whisker ca. $[-1, 3; 1, 3]$ beträgt.

Mit zunehmendem Varianzunterschied ist die Schätzung von β immer näher an Null.

Ergebnis: stationäre Systeme Für stationäre Systeme verhält sich der Johantest in allen drei Gruppen sehr ähnlich. In jeder Gruppe ist ein Effekt der Länge auf allen drei Niveaustufen klar zu erkennen (siehe Tabellen 3.16, 3.17 und 3.18). Je länger das stationäre System beobachtet wird, desto eher wird es als stationär erkannt.

Ein System wird als stationär klassifiziert, wenn beide Statistiken die kritischen Werte des Niveaus übersteigen. So verwundert es nicht, dass auf dem 0.1 Niveau stets mehr Systeme als auf dem 0.05 oder 0.01 Niveau als stationär eingestuft werden.

Doch schon bei einer Länge von 80 Datenpunkten (in der Gruppe (2,3) erst bei 90) werden auch auf dem 0.01 Niveau alle Systeme als stationär erkannt. Auf den anderen Niveaus ist das schon bei kürzeren Systemen (50-70 Datenpunkte) der Fall. Schon ab einer Länge von 40 Punkten sind die Klassifikationsraten auf dem 0.1 Niveau bei dem geforderten Wert von 950, auf dem 0.05 Niveau wird er in den Gruppen (2,3) und (3,10) nicht ganz erreicht, aber nur knapp verfehlt.

Für Systeme mit einer kürzeren Länge (also 10-30 Datenpunkte) sind die Ergebnisse der Klassifikation nicht ausreichend gut, um sie zu verwenden.

Integrierte Systeme

Tabelle 3.19: Klassifikation 1000 integrierter Systeme der Gruppe (1,1)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	459	470	71	561	404	35	735	262	3
20	765	215	20	854	142	4	951	49	0
30	801	180	19	891	105	4	974	25	1
40	796	180	24	882	113	5	981	18	1
50	804	179	17	884	111	5	972	27	1
60	801	190	9	877	119	4	972	27	1
70	805	179	16	883	112	5	963	35	2
80	803	173	24	876	121	3	962	37	1
90	776	204	20	872	123	5	972	28	0
100	782	196	22	875	122	3	964	36	0
150	765	227	8	858	139	3	955	44	1
200	793	197	10	878	122	0	964	36	0
300	811	174	15	891	104	5	966	34	0
500	835	148	17	914	85	1	975	25	0

Gruppe (1,1) In der ersten Gruppe bleibt die Klassifikationsrate für integrierte Systeme ab einer Länge von 30 Beobachtungspunkten etwa gleich (siehe Tabelle 3.19, S.177). Die Werte für kürzere Systeme sind zu schlecht, um sie zu verwenden. Sie werden im Folgenden nicht beachtet.

Bei integrierten Systemen ist das Klassifikationsergebnis anders als bei kointegrierten oder stationären Systemen auf dem 0.01 Niveau am besten, während es auf dem 0.1 Niveau am schwächsten ausfällt. Hier werden nur rund 80% der Systeme richtig klassifiziert. Auch werden 2-4% der Systeme als stationär eingestuft, was bedeutet, dass beide Statistiken die kritischen Werte überstiegen haben. Auf dem 0.05 Niveau gibt es auch noch ca. 1% Fehlklassifikationen als stationär, während dies auf dem 0.01 Niveau nur in Einzelfällen geschieht.

Auf dem 0.05 Niveau werden zwischen 585 und 914 Systeme als integriert erkannt, so dass die geforderte Marke von 950 bei keiner Länge erreicht wird. Auf dem 0.01 Niveau wird sie hingegen bereits bei 20 Datenpunkten überschritten. Es werden jedoch auch auf diesem Niveau nie mehr als 981 Systeme als integriert eingestuft. Die übrigen werden bis auf wenige Ausnahmen als kointegriert angesehen, da die erste Statistik den kritischen Wert übersteigt, die andere jedoch darunter liegt.

Tabelle 3.20: Klassifikation 1000 integrierter Systeme der Gruppe (2,3)

	Niveau 0.1			Niveau 0.05			Niveau 0.01		
T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	419	472	109	531	417	52	708	281	11
20	776	204	20	876	117	7	962	37	1
30	845	142	13	911	85	4	978	21	1
40	859	131	10	924	74	2	992	8	0
50	866	123	11	921	75	4	982	18	0
60	864	128	8	927	69	4	985	15	0
70	868	119	13	934	63	3	987	13	0
80	875	115	10	934	65	1	989	11	0
90	873	121	6	938	61	1	986	14	0
100	860	131	9	925	74	1	984	16	0
150	870	128	2	932	67	1	982	18	0
200	895	102	3	945	55	0	991	9	0
300	904	87	9	953	46	1	991	9	0
500	928	70	2	966	33	1	995	5	0

Gruppe (2,3) Von der Tendenz her zeigt sich in der zweiten Gruppe ein ähnliches Klassifikationsverhalten wie in der ersten Gruppe (siehe Tabelle 3.20, S.178). Es werden jedoch insgesamt höhere Werte erreicht.

Auf allen Niveaus ist ein leichter Effekt der Beobachtungslänge zu erkennen. Dabei ist auch hier die Klassifikation auf dem 0.01 Niveau besser als auf dem 0.05 oder gar 0.1 Niveau.

Auf dem 0.1 Niveau werden in dieser Gruppe jedoch ab einer Länge von 30 Punkten stets mehr als 840 Systeme als integriert eingeschätzt und bei 300 sogar die 900-Marke erreicht. Dennoch gibt es auch bei langen Beobachtungszeiten noch 1-2% Fehlklassifikationen als stationäre Systeme. Mehr als 900 Systeme werden auf dem 0.05 Niveau bereits bei einer Länge von 30 als integriert erkannt. Der geforderte kritische Wert von 950 wird jedoch erst bei einer Länge von 300 erzielt. Nur wenige Systeme werden als stationär eingestuft.

Auf dem 0.01 Niveau wird ab einer Länge von 40 Punkten kein System mehr für stationär gehalten. Es werden ab einer Länge von 20 mehr als 950, ab 40 Beobachtungspunkten sogar mehr als 980 Systeme richtig klassifiziert.

Tabelle 3.21: Klassifikation 1000 integrierter Systeme der Gruppe (3,10)

T	Niveau 0.1			Niveau 0.05			Niveau 0.01		
	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)
10	427	457	116	525	424	51	700	289	11
20	761	222	17	856	141	3	965	34	1
30	837	147	16	908	89	3	974	25	1
40	864	127	9	930	67	3	988	12	0
50	883	109	8	929	69	2	983	17	0
60	871	123	6	932	66	2	982	18	0
70	884	109	7	941	57	2	986	14	0
80	880	112	8	943	57	0	992	8	0
90	871	120	9	948	52	0	988	12	0
100	870	120	10	940	60	0	988	12	0
150	872	125	3	939	60	1	980	20	0
200	906	92	2	948	52	0	993	7	0
300	908	83	9	954	44	2	992	8	0
500	930	68	2	970	29	1	995	5	0

Gruppe (3,10) In der dritten Gruppe ist ein Längeneffekt nur noch minimal auszumachen (siehe Tabelle 3.21, S.179). Am stärksten noch auf dem 0.1 Niveau. Auch hier ist die Klassifikation auf dem 0.01 Niveau den anderen Ergebnissen deutlich überlegen. Ab einer Länge von 20 Datenpunkten werden mehr als 960, ab 40 dann stets mehr als 980 integrierte Systeme erkannt. Eine Fehlklassifikation als stationäres System gibt es ab 40 Datenpunkten überhaupt nicht mehr.

Die Ergebnisse auf den anderen Niveaus sind insgesamt etwas besser als in der Gruppe (2,3). Jedoch wird auch hier die geforderte 950-Marke nur auf dem 0.05 Niveau und erst ab einer Länge von 300 erreicht. Auf dem 0.1 Niveau werden ca. 88% der Systeme richtig erkannt und bei jeder Länge noch einige als stationär fehlklassifiziert. Auf dem 0.05 Niveau werden deutlich mehr als 900 Systeme als integriert eingestuft und nur einzelne als stationär.

Schätzung von β Die Schätzung von β sollte auch bei integrierten System um Null liegen, da keine Kointegration vorliegt. Und so zeigt sich auch ein ähnliches Bild der Schätzwerte wie bei stationären Systemen (siehe Tabelle D.2 im Anhang D.1, S.359) Der Median liegt in allen Gruppen um Null, Box und Whisker haben einen symmetrischen Abstand vom Median und ähnliche Werte wie für stationäre Systeme in der jeweiligen Gruppe. Ein Längeneffekt ist nicht auszumachen. Es gibt bei jeder Länge extreme Abweichungen im Minimum und Maximum der Schätzungen und eine hohe Zahl von Schätzwerten, die außerhalb des Whiskers liegen.

Ergebnis - integrierte Systeme In allen drei Gruppen wurde die kritische Marke von 950 richtig erkannten Systemen auf dem 0.01 Niveau zuverlässig bereits bei sehr kurzen Systemen erzielt. Auf dem 0.05 Niveau war dies nur in den Gruppen (2,3) und (3,10) und erst ab einer Länge von 300 der Fall. Auf dem 0.1 Niveau wurde der Wert nie erreicht. Bei integrierten Systemen sollte man also vor allem das 0.01 Niveau beachten, wohingegen die Länge kaum eine Rolle zu spielen scheint.

Ergebnis: Johantest

Der Johantest zeigt in den drei Parametergruppen ein ähnliches Klassifikationsverhalten, das sich jedoch zwischen den verschiedenen Systemtypen stark unterscheidet.

Wir haben gesehen, dass bei integrierten Systemen die Klassifikation auf dem 0.01 und 0.05 Niveau zuverlässig ist. Dabei werden schon ab einer Länge von 20 Datenpunkten mehr als 950 Systeme richtig eingestuft. Um alle Systemarten voneinander unterscheiden zu können, müssen wir also betrachten, bei welcher Länge die Klassifikation auf diesen Niveaus auch in den anderen Gruppen ausreichend genau war.

Stationäre Systeme werden auf dem 0.1 Niveau bereits bei einer Länge von 40 Beobachtungspunkten erkannt, auf dem 0.01 Niveau braucht man 20 Datenpunkte mehr. Da auch für kointegrierte Systeme die Klassifikation ab 70 Datenpunkten auf dem 0.01 sehr gut ist, sollte man diese Länge als Richtwert für eine Unterscheidung der drei Systemgruppen heranziehen. Die Schätzung des Johantests für den Kointegrationsrang r benötigt also mindestens 70 Datenpunkte und sollte auf dem 0.01 Niveau ausgewertet werden.

Schätzung für β Für kointegrierte Systeme ist die Schätzung von β in allen drei Gruppen sehr gut. Die Länge scheint dabei fast keinen Einfluss zu haben, obwohl die Streuung der Schätzwerte mit zunehmender Länge immer geringer wird.

Bei den stationären und integrierten Systemen liegt der Median für alle Längen in allen drei Gruppen nahe Null, da in diesen Systemen keine Kointegration enthalten ist. Dennoch gibt es zum Teil extreme Abweichungen von diesem Wert. Die Schätzung von β ist jedoch nur im Fall von Kointegration überhaupt interessant. Da die Schätzwerte bereits bei sehr kurzen

Reihen (20 Datenpunkte) sehr gut sind, kann man über diese Schätzmethode keine Aussage treffen, wie lang ein System sein muss, damit die Kointegrationsbeziehung richtig erkannt wird.

3.2.3 Ergebnis: Länge einer Zeitreihe

Die Auswertungen der Simulationsstudien ergaben, dass die Ergebnisse für Systeme mit weniger als 30 Datenpunkten nicht verwendbar sind. Für die meisten Auswertungsverfahren ist auch diese Länge noch zu kurz. Nur der Phillips-Perron-Test zeigt wirklich sehr gute Werte bei stationären Linearkombinationen. Bei den integrierten Systemen ist jedoch eine Länge von 200-300 Datenpunkten notwendig, um mit ihm ausreichend gute Ergebnisse zu erzielen. Dies war zudem nur in den Gruppen (2,3) und (3,10) mit unterschiedlichen Varianzen in den Reihen der Fall.

Die Alternative bei bekannter Kointegrationsbeziehung ist der Augmented-Dickey-Fuller-Test, der in allen Gruppen ab einer Länge von 100 Beobachtungen auf dem 0.05 Niveau für alle Systeme gute Werte liefert. Für ein sicheres Ergebnis sollte also bei täglicher Messung ein Zeitraum von mindestens drei Monaten berücksichtigt werden.

Der Johansentest liefert bei den kointegrierten Systemen in allen drei Gruppen auf dem 0.01 Niveau ab einer Länge von 70 Datenpunkten gute Ergebnisse. Bei stationären Systemen werden ab 60 Punkten auf dem 0.1 Niveau bereits alle Systeme richtig klassifiziert, was auf den anderen Niveaus bei 70-90 Punkten geschieht. Bei 100 Beobachtungen sind folglich in allen Gruppen auf allen Niveaus alle Systeme richtig klassifiziert. Die integrierten Systeme werden schon bei 20 Datenpunkten auf dem 0.01 Niveau gut erkannt. Auf dem 0.05 Niveau braucht man dafür etwa 300 Datenpunkte.

Der Unterschied zwischen 300 und 500 Datenpunkten ist so gering, dass es nicht notwendig ist, für eine Kointegrationsanalyse längere Reihen als 300 Datenpunkte zu erheben.

Als absolute Mindestlänge sollte man 30 Beobachtungen ansehen. Zuverlässige Ergebnisse erhält man jedoch erst ab 100 Datenpunkten. Der Johansentest kann schon bei 60-70 Punkten gut eingesetzt werden. Allgemein erhält man bessere Ergebnisse je länger man die Prozesse beobachtet. Diese Längen sind auch für die verschiedenen Parameterschätzverfahren ausreichend.

In den folgenden Simulationsstudien werden wir weiterhin die Längen 30, 60, 100 und 300 betrachten. Dabei repräsentiert die Länge 30 Systeme, die für eine Auswertung der Originaldaten zu kurz waren. Vielleicht bewirkt eine Einschränkung der Wertemenge jedoch, dass die Systemgruppen schon früher voneinander unterschieden werden können. Da für den Johansentest der Übergang von 60 zu 70 Datenpunkten entscheidend war, die Unterschiede zwischen den Ergebnissen von 70 zu 100 Datenpunkten jedoch nur gering ausfallen, nehmen wir Systeme der Länge 60 und 100 in die weitere Beobachtung auf. Ab 100 Datenpunkten

können auch die Werte des ADF-Test verwendet werden. Da der PP-Test und der Johansen-Test auf dem 0.05 Niveau für sehr lange Reihen ihr Verhalten ändern, werden wir die Systeme der Länge 300 auch weiterhin betrachten.

3.3 Messung auf Intervallskalenniveau

Es wurde bereits erwähnt, dass Überlegungen zum notwendigen Skalenniveau der Daten kein mathematisch-analytisches Problem sind (Bortz & Döring, 2006; Gaito, 1980). Deshalb soll hier in einer Simulationsstudie geprüft werden, wie viele Stufen auf einer Intervallskala notwendig sind, damit der Kointegrationsrang vom Johansen-Test noch richtig geschätzt werden kann.

Da in einer Vorstudie (siehe Tabellen im Anhang D.2.1, S.360) gezeigt werden konnte, dass der Johansen-Test auf diskretisierte (gerundete) Daten nicht mit großen Verzerrungen reagiert, können in den folgenden Kapiteln die Testergebnisse auf einer Intervallskala mit $n = 10, 7, 5$ oder 3 Stufen untereinander und mit den Originalergebnissen in der Längenuntersuchung aus Kapitel 3.2 verglichen werden.

Dazu wird ein Messvorgang mit der jeweiligen Intervallskala auf den Originaldaten simuliert. Dabei wird für jedes bivariate System der gegebene Wertebereich in n gleichmäßige Kapitele eingeteilt. Der Wertebereich jedes Systems liegt zwischen dem minimalen und maximalen angenommenen Wert der beiden Reihen. So hat jedes System eine eigene Skala. Dadurch wird die Individualität bei der Einschätzung der Ausprägung nachgestellt. Das Empfinden verschiedener Versuchspersonen kann man untereinander auch nicht objektiv vergleichen. Durch eine gemeinsame Skala wäre der Wertebereich so groß geworden, dass manche Reihen auf eine einzige Stufe reduziert worden wären. Durch die Varianzunterschiede der einzelnen Reihen, ist dies ohnehin auch auf der individuellen Skala eine Gefahr. In der Gruppe (3,10) nimmt die Zeitreihe, die einen mit Faktor 10 gewichteten integrierten Trend beinhaltet, Werte in einem viel größeren Bereich an als die zweite Reihe, die den Trend mit Faktor 3 gewichtet enthält. Da Minimum und Maximum für beide Reihen gemeinsam bestimmt werden, kann es vorkommen, dass die zweite Zeitreihe in einem Wertebereich verläuft, der auf der durch die erste Zeitreihe vorgegebenen Skala nur auf zwei Stufen abgebildet wird. In Abbildung 3.4 (S.183) ist ein kointegriertes System aus der Gruppe (3,10) der Länge 100 auf einer fünfstufigen Intervallskala dargestellt. Man sieht deutlich, dass die Zeitreihe mit der kleineren Varianz durch die simulierte Messung fast konstant wird, da die Aufteilung der Skalenstufen durch den Minimal- und Maximalwert der zweiten Zeitreihe bestimmt wird. In der Praxis würde ein so großer Varianzunterschied durch den Beurteiler vermutlich ausgeglichen. Dieser Fall wird durch die Gruppe (1,1) abgedeckt.

Da der Johansen-Test (fast) konstante Reihen nicht verarbeiten kann und auch für ein System aus (fast) identische Reihen kein Ergebnis liefert, wird in der Spalte `num` stets erfasst, für wie

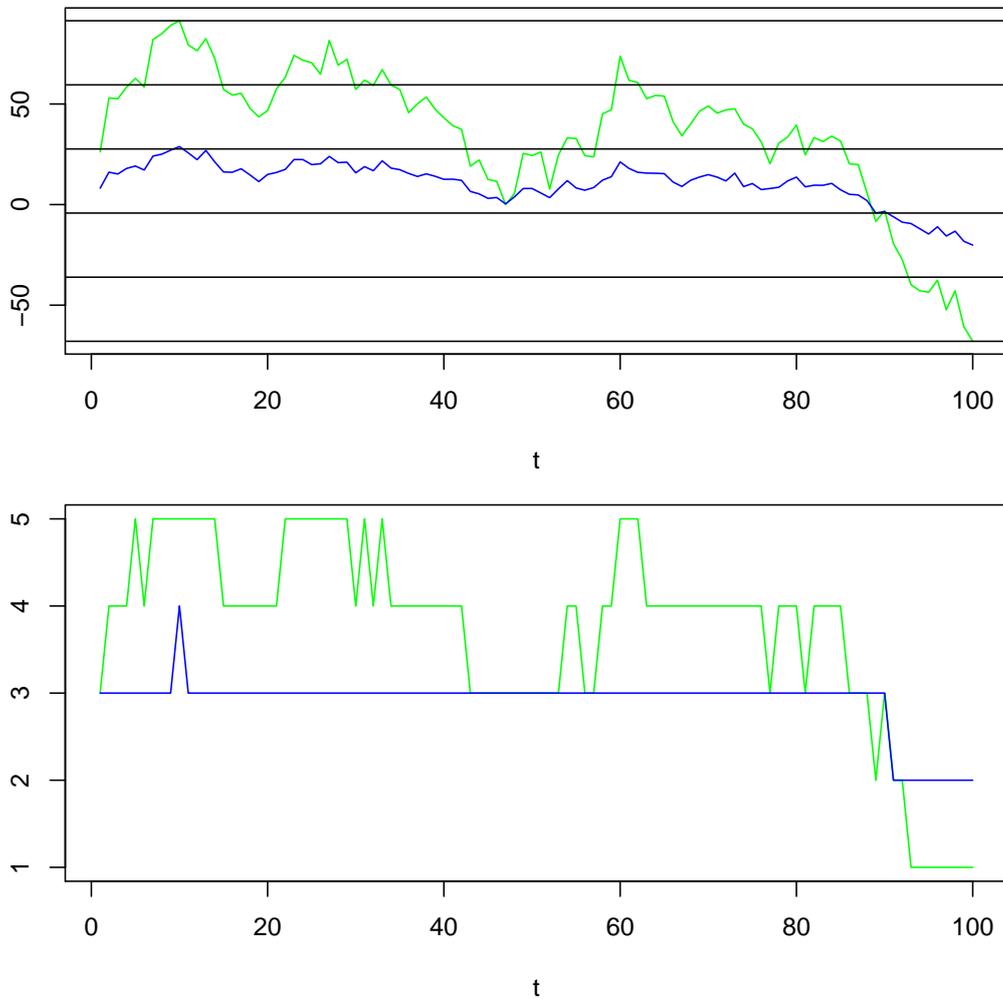


Abbildung 3.4: Kointegriertes System der Länge 100 aus der Gruppe (3,10) auf einer 5-stufigen Intervallskala. Der Wertebereich zwischen minimalem und maximalem Wert der beiden Zeitreihen wird in fünf gleichmäßige Bereiche eingeteilt. Jeder Wert innerhalb eines Bereichs wird bei der simulierten Messung auf eine Stufe der Skala abgebildet. Der blau dargestellten Zeitreihe wird durch die simulierte Messung ein fast konstanter Verlauf zuordnet, da die Stufen der Skala durch die Zeitreihe mit der größeren Varianz (in grün) bestimmt werden.

viele Systeme auf der jeweiligen Skala der Test ausgeführt werden konnte.

Das Intervall zwischen dem Minimum und Maximum der beiden Reihen jedes Systems $D = [\min, \max]$ wird jeweils in $n = 10, 7, 5$ bzw. 3 gleichgroße Teilintervall eingeteilt

$$D = I_1 \cup I_2 \cup \dots \cup I_n.$$

Jeder Wert, der im ersten Intervall I_1 liegt, wird auf die erste Stufe der Skala abgebildet usw. Dieser simulierte Messvorgang ist in Abbildung 3.4 auf Seite 183 für eine fünfstufige Skala dargestellt und wird im Folgenden für jede Skala in weiteren Abbildungen veranschaulicht. Dadurch sind alle erzeugten Skalen unipolar. Da die Bezeichnung der Stufen mit positiven Werten jedoch willkürlich festgelegt wurde, gelten die Ergebnisse in gleicher Weise für bipolare Skalen mit derselben Stufenanzahl.

Wir betrachten zunächst die Ergebnisse der Vorstudie, bevor wir die einzelnen Intervallskalen untersuchen. Als Vergleichsbasis dienen stets die Ergebnisse der Studie zur notwendigen Länge mit den unverzerrten Daten aus Kapitel 3.2.

3.3.1 Vorstudie: diskretisierte Daten

In einer Vorstudie wurde untersucht, wie sich der Johantest bei diskretisierten Daten verhält. Dazu wurden alle Messwerte nach den üblichen Regeln gerundet, wodurch die Information der Nachkommastellen verloren ging. Während der Wertebereich in den ursprünglichen Daten theoretische alle reellen Zahlen umfasste $E = \mathbb{R}$, wurde er so auf die ganzen Zahlen \mathbb{Z} eingeschränkt. Dadurch hat die simulierte Messung an Exaktheit verloren. Ohne weitere Restriktionen handelt es sich jedoch um eine Skala mit ausreichend vielen Stufen auf Intervallskalenniveau. Wie man in Abbildung 3.5 auf Seite 185 erkennt, ist die Verzerrung minimal.

Die in Anhang D.2.1 ab Seite 360 aufgeführten Ergebnisse zeigen folglich auch kaum Abweichungen zu den Originaldaten. Die größten Unterschiede gibt es bei einer Länge von $T = 30$, für die schon in der ersten Studie die Ergebnisse nicht zuverlässig waren. Bei längeren Systemen unterschieden sich die Werte der beiden Kointegrationstests sowie der Schätzungen für β in weniger als 1% der Fälle in allen Gruppen.

Basierend auf der Erkenntnis, dass der Johantest bei diskretisierten Daten angewendet werden kann, untersuchen wir nun die Messung auf Intervallskalen mit einer begrenzten Anzahl von Stufen.

3.3.2 10-stufige Intervallskala

Wir beginnen unsere Untersuchung mit einer 10-stufigen Skala. Da bei sehr vielen Stufen die Diskriminierungsfähigkeit der Versuchspersonen schnell überfordert wird (Matell & Jacoby,

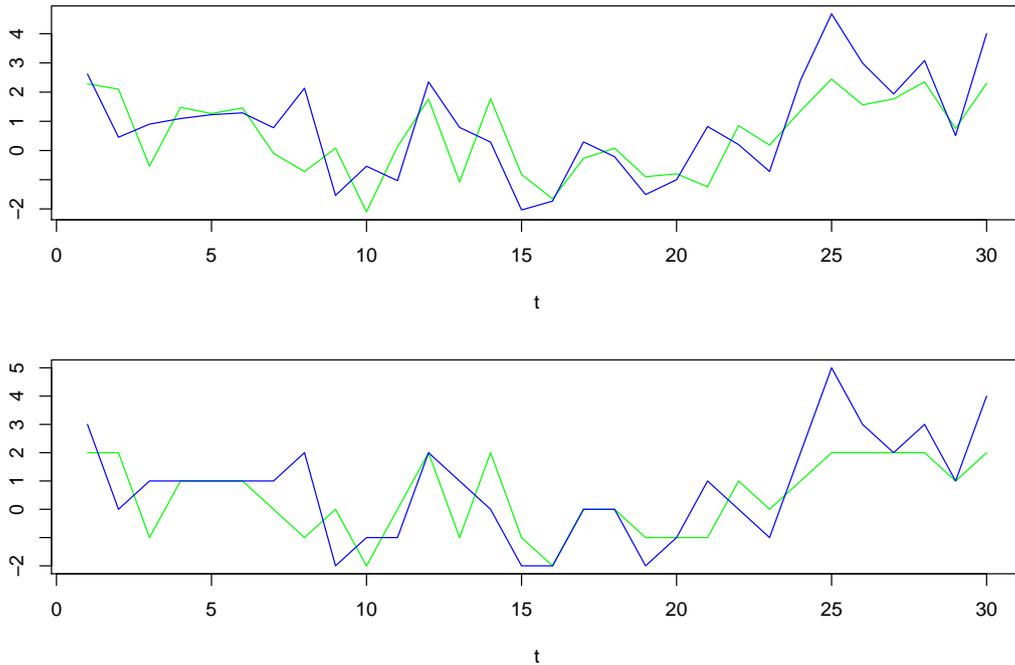


Abbildung 3.5: Kointegriertes System der Länge 30 aus der Gruppe (1,1), oben: Originaldaten, unten: diskretisierte/gerundete Werte. Der Effekt der Diskretisierung ist minimal.

1972), bietet sich diese Größe als Einstiegspunkt für die Simulationsstudien an.

In der Abbildung 3.7 auf Seite 186 ist zu sehen, wie der Verlauf des in Abbildung 3.6 (S.186) dargestellten kointegrierten Systems der Gruppe (1,1) der Länge 100 nach der simulierten Messung auf einer 10-stufigen Intervallskala verändert wird.

Der Verlauf ist noch deutlich zu erkennen und auch die Linearkombination (in schwarz) der beiden Reihen ist nur etwas „kantiger“ geworden. Die Originaldaten nehmen Werte zwischen ca. Null und -15 an, so dass alle Werte zwischen -15 und -13.5 auf die Stufe 1 der Skala abgebildet werden, Werte zwischen 13.5 und 11 erhalten den Wert 2 usw., so dass die Lage der Werte durch den Messvorgang verloren geht. Es sei noch einmal betont, dass die Vergabe von ausschließlich positiven Werten zur Darstellung der Skalenstufen jedoch willkürlich ist. Ebenso hätten in diesem speziellen Fall Werte von -1 bis -10 gewählt werden können.

Bekannter Kointegrationsvektor

Phillips-Perron-Test Beim PP-Test (siehe Tabelle 3.22, S.187) werden bei kointegrierten Systemen der **Gruppe (1,1)** bei den Längen 60, 100 und 300 alle Linearkombinationen auf der 10-stufigen Intervallskala als stationär erkannt. Bei einer Länge von 30 werden sogar 1-4 Linearkombinationen mehr als stationär erkannt als ohne Verzerrung (siehe Tabelle 3.6, S.163). Bei stationären Systemen werden ab einer Länge von 60 alle Linearkombinationen vom PP-Test als stationär erkannt. Bei einer Länge von 30 werden 1 (0.05 Niveau) bzw. 2

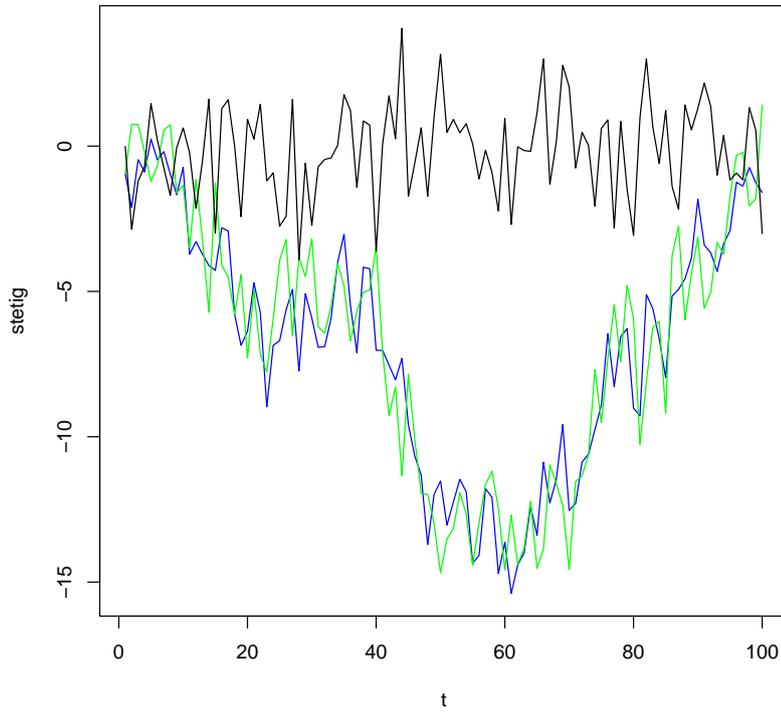


Abbildung 3.6: Kointegriertes System der Länge 100 aus der Gruppe (1,1) - Originaldaten; in schwarz die Kointegrationsbeziehung

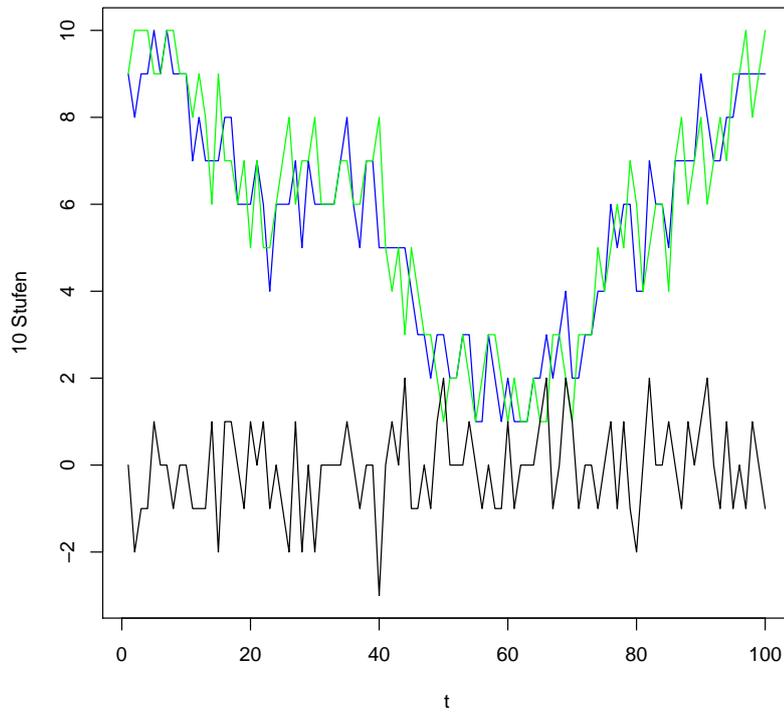


Abbildung 3.7: Kointegriertes System der Länge 100 aus der Gruppe (1,1) auf 10-stufiger Intervallskala mit positiven Werten 1-10

(0.01 Niveau) weniger erkannt. Dennoch sind dies immer noch 992 bzw. 899 Systeme.

Obwohl der PP-Test schon bei den unverzerrten Daten sehr schlecht bei der Erkennung von integrierten Systemen war, ist er durch die Transformation auf eine 10-stufige Intervallskala noch schlechter geworden. In der Gruppe (1,1) werden bei einer Länge von 300 Punkten (mehr als) doppelt so viele Linearkombinationen fälschlicherweise als stationär eingestuft. Statt etwa 300 nun mehr als 600 Systeme. Nur bei einer Länge von 30 Punkten werden in dieser Gruppe auf dem 0.01 Signifikanzniveau mehr als 800 Linearkombinationen als nicht-stationär klassifiziert, bei allen anderen Längen werden mindestens ein Drittel, zum Teil sogar zwei Drittel falsch eingeschätzt.

In der **Gruppe (2,3)** erkennt der PP-Test (siehe Tabelle 3.22, S.187 Mitte) ab Länge 60 alle 1000 stationären Linearkombinationen der kointegrierten Systeme. Bei einer Länge von 30 Datenpunkten sogar 1-5 mehr als bei den unverzerrten Daten (siehe Tabelle 3.8, S.165), wodurch er die Werte von 909, 994 bzw. 999 erreicht. Bei den stationären Systemen in dieser Gruppe hat der PP-Test nur eine leichte Abweichung bei einer Länge von 30, sonst klassifiziert er hier perfekt. Bei integrierten Systemen wird der PP-Test auch in dieser Gruppe mit zunehmender Länge schlechter und ist insgesamt nur auf dem 0.01 Niveau bei einer Länge von 30 Beobachtungspunkten zufriedenstellend.

Tabelle 3.22: Ergebnisse des PP-Tests auf 10-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	999	995	933	999	992	899	528	398	190
	60	1000	1000	1000	1000	1000	1000	644	538	356
	100	1000	1000	1000	1000	1000	1000	675	566	392
	300	1000	1000	1000	1000	1000	1000	702	619	457
(2,3)	30	999	994	909	997	994	908	251	157	40
	60	1000	1000	1000	1000	1000	1000	336	236	95
	100	1000	1000	1000	1000	1000	1000	416	312	166
	300	1000	1000	1000	1000	1000	1000	479	399	255
(3,10)	30	996	982	864	1000	994	920	264	170	56
	60	1000	999	997	1000	1000	1000	407	300	141
	100	1000	1000	1000	1000	1000	1000	509	399	233
	300	1000	1000	1000	1000	1000	1000	616	521	361

Für kointegrierte Systeme ist die Klassifikation des PP-Tests in der **Gruppe (3,10)** für die Systeme der Längen 30 und 60 leicht verändert. Auf dem 0.1 Niveau werden bei Länge 30 zwei Linearkombinationen weniger als stationär eingeschätzt; auf den anderen Niveaus werden bei Länge 60 nicht mehr alle stationären Linearkombinationen erkannt. Bei den stationären Systemen gibt es nur bei Länge 30 einige wenige Linearkombinationen, die nun

als integriert eingeschätzt werden. Bei den integrierten Systemen wird auf keinem Niveau bei keiner Länge mehr eine Klassifikationsrate von 95% erreicht. Bei den unverzerrten Daten (vgl. Tabelle 3.10, S.167) war dies auf dem 0.05 und 0.01 Niveau für alle Längen der Fall gewesen. Nun wird bei den Systemen der Länge 300 mehr als die Hälfte der Linearkombinationen auf dem 0.1 und 0.05 Niveau für stationär gehalten.

Tabelle 3.23: Ergebnisse des ADF-Tests auf 10-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	321	179	53	342	204	56	88	45	6
	60	891	767	421	886	739	407	83	42	9
	100	989	954	743	989	958	765	81	39	5
	300	1000	1000	1000	1000	1000	1000	76	40	4
(2,3)	30	317	202	54	335	198	47	81	37	3
	60	880	737	407	890	753	396	87	39	6
	100	985	955	756	987	959	782	91	48	12
	300	1000	1000	1000	1000	1000	1000	82	40	9
(3,10)	30	301	195	57	319	187	46	83	34	7
	60	814	665	324	871	748	407	91	39	6
	100	934	866	614	985	953	773	99	53	9
	300	995	987	937	1000	1000	1000	145	86	25

Augmented Dickey-Fuller-Test Der ADF-Test (siehe Tabelle 3.23, S.188) erkennt bei einer Länge von 300 in den **Gruppen (1,1) und (2,3)** noch alle stationären Linearkombinationen der kointegrierten und stationären Systeme. Bei den anderen Längen gibt es im Vergleich zu den unverzerrten Daten (siehe Tabellen 3.7, S.164 und 3.9, S.166) Abweichung bis zu 25 Systeme bei den kointegrierten und 1-10 Abweichungen bei den stationären Systemen. Dabei werden zum Teil weniger, zum Teil mehr Linearkombinationen für stationär gehalten. Weiterhin kann er erst ab einer Länge von 100 auf den Niveaus 0.05 und 0.1 sinnvoll eingesetzt werden. Bei kürzeren Reihen wird die Anzahl der stationären Linearkombinationen zu stark unterschätzt.

Bei integrierten Systemen kann man den ADF-Test auf dem 0.01 und 0.05 Niveau auch mit transformierten Daten verwenden. Nur wenige Linearkombinationen werden als stationär missklassifiziert. Auf dem 0.01 Niveau werden in den Gruppen (1,1) und (2,3) maximal 6 Systeme mehr falsch klassifiziert als bei den Originaldaten, was immer noch weniger als 1% aller Systeme ausmacht. Auf den anderen Niveaus werden maximal 12 Systeme anders eingestuft, jedoch meist mehr fehlerklassifiziert. Der ADF-Test ist bei integrierten Systemen bei einer Länge von 300 in der **Gruppe (3,10)** deutlich schlechter als bei den Originaldaten (bis zu 71 Systemen Abweichung) (vgl. Tabelle 3.11, S.168), aber auf dem 0.01 Niveau sind

die Ergebnisse noch sehr gut. Hier werden über 950, meist gar über 990 Systeme richtig als integriert erkannt. Auf dem 0.05 Niveau ist die Klassifikationsrate immer noch gut, da hier ca. 950 Systeme richtig eingeschätzt werden. Auf dem 0.1 Niveau sind es bei den Längen unter 300 Datenpunkten noch mehr als 900. Bei den langen Systemen wird durch den eingeschränkten Wertebereich die Wahrscheinlichkeit größer, dass der Test die eingeschränkte Dynamik als stationär interpretiert. Bei den kointegrierten Systemen ist das Ergebnis viel schlechter. Bei 300 Datenpunkten werden nicht mehr alle 1000 Kointegrationsbeziehungen erkannt und bei Länge 100 gibt es auf dem 0.01 Niveau mehr als 100 Linearkombinationen Abweichung (vgl. Tabelle 3.11, S.168). Für stationäre Systeme ist die Auswertung fast identisch zu den unverzerrten Daten.

Bemerkung 3.3.1 *In der Gruppe (3,10) ist der Varianzunterschied zu groß, so dass die Reihe mit der kleineren Varianz auf eine fast konstante Reihe abgebildet wird. Deshalb gibt es hier mehr Fehlklassifikationen als in den anderen Gruppen. Wahrscheinlich ist es allgemein besser, die Reihen auf unterschiedlichen Skalen zu messen. Bei der Konstruktion der 10 Stufen werden Minimum und Maximum über den gemeinsamen Wertebereich beider Reihen bestimmt (siehe Abbildung 3.4, S.183).*

Johansentest

Tabelle 3.24: Klassifikation 1000 kointegrierter Systeme auf 10-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	284	616	100	446	517	37	744	250	6	1000
	60	0	875	125	11	915	74	65	917	18	1000
	100	0	863	137	0	927	73	0	977	23	1000
	300	0	855	145	0	920	80	0	991	9	1000
(2,3)	30	262	650	88	423	545	32	750	245	5	1000
	60	2	891	107	7	945	48	68	918	14	1000
	100	0	903	97	0	951	49	0	986	14	1000
	300	0	905	95	0	956	44	0	997	3	1000
(3,10)	30	277	630	93	418	537	45	705	291	4	1000
	60	24	876	100	51	898	51	188	802	10	1000
	100	5	900	95	11	941	48	43	943	14	1000
	300	0	910	90	1	956	43	2	992	6	1000

Kointegrierte Systeme Bei den kointegrierten Systemen wurde keine Reihe so verzerrt, dass der Johansentest nicht mehr ausgeführt werden konnte. Zur Auswertung können also in allen drei Gruppen alle 1000 Systeme herangezogen werden.

Bei einer Länge von 30 Datenpunkten gibt es in der **Gruppe (1,1)** auf allen Niveaus einige Abweichungen von der Klassifikation der unverzerrten Daten (siehe Tabelle 3.12, S.170). Jedoch werden maximal 22 Systeme anders klassifiziert. Bei dieser Länge wird jedoch weiterhin ein Großteil der Systeme falsch eingestuft.

Bei den Längen 60 und 100 gibt es kaum Abweichungen zur ursprünglichen Klassifikation. Maximal 5 Systeme werden anders zugeordnet, was weniger als 1% ausmacht. Bei einer Länge von 300 Beobachtungspunkten werden 19, 11 bzw. 1 Systeme statt als kointegriert nun als stationär eingestuft. Auch diese Abweichung ist bei 1000 Wiederholungen sehr gering.

Das Klassifikationsverhalten hat sich auch kaum verändert. Ab einer Länge von 60 werden mehr als 900 Systeme in allen Gruppen auf dem 0.05 Niveau als kointegriert erkannt. Der Längeneffekt in allen Niveaus bleibt deutlich zu erkennen. Auf dem 0.01 Niveau werden bei einer Länge von 300 fast alle (mehr als 99%) Systeme als kointegriert erkannt.

In der **Gruppe (2,3)** werden nur bei der Länge 30 mehr als 10 Systeme anders klassifiziert als bei unverzerrten Daten. Der Zeiteffekt bleibt klar erkennbar. Ab einer Länge von 100 Datenpunkten gibt es kein als integriert eingestuftes System mehr und auf dem 0.01 Niveau werden stets mehr Systeme richtig erkannt als auf 0.05 oder 0.1. Durch die geringe Abweichungszahl bleibt auch das Klassifikationsverhalten gleich. Ab einer Länge von 60 bekommt man auf dem 0.01 und 0.05 Niveau gute, ab 100 Beobachtungspunkten sogar sehr gute Ergebnisse.

Auch in der **Gruppe (3,10)** ist der Längeneffekt in allen Niveaus zu erkennen. Die Zahl der Abweichungen ist jedoch viel größer als in den anderen beiden Gruppen. Bei einer Länge von 300 Punkten sind dies nur 1-2 Systeme. Am auffälligsten ist die Abweichung auf dem 0.01 Niveau. Hier werden bei einer Länge von 60 Punkten 113 kointegrierte und 4 zuvor als stationär eingestufte Systeme nun als integriert klassifiziert. Denselben Effekt hat man auch auf den anderen Niveaus und auch bei der Länge 100. Jedoch nicht so stark wie bei Länge 60 auf dem 0.01 Niveau. Diese großen Abweichungen in der Gruppe (3,10) erklären sich durch die Konstruktion der Skala, die sich vor allem nach der Reihe mit der größeren Varianz richtet. Da ist es schon bemerkenswert, dass ab einer Länge von 100 mehr als 90% als kointegriert erkannt werden.

Die **Schätzungen für β** werden in keiner der drei Gruppen spürbar beeinflusst (siehe Tabelle 3.25, S.191). Alle Statistiken des Box-Plots sind bis auf die erste (meist sogar bis auf die zweite) Nachkommastelle identisch zu den unverzerrten Schätzungen. Lediglich bei Minimum und Maximum gibt es einige größere Abweichungen. Diese sind jedoch auch in der Gruppe (3,10) nicht besonders groß, obwohl es auf dem 0.01 Niveau einige Klassifikationen als integriert gab.

stationäre Systeme Auch bei den stationären Systemen (siehe Tabelle 3.26, S.191) wurden für die Auswertung alle 1000 Systeme in allen Gruppen berücksichtigt. Man kann einen

Tabelle 3.25: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf 10-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-34,341	-1,486	-1,137	-1,007	-0,896	-0,537	19,432	69
	60	-1,779	-1,198	-1,056	-1,004	-0,958	-0,821	-0,377	57
	100	-1,435	-1,125	-1,034	-1,003	-0,974	-0,885	-0,691	40
	300	-1,073	-1,041	-1,010	-0,999	-0,988	-0,957	-0,936	25
(2,3)	30	-7,220	-0,836	-0,718	-0,676	-0,639	-0,523	2,932	71
	60	-4,452	-0,747	-0,692	-0,672	-0,654	-0,599	-0,490	44
	100	-0,854	-0,724	-0,686	-0,672	-0,658	-0,619	-0,570	26
	300	-0,707	-0,696	-0,676	-0,670	-0,663	-0,644	-0,633	12
(3,10)	30	-3,285	-0,395	-0,328	-0,306	-0,282	-0,216	0,126	35
	60	-0,417	-0,361	-0,317	-0,302	-0,287	-0,244	-0,229	10
	100	-0,367	-0,351	-0,314	-0,302	-0,289	-0,251	-0,234	12
	300	-0,349	-0,345	-0,314	-0,303	-0,293	-0,262	-0,253	18

Tabelle 3.26: Klassifikation 1000 stationärer Systeme auf 10-stufiger Intervallskala

(a,b)	T	Niveau 0.1			Niveau 0.05			Niveau 0.01			num
		I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	
(1,1)	30	89	96	815	204	188	608	552	258	190	1000
	60	0	0	1000	0	0	1000	2	8	990	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	93	102	805	210	208	582	560	252	188	1000
	60	0	0	1000	0	1	999	6	12	982	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	94	101	805	220	207	573	550	269	181	1000
	60	0	0	1000	0	0	1000	6	7	987	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

leichten Längeneffekt erkennen, der allerdings durch die Qualität der Klassifikation kaum auszumachen ist. Ab einer Länge von 100 Punkten erzielt der Test in allen Gruppen auf allen Niveaus eine perfekte Klassifikation. So gibt es auch nur bei den Längen 30 und 60 Abweichungen zu den ursprünglichen Daten (vgl. Tabellen 3.16, S.174, 3.17, S.175 und 3.18, S.175). Diese sind bei einer Länge von 30 Punkten deutlich größer als bei 60. Es werden jedoch nie mehr als 24 Systeme anders klassifiziert. Die Werte für β streuen weiterhin um Null (siehe Tabelle D.7, S.364). Auch die anderen Boxplot-Statistiken haben keine große Abweichung.

Tabelle 3.27: Klassifikation 1000 integrierter Systeme auf 10-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	797	177	26	882	110	8	970	29	1	1000
	60	765	224	11	858	138	4	965	34	1	1000
	100	730	245	25	823	171	6	944	56	0	1000
	300	598	363	39	714	273	13	862	137	1	1000
(2,3)	30	822	164	14	901	96	3	979	20	1	1000
	60	845	144	11	915	81	4	982	18	0	1000
	100	809	176	15	891	107	2	960	40	0	1000
	300	714	269	17	806	189	5	925	75	0	1000
(3,10)	30	765	168	14	841	100	6	915	31	1	947
	60	787	188	12	868	115	4	945	42	0	987
	100	707	273	9	816	169	4	917	72	0	989
	300	551	420	21	644	343	5	793	199	0	992

Integrierte Systeme Bei den integrierten Systemen (siehe Tabelle 3.27, S.192) ist die Verzerrung in der Gruppe (3,10) so groß, dass der Johansentest bereits nicht mehr für alle Systeme ausgeführt werden kann, da eine der Reihen nun als nahezu konstant betrachtet wird. Auffällig ist zudem, dass die Klassifikation bei einer Länge von 300 auf allen Niveaus besonders schlecht ist. Auch generell sind die Abweichungen bei den integrierten Systemen deutlich größer als in den anderen Gruppen. Dabei bleibt der Niveaueffekt bestehen, dass auf dem 0.01 Niveau die Klassifikation noch am besten ist. Hier werden weiterhin sehr gute Werte mit mehr als 900 richtig klassifizierten Systemen. In den Gruppen (1,1) und (2,3) sogar 940-990 Systemen erzielt. Es ist deutlich zu erkennen, dass die Abweichungen mit der Länge zunehmen. Der Trend geht dahin, integrierte Systeme fälschlicherweise als kointegriert zu klassifizieren. Dies war zu erwarten, da durch die Einführung der Stufen Informationen über den Trend verloren gehen. Und deshalb eine Schwankung eher als gemeinsam abgebildet wird. Der Effekt, dass die Varianz gegen unendlich strebt, wird durch die eingeschränkte Anzahl von Werten, gehemmt. Man sieht auch, dass durch die Länge der Reihen verstärkt der Eindruck eines Common Trends entsteht, wenn die Reihen integriert sind. Die Boxplot-

Statistiken für die Schätzung von β (siehe Tabelle D.7, S.364) haben sich kaum verändert. Minimum und Maximum nehmen ganz unterschiedliche extreme Werte an. Der Median liegt immer noch um Null.

Ergebnis: 10-stufige Intervallskala

Bekannter Kointegrationsvektor Der Phillips-Perron-Test erkennt nach Transformation die integrierten Linearkombinationen nicht mehr, sondern schätzt den Großteil fälschlicherweise als stationär, die Systeme folglich als kointegriert ein. Der ADF-Test hingegen zeigt kaum Abweichungen zu den nicht-transformierten Daten. Dies bedeutet jedoch auch, dass er erst ab einer Länge von 100 Datenpunkten zuverlässige Ergebnisse bringt. Von der Verwendung des PP-Test bei Daten auf einer 10-stufigen Skala muss abgeraten werden.

Johansen-Test Bei den kointegrierten und stationären Systemen gibt es kaum Abweichungen in der Klassifikation durch den Johansen-Test zu den unverzerrten Daten. Die Systeme der Längen 100 und 300 werden fast alle genauso eingeschätzt, wie bei den Originaldaten. Die Einschränkung des Wertebereichs bedeutet jedoch einen Informationsverlust über den Verlaufs eines integrierten Trends. Wir hatten in Kapitel 2.1.5 integrierte Prozesse unter anderem dadurch charakterisiert, dass die Varianz mit zunehmender Länge des Prozesses immer größer wird. Durch die endliche Wertemenge kann diese Eigenschaft nicht mehr abgebildet werden. So wirkt sich bei langen Zeitreihen ein Ausbleiben starker Schwankungen auf die Klassifikation aus. Die Abbildung verschiedener integrierter Zeitreihen verläuft auf 10 Stufen so ähnlich, dass die Systeme fälschlicherweise für kointegriert gehalten werden.

3.3.3 7-stufige Intervallskala

Für die meisten Fragestellungen ist eine gerade Anzahl von Skalenstufen ungeeignet. Häufig wird in psychologischen Fragebögen eine siebenstufige Ratingskala eingesetzt. Die folgenden Ergebnisse werden zeigen, ob man auf einer solchen Skala integrierte und kointegrierte Prozesse noch unterscheiden kann.

Bekannter Kointegrationsvektor

Phillips-Perron-Test In den Gruppen (1,1) und (2,3) werden für die **kointegrierten Systeme** außer bei Länge 30 auf allen Niveaus alle 1000 Linearkombinationen als stationär erkannt (siehe Tabelle 3.29, S.195). Bei einer Länge von 30 Punkten sind es jedoch auch noch stets über 900, auf dem 0.05 und 0.1 Niveau sogar mehr als 989 richtig eingestufte Linearkombinationen. So gibt es auch nur bei Länge 30 Abweichungen von den unverzerrten Daten. In der dritten Gruppe (3,10) werden bei kointegrierten Systemen der Länge 300 auch

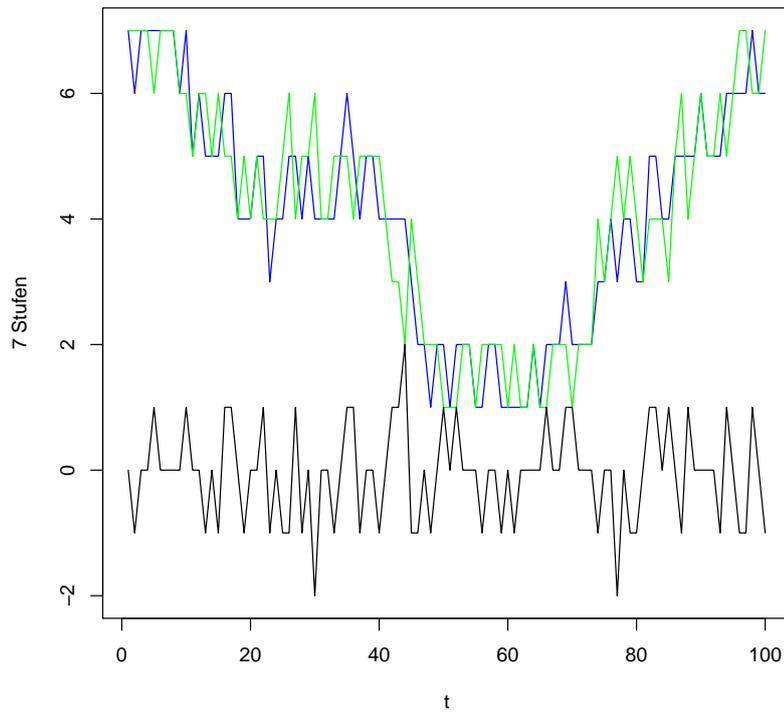


Abbildung 3.8: Kointegriertes System der Länge 100 aus der Gruppe (1,1) auf einer 7-stufigen Intervallskala

Tabelle 3.28: Ergebnisse des PP-Tests auf 7-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	998	992	911	999	992	904	586	443	221
	60	1000	1000	1000	1000	1000	1000	718	625	411
	100	1000	1000	1000	1000	1000	1000	750	671	514
	300	1000	1000	1000	1000	1000	1000	846	767	634
(2,3)	30	999	992	900	999	989	904	311	212	69
	60	1000	1000	1000	1000	1000	1000	446	335	172
	100	1000	1000	1000	1000	1000	1000	548	433	262
	300	1000	1000	1000	1000	1000	1000	668	569	411
(3,10)	30	974	934	737	998	991	909	322	196	66
	60	996	990	967	1000	1000	1000	500	394	204
	100	1000	998	990	1000	1000	1000	585	483	304
	300	1000	1000	1000	1000	1000	1000	745	656	505

alle Linearkombinationen perfekt klassifiziert. Bei kürzeren Zeitreihen werden jedoch einige nicht mehr als stationär erkannt. Dabei ist die Abweichungsrate auf dem 0.01 Niveau am größten und damit die Klassifikationsleistung am schlechtesten. Mit der Länge nimmt auch die Anzahl der richtig erkannten stationären Linearkombinationen zu. Und nur bei einer Länge von 30 Datenpunkten werden weniger als 950 Systeme richtig eingeschätzt. Dennoch ist ein erster Effekt durch die Verringerung der Stufenanzahl zu erkennen. Der Effekt ist in der Gruppe (3,10) aus den oben genannten Gründen (siehe Bemerkung 3.3.1, S.189) am stärksten.

Der Phillips-Perron-Test erkennt auch die stationären Linearkombinationen in allen Gruppen von **stationären Systemen** ab einer Länge von 60 auf allen Niveaus perfekt. So gibt es auch nur bei einer Länge von 30 leichte Abweichungen von den Originalwerten in den Tabellen 3.6, 3.8 und 3.10 (S.163, 165 und 167). Dennoch werden ähnlich wie bei kointegrierten Systemen stets mehr als 900, auf dem 0.05 und 0.1 Niveau sogar ca. 99% der Systeme richtig eingestuft.

Den Informationsverlust durch die 7-stufige Intervallskala merkt man für **integrierte Systeme** deutlich, während der ADF-Test darauf kaum reagiert. Wie schon auf der 10-stufigen Intervallskala versagt der PP-Test total. Bis zu 846 Linearkombinationen werden fälschlicherweise als stationär eingeschätzt. Auf dem 0.01 Niveau sind es immer noch bis zu 634, wobei mit steigender Länge immer mehr Linearkombinationen für stationär gehalten werden. Die Abweichungen werden mit zunehmender Länge so auch immer größer.

Tabelle 3.29: Ergebnisse des ADF-Tests auf 7-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	321	189	36	338	200	47	84	38	7
	60	893	755	404	880	745	409	96	41	7
	100	982	949	761	990	962	774	100	48	6
	300	1000	1000	1000	1000	1000	1000	113	59	13
(2,3)	30	290	164	46	317	177	49	87	44	14
	60	853	712	376	871	728	393	95	43	9
	100	982	949	743	987	953	762	91	42	9
	300	1000	1000	994	1000	1000	1000	114	67	17
(3,10)	30	298	171	47	319	167	45	76	38	6
	60	691	521	235	877	741	400	113	58	16
	100	798	677	352	990	964	775	124	63	21
	300	948	904	701	1000	1000	1000	179	104	35

Augmented Dickey-Fuller-Test Der ADF-Test hatte schon bei den unverzerrten Daten nur bei einer Länge von 300 alle 1000 Linearkombinationen der **kointegrierten Systeme**

als stationär erkannt (vgl. Tabellen 3.7, 3.9 und 3.11). Dies gelingt auf der 7-stufigen Intervallskala nur in der Gruppe (1,1) noch auf allen Niveaus (siehe Tabelle 3.29, S.195). In der Gruppe (2,3) gilt dies immerhin noch auf den Niveaus 0.1 und 0.05; auf dem 0.01 Niveau sind 994 richtig klassifizierte Systeme immer noch eine gute Erfolgsrate. In der Gruppe (3,10) werden überhaupt nur noch bei einer Länge von 300 auf dem 0.1 Niveau nahezu 95% (948) stationäre Linearkombinationen erkannt. Während es in den anderen beiden Gruppen Abweichungen von 20-40 Linearkombinationen zu den unverzerrten Daten aus dem Kapitel zur Länge einer Zeitreihe gibt, die nicht als stationär eingestuft werden, sind es in dieser Gruppe bis zu fast 400 (0.01 Länge 100), die nicht erkannt werden, so dass bei einer Länge von 100 und darunter auf keinem Niveau eine ausreichend gute Erkennungsrate erreicht wird.

Beim ADF-Test gibt es in der Gruppe der **stationären Systeme** nur geringe Abweichungen zu den Originaldaten. Es werden sogar einige Linearkombinationen mehr als stationär eingestuft, als dies bei den unverzerrten Daten der Fall war. Die Abweichung beträgt jedoch maximal 30 Systeme, so dass sich das Klassifikationsverhalten kaum verändert hat. Auch bei einer 7-stufigen Intervallskala ist der ADF-Test erst ab einer Länge von 100 auf dem 0.05 Niveau zuverlässig.

Der ADF-Test klassifiziert bei den **integrierten Systemen** wie der PP-Test mehr Linearkombinationen als stationär, als dies bei den unverzerrten Daten der Fall war. Dennoch sind die Abweichungen maximal 40 (auf dem 0.1 Niveau bei Länge 300), so dass sich das Klassifikationsverhalten kaum ändert. Dennoch ist auch auf der 7-stufigen Skala ein Längeneffekt zu erkennen. Mit zunehmender Länge wird in immer mehr Linearkombinationen die Einheitswurzel nicht erkannt. Auf dem 0.01 Niveau ist die Klassifikationsrate jedoch in allen Gruppen noch sehr gut und auf dem 0.05 Niveau nur bei einer Länge von 300 Datenpunkten nicht ausreichend. Auch dieser Effekt ist leicht zu erklären. Je länger die Reihen und je kleiner der Wertebereich ist, desto eher wird der Verlauf als regelmäßig interpretiert und die Integration nicht erkannt.

Johansenstest

Kointegrierte Systeme Die kointegrierten Systeme (siehe Tabelle 3.30, S.197) wurden durch die Abbildung auf die 7-stufige Skala noch nicht so stark verzerrt, so dass der Johansenstest bei allen Systemen der Gruppen (1,1) und (2,3) ausgeführt werden konnte und nur bei 8 der Länge 30 in der Gruppe (3,10) nicht. Die Klassifikation ist auf dem 0.01 Niveau immer noch im Allgemeinen am besten. Auf dem 0.1 Niveau wird für keine Länge in keiner Gruppe eine richtige Klassifikationsrate von mehr als 90% erreicht. Die Abweichungen zu den unverzerrten Daten (vgl. Tabellen 3.12, 3.13) sind in den Gruppen (1,1) und (2,3) noch relativ gering, maximal 40 Systeme werden anders klassifiziert. Dabei werden meist kointegrierte Systeme nun als stationär eingeschätzt. In der Gruppe (3,10) gibt es besonders auf dem 0.01 Niveau große Abweichungen (vgl. Tabelle 3.14). Bis zu 371 kointegrierte Systeme (Länge 60, 0.01 Niveau) werden fälschlicherweise als integriert klassifiziert. Dennoch werden

auf dem 0.01 Niveau bei einer Länge von 300 Datenpunkten sehr gute Werte erreicht. Auch die Klassifikation bei Länge 100 ist bis auf die Gruppe (3,10) noch sehr zuverlässig. Der Informationsverlust beim Übergang von 10 auf 7 Stufen bewirkt eine weitere Verschlechterung der Klassifikationsleistung, die insbesondere in der dritten Gruppe bemerkbar ist. Die Schätzungen für β werden hingegen durch die verwendete Skala kaum beeinflusst (siehe Tabelle 3.31, S.198). Der Median liegt in allen drei Gruppen für alle Längen sehr nahe am wirklichen Wert. Da es durch die Einschränkung der Skalenstufen vermehrt zu Fehlklassifikationen kommt, werden auch Box und Whisker etwas breiter, da die Schätzwerte für integrierte Systeme fast beliebige Werte annehmen können.

Tabelle 3.30: Klassifikation 1000 kointegrierter Systeme auf 7-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	273	621	106	430	530	40	763	232	5	1000
	60	1	872	127	9	922	69	77	905	18	1000
	100	0	854	146	0	922	78	0	978	22	1000
	300	0	834	166	0	907	93	0	981	19	1000
(2,3)	30	262	651	87	418	549	33	730	267	3	1000
	60	3	904	93	13	939	48	93	894	13	1000
	100	0	894	106	0	954	46	1	989	10	1000
	300	0	888	112	0	950	50	0	994	6	1000
(3,10)	30	356	555	81	532	432	28	765	224	3	992
	60	93	819	88	154	805	41	371	618	11	1000
	100	28	880	92	51	906	43	169	819	12	1000
	300	2	880	118	6	947	47	18	976	6	1000

Stationäre Systeme Der Johantest konnte bei fast allen stationären Systemen auf der 7-stufigen Intervallskala ausgeführt werden (siehe Tabelle 3.32, S.198). Lediglich 19 Systeme der Länge 30 und 2 Systeme der Länge 60 aus der Gruppe (3,10) erfüllten die Voraussetzungen an eine Testausführung nicht mehr. Ab einer Länge von 100 ist die Klassifikation auf allen Niveaus in allen Gruppen perfekt. Nur bei Länge 30 reicht die Klassifikationsquote nicht aus. Auch gibt es nur hier größere Abweichungen. Bei allen anderen Längen werden maximal 4 Systeme anders klassifiziert als bei den unverzerrten Daten (vgl. Tabellen 3.16, 3.17 & 3.18, S.174 und 175). Auf 7 Stufen gibt es ein ganz ähnliches Bild wie bei einer Skala mit 10 Werten.

Integrierte Systeme Für integrierte Systeme (siehe Tabelle 3.33, S.199) ist die Klassifikation nur auf dem 0.01 Niveau noch ausreichend gut, obwohl auch auf den anderen Niveaus in allen Gruppen die meisten integrierten Systeme als solche erkannt werden. Stets ist die Erkennungsrate bei Länge 300 am schlechtesten. Der Effekt wirkt sich nun jedoch auch bereits

Tabelle 3.31: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf 7-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-7,251	-1,516	-1,143	-1,005	-0,894	-0,527	15,865	77
	60	-1,790	-1,217	-1,057	-1,005	-0,951	-0,795	-0,401	56
	100	-1,401	-1,136	-1,038	-1,001	-0,971	-0,876	-0,538	37
	300	-1,076	-1,052	-1,013	-0,999	-0,987	-0,948	-0,924	13
(2,3)	30	-44,885	-0,871	-0,730	-0,683	-0,636	-0,498	10,162	56
	60	-0,864	-0,780	-0,704	-0,678	-0,652	-0,574	-0,480	27
	100	-0,833	-0,744	-0,694	-0,674	-0,658	-0,606	-0,581	21
	300	-0,748	-0,722	-0,687	-0,674	-0,663	-0,628	-0,617	7
(3,10)	30	-5,110	-0,453	-0,345	-0,305	-0,272	-0,162	2,538	39
	60	-0,537	-0,425	-0,336	-0,306	-0,276	-0,193	-0,001	9
	100	-0,507	-0,405	-0,332	-0,307	-0,282	-0,211	-0,102	10
	300	-0,411	-0,411	-0,336	-0,310	-0,285	-0,208	-0,203	2

bei Länge 100 aus. Und selbst auf dem 0.01 Niveau gibt es in der Gruppe (1,1) bei dieser Länge Systeme, die sogar als stationär fehlklassifiziert werden.

In der Gruppe (3,10) ist die Verzerrung so groß, dass der Johantest für einige Systeme gar nicht mehr ausgeführt werden kann. Dies gilt auch für zwei Systeme der Länge 30 aus der Gruppe (2,3). Diese Systeme werden also nicht als integriert erkannt, da eine Reihe nahezu konstant abgebildet wird und somit vom Test als stationär interpretiert wird. Dieser Effekt tritt dadurch auf, dass die Skalenstufen aus dem Wertebereich beider Zeitreihen gemeinsam bestimmt wurde.

Tabelle 3.32: Klassifikation 1000 stationärer Systeme auf 7-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	90	94	816	204	200	596	521	291	188	1000
	60	0	0	1000	0	1	999	5	8	987	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	83	92	825	206	194	600	538	277	185	1000
	60	0	0	1000	0	1	999	7	12	981	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	99	89	793	214	188	579	527	268	186	981
	60	0	0	998	0	0	998	5	10	983	998
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Tabelle 3.33: Klassifikation 1000 integrierter Systeme auf 7-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	777	205	18	869	127	4	970	30	0	1000
	60	729	253	18	825	169	6	941	57	2	1000
	100	633	334	33	768	222	10	911	88	1	1000
	300	417	505	78	543	426	31	743	253	4	1000
(2,3)	30	821	168	8	896	95	6	972	24	1	997
	60	820	170	10	901	96	3	976	24	0	1000
	100	756	230	14	850	148	2	960	40	0	1000
	300	574	385	41	680	303	17	844	156	0	1000
(3,10)	30	673	162	7	741	99	2	818	24	0	842
	60	705	194	13	777	131	4	864	48	0	912
	100	626	294	12	727	203	2	860	72	0	932
	300	464	444	54	558	381	23	714	248	0	962

Ergebnis: 7-stufige Intervallskala

Bekannter Kointegrationsvektor Auf einer 7-stufigen Intervallskala erkennt der Phillips-Perron-Test viel zu wenig integrierte Linearkombinationen, so dass er für eine Unterscheidung von integrierten und kointegrierten Systemen nicht geeignet ist. Der Augmented Dickey-Fuller-Test zeigt nur bei den stationären Systemen geringe Abweichungen. Der Informationsverlust durch die Einschränkung des Wertebereichs macht sich vor allem in der Gruppe (3,10) auch schon bei den kointegrierten Systemen bemerkbar. Bei 300 Datenpunkten werden, wie in der Gruppe (2,3) auf dem 0.01 Niveau ebenfalls, nicht mehr alle Linearkombinationen als stationär erkannt. Wie auf der 10-stufigen Intervallskala werden die integrierten Systeme der Länge 300 nicht mehr als solche erkannt.

Johansentest Insgesamt schneidet der Johansentest auf der 7-stufigen Intervallskala am besten für eine Länge von 60-100 Beobachtungspunkten ab. Durch den Informationsverlust kommt es vermehrt zu Fehlklassifikationen. In der Gruppe (3,10) kann der Test für erste stationäre, integrierte und einzelne kointegrierte Systeme bereits nicht mehr ausgeführt werden, da eine der Reihen als fast konstant auf eine der sieben Stufen abgebildet wird. Auch bei kointegrierten und integrierten Systemen hat sich die Klassifikation insbesondere in dieser Gruppe und vor allem bei den längsten Systemen verschlechtert. Die Schätzung von β wird jedoch kaum beeinflusst. Für stationäre und integrierte Systeme liegen die Schätzwerte symmetrisch um Null und die Werte der Boxplot-Statistiken verändert sie sich kaum (siehe Tabelle D.8 im Anhang auf Seite D.8).

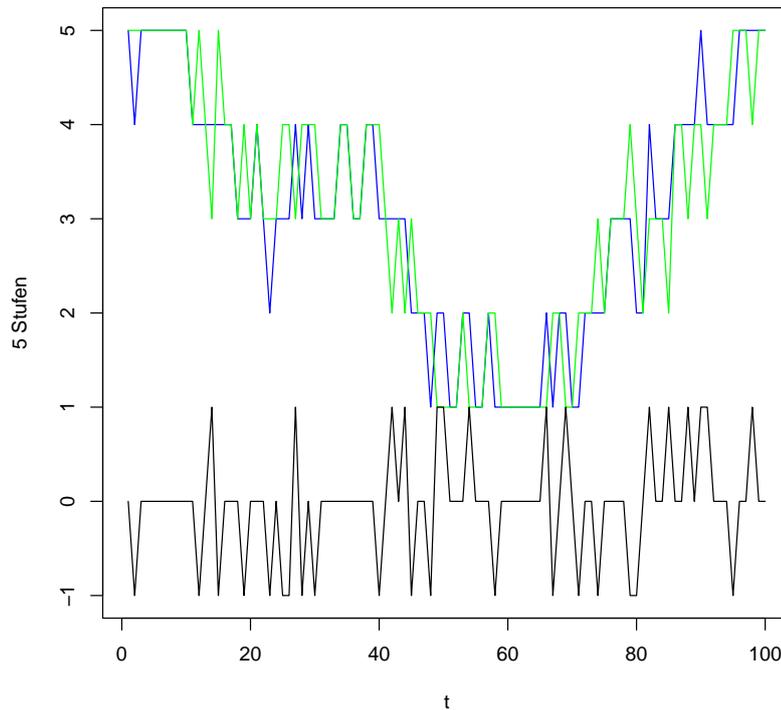


Abbildung 3.9: Kointegriertes System der Länge 100 aus der Gruppe (1,1) auf 5-stufiger Intervallskala

3.3.4 5-stufige Intervallskala

Die meisten psychologischen Ratingskalen unterscheiden fünf Merkmalsausprägungen. Deshalb sind die Ergebnisse dieses Kapitels von besonderem Interesse.

Bekannter Kointegrationsvektor

Phillips-Perron-Test In den Gruppen (1,1) und (2,3) gibt es beim PP-Test für kointegrierte und stationäre Systeme kaum Veränderungen zu den unverzerrten Daten (vgl. Tabellen 3.6, S.163, 3.8, S.165). Ab Länge 60 werden (fast) alle stationären Linearkombinationen richtig erkannt. Aber auch bei den integrierten Systemen werden viele, oft sogar die meisten, Linearkombinationen als stationär eingestuft. In der Gruppe (3,10) ist die Verzerrung der Daten so groß, dass viele Linearkombinationen bei kointegrierten Systemen nicht mehr als stationär erkannt werden. Das sind sehr starke Abweichungen von den unverzerrten Daten (siehe Tabelle 3.10, S.167) und auch noch schlechtere Ergebnisse als auf einer 10- oder 7-stufigen Skala (vgl. Tabellen 3.22 und 3.28).

Augmented Dickey-Fuller-Test Während der ADF-Test (siehe Tabelle 3.35, S.202) in der Gruppe (1,1) bei den kointegrierten Systemen noch kaum Abweichungen zu den unverzerrten

Tabelle 3.34: Ergebnisse des PP-Tests auf 5-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	999	995	929	999	995	897	621	500	264
	60	1000	1000	1000	1000	1000	1000	805	707	536
	100	1000	1000	1000	1000	1000	1000	845	769	630
	300	1000	1000	1000	1000	1000	1000	930	896	791
(2,3)	30	997	982	881	998	989	909	376	258	111
	60	1000	1000	999	1000	1000	1000	567	439	254
	100	1000	1000	1000	1000	1000	1000	653	549	377
	300	1000	1000	1000	1000	1000	1000	798	698	551
(3,10)	30	876	789	588	999	995	918	369	252	106
	60	961	930	837	1000	1000	1000	559	427	244
	100	991	978	931	1000	1000	1000	640	534	359
	300	996	989	978	1000	1000	1000	787	719	569

Daten zeigt (vgl. Tabelle 3.7, S.164), macht sich die Verzerrung bereits in der Gruppe (2,3) deutlich bemerkbar, so dass auch bei 300 Beobachtungspunkten keine perfekte Erkennquote mehr erreicht wird. Die Verzerrung in der Gruppe (3,10) ist hingegen so groß, dass die viele stationären Linearkombinationen bei kointegrierten Systemen nicht mehr erkannt werden. Selbst bei einer Länge von 300 Datenpunkten werden noch auf dem 0.1 Niveau mehr als 20% falsch eingestuft. Die Abweichungen zu den Werten der Originaldaten (vgl. Tabellen 3.7, 3.9 & 3.11) bei den stationären Systemen sind in allen Gruppen so gering, dass der Test ab einer Länge von 100 noch zur Erkennung der Stationarität eingesetzt werden kann. Für integrierte Systeme sind die Abweichungen des ADF-Tests schon größer. Am stärksten ist der Effekt auf dem 0.1 Niveau zu merken. Außerdem nimmt die Größe der Abweichungen mit der Länge der Systeme zu. Auch zwischen den drei Gruppen unterscheidet sich die Häufigkeit der Abweichung. Je größer der Varianzunterschied, desto häufiger kommt es zu Fehlklassifikationen. Dennoch ist der Test auf dem 0.01 Niveau noch immer sehr gut und liefert auch auf dem 0.05 Niveau, dafür dass ein Verlauf nur auf 5 Punkten unterschieden wird, gute Ergebnisse. Der ADF-Test kann also bei einer Länge von 100 in den ersten beiden Gruppen zur Klassifikation eingesetzt werden. In der dritten Gruppe reicht die Klassifikationsleistung wegen der starken Varianzunterschiede jedoch nicht mehr aus.

Johansentest

Kointegrierte Systeme Der Johansentest konnte für alle kointegrierten Systeme der Gruppe (1,1) bis auf vier der Länge 30 auch auf einer 5-stufigen Intervallskala ausgeführt werden (siehe Tabelle 3.36, S.202). Ab einer Länge von 60 Datenpunkten werden auf allen Niveaus

Tabelle 3.35: Ergebnisse des ADF-Tests auf 5-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	317	187	51	340	206	59	95	52	9
	60	879	764	416	890	754	413	102	54	12
	100	993	957	761	988	951	767	116	60	11
	300	1000	1000	1000	1000	1000	1000	144	82	21
(2,3)	30	308	177	46	328	199	50	100	59	7
	60	832	686	328	889	743	396	99	54	11
	100	946	889	632	987	958	777	115	73	17
	300	993	987	958	1000	1000	1000	159	94	29
(3,10)	30	216	113	13	325	194	45	100	61	20
	60	500	330	111	869	750	407	137	67	17
	100	588	437	187	991	960	782	137	80	20
	300	797	659	394	1000	1000	1000	218	137	59

Tabelle 3.36: Klassifikation 1000 kointegrierter Systeme auf 5-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	274	618	104	428	519	49	738	252	6	996
	60	1	867	132	8	917	75	73	909	18	1000
	100	0	850	150	0	911	89	0	975	25	1000
	300	0	781	219	0	867	133	0	966	34	1000
(2,3)	30	299	611	90	468	498	34	768	230	2	1000
	60	9	893	98	24	923	53	147	840	13	1000
	100	1	894	105	4	942	54	21	966	13	1000
	300	0	835	165	0	922	78	1	988	11	1000
(3,10)	30	364	483	59	479	399	28	675	228	3	906
	60	290	623	79	395	561	36	642	342	8	992
	100	139	772	89	223	734	43	451	537	12	1000
	300	29	799	172	48	865	87	125	859	16	1000

Tabelle 3.37: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf 5-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-17,718	-1,514	-1,141	-1,003	-0,891	-0,539	60,701	78
	60	-1,990	-1,232	-1,063	-1,002	-0,950	-0,792	-0,201	43
	100	-1,501	-1,135	-1,035	-0,999	-0,967	-0,870	-0,708	40
	300	-1,087	-1,057	-1,013	-0,998	-0,984	-0,944	-0,899	19
(2,3)	30	-3,313	-0,919	-0,753	-0,697	-0,639	-0,470	6,425	62
	60	-1,107	-0,825	-0,722	-0,688	-0,652	-0,552	-0,376	20
	100	-0,822	-0,789	-0,712	-0,687	-0,660	-0,581	-0,555	13
	300	-0,780	-0,769	-0,709	-0,686	-0,664	-0,599	-0,528	8
(3,10)	30	-2,339	-0,495	-0,370	-0,326	-0,286	-0,161	1,753	55
	60	-0,804	-0,462	-0,360	-0,323	-0,289	-0,185	-0,105	46
	100	-0,623	-0,465	-0,358	-0,322	-0,286	-0,179	-0,060	33
	300	-0,509	-0,479	-0,369	-0,332	-0,294	-0,185	-0,150	11

die meisten kointegrierten Systeme richtig klassifiziert, stets mehr als 78%. Ab 100 Datenpunkten gibt es hier keine Fehlklassifikation als integriert mehr. Die Abweichungen zu Tabelle 3.12 in dieser Gruppe sind nur bei einer Länge von 300 nennenswert. Hier werden auf dem 0.1 Niveau 219 Systeme, auf 0.05 Niveau 133 und auf dem 0.01 Niveau immer noch 34 Systeme als stationär klassifiziert. Auf dem 0.01 Niveau liefert der Test bei den Längen 100 und 300 sehr gute Werte. Ähnlich verhält es sich in der Gruppe (2,3), wobei hier auf dem 0.01 Niveau auch bei den Längen 60 und 100 größere Abweichungen zu spüren sind (vgl. Tabelle 3.13, S.171). Hier werden einige Systeme als integriert statt kointegriert klassifiziert. Und auch bei 300 wird auf dem 0.01 Niveau noch ein System als integriert missklassifiziert. Dennoch kann der Test auf dem 0.01 auch hier bei den Längen 100 und 300 gut eingesetzt werden. In der Gruppe (3,10) kommt es zu so großen Abweichungen in der Klassifikation (vgl. Tabelle 3.14), dass sie eigentlich nur noch bei Länge 300 gelingt. Dies jedoch auch nicht auf dem gewünschten Prozentsatz. Maximal 865 Systemen werden als kointegriert erkannt (auf Niveau 0.05). Das Problem in dieser Gruppe ist, dass der Trend der Reihe mit der kleinen Varianz auf 5 Stufen nicht mehr abgebildet werden kann bzw. dass der gemeinsame Trend in unterschiedlicher Stärke nicht mehr erkannt werden kann.

Dennoch sind die Schätzungen für den Kointegrationsparameter β sehr nahe an den wahren Werten (siehe Tabelle 3.37, S.203). Es ist jedoch eine leicht Verschiebung in den Gruppen (2,3) mit $\beta = 0.666$ und (3,10) mit $\beta = 0.3$ zu sehen. Dies lässt sich damit erklären, dass durch die Einschränkung der Wertemenge auf fünf Stufen auch in den Schätzprozeduren nicht mehr alle Parameterwerte für eine Kointegrationsbeziehung in Frage kommen. Die Werte von Box und Whisker liegen auch etwas weiter vom Median entfernt als etwa auf der 7-stufigen Intervallskala.

Tabelle 3.38: Klassifikation 1000 stationärer Systeme auf 5-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	88	99	813	201	184	615	539	268	193	1000
	60	0	0	1000	0	1	999	1	13	986	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	88	106	806	205	201	594	540	271	189	1000
	60	0	0	1000	0	0	1000	5	16	979	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	74	86	657	188	159	470	462	209	146	817
	60	0	1	926	0	1	926	6	15	906	927
	100	0	0	974	0	0	974	0	0	974	974
	300	0	0	998	0	0	998	0	0	998	998

Stationäre Systeme Bei stationären Systemen werden ab einer Länge von 100 alle Systeme auf allen Niveaus richtig klassifiziert (siehe Tabelle 3.38, S.204). In der Gruppe (3,10) konnte der Johansentest jedoch nicht bei allen Systemen durchgeführt werden. Die Abweichungen sind so gering, dass der Test ab einer Länge von 60 auch auf einer 5-stufigen Skala nahezu perfekte Klassifikationsergebnisse bei den stationären Systemen bringt (vgl. Tabellen 3.16, 3.17 und 3.18).

Tabelle 3.39: Klassifikation 1000 integrierter Systeme auf 5-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	764	207	27	861	131	6	953	45	0	998
	60	670	306	24	794	201	5	930	70	0	1000
	100	540	408	52	672	303	25	860	139	1	1000
	300	254	546	200	374	536	90	585	400	15	1000
(2,3)	30	764	187	8	848	109	2	932	27	0	959
	60	760	215	13	849	133	6	946	42	0	988
	100	662	298	31	783	205	3	906	85	0	991
	300	424	468	100	528	430	34	726	259	7	992
(3,10)	30	528	118	4	573	75	2	627	23	0	650
	60	579	174	10	643	116	4	721	42	0	763
	100	494	261	30	590	186	9	703	80	2	785
	300	361	392	113	453	361	52	592	265	9	866

Integrierte Systeme Auf der 5-stufigen Skala konnte der Johantest auch nicht mehr bei allen integrierten Systemen aus der Gruppe (2,3) durchgeführt werden (siehe Tabelle 3.39, S.204). Dennoch bei mehr als 95%. In der Gruppe (1,1) wurden zwei Systeme der Länge 30 nicht berücksichtigt. In der Gruppe (3,10) war dies nur bei 650-866 Systemen möglich. Auf dem 0.01 Niveau werden die meisten Systeme richtig klassifiziert. Bei 300 Datenpunkten ist die Klassifikationsrate allerdings auch hier nicht ausreichend gut. Wenn man die nicht berücksichtigten Systeme miteinbezieht, ist das Ergebnis in dieser Gruppe so schlecht, dass der Test nicht mehr sinnvoll eingesetzt werden kann.

Ergebnis: 5-stufige Intervallskala

Die Messungen auf einer fünfstufigen Intervallskala ist nur dann zuverlässig, wenn die beiden Zeitreihen eines Systems keinen zu großen Varianzunterschied aufweisen. In der Praxis bedeutet das, dass in einem solchen Fall für jede Zeitreihe eine eigene Skala verwendet werden sollte. Der ADF-Test und der Johantest können bei kleineren Varianzunterschieden bei einer Länge von 100 Beobachtungen noch sehr gut eingesetzt werden, während der PP-Test scheitert. Bei zu großen Längen kommt es ebenfalls zu einer Verzerrung der Klassifikation, da integrierte Trends nicht mehr wiedergegeben werden können.

Bekannter Kointegrationsvektor Die Verzerrung durch die simulierte Messung bewirkt nun in der Gruppe (3,10), dass der Johantest nicht nur die integrierten, sondern auch nicht mehr alle kointegrierten Systeme erkennen kann. Der Augmented Dickey-Fuller-Test zeigt in den ersten beiden Gruppen auf dem 0.01 Niveau ausreichend gute Ergebnisse und stuft auch in der Gruppe (3,10) die meisten Linearkombinationen richtig ein. Es sind in dieser Gruppe jedoch zu wenig, um den Test zuverlässig einzusetzen.

Johantest In den Gruppen (1,1) und (2,3) kann der Johantest auch auf der 5-stufigen Skala bei einer Länge von 60-100 eingesetzt werden, wenn man sich an den Ergebnissen auf dem 0.01 Niveau hält. Aber auch in diesen Gruppen war die Verzerrung vereinzelt so groß, dass der Test nicht mehr ausgeführt werden konnte. In der Gruppe (3,10) werden die meisten Systeme noch richtig klassifiziert, aber die Verzerrung ist so groß, dass die Zeitreihe mit der kleineren Varianz häufig fast konstant auf die 5-stufige Skala abgebildet wird, so dass der Test entweder nicht angewandt werden kann oder der gemeinsame Trend nicht mehr erkannt werden kann.

3.3.5 3-stufige Intervallskala

Abschließend betrachten wir noch den Extremfall einer 3-stufigen Intervallskala. Hier gibt es nur noch die Unterscheidung *positiv - neutral - negativ*. Wir prüfen, ob diese Information

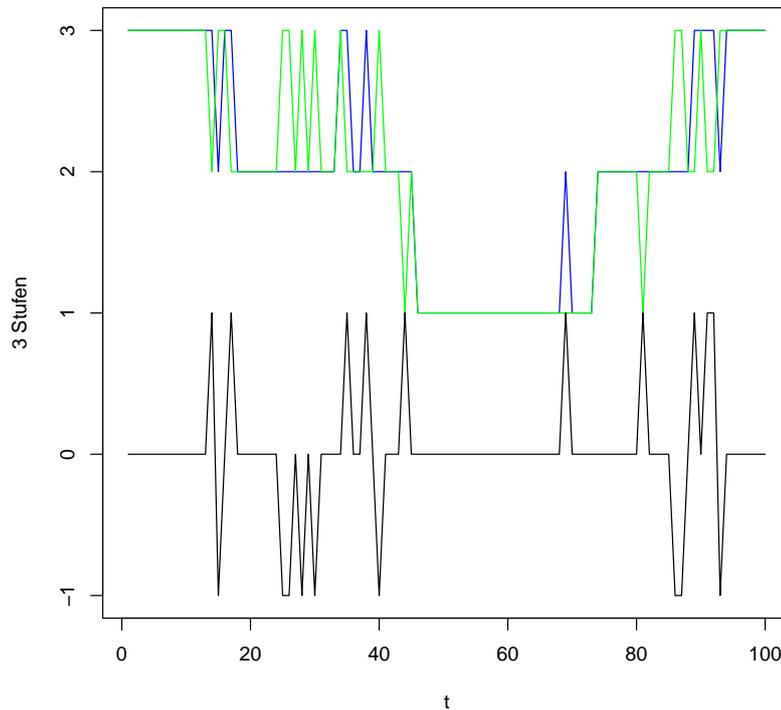


Abbildung 3.10: Kointegriertes System und Kointegrationsbeziehung der Länge 100 aus der Gruppe (1,1) auf 3-stufiger Intervallskala

ausreicht, um einen integrierten und eventuell gemeinsamen Anteil in den Systemen aufzudecken. Die Abbildung 3.10 zeigt ein kointegriertes System der Länge 100 aus der Gruppe (1,1) auf der 3-stufigen Intervallskala. Die Einschränkung der Werte auf eine 3-stufige Intervallskala bedeutet einen enormen Informationsverlust. Im Folgenden soll überprüft werden, ob bei einem so kleinen Wertebereich der stochastische Trend von integrierten und kointegrierten Systemen überhaupt noch von dem stationären Verlauf zu unterscheiden ist.

Bekannter Kointegrationsvektor

Phillips-Perron-Test Der Phillips-Perron-Test (siehe Tabelle 3.40, S.207) stuft alle Linearkombination bei den stationären Systemen der aller Gruppen ab einer Länge von 60 als stationär ein, und auch die meisten bei kointegrierten Systemen und viele der integrierten Systeme. Es gibt also noch eine Unterscheidung zwischen stationären und nicht-stationären Linearkombinationen. In den beiden Gruppen mit Varianzunterschied werden einige Kointegrationsbeziehungen nicht mehr als stationär erkannt. Dieser Effekt ist in der Gruppe (3,10) deutlich größer als in der Gruppe (2,3). Dennoch liefert der PP-Test für stationäre Linearkombinationen sehr gute Ergebnisse, versagt aber bei solchen mit Einheitswurzel.

Tabelle 3.40: Ergebnisse des PP-Tests auf 3-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	997	990	909	999	984	904	699	576	334
	60	1000	1000	1000	1000	1000	1000	699	576	334
	100	1000	1000	1000	1000	1000	1000	893	843	737
	300	1000	1000	1000	1000	1000	1000	965	940	883
(2,3)	30	967	939	777	998	994	920	798	698	551
	60	999	995	970	1000	1000	1000	696	574	363
	100	1000	996	987	1000	1000	1000	756	663	484
	300	1000	1000	994	1000	1000	1000	869	810	690
(3,10)	30	640	530	333	995	988	918	410	294	136
	60	783	709	560	1000	1000	1000	596	488	294
	100	868	819	692	1000	1000	1000	642	549	389
	300	937	911	834	1000	1000	1000	793	726	588

Augmented Dickey-Fuller-Test Für den ADF-Test (siehe Tabelle 3.41, S.208) gibt es in der Gruppe (1,1) nur leichte Abweichungen von den unverzerrten Daten (vgl. Tabelle 3.7, S.164). Ab einer Länge von 60 werden maximal 20 Linearkombinationen weniger als stationär erkannt. Auch bei den stationären Systemen in dieser Gruppen werden maximal 25 nicht mehr erkannt. Und bei den integrierten Systemen ist die Abweichung auf dem 0.01 Niveau auch gering, so dass der ADF-Test in dieser Gruppe noch so gut ist, dass er ab einer Länge von 100 verwendet werden kann. In der Gruppe (2,3) sind die Abweichungen bei kointegrierten und integrierten Systemen schon größer (vgl. Tabelle 3.11, S.168). Am meisten fällt auf, dass die stationäre Linearkombination von kointegrierten Systemen nicht mehr erkannt wird. Auf dem 0.01 Niveau werden noch immer nahezu 95% der integrierten Systeme mit Einheitswurzel erkannt. Auch in der Gruppe (3,10) ist der ADF-Test für integrierte Systeme auf dem 0.01 Niveau noch sehr gut. Allerdings gibt es hier massive Probleme beim Erkennen der Kointegrationsbeziehung bei kointegrierten Systemen, während die stationären noch gut eingeschätzt werden. Deshalb ist der ADF-Test in dieser Gruppe zur Unterscheidung von kointegrierten und integrierten Systemen nicht mehr geeignet.

3-stufige Intervallskala - Johantest

Kointegrierte Systeme Bei der Ausführung des Johantests für kointegrierte Systeme auf einer dreistufigen Skala verändert die Verzerrung in allen Gruppen deutlich die Klassifikationsergebnisse (siehe Tabelle 3.42, S.209). Nur in der Gruppe (2,3) konnte der Test für alle Systeme durchgeführt werden. In der ersten Gruppe gab es zu viele Systeme, deren beiden Reihen als (fast) identisch abgebildet wurden. In der dritten Gruppe wurde bei vielen

Tabelle 3.41: Ergebnisse des ADF-Tests auf 3-stufiger Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	329	203	61	331	195	53	99	53	16
	60	870	742	413	877	764	404	99	53	16
	100	969	942	754	986	958	779	151	74	15
	300	1000	1000	998	1000	1000	1000	228	134	35
(2,3)	30	259	159	41	314	187	47	159	94	29
	60	685	508	250	885	753	408	142	83	23
	100	775	646	340	983	955	771	166	85	31
	300	920	848	669	1000	1000	1000	260	159	52
(3,10)	30	142	86	17	298	168	42	94	53	19
	60	239	149	54	874	750	398	161	96	33
	100	309	219	82	988	954	776	159	93	30
	300	517	403	191	1000	1000	1000	261	171	63

Systemen die Zeitreihe mit der kleineren Varianz zu einer (fast) konstanten Reihe transformiert. In den meisten Fällen werden zwar die meisten Systeme als kointegriert klassifiziert. Der Zielwert von 950 wird jedoch nirgends überschritten. Auf einer dreistufigen Skala kann der Johantest bei kointegrierten Systemen nicht eingesetzt werden.

Die Schätzungen von β (siehe Tabelle 3.43, S.209) hingegen sind in der ersten Gruppe erstaunlich gut und in den anderen beiden Gruppen zwar durch die Verzerrung verschoben, dennoch liegen Box und Whisker symmetrisch um den Median. Die Varianz der Schätzungen ist für den Grad der Verzerrung sehr gering.

Stationäre Systeme In den Gruppen (1,1) und (2,3) kann der Johantest bei fast allen stationären Systemen noch ausgeführt werden, in der Gruppe (3,10) gelingt dies bei weniger als 30% der Systeme. Jedoch werden alle Systeme ab einer Länge von 60 Datenpunkten, für die der Test ausgeführt wurde, bis auf wenige Ausnahmen auf dem 0.01 Niveau bei einer Länge von 60 als stationär richtig klassifiziert (siehe Tabelle 3.44, S.210).

Integrierte Systeme Auch bei den integrierten Systemen ist die Verzerrung deutlich zu merken (siehe Tabelle 3.45, S.210). Mit steigendem Varianzunterschied in den drei Gruppen konnten immer weniger Systeme im Johantest verwertet werden. In der Gruppe (3,10) waren es nur noch 319-527. Obwohl die meisten Systeme, für die der Test ausgeführt wurde, richtig klassifiziert wurden, sind die Ergebnisse zu schlecht, um den Johantest auf einer dreistufigen Skala zur Klassifikation einzusetzen.

Tabelle 3.42: Klassifikation 1000 kointegrierter Systeme auf 3-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	249	531	130	395	451	64	688	215	7	910
	60	5	796	168	14	867	88	80	868	21	969
	100	0	751	227	0	853	125	1	932	45	978
	300	0	557	437	0	705	289	0	871	123	994
(2,3)	30	386	512	102	523	429	48	792	197	11	1000
	60	111	777	112	174	773	53	366	621	13	1000
	100	59	799	142	79	851	70	201	782	17	1000
	300	8	703	289	13	815	172	42	900	58	1000
(3,10)	30	208	150	27	260	114	11	327	58	0	385
	60	208	292	33	279	235	19	396	133	4	533
	100	153	388	69	213	369	28	329	277	4	610
	300	51	462	199	70	511	131	134	537	41	712

Tabelle 3.43: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf 3-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-4,152	-1,604	-1,161	-1,004	-0,862	-0,415	20,892	78
	60	-1,996	-1,278	-1,077	-1,010	-0,943	-0,742	-0,532	58
	100	-1,550	-1,164	-1,042	-1,000	-0,960	-0,841	-0,665	45
	300	-1,117	-1,057	-1,015	-1,000	-0,986	-0,944	-0,872	50
(2,3)	30	-3,092	-1,063	-0,795	-0,698	-0,613	-0,341	11,426	58
	60	-1,334	-0,998	-0,781	-0,700	-0,626	-0,422	-0,257	12
	100	-1,029	-0,973	-0,773	-0,700	-0,637	-0,436	-0,373	7
	300	-1,048	-0,937	-0,768	-0,698	-0,634	-0,442	-0,429	3
(3,10)	30	-1,456	-0,821	-0,591	-0,525	-0,434	-0,204	5,424	24
	60	-1,073	-0,875	-0,589	-0,516	-0,366	-0,086	0,000	3
	100	-1,079	-0,907	-0,595	-0,519	-0,353	-0,059	-0,059	1
	300	-1,373	-0,950	-0,584	-0,487	-0,271	-0,027	-0,027	1

Tabelle 3.44: Klassifikation 1000 stationärer Systeme auf 3-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	83	104	809	220	186	590	541	261	194	996
	60	0	0	1000	0	0	1000	5	13	982	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	61	125	768	158	209	587	522	272	160	954
	60	0	0	997	0	0	997	3	9	985	997
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	21	33	254	60	66	182	159	103	46	308
	60	0	0	280	0	0	280	1	2	277	280
	100	0	0	267	0	0	267	0	0	267	267
	300	0	0	285	0	0	285	0	0	285	285

Tabelle 3.45: Klassifikation 1000 integrierter Systeme auf 3-stufiger Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	592	200	35	684	133	10	788	39	0	827
	60	592	200	35	684	133	10	788	39	0	827
	100	375	447	133	496	404	55	709	242	4	955
	300	130	459	381	207	510	253	391	489	90	970
(2,3)	30	424	468	100	528	430	34	726	259	7	992
	60	578	232	24	667	163	4	760	73	1	834
	100	477	308	64	567	253	29	723	120	6	849
	300	241	416	263	332	428	160	520	361	39	920
(3,10)	30	254	63	2	278	40	1	305	14	0	319
	60	309	116	15	359	78	3	411	29	0	440
	100	300	152	27	350	119	10	428	51	0	479
	300	161	216	150	210	222	95	317	182	28	527

Ergebnis: 3-stufige Intervallskala

Die Unterscheidung von nur drei Stufen ist zu gering, um einen stochastischen Trend zu erkennen.

Bekannter Kointegrationsvektor Der Phillips-Perron-Test versagt auf einer 3-stufigen Intervallskala, während der ADF-Test in der Gruppe (1,1) noch erstaunlich gute Ergebnisse liefert. Hier gibt es bei Länge 100 kaum Abweichungen zu den Originaldaten, obwohl der Wertebereich nun auf drei Stufen eingeschränkt wurde. In den anderen beiden Gruppen, in denen die Zeitreihen der Systeme Varianzunterschiede aufweisen, kann er jedoch nicht mehr eingesetzt werden.

Johansentest Der Johansentest zeigt nur für stationäre Systeme auf dieser Skala noch gute Werte. Für kointegrierte und integrierte Systeme ist die Verzerrung zu groß, um sie zuverlässig zu unterscheiden. In vielen Fällen konnte der Test nicht mehr ausgeführt werden, da entweder beide Reihen auf der Skala nun einen identischen Verlauf zeigten oder eine der beiden Reihen konstant abgebildet wurde.

3.3.6 Ergebnis: Messung auf Intervallskalenniveau

Der Phillips-Perron-Test kann zur Unterscheidung von stationären und nicht-stationären Linearkombinationen auf einer Intervallskala mit 10 Stufen bereits nicht mehr eingesetzt werden. Der Augmented Dickey-Fuller-Test zeigt hingegen bis zu einer 5-stufigen Intervallskala noch gute Ergebnisse. Auch der Johansentest klassifiziert auf fünf Stufen noch sehr genau.

Allerdings ist ein deutlicher Unterschied in den Gruppen zu erkennen. Mit steigendem Varianzunterschied wird eine Klassifikation immer schwieriger. Dies weist daraufhin, dass die Skala für jedes Merkmal so gewählt werden sollte, dass der jeweilige Wertebereich durch die Skalenwerte gut abgedeckt wird.

Insbesondere bei den längsten Systemen mit 300 Datenpunkten gab es vermehrt Probleme, den stochastischen Trend zu erkennen, da durch den eingeschränkten Wertebereich die Eigenschaft der wachsenden Varianz nicht mehr abgebildet werden kann.

Auf einer Skala mit nur drei Punkten versagen beide Kointegrationsverfahren. Für Daten, die auf einer Intervallskala mit 7-10 Skalenstufen gemessen wurden, können beide Verfahren eingesetzt werden.

Da der Phillips-Perron-Test bereits auf einer 10-stufigen Intervallskala nicht mehr zuverlässig eingesetzt werden kann und bereits bei den unverzerrten Daten Schwierigkeiten bei der Unterscheidung von integrierten und kointegrierten Systemen hatte, wird er bei den folgenden

Simulationen nicht mehr ausgewertet. Es sei noch erwähnt, dass die Ergebnisse des Phillips-Perron-Tests auch in den weiteren Simulationsstudie ebenso unzureichend waren, wie hier für die intervallskalierten Messungen berichtet.

3.4 Ceiling- und Flooreffekte

Wenn die gewählte Skala den Wertebereich des zu messenden Merkmals nicht vollständig abdeckt, kann es zu sogenannten Decken- oder Bodeneffekten kommen (Bortz, 2005). Nämlich dann, wenn die extremen Werte der Skala bereits für Merkmalsausprägungen verwendet wurden, die deutlich schwächer sind als die stärksten Ausprägungen. Werden dabei Werte von unten abgeschnitten, spricht man von einem Boden- bzw. *Flooreffekt*. Dadurch ist die Differenzierung im unteren Bereich der Skala eingeschränkt, während bei einem Decken- bzw. *Ceilingeffekt* dies für den oberen Wertebereich gilt.

Ein Intelligenztest kann den IQ nur bis zu einem Wert diskriminieren, mit dem Testpersonen alle Fragen des Tests richtig beantworten können, z.B. 140. Eine Unterscheidung der Intelligenzleistung von Personen mit höherem IQ ist nicht mehr möglich. Außer diesem Deckeneffekt liegt zusätzlich auch ein Bodeneffekt vor, wenn man ein Mindestmaß an Intelligenz besitzen muss, um wenigstens eine richtige Aufgabe zu erzielen. Die Intelligenz von Menschen, die den Test mit Null Punkten absolvieren, kann nicht differenziert eingestuft werden (Krauth, 1995).

Für die folgenden Simulationsstudien werden die Zeitreihen der Systeme auf verschiedenen Intervallskalen „gemessen“, wobei einige der unteren bzw. oberen Stufen zusammengefasst werden. Dadurch ist die Diskriminierung an den Randbereichen stark eingeschränkt. Wir testen, ob der ADF-Test und der Johansenstest bei Daten, die auf einer solchen Skala gemessen wurden, die integrierten Anteile der Reihen erkennen und kointegrierte von Systemen ohne gemeinsamen Trend unterscheiden können.

3.4.1 Von einer Seite beschränkte Skalen

Für die ersten beiden Simulationsstudien zu Ceiling- und Flooreffekten wurden zwei Skalen konstruiert, von denen eine die oberen 30% des möglichen Wertebereichs auf eine Stufe zusammenfasst und die andere die unteren 30%. Die übrigen Stufen sind identisch mit den Stufen der 10-stufigen Intervallskala aus Kapitel 3.3. Die von oben beschränkte Skala ordnet allen Werten, die auf der 10-stufigen Skala die Werte 7, 8, 9, 10 hatten, nun den Wert 7 zu. So entsteht eine Intervallskala, die den oberen Wertebereich nicht mehr diskriminieren kann (siehe Abbildung 3.11, S.213). Ähnlich wurde die von unten beschränkte Skala konstruiert. Hier werden die Stufen 1, 2, 3, 4 zu einer Stufe zusammengefasst. Da so die Stufen 1, 2, 3 der ursprünglich 10-stufigen Skala wegfallen, entsteht eine 7-stufige Skala, die den unte-

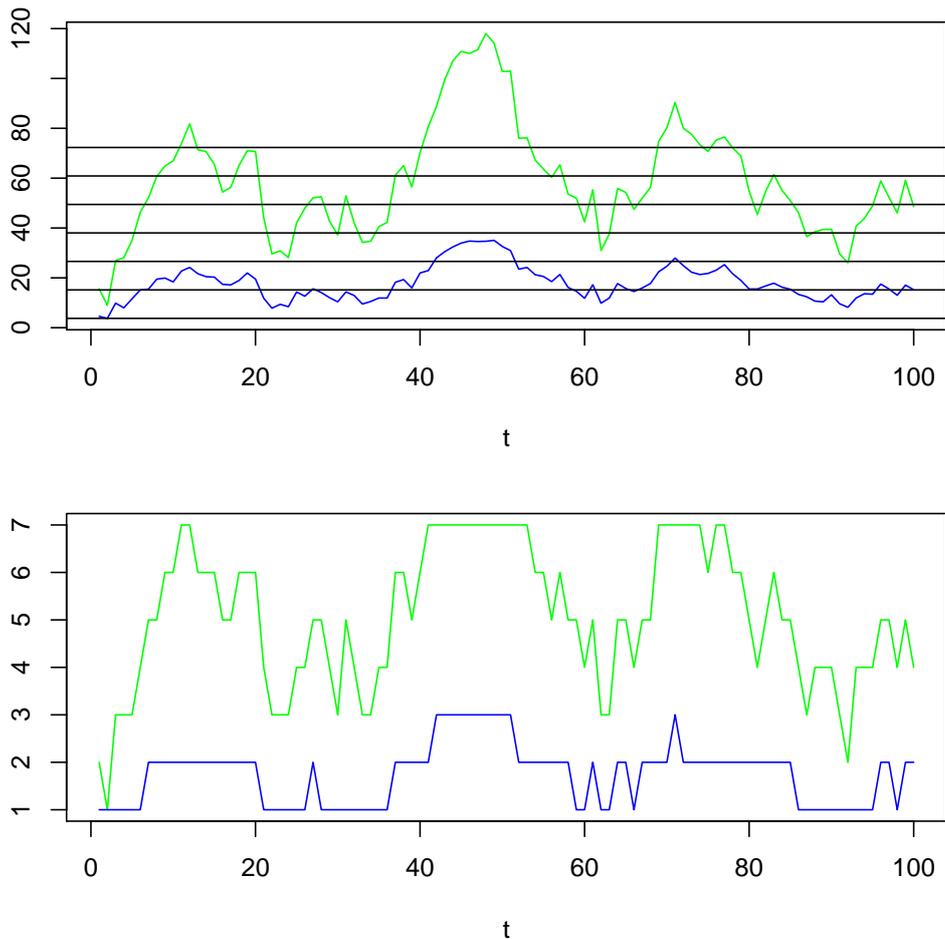


Abbildung 3.11: Konstruktion der Stufen der von oben beschränkten Intervallskala, hier am Beispiel eines kointegriertes Systems der Länge 100 aus der Gruppe (3,10).

ren Wertebereich nicht ausreichend unterscheiden kann (siehe Abbildung 3.12, S.214). Die verwendeten Funktionen werden in Anhang D.6 (S.394) beschrieben.

Da sich die Ergebnisse der beiden Simulationen kaum unterscheiden, werden sie hier gemeinsam beschrieben, wobei der Schwerpunkt der Darstellung bei der von oben beschränkten Skala liegt. Die entsprechenden Tabellen für die von unten beschränkte Skala sind im Anhang D.3.3 zu finden.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Der ADF-Test (siehe Tabelle 3.46, S.215) zeigt in der ersten Gruppe nur sehr geringe Abweichungen von den Originaldaten (vgl. Tabelle 3.7, S.164). Es werden maximal 30 Linearkombination anders eingeschätzt wird als im unverzerrten Fall. Auch für integrierte Systeme ist der Test weiterhin gut. Auf dem 0.01 und dem 0.05 Niveau werden bei jeder Länge mehr als 950 Linearkombinationen als nicht-stationär erkannt. Auf dem 0.1 Niveau trifft dies immer noch für mehr als 900 Systeme zu.

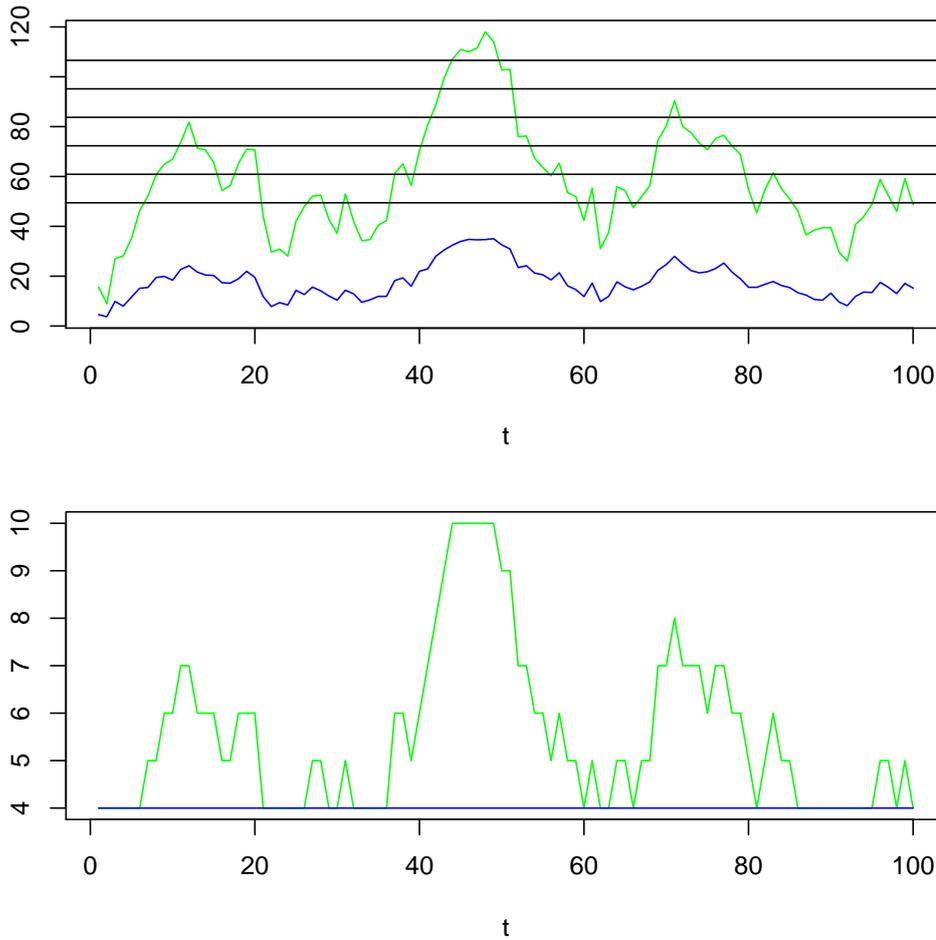


Abbildung 3.12: Konstruktion der Stufen der von unten beschränkten Intervallskala, hier am Beispiel eines kointegriertes System der Länge 100 aus der Gruppe (3,10). Der Verlauf nach dieser simulierten Messung unterscheidet sich stark im Vergleich zur simulierten Messung auf der von oben beschränkten Skala (vgl. Abbildung 3.11).

Während der ADF-Test in der Gruppe (2,3) bei den integrierten und auch bei den stationären Systemen kaum Abweichungen von seinem Verhalten bei unverzerrten Daten zeigt (vgl. Tabelle 3.9, S.166), werden auf allen Niveaus bei allen Längen nicht genügend Kointegrationsbeziehungen als stationär erkannt. Bei den unverzerrten Daten werden ab einer Länge von 90 auf dem 0.05 Niveau und ab einer Länge von 150 auf allen Niveaus mehr als 950 Linearkombinationen als stationär eingestuft. Bei einer Länge von 300 war die Klassifikation auf allen Niveaus perfekt. Dies ist ein Effekt, der sich in der Gruppe (1,1) noch nicht angedeutet hat. Man erkennt zwar weiterhin einen Effekt der Länge auf allen drei Niveaus, aber auf keinem Niveau wird annähernd ein Wert wie bei unverzerrten Daten erzielt.

Dieser Effekt setzt sich in der Gruppe (3,10) weiter noch fort. Hier werden maximal 767 bzw. 817 (bei Länge 300 auf dem 0.1 Niveau) Kointegrationsbeziehungen erkannt. Die Abweichungen zu den unverzerrten Daten betragen bis zu 500 Systeme (vgl. Tabelle 3.11, S.168). Die Ergebnisse für stationäre und integrierte Systeme weichen hingegen nur minimal ab.

Tabelle 3.46: Ergebnisse des ADF-Tests auf von oben beschränkter Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	339	203	59	331	210	45	81	44	6
	60	890	755	410	887	760	401	87	48	9
	100	985	947	732	985	955	774	94	42	5
	300	1000	1000	1000	1000	1000	1000	81	40	6
(2,3)	30	208	128	26	334	199	46	84	42	7
	60	570	404	167	875	758	403	86	46	10
	100	663	531	309	985	957	775	92	47	8
	300	882	799	620	1000	1000	1000	95	45	8
(3,10)	30	184	112	28	324	189	47	88	46	8
	60	454	322	119	872	749	400	85	44	14
	100	614	500	250	985	948	770	96	46	9
	300	767	718	579	1000	1000	1000	130	80	26

Klassifikation mit Johantest

Kointegrierte Systeme In den Gruppen (1,1) und (2,3) konnte der Johantest (siehe Tabelle 3.47, S.216) bei allen (bis auf zwei der Länge 30 in der Gruppe (2,3)) Systemen durchgeführt werden, in der Gruppe (3,10) war die Verzerrung durch die beschränkte Skala so groß, dass die Voraussetzungen für die Testdurchführung nicht bei allen Systemen gegeben waren. Ein Beispiel ist das kointegrierte System der Länge 100 auf der nach unten beschränkten Skala in Abbildung 3.12 auf Seite 214. Da die untere Zeitreihe durch die simulierte Messung konstant wird, kann der Johantest sie nicht verarbeiten. Gleiches gilt auf

Tabelle 3.47: Klassifikation 1000 kointegrierter Systeme auf nach oben beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	257	622	121	394	543	63	668	318	14	1000
	60	4	836	160	14	899	87	84	884	32	1000
	100	0	844	156	0	905	95	1	960	39	1000
	300	0	828	172	0	894	106	0	964	36	1000
(2,3)	30	431	477	90	581	382	35	788	206	4	998
	60	167	732	101	257	696	47	463	525	12	1000
	100	93	804	103	136	812	52	249	738	13	1000
	300	2	880	118	12	927	61	32	958	10	1000
(3,10)	30	313	283	76	405	237	30	550	117	5	672
	60	156	489	75	247	434	39	420	290	10	720
	100	70	578	93	116	574	51	249	477	15	741
	300	11	641	101	17	682	54	38	700	15	753

der nach oben beschränkten Skala z.B. für das kointegrierte System aus Abbildung 3.13 auf Seite 217. Bei den unbeschränkten Intervallskalen aus dem letzten Kapitel trat dieses Phänomen erst bei 3 Punkten so deutlich auf (vgl. Tabelle 3.41, S.208). Durch das Abschneiden einiger Werte in den Randbereichen kommt es noch häufiger dazu, dass die Zeitreihe mit der geringeren Varianz auf eine (fast) konstante Reihe abgebildet wird.

Auf allen Niveaus ist in allen Gruppen ein Längeneffekt zu beobachten. In der Gruppe (1,1) ist das Klassifikationsverhalten beinahe wie bei den unverzerrten Daten (vgl. Tabelle 3.12, S.170). Auf dem 0.01 Niveau ist die Klassifikation am besten und übersteigt ab einer Länge von 100 die 950-Marke, was bei den anderen beiden Niveaus ausbleibt, wobei auf dem 0.05 Niveau mehr als 900 Systeme richtig als kointegriert erkannt werden. Auf dem 0.1 Niveau gibt es bei jeder Länge ca. 12-17% Fehler. In dieser Gruppen werden ab einer Länge von 100 keine Systeme mehr als integriert fehlklassifiziert.

Dies ist in den anderen beiden Gruppen nicht mehr der Fall. Im Gegensatz zu den unverzerrten Daten (vgl. Tabelle 3.13, S.171 und 3.14, S.171) werden auch einige Systeme als integriert interpretiert. Dabei steigt die Anzahl der Fehlklassifikation mit abnehmendem Niveau, sinkt jedoch mit zunehmender Länge auf allen Niveaus in beiden Gruppen auf beiden Skalen. In der Gruppe (2,3) wird auf dem 0.01 Niveau bei einer Länge von 300 noch die 950-Marke erreicht. Hier sind die Abweichungen in der Klassifikation von den unverzerrten Daten noch unbedeutend, während sie insbesondere bei der Länge 60 dazuführen, dass auf dem 0.01 Niveau nun nur noch 525 Systeme als kointegriert klassifiziert werden im Gegensatz zu 917 bei den unverzerrten Daten.

In der Gruppe (3,10) kann der Test nicht nur nicht für alle Systeme durchgeführt werden,

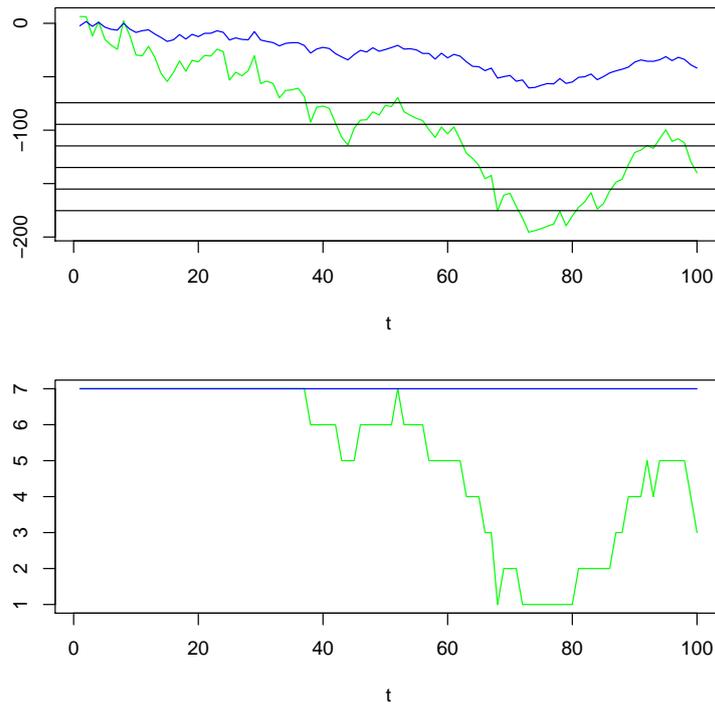


Abbildung 3.13: Kointegriertes System der Länge 100 auf der nach oben beschränkten Skala, für das der Johantest nicht ausgeführt werden kann, da eine Zeitreihe durch die simulierte Messung konstant abgebildet wird.

die Verzerrung ist auch so groß, dass die 900-Marke nur auf der nach unten beschränkten Skala bei Zeitreihen der Länge 300 auf dem 0.01 Niveau knapp erreicht wird (902). Zwar werden weiterhin die meisten Systeme als kointegriert eingestuft, die Anzahl der als integriert klassifizierten Systeme ist jedoch so groß, dass man den Test nicht mehr zuverlässig einsetzen kann.

Für die Schätzung von β (siehe Tabelle 3.48, S.218) sind die Abweichung von der Schätzung bei unverzerrten Daten sehr gering. In allen Systemen bei Länge 300 wird der Kointegrationskoeffizient auf einen Wert im Intervall $[-1,011; -0,988]$ geschätzt, nur 24 bzw. 30 Werte liegen weiter als 0,046 von -1 entfernt. Und selbst bei 30 Datenpunkten liegen die Werte von über 900 Schätzungen zwischen $-1,51$ und $-0,52$.

In der Gruppe (2,3) merkt man eine systematische Verschiebung des Medians bei allen Längen. So verschieben sich auch die anderen Box-Plot-Statistiken mit. Mit zunehmender Länge der Reihen werden sowohl Box als auch Whisker immer enger. Die Verzerrung bewirkt, dass der Faktor nicht mehr richtig erkannt werden kann. Auch in der Gruppe (3,10) hat sich der Median der Schätzungen verschoben und liegt nun bei ca. $-0,37$. Auch hier werden die Schätzungen immer genauer je mehr Datenpunkte berücksichtigt werden.

Stationäre Systeme Die Verzerrung durch die beschränkten Skalen wirkt sich auf die Klassifikation der stationären Systeme kaum aus (siehe Tabelle 3.49, S.219). Allerdings kann der

Tabelle 3.48: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf von oben beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-12,997	-1,520	-1,137	-0,997	-0,881	-0,536	16,692	95
	60	-7,985	-1,212	-1,061	-1,004	-0,955	-0,798	-0,392	59
	100	-1,414	-1,126	-1,033	-1,001	-0,970	-0,876	-0,717	43
	300	-1,087	-1,044	-1,011	-1,000	-0,988	-0,954	-0,939	24
(2,3)	30	-13,218	-1,338	-0,873	-0,715	-0,561	-0,102	6,918	71
	60	-6,286	-1,259	-0,845	-0,714	-0,569	-0,280	58,821	28
	100	-2,828	-1,204	-0,830	-0,721	-0,574	-0,339	1,011	23
	300	-2,076	-1,129	-0,820	-0,717	-0,563	-0,340	-0,340	4
(3,10)	30	-7,718	-0,619	-0,447	-0,390	-0,331	-0,164	49,701	75
	60	-4,427	-0,552	-0,422	-0,380	-0,328	-0,188	1,316	80
	100	-4,847	-0,523	-0,402	-0,373	-0,319	-0,194	2,423	107
	300	-0,716	-0,488	-0,395	-0,368	-0,324	-0,219	-0,018	127

Johansentest in der Gruppe (3,10) nur bei 300 Datenpunkten für alle 1000 Systeme ausgeführt werden. Bei Länge 100 gibt es ein System mit einer (fast) konstanten Zeitreihe, bei 60 Datenpunkten 2-3 und bei einer Länge von 30 Beobachtungen 18.

In allen Gruppen werden auf allen Niveaus wie bei den unverzerrten Daten ab einer Länge von 100 alle Systeme, für die der Test ausgeführt wurde, als stationär erkannt. Auf dem 0.01 Niveau und vereinzelt auch auf dem 0.05 Niveau gibt es bei Länge 60 noch einige Systeme, die als integriert klassifiziert werden. Dies war jedoch auch bei den Rohdaten der Fall (vgl. Tabelle 3.16, 3.17). Die Schätzungen für β liegen in allen Gruppen um Null herum, wie in den Tabellen D.11 und D.15 im Anhang zu sehen ist.

Integrierte Systeme Bei den integrierten Systemen konnte der Johansentest in keiner Gruppe auf keiner der beiden Skalen für alle 1000 Systeme durchgeführt werden. Mit zunehmender Länge sind dies jedoch immer mehr und in den Gruppen (1,1) und (2,3) stets mehr als 782, meist annähernd 900, nur in der Gruppe (3,10) wurde der Test je nach Länge nur bei 624-730 Systemen ausgeführt.

Im Allgemeinen werden mit zunehmender Länge immer weniger Systeme als integriert erkannt. Die meisten falsch klassifizierten Systeme werden als kointegriert angesehen, einige auch als stationär. Dies ist auf dem 0.01 Niveau jedoch bei maximal 8 Systemen der Fall.

Dabei ist die Klassifikationsquote auf dem 0.01 Niveau stets am besten. Hier werden bei den Längen bis 100 in allen Gruppen mehr als 90% der Systeme, für die der Test durchgeführt wurde, als integriert klassifiziert. Auf den anderen Niveaus sind dies über 80% bzw. über 70%. Bei Länge 300 wurde zwar in allen Gruppen für die meisten Systeme der Johansentest ausgeführt, aber es kommt prozentual auch zu den größten Missklassifikationen als kointe-

Tabelle 3.49: Klassifikation 1000 stationärer Systeme auf nach oben beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	81	94	825	202	183	615	542	254	204	1000
	60	0	0	1000	0	1	999	2	12	986	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	89	98	813	205	209	586	558	260	182	1000
	60	0	0	1000	0	1	999	5	9	986	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	84	94	804	212	200	570	540	264	178	982
	60	0	0	998	0	1	997	8	6	984	998
	100	0	0	999	0	0	999	0	0	999	999
	300	0	0	1000	0	0	1000	0	0	1000	1000

grierte Systeme. Die Schätzungen für β liegen auch hier in allen Gruppen um Null herum (siehe Tabelle D.11, S.368 und Tabelle D.15, S.372).

Ergebnis: von einer Seite beschränkte Skalen

Augmented Dickey-Fuller-Test Der ADF-Test zeigt nur in der Gruppe (1,1) kaum Veränderung durch die beschränkte Skala. In den anderen Gruppen werden vor allem die stationären Linearkombinationen der kointegrierten Systeme nicht mehr erkannt und auch bei den integrierten Systemen größerer Länge gibt es in der Gruppe (3,10) Probleme.

Johansentest Nur die stationären Systeme werden vom Johansentest in allen Gruppen noch zuverlässig ab einer Länge von 60 Datenpunkten erkannt. Für die Systeme mit integriertem Anteil bewirkt die Beschränkung des oberen Wertebereichs vor allem in den Gruppen mit unterschiedlicher Gewichtung des Trends starke Abweichungen. Für kointegrierte Systeme wird nur noch auf dem 0.01 Niveau in den ersten beiden Gruppen die 950-Marke erreicht. Zwar werden noch die meisten integrierten Systeme als solche erkannt. Der Test konnte jedoch bei einem großen Prozentsatz (bis 35%) nicht ausgeführt werden.

3.4.2 Bis zur Hälfte beschränkte Skala

Für eine weitere Simulationsstudie wurde eine Skala konstruiert, die noch weitere Stufen im oberen Wertebereich zu einer zusammenfasst, so dass die oberen 50% aller Werte nicht

Tabelle 3.50: Klassifikation 1000 integrierter Systeme auf nach oben beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	598	214	36	691	143	14	790	57	1	848
	60	586	278	41	677	215	13	810	90	5	905
	100	536	321	62	639	258	22	775	142	2	919
	300	435	412	105	527	380	45	681	263	8	952
(2,3)	30	589	165	28	654	119	9	739	40	3	782
	60	625	203	20	690	151	7	784	64	0	848
	100	608	235	18	682	167	12	774	84	3	861
	300	528	330	50	617	275	16	742	163	3	908
(3,10)	30	468	134	22	528	90	6	592	32	0	624
	60	511	144	16	567	99	5	631	40	0	671
	100	464	212	21	543	145	9	611	84	2	697
	300	368	309	53	439	272	19	576	152	2	730

unterschieden werden. Die Abbildung 3.14 auf Seite 221 zeigt ein kointegriertes System der Länge 100 auf dieser Skala.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Der ADF-Test kann in der ersten Gruppe ab einer Länge von 100 auf dem Niveau 0.05 noch gut eingesetzt werden (siehe Tabelle 3.51, S.222). Er erkennt mehr als 930 bzw. 950 stationäre Linearkombinationen bei kointegrierten bzw. stationären Systemen und auch ca. 950 integrierte Linearkombinationen bei den integrierten Systemen. Die Zahlen sind ähnlich wie im unverzerrten Fall (vgl. Tabelle 3.7, S.164), nur bei den integrierten Systemen werden auf diesem Niveau ca. 20-30 Linearkombinationen nun falsch interpretiert.

In der Gruppe (2,3) wirkt sich die Verzerrung durch das Abschneiden der oberen 50% des Wertebereichs sehr stark aus. Hier werden für die kointegrierten Systeme stets weniger als 500 stationäre Linearkombinationen erkannt. Für die stationären und integrierten Systeme ergeben sich kaum Abweichungen zum Verhalten bei den unverzerrten Daten (vgl. Tabelle 3.9, S.166).

Auch der ADF-Test versagt in Gruppe (3,1) bei den kointegrierten Systemen total und erkennt maximal 40% der Linearkombinationen als stationär. Trotzdem kann der Test in dieser Gruppe keine befriedigende Unterscheidung zwischen kointegrierten und integrierten Systemen leisten. Der Informationsverlust durch die Verzerrung ist einfach zu groß.

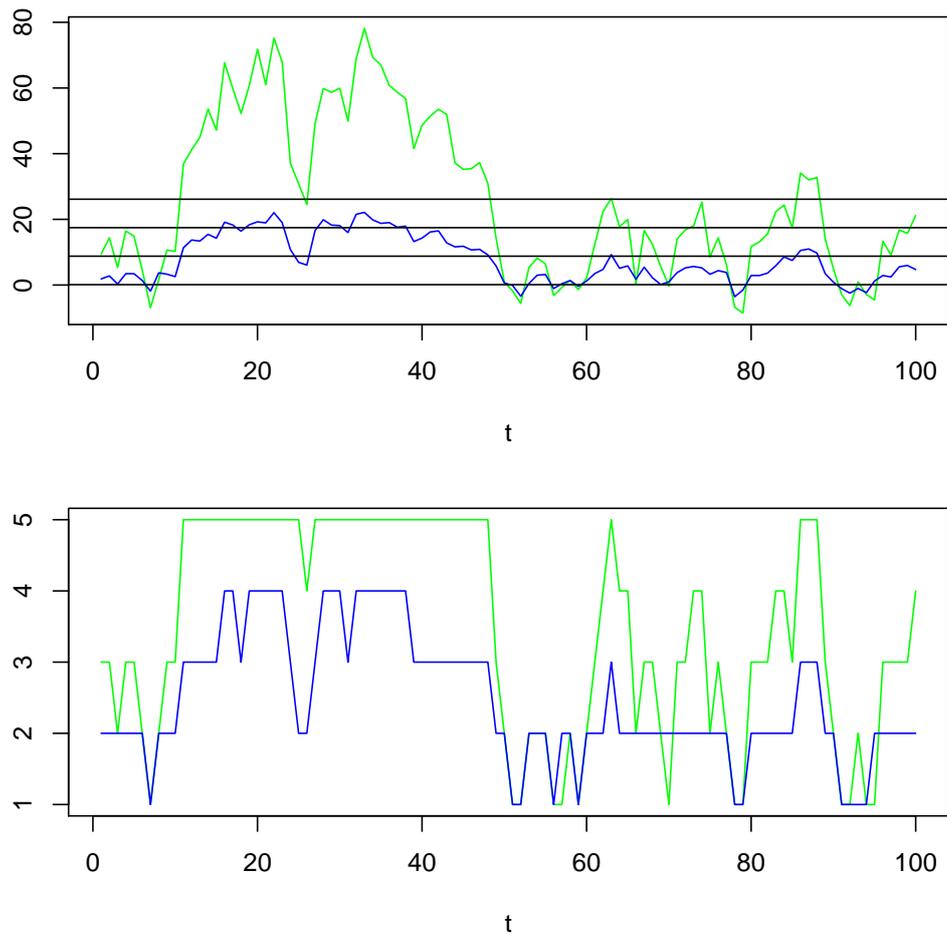


Abbildung 3.14: Konstruktion der Stufen der von oben bis zur Hälfte beschränkten Intervallskala, hier am Beispiel eines kointegriertes Systems der Länge 100 aus der Gruppe (3,10).

Tabelle 3.51: Ergebnisse des ADF-Tests auf bis zur Hälfte von oben beschränkter Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	370	250	101	322	183	46	98	52	19
	60	892	758	444	891	766	403	121	63	23
	100	971	932	727	990	956	758	107	55	11
	300	1000	1000	1000	1000	1000	1000	103	54	19
(2,3)	30	139	67	15	337	207	47	91	45	9
	60	319	236	105	885	746	393	100	56	21
	100	345	260	129	987	954	778	112	64	14
	300	495	392	272	1000	1000	1000	109	51	16
(3,10)	30	103	62	14	337	203	48	97	52	14
	60	161	109	34	891	753	393	95	49	13
	100	220	132	51	988	959	763	113	56	14
	300	394	295	158	1000	1000	1000	150	79	24

Tabelle 3.52: Klassifikation 1000 kointegrierter Systeme auf bis zur Hälfte nach oben beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	0	860	140	0	934	66	0	990	10	1000
	60	6	693	301	24	765	211	92	808	100	1000
	100	0	667	333	0	775	225	1	887	112	1000
	300	0	626	374	0	740	260	0	869	131	1000
(2,3)	30	9	912	79	15	954	31	43	954	3	1000
	60	106	592	186	162	616	106	299	547	38	884
	100	62	658	213	94	701	138	194	696	43	933
	300	12	716	270	19	801	178	46	884	68	998
(3,10)	30	73	752	72	90	779	28	143	749	5	897
	60	197	179	89	237	186	42	315	140	10	465
	100	140	250	94	179	243	62	274	192	18	484
	300	24	322	150	59	338	99	114	346	36	496

Klassifikation mit Johantest

Kointegrierte Systeme Nur in der Gruppe (1,1) konnte der Johantest (siehe Tabelle 3.52, S.222) für alle Längen bei allen 1000 Systemen durchgeführt werden. Auffällig ist, dass in der Gruppe (3,10) weniger als die Hälfte aller kointegrierten Systeme die Voraussetzungen an den Test erfüllen. Die anderen werden durch die simulierte Messung nahezu identisch abgebildet oder eine der Zeitreihen zeigt auf der Skala einen konstanten Verlauf.

In den Gruppen (2,3) und (3,10) ist ein klarer Längeneffekt auf allen drei Niveaus zu erkennen. Je länger die Zeitreihen werden, desto mehr Systeme werden als kointegriert klassifiziert. In der Gruppe (1,1) ist dieser Effekt nicht zu erkennen. Hier ist dafür der Effekt des Niveaus besser zu erkennen als in den anderen Gruppen. Nur in dieser Gruppe werden keine Systeme mehr als integriert eingestuft, wenn die Beobachtungszeit lang genug ist.

In den Gruppen (1,1) und (2,3) wird der Großteil (über 70%, maximal 94% in Gruppe (1,1) Länge 100 auf 0.01 Niveau) der Systeme richtig als kointegriert klassifiziert. In der Gruppe (3,10) ist dies nur bei Länge 300 der Fall. Hier wird ein ebenso großer oder teils sogar größerer Anteil als integriertes System klassifiziert. Eine Erkennungsrate von 95% wird jedoch nirgends erreicht.

Tabelle 3.53: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf bis zur Hälfte von oben beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-1,060	-1,032	-1,007	-1,000	-0,990	-0,965	-0,945	33
	60	-9,309	-1,383	-1,106	-1,010	-0,920	-0,650	3,500	83
	100	-9,726	-1,228	-1,060	-1,002	-0,946	-0,780	4,779	53
	300	-1,157	-1,078	-1,019	-0,999	-0,979	-0,921	-0,878	31
(2,3)	30	-0,921	-0,921	-0,800	-0,754	-0,698	-0,558	-0,513	3
	60	-20,265	-1,804	-0,939	-0,636	-0,352	0,348	19,290	45
	100	-22,145	-1,781	-0,904	-0,618	-0,310	0,506	119,117	32
	300	-7,221	-1,853	-0,899	-0,557	-0,260	0,157	6,353	13
(3,10)	30	-1,877	-0,489	-0,382	-0,334	-0,175	-0,010	-0,010	1
	60	-50,548	-1,362	-0,805	-0,606	-0,434	0,113	37,116	61
	100	-128,341	-1,075	-0,716	-0,578	-0,430	-0,034	40,561	52
	300	-50,430	-0,935	-0,656	-0,564	-0,462	-0,171	7,277	54

Schon bei der einfach nach oben beschränkten Skala hatte sich die Schätzungen von β in den Gruppen (2,3) und (3,10) verschoben. Auf dieser Skala tritt dieser Effekt erneut auf (siehe Tabelle 3.53, S.223). Die Verschiebung ist hier sogar noch stärker als auf den anderen beschränkten Skalen. Die Boxplot-Statistiken haben einen gleichmäßig und symmetrisch größeren Abstand zum Median als auf der anderen Skala.

Tabelle 3.54: Klassifikation 1000 stationärer Systeme auf bis zur Hälfte nach oben beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	72	85	785	200	191	551	535	242	165	942
	60	0	3	986	0	4	985	9	22	958	989
	100	0	0	997	0	0	997	0	0	997	997
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	70	82	711	170	173	520	514	202	147	863
	60	0	1	945	0	4	942	5	15	926	946
	100	0	0	972	0	1	971	0	1	971	972
	300	0	0	996	0	0	996	0	0	996	996
(3,10)	30	38	45	314	93	70	234	233	89	75	397
	60	0	1	474	0	2	473	2	9	464	475
	100	0	0	526	0	0	526	0	0	526	526
	300	0	0	633	0	0	633	0	0	633	633

Stationäre Systeme Bei den stationären Systemen konnte der Johantest (siehe Tabelle 3.54, S.224) selbst in der ersten Gruppe nicht für alle Längen durchgeführt werden, was auf der einfach nach oben beschränkten Skala noch für alle Gruppen der Fall war. Wie bei den kointegrierten Systemen wurden in der Gruppe (3,10) die wenigsten Tests ausgeführt. In der zweiten Gruppe gibt es deutlich weniger Systeme, für die der Test nicht ausgeführt werden konnte. Es wurden mehr ca. 97% ab Länge 100 alle Systeme getestet.

Ab einer Länge von 60 ist die Klassifikation in allen Gruppen auf dem 0.1 Niveau nahezu perfekt. Bis auf einzelne Ausnahmen werden alle getesteten Systeme als stationär klassifiziert. Auf dem 0.05 Niveau werden 2-4 Systeme als kointegriert missklassifiziert; auf dem 0.01 gibt es etwa 10-30 Fehlklassifikationen, einige jedoch gar als integriertes System. Dies war auch schon auf der nach oben beschränkten Skala zu sehen und hat sich nur minimal verstärkt. Insgesamt ist die Klassifikationsleistung bei den stationären Systemen jedoch auch bei nur 50% der Informationen noch sehr gut, wenn der Test denn ausgeführt werden konnte. Die Schätzungen für β liegen um Null, zeigen jedoch zum Teil sehr starke Ausreißer (siehe Tabelle D.12, S.369).

Integrierte Systeme Bei den integrierten Systemen (siehe Tabelle 3.55, S.225) konnte der Johantest in allen Gruppen bei nur ca. 50-70% (in der Gruppe (3,10) bei Länge 30 und 60 sogar noch etwas weniger) ausgeführt werden. In allen Gruppen wird der Großteil als integriert klassifiziert, wobei die Raten auf dem 0.01 Niveau stets am besten sind. Hier werden auch nur sehr wenige Systeme als stationär fehlklassifiziert. In der Gruppe (1,1) werden bei Länge 300 auf dem 0.1 Niveau fast genauso viele Systeme als kointegriert wie als integriert eingestuft. In keiner Gruppe wird die Marke von 95% richtig klassifizierten Systemen erreicht.

Tabelle 3.55: Klassifikation 1000 integrierter Systeme auf bis zur Hälfte nach oben beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	289	154	52	348	122	25	423	65	7	495
	60	265	222	78	330	194	41	439	118	8	565
	100	276	249	94	333	233	53	438	166	15	619
	300	172	329	184	221	346	118	331	323	31	685
(2,3)	30	543	330	71	625	281	38	772	167	5	944
	60	327	157	35	377	123	19	442	72	5	519
	100	336	197	56	387	171	31	464	119	6	589
	300	275	271	109	331	278	46	437	210	8	655
(3,10)	30	288	104	18	320	77	13	375	32	3	410
	60	304	138	30	356	105	11	409	59	4	472
	100	312	154	57	360	140	23	426	91	6	523
	300	203	231	112	276	211	59	380	149	17	546

Dies war auf der einfach beschränkten Skala noch der Fall. Die Schätzungen für β streuen um Null, jedoch stärker als im unverzerrten Fall (siehe Tabelle 3.53, S.223).

Ergebnis: Messung auf bis zur Hälfte beschränkter Skala

Wenn nur die Hälfte des Wertebereichs berücksichtigt werden kann, ist der Einsatz von Kointegrationsverfahren nicht möglich.

Augmented Dickey-Fuller-Test Nur bei den Systemen der Länge 300 aus der Gruppe (1,1) erreicht der ADF-Test noch gute Ergebnisse. Hier werden alle stationären und auf dem 0.05 Niveau noch 946 integrierte erkannt. In den übrigen Gruppen kann der Test bei keiner Länge eingesetzt werden.

Johansen-Test Insgesamt setzen sich die Effekte bei den kointegrierten Systemen aus der vorigen Studie fort, die schon bei der beschränkten Skala beobachtet werden konnten. Je mehr Informationen man abschneidet, desto schlechter werden in allen Gruppen die Klassifikationsergebnisse, desto mehr Systeme lassen den Johansen-Test gar nicht mehr zu. Dabei sind die Auswirkungen mit zunehmender Varianzunterschied größer.

Für stationäre Systeme bewirkt der beschränkte Wertebereich, dass sie nicht immer die Voraussetzungen des Tests erfüllen, da einige Zeitreihen nach der Messung einen konstanten Verlauf zeigen. Die Systeme, für die der Test ausgeführt werden konnte, werden ab einer

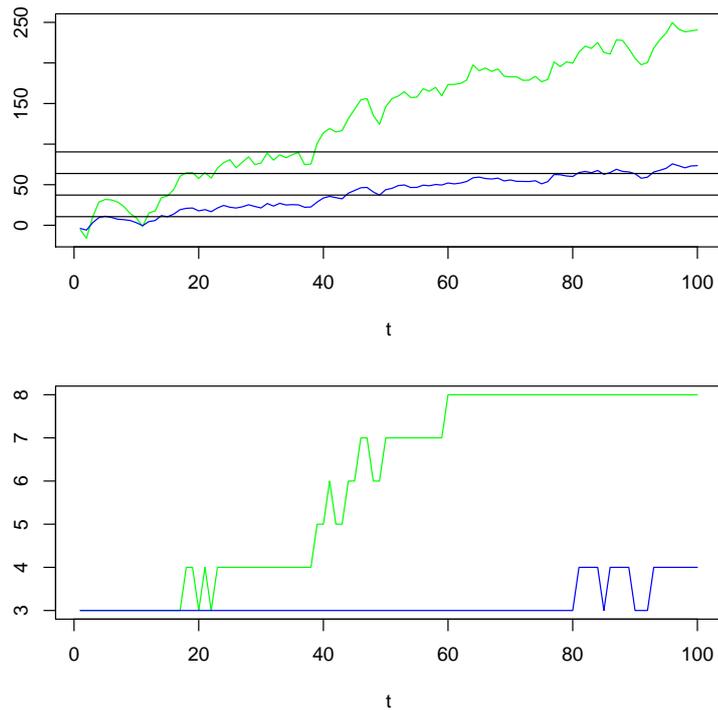


Abbildung 3.15: Kointegriertes Systeme der Länge 100 aus der Gruppe (3,10), für das die Kointegration nach der Messung auf der von beiden Seiten beschränkten Skala nicht mehr erkannt wird. Es ist deutlich zu sehen, dass eine der beiden Zeitreihen dann konstant ist, wenn die andere steigt. Der gemeinsame Trend ist nicht mehr zu erkennen.

Länge von 100 Datenpunkten perfekt klassifiziert. Bei den integrierten Systemen konnte der Johansenstest nur bei etwa jedem zweiten Systeme durchgeführt werden.

3.4.3 Von beiden Seiten beschränkt

In einer weiteren Simulationsstudie wurden gleichzeitig Ceiling- und Flooreffekt provoziert. Die verwendete Skala fasst sowohl die unteren als auch die oberen 30% des jeweiligen Wertebereichs zu einer Stufe zusammen, so dass eine 6-stufige Skala entsteht. Für das kointegrierte System in der Abbildung 3.15 auf Seite 226 führt das dazu, dass die Kointegration nicht mehr erkannt wird, während sie in dem Systeme, das in Abbildung 3.16 auf Seite 227 dargestellt ist, erhalten bleibt.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

In der Gruppe (1,1) sind die Ergebnisse des ADF-Tests (siehe Tabelle 3.56, S.227) noch sehr nah an denen bei unverzerrten Daten (vgl. Tabelle 3.7, S.164). In den anderen beiden Gruppen können jedoch deutlich weniger kointegrierte Systeme richtig klassifiziert werden. Für die stationären Systeme ist die Klassifikation bei einer Länge von 300 Datenpunkten immer

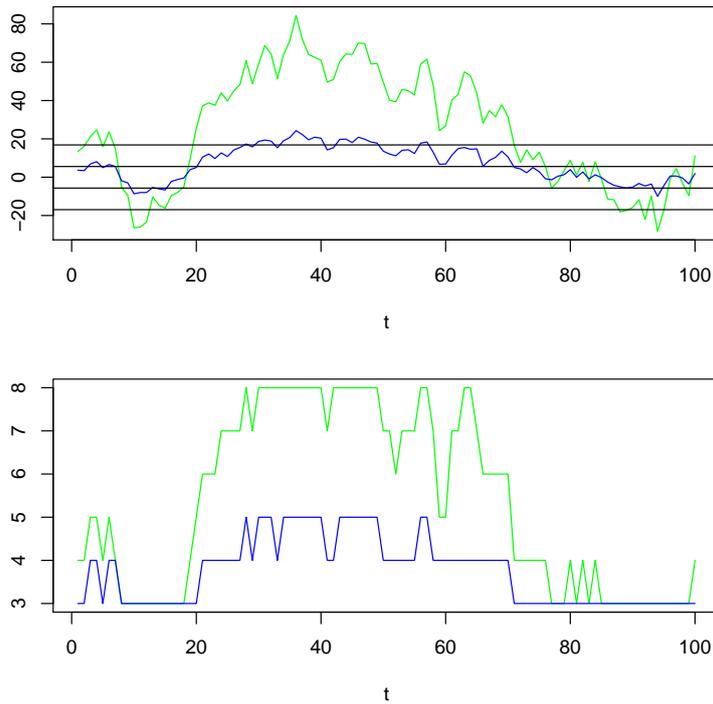


Abbildung 3.16: Kointegriertes Systeme der Länge 100 aus der Gruppe (3,10), für das die Kointegration nach der Messung auf der von beiden Seiten beschränkten Skala weiterhin erkannt wird.

Tabelle 3.56: Ergebnisse des ADF-Tests auf von unten beschränkter Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	344	197	52	335	213	54	86	54	13
	60	871	740	418	885	741	416	102	53	19
	100	975	931	731	991	956	770	109	54	12
	300	1000	1000	1000	1000	1000	1000	83	44	10
(2,3)	30	168	100	20	321	204	52	104	52	16
	60	365	233	79	875	751	393	115	45	15
	100	427	307	132	985	957	777	95	58	18
	300	620	484	270	1000	1000	1000	99	54	9
(3,10)	30	132	71	19	332	193	45	106	63	15
	60	297	186	59	873	744	405	95	49	17
	100	407	287	109	985	953	770	98	56	14
	300	611	519	338	1000	1000	1000	109	73	24

noch perfekt und auch bei einer Länge 100 ausreichend gut. Auf dem 0.05 Niveau werden in allen Gruppen ca. 95% integrierte Linearkombinationen als solche erkannt. Dennoch kann der Test auf dieser Skala nicht mehr zur Unterscheidung von kointegrierten und integrierten Systemen eingesetzt werden.

Klassifikation mit Johantest

Tabelle 3.57: Klassifikation 1000 kointegrierter Systeme auf von beiden Seiten beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	266	647	87	415	536	49	732	266	2	1000
	60	5	886	109	13	928	59	81	903	16	1000
	100	0	872	128	0	933	67	0	982	18	1000
	300	0	860	140	0	934	66	0	990	10	1000
(2,3)	30	495	448	56	611	362	26	835	158	6	999
	60	220	701	79	325	640	35	562	433	5	1000
	100	128	791	81	184	782	34	343	646	11	1000
	300	9	912	79	15	954	31	43	954	3	1000
(3,10)	30	270	331	38	355	272	12	510	128	1	639
	60	183	531	40	253	482	19	386	365	3	754
	100	125	635	54	179	612	23	283	525	6	814
	300	73	752	72	90	779	28	143	749	5	897

Kointegrierte Systeme In den Gruppen (1,1) und (2,3) konnte der Johantest (siehe Tabelle 3.57, S.228) nur bei einem System der Länge 30 nicht ausgeführt werden; bei allen anderen Längen wurde der Test bei allen 1000 Systemen ausgeführt. In der Gruppe (3,10) wurden bei Länge 30 nur 639 Systeme getestet, mit zunehmener Länge sind es mehr, bis bei Länge 300 897 Systeme getestet wurden.

In der Gruppe (1,1) ist ein Effekt der Länge beim Klassifikationsverhalten nicht so deutlich ausgeprägt wie in den anderen Gruppen. Auf dem 0.1 Niveau werden bei Länge 100 und 300 sogar etwas weniger Systeme richtig klassifiziert als bei Länge 60. Sonst nimmt die Anzahl der richtig klassifizierten Systeme mit der Länge zu.

In der Gruppe (1,1) werden auf dem 0.01 Niveau bei einer Länge von 100 mehr als 98% der Systeme richtig klassifiziert, auf dem 0.05 Niveau sind dies noch 93%. In der Gruppe (2,3) wird ein Wert über 90% nur bei Länge 300 erreicht; in der Gruppe (3,10) werden stets weniger Systeme richtig klassifiziert.

In den Gruppe (2,3) und (3,10) hat sich die Schätzung von β verschoben, während der Median in Gruppe (1,1) bei -1.00 liegt.

Tabelle 3.58: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf von beiden Seiten beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-6,408	-1,413	-1,113	-1,003	-0,911	-0,614	221,424	99
	60	-1,983	-1,166	-1,045	-1,003	-0,964	-0,850	-0,414	78
	100	-1,426	-1,091	-1,022	-1,000	-0,976	-0,908	-0,660	58
	300	-1,060	-1,032	-1,007	-1,000	-0,990	-0,965	-0,945	33
(2,3)	30	-202,872	-1,134	-0,875	-0,786	-0,698	-0,440	19,762	73
	60	-2,277	-1,003	-0,835	-0,773	-0,709	-0,528	0,298	21
	100	-1,125	-0,983	-0,818	-0,766	-0,708	-0,556	-0,449	6
	300	-0,921	-0,921	-0,800	-0,754	-0,698	-0,558	-0,513	3
(3,10)	30	-1,321	-0,648	-0,436	-0,381	-0,287	-0,079	3,109	13
	60	-3,453	-0,623	-0,404	-0,367	-0,239	-0,032	1,181	3
	100	-1,075	-0,518	-0,389	-0,348	-0,218	-0,027	1,843	6
	300	-1,877	-0,489	-0,382	-0,334	-0,175	-0,010	-0,010	1

Tabelle 3.59: Klassifikation 1000 stationärer Systeme auf von beiden Seiten beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	87	91	822	197	179	624	548	256	196	1000
	60	0	0	1000	0	1	999	4	9	987	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	91	103	806	194	186	620	527	278	195	1000
	60	0	0	1000	1	0	999	5	8	987	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	77	103	820	211	194	595	538	290	172	1000
	60	0	0	1000	0	0	1000	7	7	986	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Stationäre Systeme Für die stationären Systeme konnte der Test in jeder Gruppe bei jeder Länge für alle Systeme ausgeführt werden (siehe Tabelle 3.59, S.229). Auf dem 0.1 Niveau werden ab einer Länge von 60 alle Systeme als stationär erkannt. Auf dem 0.05 Niveau gibt es nur in der Gruppe (1,1) genau ein System, das als kointegriert statt als stationär klassifiziert wird. Auf dem 0.01 Niveau sind es in allen Gruppen 8-9 zum Teil auch als integriert eingestufte Systeme. Die Schätzungen für β streuen um Null.

Tabelle 3.60: Klassifikation 1000 integrierter Systeme auf von beiden Seiten beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	645	244	49	752	159	27	877	55	6	938
	60	633	288	53	717	236	21	851	115	8	974
	100	551	372	61	670	288	26	823	149	12	984
	300	453	425	109	568	362	57	733	246	8	987
(2,3)	30	621	196	25	691	134	17	777	62	3	842
	60	659	204	35	731	151	16	820	74	4	898
	100	617	253	38	688	195	25	799	99	10	908
	300	543	330	71	625	281	38	772	167	5	944
(3,10)	30	377	101	12	420	66	4	455	34	1	490
	60	435	138	12	489	90	6	544	39	2	585
	100	422	171	11	480	117	7	548	56	0	604
	300	322	295	26	396	234	13	486	155	2	643

Integrierte Systeme In keiner Gruppe der integrierten Systeme konnten alle 1000 Systeme getestet werden (siehe Tabelle 3.60, S.230). Während es in der Gruppe (1,1) noch bei 938-987 Systemen möglich ist, kann der Test nur noch bei 842-944 Systemen in Gruppe (2,3) und bei 490-643 Systemen in Gruppe (3,10) durchgeführt werden. Je länger die Reihen, desto mehr Systeme konnten getestet werden.

Bei der Klassifikation bewirkt die Zunahme der Länge, dass immer weniger Systeme richtig klassifiziert werden, obwohl der Test sogar bei mehr Systemen durchgeführt wird. In allen Gruppen werden auf allen Niveaus die meisten Systeme richtig als integriert klassifiziert. Doch selbst auf dem 0.01 Niveau liegen die Klassifikationsraten nur bei knapp 90%, nirgends werden 95% der Systeme richtig klassifiziert. Die Schätzungen von β liegen um Null.

Ergebnis: von beiden Seiten beschränkte Skala

Die Beschränkung des Wertebereichs wirkt sich bei beiden Kointegrationsverfahren nur bei den stationären Systemen kaum aus. Für die anderen Systeme scheitert insbesondere in den

Gruppen mit unterschiedlicher Gewichtung des Trends die Ausführung des Johansen-Test oder die Klassifikation der Linearkombinationen der kointegrierten Systeme.

3.4.4 Ergebnis: Messung mit Ceiling- und Flooreffekten

Augmented Dickey-Fuller-Test Der Augmented Dickey-Fuller-Test hat auf allen Skalen bei Länge 300 alle stationären Linearkombinationen der stationären Systeme erkannt. In den Gruppen mit unterschiedlicher Varianz gab es jedoch zunehmend Probleme beim Erkennen der Kointegrationsbeziehung, so dass er zur Unterscheidung von kointegrierten und integrierten Systemen nicht eingesetzt werden kann.

Klassifikation mit Johansen-Test Durch die Beschränkung des Wertebereichs kann ein stochastischer Trend in den Systemen nicht mehr ausreichend gut abgebildet werden. Das führt dazu, dass der Johansen-Test insbesondere bei unterschiedlicher Gewichtung der integrierten Anteile nicht mehr ausgeführt werden kann. Einzig die Klassifikation der stationären Systeme ist auch den beschränkten Skalen zuverlässig.

Der Decken- bzw. Bodeneffekt bewirkt, dass die Informationen über integrierte Anteile durch die Messung verloren gehen.

3.5 Messung auf Ordinalskalenniveau

Da man in der psychologischen Forschung nicht immer auf intervallskalierte Daten zurückgreifen kann, soll hier anhand von vier konstruierten Ordinalskalen getestet werden, ob die Information der Rangreihe ausreichend ist, um Kointegration zu erkennen. Der integrierte Anteil der Zeitreihen bewirkt einen dynamischen Verlauf. Es ist nun zu prüfen, ob die Information der Richtung entscheidender ist als der tatsächliche Wert dieser Veränderung.

3.5.1 Ordinalskala 1: „symmetrisch um Median gewichtet“

Die erste Ordinalskala teilt den Wertebereich symmetrisch um seinen Mittelwert in sieben Bereiche. Alle Werte in einem Bereich werden auf dieselbe Stufe der Skala abgebildet (Messvorgang). Der mittlere Bereich entspricht der Stufe 4. Er ist relativ klein gewählt. Die Breite der Bereiche nimmt mit dem Abstand zum Mittelwert zu. So entsteht eine Skala mit sieben Stufen, deren Randbereiche Floor- und Ceilingeffekte aufweisen sollten. Hier werden jedoch anders als bei den beschränkten Skalen nur jeweils 20% zu einer Stufe zusammengefasst (siehe Abbildung 3.17, S.232). Die exakten Stufenabstände können der Beschreibung in Anhang D.6 ab Seite 394 entnommen werden.

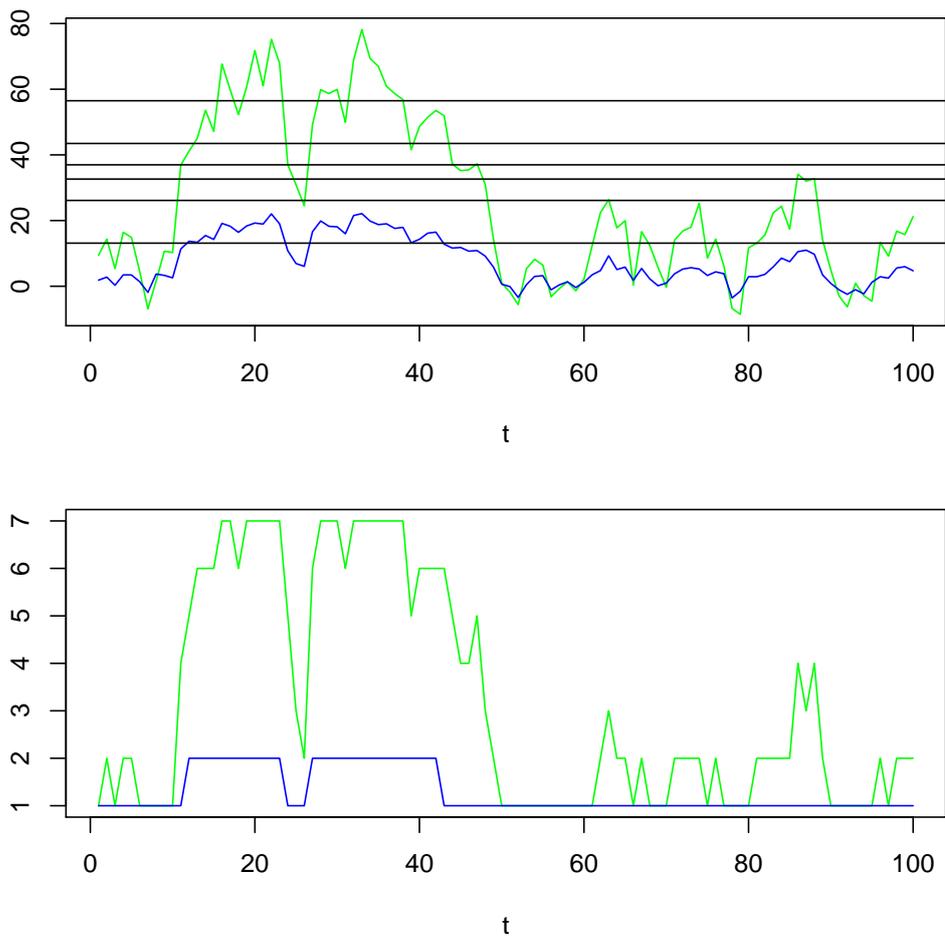


Abbildung 3.17: Konstruktion der ersten Ordinalskala (7 Stufen): Die Stufenabstände werden mit Abstand zum Median größer.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Tabelle 3.61: Ergebnisse des ADF-Tests auf der Ordinalskala 1: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	325	194	56	340	217	57	88	46	11
	60	868	735	427	885	736	411	104	61	18
	100	983	943	749	988	955	771	114	68	15
	300	1000	1000	1000	1000	1000	1000	91	58	14
(2,3)	30	183	115	29	345	201	48	83	42	16
	60	335	217	68	878	735	393	108	53	14
	100	360	251	99	983	956	755	106	65	22
	300	577	446	214	1000	1000	1000	88	48	13
(3,10)	30	136	79	18	356	209	50	92	52	14
	60	301	182	73	871	745	416	92	57	10
	100	354	251	102	988	942	760	114	68	19
	300	544	394	197	1000	1000	1000	105	62	20

Gruppe (1,1) Der ADF-Test (siehe Tabelle 3.61, S.233) zeigt in der Gruppe (1,1) kaum eine Veränderung in seinem Klassifikationsverhalten im Vergleich zu den unverzerrten Daten (vgl. Tabelle 3.7, S.164). Lediglich bei den integrierten Systemen werden mehr als 20 Linearkombinationen anders eingeschätzt. Dennoch klassifiziert der Test ab einer Länge von 100 auf dem 0.05 in jeder Gruppe mehr als 930 Linearkombinationen auch nach dieser Verzerrung richtig. Das Verhalten ist vergleichbar und auch zahlenmäßig ganz ähnlich zu dem auf der 7-stufigen Intervallskala (siehe Tabelle 3.29, S.195) oder der von beiden Seiten beschränkten Skala (siehe Tabelle 3.56, S.227). Die Rate der richtig erkannten integrierten Linearkombinationen hat sich um ca. 10 Systeme verschlechtert.

Gruppe (2,3) In der Gruppe (2,3) verhält sich der ADF-Test wie bei der von beiden Seiten beschränkten Skala (vgl. Tabelle 3.56, S.227). Nur bei den stationären Systemen ist die Verzerrung nicht zu bemerken. Bis zu einer Länge von 100 Datenpunkten zeigen sich leichte Abweichungen bis maximal 19 Systeme, deren Linearkombination anders eingeschätzt wird als bei den Originaldaten (vgl. Tabelle 3.9, S.166). Bei einer Länge von 300 ist das Ergebnis für stationäre Systeme perfekt. Der ADF-Test erkennt bis zu 25 integrierte Linearkombinationen weniger als im unverzerrten Fall. Dennoch ist die Erkennrate so hoch, dass der Test auf dem 0.05 oder 0.01 in dieser Gruppe bei jeder Länge eingesetzt werden könnte. Dies ist jedoch durch die schlechten Ergebnisse bei den kointegrierten Systemen ausgeschlossen. Nur bei einer Länge von 300 werden mehr als 500 stationäre Linearkombinationen auf dem 0.1

Niveau erkannt. Auf dem 0.01 Niveau sind es bei dieser Länge nur 214, wobei im unverzerrten Fall alle 1000 und auf der 7-stufigen Intervallskala noch 994 Linearkombinationen waren (vgl. Tabelle 3.29, S.195). Das Verhalten ist jedoch dasselbe wie bei der von beiden Seiten beschränkten Intervallskala (vgl. Tabelle 3.56, S.227). Hier wurden bei einer Länge von 300 270 (0.01 Niveau) bzw. 620 (0.1 Niveau) stationäre Linearkombinationen erkannt. Bei den kürzeren Systemen sind es auf jedem Niveau noch weniger, wobei auf dem 0.1 Niveau stets die meisten Linearkombinationen als stationär eingestuft werden.

Gruppe (3,10) Die Effekte setzen sich auch in der Gruppe (3,10) fort. Die Werte sind vergleichbar mit denen aus der Gruppe (2,3) und bei kointegrierten Systemen eher noch schlechter. Kaum eine Auswirkung hat die Verzerrung auf die Klassifikation von stationären Systemen und bei den integrierten Systemen schneidet der Test noch ziemlich gut ab.

Johansentest

Tabelle 3.62: Klassifikation 1000 kointegrierter Systeme auf der Ordinalskala 1

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	249	665	86	411	547	42	708	287	5	1000
	60	10	877	113	17	926	57	89	894	17	1000
	100	0	871	129	0	930	70	1	980	19	1000
	300	0	865	135	0	929	71	0	985	15	1000
(2,3)	30	481	453	66	608	370	22	818	178	4	1000
	60	240	686	74	340	625	35	546	445	9	1000
	100	180	741	79	240	726	34	365	627	8	1000
	300	37	889	74	61	910	29	138	859	3	1000
(3,10)	30	395	346	55	491	287	18	670	122	4	796
	60	258	608	65	368	537	26	581	343	7	931
	100	196	713	74	266	689	28	427	548	8	983
	300	77	839	84	97	870	33	174	822	4	1000

Kointegrierte Systeme Für die kointegrierten Systeme konnte der Johansentest in den Gruppen (1,1) und (2,3) für alle 1000 Systeme bei allen Längen ausgeführt werden (siehe Tabelle 3.62, S.234). In der Gruppe (3,10) gelangt dies nur bei einer Länge von 300. Bei 30 Beobachtungspunkten waren es nur 796, bei 60 schon 931 und bei 100 Zeitpunkten 983 Systeme.

In den Gruppen (2,3) und (3,10) werden auf allen Niveaus mit zunehmender Länge immer mehr Systeme als kointegriert klassifiziert. In der Gruppe (1,1) ist dieser Effekt nicht zu erkennen.

In der Gruppe (1,1) werden auf dem 0.01 Niveau bei den Längen 100 und 300 mehr als 98% der Systeme richtig klassifiziert, bei 60 Datenpunkten sind es noch knapp 90%. Auf dem 0.05 Niveau liegt der Anteil bei ca. 93% ab einer Länge von 60 und auf dem 0.1 immer noch bei ca. 87%. Bei den Längen 100 und 300 wird nur auf dem 0.01 ein System als integriert klassifiziert, alle anderen Fehlklassifikationen bewerten das System als stationär.

In der Gruppe (2,3) werden bei den Längen 60 und 100 stets auf dem 0.1 Niveau mehr Systeme richtig klassifiziert als auf dem 0.05 Niveau und noch einmal mehr als auf dem 0.01 Niveau; bei einer Länge von 300 zeigt sich dieser Unterschied nicht so deutlich. Auf dem 0.01 werden bei Länge 60 mehr Systeme als integriert als als kointegriert klassifiziert. Sonst werden mindestens 62%, bei Länge 300 mehr als 85% richtig klassifiziert.

In der Gruppe (3,10) liegt die Klassifikationsrate bei einer Länge von 300 zwischen 82%-87%. Auch hier werden bei Länge 60 auf dem 0.01 Niveau mehr Systeme als integriert klassifiziert als als kointegriert erkannte werden. Sonst liegen die Raten bei 60-70%.

Ein zuverlässiges Ergebnis kann man also nur in der Gruppe (1,1) auf dem 0.01 Niveau bei einer Länge ab 100 erwarten. Hier zeigt sich noch ein ähnliches Bild wie bei den unverzerrten Daten mit maximal 23 Systemen Unterschied auf den Längen ab 60 (vgl. Tabelle 3.7, S.164).

Tabelle 3.63: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf der Ordinalskala 1

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-7,208	-1,407	-1,108	-1,005	-0,905	-0,601	23,585	105
	60	-1,986	-1,168	-1,045	-1,002	-0,963	-0,842	-0,377	62
	100	-1,459	-1,092	-1,025	-1,000	-0,979	-0,912	-0,635	51
	300	-1,062	-1,032	-1,007	-0,999	-0,991	-0,966	-0,938	33
(2,3)	30	-7,145	-1,109	-0,865	-0,792	-0,701	-0,463	3,903	76
	60	-7,462	-1,001	-0,836	-0,786	-0,725	-0,559	0,916	43
	100	-11,194	-0,954	-0,820	-0,778	-0,728	-0,591	-0,300	35
	300	-1,054	-0,921	-0,805	-0,774	-0,721	-0,597	-0,491	25
(3,10)	30	-3,792	-0,694	-0,404	-0,268	-0,195	0,111	3,256	20
	60	-55,383	-0,635	-0,373	-0,246	-0,186	-0,024	0,340	6
	100	-0,677	-0,560	-0,338	-0,233	-0,185	-0,034	-0,034	3
	300	-0,658	-0,533	-0,322	-0,225	-0,181	-0,011	-0,011	2

In der Gruppe (1,1) wird β um -1 geschätzt. Dabei zeigt der Median auf zwei Nachkommastellen keine Abweichung von der Schätzung im unverzerrten Fall (vgl. Tabelle 3.15, S.173). Die Statistiken des Boxplots haben symmetrisch etwas mehr Abstand zum Median bekommen. Dies ist mit zunehmender Länge stärker. In der Gruppe (2,3) hat sich der Median für die β Schätzung um 0,11 verschoben und liegt nun bei 0,77. Die Boxplot-Statistiken haben die Verschiebung mitgemacht und außerdem mehr Abstand zum Median gewonnen. Die Schätzung streuen also etwas stärker als im unverzerrten Fall. Ähnliches lässt sich auch in

der Gruppe (3,10) beobachten. Hier nimmt die Verschiebung mit der Länge der Reihen noch zu. Bei 30 Datenpunkten liegt der Median bei 0,26, bei 300 bereits bei 0,22 statt 0,3 wie im ursprünglichen Fall.

Stationäre Systeme Für die stationären Systeme (siehe Tabelle 3.64, S.236) konnte der Johansentest nur in der Gruppe (3,10) bei einer Länge von 30 für drei Systeme nicht ausgeführt werden. Sonst wurden alle Systeme berücksichtigt.

In allen Gruppen werden auf dem 0.1 Niveau alle Systeme ab einer Länge von 60 als stationär erkannt, dies trifft bis auf ein System in der Gruppe (1,1) bei Länge 60 auch auf dem 0.05 Niveau zu. Auf dem 0.01 Niveau gibt es in allen drei Gruppen bei Länge 60 einige (maximal 14) Systeme, die nicht als stationär klassifiziert werden, ab Länge 100 werden auch hier alle Systeme erkannt.

In allen Gruppen liegen die Schätzungen von β um Null, obwohl die Streuung größer ist als im unverzerrten Fall.

Tabelle 3.64: Klassifikation 1000 stationärer Systeme auf der Ordinalskala 1

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	87	85	828	192	197	611	533	260	207	1000
	60	0	0	1000	0	1	999	4	14	982	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	84	104	812	200	207	593	531	276	193	1000
	60	0	0	1000	0	0	1000	9	7	984	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	90	106	801	195	197	605	527	280	190	997
	60	0	0	1000	0	0	1000	7	10	983	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Integrierte Systeme In keiner Gruppe konnte der Johansentest (siehe Tabelle 3.65, S.237) bei jeder Länge bei allen Systemen ausgeführt werden. Dabei steigt die Zahl der getesteten Systeme mit der Länge, wobei in der Gruppe (1,1) die meisten (mehr als 977) und in der Gruppe (3,10) mit 590-752 die wenigsten Systeme berücksichtigt werden konnten.

Wenn man die Länge 30 außer Acht lässt werden im Allgemeinen mit zunehmender Länge weniger Systeme als integriert klassifiziert, obwohl der Test bei mehr Systemen ausgeführt

wird. Die meisten Fehlspezifikationen sind als kointegrierte Systeme. Jedoch werden nirgends mehr Systeme als kointegriert klassifiziert als als integriert.

Die beste Klassifikationsrate wird auf dem 0.01 Niveau erreicht. Hier liegt die Rate bei Länge 60-100 um die 90%. Die Ergebnisse sind jedoch so schlecht, dass der Test nicht zuverlässig zwischen integrierten und kointegrierten Systemen unterscheiden kann. Auch hier schwanken die Werte der β Schätzung um Null (siehe Tabelle D.17, S.374), jedoch mit größerer Streuung als im unverzerrten Fall (vgl. Tabelle D.2, S.359).

Tabelle 3.65: Klassifikation 1000 integrierter Systeme auf der Ordinalskala 1

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	682	243	52	794	160	23	903	69	5	977
	60	635	317	43	747	230	18	883	109	3	995
	100	546	379	75	670	295	35	850	145	5	1000
	300	452	435	111	557	386	55	723	261	14	998
(2,3)	30	647	222	31	729	155	16	829	67	4	900
	60	693	232	32	779	165	13	881	74	2	957
	100	645	273	44	728	212	22	853	102	7	962
	300	579	327	71	668	274	35	805	166	6	977
(3,10)	30	453	124	13	502	86	2	559	31	0	590
	60	510	149	14	576	91	6	632	40	1	673
	100	480	225	18	540	173	10	638	82	3	723
	300	378	341	33	439	297	16	559	189	4	752

Ergebnis: Ordinalskala 1

Augmented Dickey-Fuller-Test Auf dieser Ordinalskala kann ein Test bei bekanntem Kointegrationsvektor nur bei gleicher Varianz der beiden Reihen durchgeführt werden.

Johansenstest Der Johansenstest zeigt auf dieser Ordinalskala ein ähnliches Verhalten wie bei der simulierten Messung mit der von beiden Seiten beschränkten Skala. Insbesondere die integrierten Systeme können nicht mehr ausreichend gut klassifiziert werden, während es bei den stationären Systemen kaum Abweichungen gibt.

3.5.2 Ordinalskala 2: „symmetrisch mit Differenzierung der Randbereiche“

Auch die zweite Ordinalskala teilt den Wertebereich symmetrisch um den Mittelwert in sieben Bereiche ein, die auf die sieben Stufen der Skala abgebildet werden. Jedoch wird hier

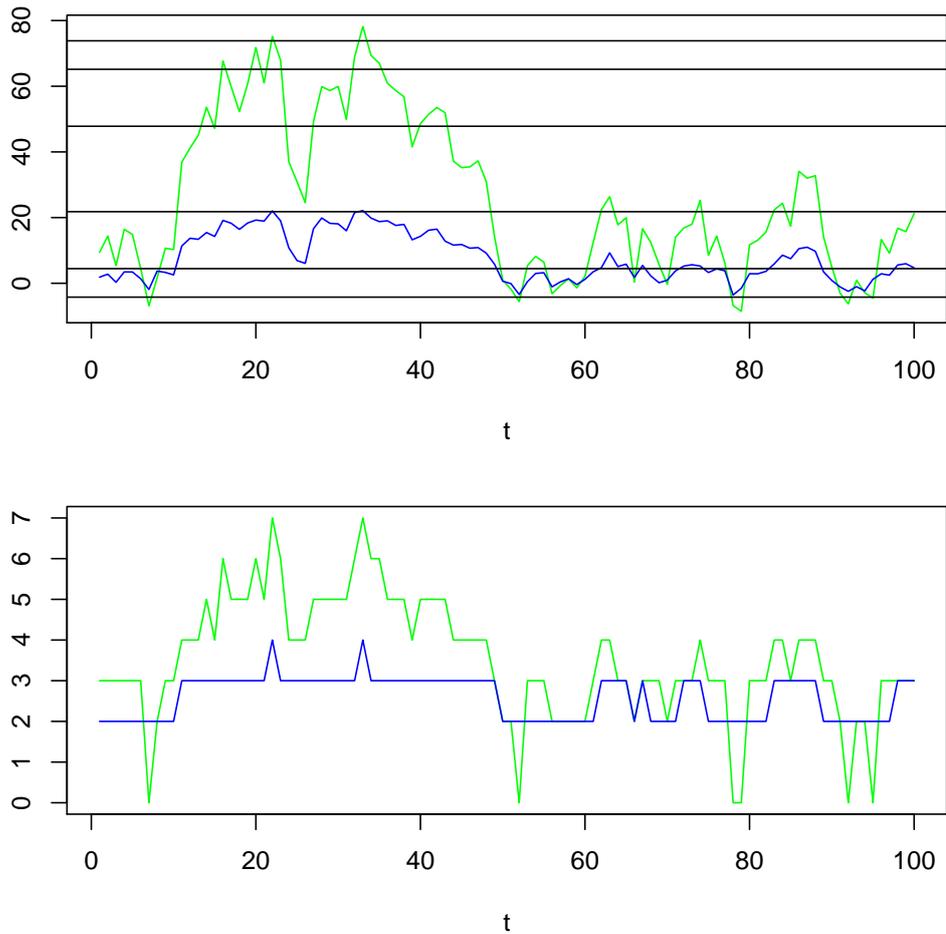


Abbildung 3.18: Konstruktion der zweiten Ordinalskala (7 Stufen): Die Stufenabstände werden mit Abstand zum Median kleiner.

in den Randbereichen stärker differenziert, so dass eher im mittleren Bereich Informationen verloren gehen und keine Ceiling- oder Flooreffekte auftreten sollten (siehe Abbildung 3.18, S.238).

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Der ADF-Test (siehe Tabelle 3.66, S.239) zeigt in der ersten Gruppe nur wenig Abweichung in seinem Klassifikationsverhalten (vgl. Tabelle 3.7, S.164). Zwar werden bei den kointegrierten Systemen etwas weniger stationäre Linearkombinationen erkannt, maximal 32 bei Länge 60 auf 0.1 Niveau und bei Länge 300 nicht mehr auf allen Niveaus alle 1000. Dennoch ist die Klassifikationsrate bei Länge 300 nahezu perfekt und bei Länge 100 auf dem 0.05 fast erreicht.

Für stationäre Systeme zeigt der ADF-Test bei Länge 300 keine Abweichung und erkennt alle 1000 Linearkombinationen als stationär, bei Länge 100 sind es maximal 8 nicht mehr erkannte und auf den Längen 30 und 60 beträgt die Differenz 2-36 Systeme. Bei den integrierten Systemen ist die Verzerrung der Daten deutlicher zu bemerken. Bei jeder Länge

Tabelle 3.66: Ergebnisse des ADF-Tests auf der Ordinalskala 2: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	310	177	46	310	185	44	98	46	12
	60	849	729	403	894	744	403	149	76	18
	100	971	942	752	985	952	769	140	59	9
	300	1000	999	999	1000	1000	1000	210	108	30
(2,3)	30	216	114	29	317	176	44	100	56	9
	60	550	407	163	884	753	384	139	68	14
	100	682	522	239	991	962	768	137	65	14
	300	922	867	691	1000	1000	1000	231	130	46
(3,10)	30	187	106	25	310	158	40	95	52	21
	60	414	283	108	882	735	392	141	88	16
	100	486	370	149	991	965	761	157	86	25
	300	734	620	368	1000	1000	1000	249	156	52

und jedem Niveau werden mehr Linearkombinationen für stationär gehalten, als dies bei den unverzerrten Daten der Fall war. Auf dem 0.01 Niveau ist die Erkennquote jedoch noch bei über 970, bei Länge 100 sogar bei über 990. Die Abweichungen sind auf dem 0.1 Niveau im Allgemeinen und bei der Länge 300 im Speziellen am größten. Auf dem 0.1 Niveau werden 210 Systeme mit 300 Beobachtungspunkten als stationär eingestuft, dies sind 137 mehr als im unverzerrten Fall. So erreicht die Klassifikationsrate auf dem 0.05 Niveau die 950-Marke nicht, obwohl mehr als 90% (bei Länge 300 nur 89%) richtig eingeschätzt werden. Über alle Systemgruppen gesehen liefert der Test bei Länge 300 auf dem 0.01 die besten Werte. Hier erreicht er in allen Gruppen die 950-Marke. Auch auf dem 0.05 Niveau sind die Ergebnisse bei Länge 100 (und 300) zufriedenstellend, obwohl sie die Zielmarke nicht erreichen.

Auch in der Gruppe (2,3) wirkt sich die Verzerrung bei kointegrierten Systemen so stark auf den ADF-Test aus, dass maximal 922 stationäre Linearkombinationen erkannt werden (Länge 300, Niveau 0.1). Bei einer Länge von 100 sind es maximal 682, auf dem 0.05 knapp über die Hälfte aller Systeme. Die Abweichungen bei den integrierten Systemen sind auch deutlich, aber nicht so gravierend wie bei den kointegrierten Systemen. Nur für die stationären Systeme ändert sich die Einschätzung kaum (vgl. Tabelle 3.9, S.166).

Die in der Gruppe (2,3) beobachteten Effekte setzen sich in Gruppe (3,10) fort. Bei den stationären Systemen zeigen sich kaum Abweichungen (vgl. Tabelle 3.11, S.168), während die Auswirkung der Verzerrung insbesondere bei den kointegrierten Systemen zu bemerken ist. Hier erkennt der ADF-Test noch maximal 734 stationäre Linearkombinationen. Bei den integrierten Systemen sind die Auswirkungen etwas stärker noch als in der zweiten Gruppe, so dass der Test auf dem 0.05 bei keiner Länge mehr als 950 Linearkombinationen als integriert erkennt. Und auch auf dem 0.01 Niveau gibt es bei Länge 300 erste Probleme.

Johansentest

Tabelle 3.67: Klassifikation 1000 kointegrierter Systeme auf der Ordinalskala 2

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	278	590	132	426	518	56	724	272	4	1000
	60	5	753	242	10	843	147	82	881	37	1000
	100	0	717	283	0	834	166	0	932	68	1000
	300	0	497	503	0	653	347	0	861	139	1000
(2,3)	30	348	534	117	511	433	55	761	231	7	999
	60	74	728	198	131	762	107	314	660	26	1000
	100	15	743	242	32	824	144	100	849	51	1000
	300	1	578	421	1	722	277	7	880	113	1000
(3,10)	30	420	338	80	544	262	32	713	124	1	838
	60	241	509	148	351	462	85	567	312	19	898
	100	151	581	193	226	587	112	389	491	45	925
	300	22	567	379	52	688	228	119	757	92	968

Kointegrierte Systeme Auf dieser Skala kann der Johansentest für kointegrierte und stationäre Systeme für alle bis auf ein System ausgeführt werden (Gruppe (2,3), Länge 30). In der Gruppe (3,10) ist dies nur für 838-968 Systeme der Fall (siehe Tabelle 3.67, S.240). Die Ergebnisse bei einer Länge von 30 Datenpunkten sind jedoch so schlecht, dass sie aus der folgenden Betrachtung ausgenommen sind.

In der Gruppe (1,1) wird ab einer Länge von 100 kein System als integriert missklassifiziert. Dennoch wird nur eine maximale Klassifikationsrate von 93% erreicht. Bei einer Länge von 300 Datenpunkten werden auf allen Niveaus deutlich weniger Systeme richtig klassifiziert als dies bei 100 der Fall war. In der Gruppe (2,3) werden maximal 880 kointegrierte Systeme erkannt (Länge 300, Niveau 0.01). Auch bei Länge 300 werden vereinzelt Systeme als integriert missklassifiziert; bei 100 sind dies 15-100; bei 60 sogar bis 314 Systeme. Noch schlechter ist das Klassifikationsverhalten in der Gruppe (3,10). Hier werden bei Länge 60 mehr der getesteten Systeme als integriert eingestuft denn als kointegriert. In dieser Gruppe wird eine maximale Klassifikationsrate von 78% (Länge 300, Niveau 0.01) erreicht. Es werden also insgesamt viel zu wenig kointegrierte Systeme richtig klassifiziert.

In der Gruppe (1,1) wird β um -1 herum geschätzt (siehe Tabelle 3.68, S.241), wobei sich der Median auf zwei Nachkommastellen zum unverzerrten Fall nicht verändert hat. Die Statistiken des Boxplots haben den Abstand zum Median im Vergleich vergrößert, was jedoch mit zunehmender Länge schwächer ausgeprägt ist. Bei den anderen beiden Gruppen ist die gesamte „Box“ bei allen Längen gleichmäßig verschoben. Diese Verschiebung ist als Effekt der Skala zu werten.

Tabelle 3.68: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf der Ordinalskala 2

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-175,057	-1,795	-1,201	-1,004	-0,801	-0,209	65,669	66
	60	-6,435	-1,448	-1,128	-1,010	-0,914	-0,612	-0,413	37
	100	-1,765	-1,295	-1,077	-1,004	-0,930	-0,715	-0,603	28
	300	-1,220	-1,121	-1,030	-0,999	-0,969	-0,878	-0,878	8
(2,3)	30	-5,412	-0,880	-0,629	-0,539	-0,457	-0,206	6,149	63
	60	-1,116	-0,796	-0,614	-0,548	-0,492	-0,315	1,438	48
	100	-0,951	-0,741	-0,604	-0,556	-0,513	-0,379	0,015	65
	300	-0,875	-0,727	-0,616	-0,573	-0,539	-0,434	-0,165	50
(3,10)	30	-15,046	-0,654	-0,416	-0,323	-0,256	-0,023	1,970	92
	60	-2,431	-0,598	-0,399	-0,322	-0,264	-0,068	1,826	64
	100	-1,573	-0,563	-0,392	-0,334	-0,277	-0,107	0,144	81
	300	-1,027	-0,582	-0,406	-0,341	-0,288	-0,113	0,048	87

Tabelle 3.69: Klassifikation 1000 stationärer Systeme auf der Ordinalskala 2

(a,b)	T	Niveau 0.1			Niveau 0.05			Niveau 0.01			num
		I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	
(1,1)	30	100	112	788	214	197	589	539	284	177	1000
	60	0	3	997	0	3	997	3	17	980	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	90	97	799	210	200	576	563	250	173	986
	60	0	0	1000	0	1	999	9	14	977	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	38	51	325	89	86	239	248	99	67	414
	60	0	0	416	0	1	415	3	7	406	416
	100	0	0	439	0	0	439	0	0	439	439
	300	0	0	517	0	0	517	0	0	517	517

Stationäre Systeme Die Messung auf dieser Ordinalskala wirkt sich auch auf die stationären Systeme (siehe Tabelle 3.69, S.241) stärker aus als dies bei der ersten Ordinalskala der Fall war (vgl. Tabelle 3.64, S.236). In den Gruppen (1,1) und (2,3) werden ab Länge 60 alle 1000 Systeme vom Johantest berücksichtigt; in der Gruppe (3,10) sind dies jedoch nur bei Länge 300 mehr als 50% (517 Systeme).

Die Klassifikation in den ersten beiden Gruppen ist ähnlich gut wie für die erste Ordinalskala. Nur bei Länge 60 werden nicht alle Systeme als stationär erkannt. Es gibt bis zu 23 missklassifizierte Systeme.

Auch in der Gruppe (3,10) werden nur bei Länge 60 nicht alle getesteten Systeme als stationär klassifiziert. Auf dem 0.01 Niveau sind es 8 Systeme, auf 0.05 nur eins. Die Werte für die β Schätzung liegen um Null (siehe Tabelle D.17, S.374).

Tabelle 3.70: Klassifikation 1000 integrierter Systeme auf der Ordinalskala 2

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	738	225	35	846	144	8	957	40	1	998
	60	612	336	52	747	239	14	901	97	2	1000
	100	437	438	125	576	381	43	819	178	3	1000
	300	132	480	388	226	562	212	475	473	52	1000
(2,3)	30	772	191	17	856	120	4	943	36	1	980
	60	739	241	15	839	150	6	944	51	0	995
	100	614	317	68	740	234	25	894	105	0	999
	300	298	478	224	439	458	103	675	312	13	1000
(3,10)	30	673	153	17	745	91	7	810	33	0	843
	60	676	205	28	769	134	6	854	55	0	909
	100	577	286	62	687	211	27	821	103	1	925
	300	330	408	214	428	415	109	629	296	27	952

Integrierte Systeme Bei den integrierten Systemen konnte der Johantest (siehe Tabelle 3.70, S.242) in jeder Gruppe bei mehr Systemen durchgeführt werden als auf der ersten Ordinalskala (siehe Tabelle 3.65, S.237). In der Gruppe (1,1) sind die Voraussetzungen nur bei zwei Systemen der Länge 30 nicht erfüllt; in der Gruppe (2,3) erfüllen 980-1000 Systeme die Anforderungen; in der Gruppe (3,10) kann der Test noch bei 843-952 Systemen durchgeführt werden, wobei die Anzahl mit der Länge der Reihen steigt. Auf dem 0.01 Niveau werden in allen Gruppen die meisten Systeme richtig eingestuft. Dies ist schon auf dem 0.05 Niveau in der Gruppe (1,1) nicht mehr der Fall. Bei einer Länge von 300 werden auf dem 0.1 Niveau in allen drei Gruppen mehr Systeme als kointegriert als als integriert klassifiziert. Eine Klassifikationsrate von über 95% wird nur bei einer Länge von 30 auf dem 0.01 Niveau erreicht; dies jedoch in allen Gruppen. Insgesamt kann der Johantest auf dieser Skala

jedoch nicht genug integrierte Systeme erkennen. Zur Unterscheidung von kointegrierten und integrierten Systemen ist er durch die verzerrte Messung nicht mehr in der Lage.

Ergebnis: Messung auf Ordinalskala 2

Augmented Dickey-Fuller-Test Die Verzerrung ist so groß, dass ein Test bei bekanntem Kointegrationsvektor nicht durchgeführt werden kann. In keiner Gruppe gibt es eine Länge, bei der jeweils 950 Systeme aus jeder Gruppe richtig erkannt werden. In der Gruppe (1,1) werden zwar noch fast alle stationären Linearkombinationen richtig eingestuft. Hier ist das Ergebnis für die integrierten Systeme jedoch zu gering. In den anderen beiden Gruppen ist die Erkennrate der Kointegrationsbeziehung bei allen Längen auf allen Niveaus zu gering.

Johansentest Der Johansentest erreicht auf dieser Ordinalskala nur bei den stationären Systemen die 950-Marke. Selbst in der Gruppe (1,1) können nicht genügend kointegrierte Systeme korrekt klassifiziert werden. Gleiches gilt für die integrierten Systeme. In der Gruppe (3,10) kann der Test für jede Systemgruppe nicht immer ausgeführt werden.

3.5.3 Ordinalskala 3: „wachsende Stufenabstände“

Auf dieser Skala werden auf 10 Stufen die Werte im unteren Bereich stärker differenziert als im oberen Bereich (siehe Abbildung 3.19, S.244). Durch das monotone Wachsen der Abstände kann es zu einem Deckeneffekt im oberen Wertebereich kommen, da 20% der oberen Werte auf eine gemeinsame Stufe abgebildet werden. Deshalb ist ein Vergleich mit der 10-stufigen Intervallskala und der nach oben beschränkten Skala sinnvoll.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Der ADF-Test (siehe Tabelle 3.71, S.244) zeigt in der ersten Gruppe die größten Abweichungen bei der Klassifikation im Vergleich zu den unverzerrten Daten bei den integrierten Systemen. Hier werden maximal 46 Linearkombinationen mehr als stationär eingestuft (Länge 60, Niveau 0.1). Auf dem 0.01 Niveau sind es nicht mehr als 8. Die Werte bei den integrierten Systemen sind ebenfalls schlechter als auf der 10-stufigen Intervallskala (vgl. Tabelle 3.23, S.188) oder der nach oben beschränkten Skala mit 7 Stufen (vgl. Tabelle 3.46, S.215). So erreicht der Test bei einer Länge von 300 auf dem 0.05 Niveau nun nicht mehr die 950-Marke und auch bei den kointegrierten Systemen werden bei Länge 100 auf diesem Niveau nicht mehr ausreichend viele Linearkombinationen richtig erkannt. Dennoch sind die Werte so gut, dass ein Einsatz des Tests auf dem 0.05 ab Länge 100 und auf dem 0.01 ab Länge 300 in Ordnung möglich ist.

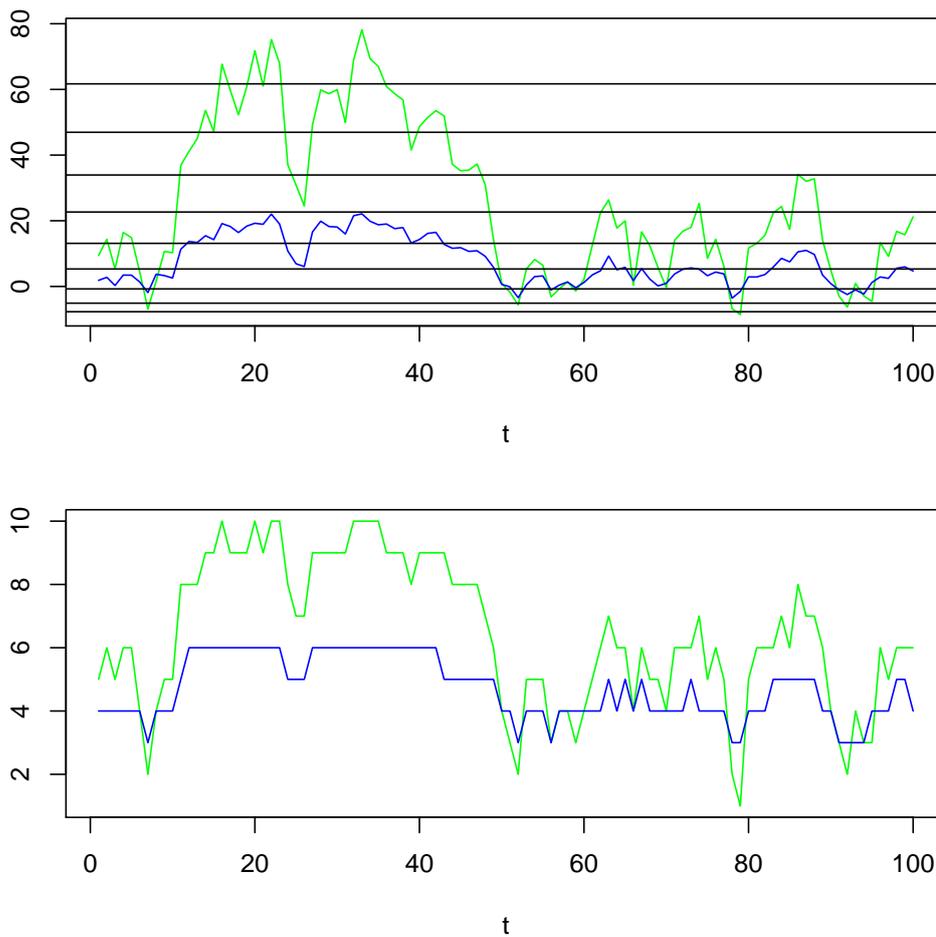


Abbildung 3.19: Konstruktion der dritten Ordinalskala (10 Stufen): Die Stufenabstände werden zunehmend größer.

Tabelle 3.71: Ergebnisse des ADF-Tests auf der Ordinalskala 3: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	325	223	71	311	189	50	87	53	14
	60	868	734	416	893	746	389	117	67	14
	100	985	941	760	991	962	771	100	42	6
	300	1000	1000	1000	1000	1000	1000	107	57	11
(2,3)	30	214	125	40	327	197	45	101	56	15
	60	593	433	179	893	754	403	102	57	14
	100	736	611	326	991	959	781	109	55	12
	300	955	926	798	1000	1000	1000	111	60	14
(3,10)	30	210	129	20	342	201	56	89	50	17
	60	499	352	127	864	736	388	100	51	12
	100	570	418	207	988	956	780	101	53	10
	300	815	722	523	1000	1000	1000	156	93	22

In der Gruppe (2,3) reagiert der Test nur bei den stationären Systemen kaum auf die Verzerrung. Die Ergebnisse sind schlechter als auf der 10-stufigen Intervallskala und der nach oben beschränkten Skala. So werden auf dem 0.05 Niveau nur noch ca. 940 integrierte Linearkombinationen erkannt. Auf dem 0.01 Niveau sind es allerdings bei allen Längen mehr als 980. Doch bei den kointegrierten Systemen betragen die Abweichungen zum unverzerrten Fall (vgl. Tabelle 3.9, S.166) bei den Längen 60 und 100 mehr als 200 Systeme und auch bei Länge 300 sind sie so groß, dass nur noch bei Länge 300 auf dem 0.1 Niveau die 950-Marke erreicht wird. Bei dieser Länge werden auf dem 0.01 Niveau weniger als 800 Linearkombinationen als stationär erkannt. Auf der 10-stufigen Intervallskala waren dies bei dieser Länge auf jedem Niveau 1000 (vgl. Tabelle 3.23, S.188). Die Fehlklassifikation bei den kointegrierten Systemen ist jedoch nicht so stark ausgeprägt, wie dies auf der nach oben beschränkten Skala der Fall war (vgl. Tabelle 3.46, S.215). Die Auswirkungen der Verzerrung sind jedoch so groß, dass der Test auf keinem Niveau in allen Gruppen so gute Werte liefert, dass er zur Unterscheidung von kointegrierten und integrierten Systemen eingesetzt werden könnte.

Der Trend aus der zweiten Gruppe setzt sich in der Gruppe (3,10) fort. Bei den stationären Systemen sind die Abweichungen des Tests vernachlässigbar (vgl. Tabelle 3.11, S.168). Bei den integrierten Systemen ist das Ergebnis bei 300 Datenpunkten etwas schlechter. Am gravierendsten sind jedoch die Unterschiede bei den kointegrierten Systemen im Vergleich zu den unverzerrten Daten. Der ADF-Test erkennt noch maximal 815 (Länge 300, Niveau 0.1) Linearkombinationen als stationär, meist jedoch weniger als die Hälfte oder knapp darüber. In dieser Gruppe (3,10) war der Effekt der 10 Stufen bereits auf der Intervallskala bei den kointegrierten Systemen zu bemerken. Die hier erzielten Werte sind immer noch etwas besser als die Ergebnisse auf der nach oben beschränkten Skala.

Johansentest

Kointegrierte Systeme Auf der dritten Ordinalskala kann der Johansentest (siehe Tabelle 3.72, S.246) nur für 23 kointegrierte Systeme der Länge 30 (20 Systeme) bzw. 60 (3 Systeme) in der Gruppe (3,10) nicht ausgeführt werden, da eine der Zeitreihen durch die simulierte Messung einen konstanten Verlauf erhält. Alle anderen Systeme jeder Länge werden in jeder Gruppe berücksichtigt. Im Allgemeinen werden mit zunehmender Länge mehr Systeme als kointegriert interpretiert. Bei einer Länge von 300 ist die Klassifikationsrate auf dem 0.01 Niveau besser als auf dem 0.05 Niveau und noch größer als auf dem 0.1 Niveau; in den Gruppen (1,1) und (2,3) trifft dies auch auf die Länge 100 zu. Hier wird die 950-Marke auf dem 0.01 Niveau erreicht (in Gruppe (2,3) bei Länge 100 mit 949 leicht verfehlt). In der Gruppe (3,10) werden auf allen Niveaus bei allen Längen viele Systeme als integriert fehlklassifiziert. Dies geschieht in den anderen Gruppen nur in wenigen Ausnahmefällen. Bei einer Länge von 300 in beiden Gruppen gar nicht mehr. Die missklassifizierten Systeme werden als stationär angesehen. In der Gruppe (3,10) kann der Johansentest auf dieser Skala nicht genug kointegrierte Systeme erkennen. In den anderen Gruppen erzielt er bei den

Tabelle 3.72: Klassifikation 1000 kointegrierter Systeme auf der Ordinalskala 3

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	258	624	118	399	541	60	689	301	10	1000
	60	5	824	171	14	881	105	84	881	35	1000
	100	0	818	182	0	892	108	0	955	45	1000
	300	0	790	210	0	873	127	0	955	45	1000
(2,3)	30	286	601	113	443	507	50	718	269	13	1000
	60	28	831	141	56	860	84	175	800	25	1000
	100	1	850	149	1	915	84	16	949	35	1000
	300	0	838	162	0	898	102	0	976	24	1000
(3,10)	30	382	486	112	509	415	56	774	197	9	980
	60	137	722	138	216	708	73	423	546	28	997
	100	84	761	155	125	796	79	230	735	35	1000
	300	22	819	159	38	865	97	65	911	24	1000

Längen 100 und 300 auf dem 0.01 Niveau ausreichend gute Werte.

Tabelle 3.73: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf der Ordinalskala 3

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-12,348	-1,691	-1,191	-0,997	-0,85	-0,343	13,841	70
	60	-2,189	-1,315	-1,085	-1,003	-0,927	-0,693	-0,361	34
	100	-1,476	-1,19	-1,050	-1,003	-0,957	-0,824	-0,745	28
	300	-1,141	-1,068	-1,017	-1,000	-0,983	-0,933	-0,913	13
(2,3)	30	-3,444	-1,025	-0,726	-0,612	-0,524	-0,227	155,644	33
	60	-1,258	-0,955	-0,714	-0,609	-0,549	-0,352	5,341	6
	100	-0,943	-0,925	-0,704	-0,609	-0,553	-0,438	-0,438	1
	300	-0,861	-0,861	-0,709	-0,614	-0,563	-0,454	-0,454	0
(3,10)	30	-1,539	-0,621	-0,378	-0,275	-0,214	0,005	1,076	62
	60	-7,132	-0,564	-0,358	-0,269	-0,220	-0,052	0,726	59
	100	-0,917	-0,558	-0,360	-0,272	-0,225	-0,050	-0,050	45
	300	-0,616	-0,557	-0,361	-0,277	-0,230	-0,101	-0,101	24

In der Gruppe (1,1) wird β im Bereich um -1 geschätzt (siehe Tabelle 3.73, S.246), mit einer Abweichung im Median von weniger als zwei Nachkommastellen im Vergleich zum unverzerrten Fall. Die Streuung ist etwas größer, nimmt mit der Länge der Zeitreihen jedoch ab. Ebenso wird die Abweichung mit zunehmender Länge geringer. In den anderen beiden Gruppen kommt es zu einer verschobenen Schätzung. In der Gruppe (2,3) liegt der Median um ca. 0,05 auf -0,6 verschoben. Die Boxplot-Statistiken haben diese Verschiebung mitvollzogen, ihr Abstand zum Median hat sich jedoch zusätzlich vergrößert. Gleiches gilt für die Gruppe

(3,10). Hier wird der Median nun bei allen Längen auf etwas -0,27 geschätzt.

Tabelle 3.74: Klassifikation 1000 stationärer Systeme auf der Ordinalskala 3

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	92	96	812	207	181	612	552	274	174	1000
	60	0	2	998	0	3	997	5	9	986	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	88	96	816	205	205	590	555	260	185	1000
	60	0	0	1000	0	2	998	3	9	988	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	90	108	787	228	198	559	556	249	180	985
	60	0	0	998	0	1	997	2	9	987	998
	100	0	0	999	0	0	999	0	0	999	999
	300	0	0	1000	0	0	1000	0	0	1000	1000

Stationäre Systeme Für stationäre Systeme (siehe Tabelle 3.74, S.247) konnte der Johansen-Test nur in der Gruppe (3,10) für 18 Systeme nicht ausgeführt werden; 15 bei Länge 30 und 2 bzw. 1 bei den Längen 60 und 100.

Bei den Längen 100 und 300 werden alle getesteten Systeme in allen Gruppen als stationär klassifiziert. Nur auf dem 0.01 Niveau gibt es in allen Gruppen auch als integriert missklassifizierte Systeme bei Länge 60. Es werden in keiner Gruppe jedoch mehr als 10 Systeme falsch eingestuft. Die Schätzungen für den Kointegrationskoeffizienten liegen um Null (siehe Tabelle D.19, S.376).

Integrierte Systeme Bei den integrierten Systemen in der Gruppe (1,1) hat nur ein System der Länge 30 die Anforderungen des Johansen-Test nicht erfüllt (siehe Tabelle 3.75, S.248). In der Gruppe (2,3) waren es bei dieser Länge 11 und bei allen anderen Längen 4 Systeme. In der Gruppe (3,10) wurde der Test bei 837-944 Systemen durchgeführt. Mit zunehmender Länge werden immer weniger Systeme richtig als integriert erkannt. Die Klassifikationsrate ist in allen Gruppen auf dem 0.01 Niveau besser als auf den anderen beiden, wie es auch bei den unverzerrten Daten der Fall war. In allen Gruppen wird auf dem 0.01 bei allen Länge bis auf 300 eine Klassifikationsrate von mehr als 90% erreicht. Aber nur in der Gruppe (2,3) übersteigt sie auch die 95%-Marke. Die Klassifikationsergebnisse sind zwar besser als auf den anderen Ordinalskalen, insgesamt jedoch nicht zuverlässig genug. Die β Schätzwerte streuen um Null.

Tabelle 3.75: Klassifikation 1000 integrierter Systeme auf der Ordinalskala 3

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	768	206	25	857	133	9	965	32	2	999
	60	724	243	33	816	173	11	943	55	2	1000
	100	633	319	48	745	238	17	907	92	1	1000
	300	447	469	84	561	409	30	755	242	3	1000
(2,3)	30	778	190	21	862	118	9	948	41	0	989
	60	820	162	14	887	105	4	964	32	0	996
	100	762	212	22	845	147	4	952	43	1	996
	300	602	366	28	703	287	6	859	135	2	996
(3,10)	30	672	152	13	744	86	7	809	28	0	837
	60	708	165	16	780	104	5	855	34	0	889
	100	636	247	21	732	162	10	845	56	3	904
	300	501	392	51	598	333	13	739	204	1	944

Ergebnis: Messung auf Ordinalskala 3

Augmented Dickey-Fuller-Test Die hier beobachteten Effekte sind eher auf die Gewichtung der Werte als auf die Anzahl der Stufen zurückzuführen. Jedoch ist der Ceilingeffekt in den Gruppen (2,3) und (3,10) so stark, dass der Test nicht verwendet werden kann. Die Ergebnisse in der Gruppe (1,1) sind noch relativ gut, erfüllen das Kriterium von 950 richtig klassifizierten Systemen für alle Systemgruppen jedoch nur bei Länge 300 auf dem 0.01 Niveau.

Johansen-Test Auf dieser Ordinalskala schneidet der Johansen-Test viel besser ab als auf den in den vorigen Simulationsstudien betrachteten Ordinalskalen. Der Test kann deutlich häufiger ausgeführt werden und auf dem 0.01 Niveau zeigt er bei Länge 100 bei allen Systemarten gute Klassifikationsergebnisse, erreicht das geforderte Kriterium jedoch nicht ganz. In der Gruppe (2,3) fehlt dafür jedoch nur ein kointegriertes System.

3.5.4 Ordinalskala 4: „unsystematisch variiertes Stufenabstand“

Bei der Konstruktion der vierten Ordinalskala gibt es zwei Bereiche, in denen die Differenzierung sehr fein ist. Dafür werden an den Randbereichen jeweils 20% der Werte auf eine Stufe abgebildet (siehe Abbildung 3.20, S.249). Da es 10 Stufen gibt, bietet sich ein Vergleich mit der 10 stufigen Intervallskala und der eben betrachteten Ordinalskala mit 10 Stufen, sowie wegen der zu erwarteten Randeffekte mit der von beiden Seiten beschränkten Skala an.

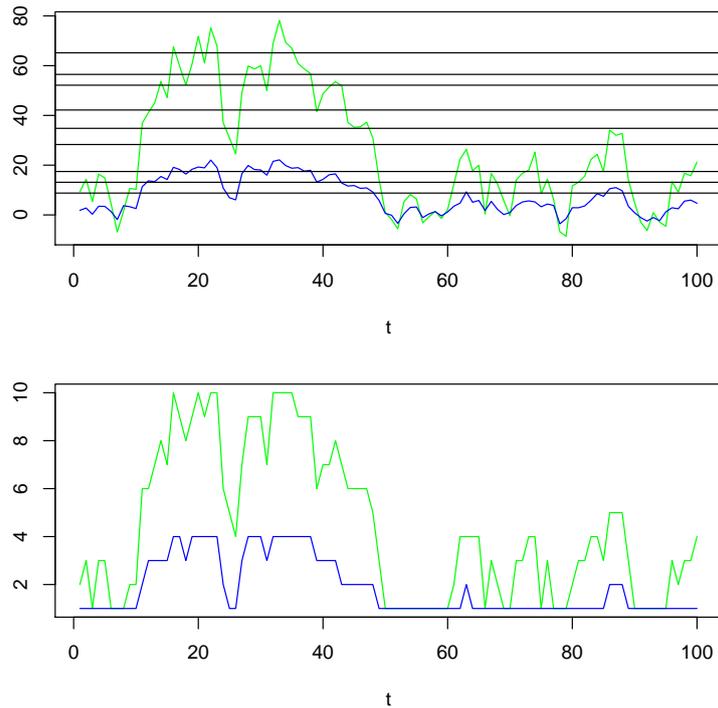


Abbildung 3.20: Konstruktion der vierten Ordinalskala (10 Stufen): Die Randbereiche werden kaum differenziert, im mittleren Bereich sind die Stufen unterschiedlich groß gewählt.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Der ADF-Test (siehe Tabelle 3.76, S.250) zeigt in der **Gruppe (1,1)** kaum Abweichungen. Die maximale Abweichung von den unverzerrten Daten liegt bei 23 integrierten Systeme der Länge 100 auf dem 0.1 Niveau (vgl. Tabelle 3.7, S.164). Bei den kointegrierten Systemen verfehlt er bei Länge 100 auf 0.05 Niveau nur knapp die 950-Marke (946), die er bei den stationären und integrierten Systemen erreicht. Er kann in dieser Gruppe ab einer Länge von 100 auf dem 0.05 und ab 300 auch auf dem 0.01 Niveau eingesetzt werden, ähnlich wie im unverzerrten Fall. Die Ergebnisse sind vergleichbar mit denen auf der anderen 10-stufigen Ordinalskala (vgl. Tabelle 3.71, S.244).

Ähnlich wie auf der ersten 10-stufigen Ordinalskala verändert sich das Verhalten des Tests in der **Gruppe (2,3)** nur für stationäre Systeme nicht. Die Auswirkungen der Verzerrung sind bei den kointegrierten Systemen deutlicher zu bemerken als in der Gruppe (1,1). Der ADF-Test reagiert in dieser Gruppe bei den kointegrierten Systemen noch stärker als in der Gruppe (1,1). So erkennt er auch bei Länge 300 nicht alle 1000 stationären Linearkombinationen. Dennoch sind es hier mehr als 940, was auf der ersten Ordinalskala nicht der Fall war. Doch die Werte bei Länge 100 sind nicht ausreichend, um den Test zur Unterscheidung von kointegrierten und integrierten Systemen einzusetzen, obwohl die Ergebnisse für integrierte Systeme trotz Verzerrung sehr gut sind. Auch sind sie viel besser als auf der von beiden Seiten beschränkten Skala.

Tabelle 3.76: Ergebnisse des ADF-Tests auf der Ordinalskala 4: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	313	183	52	333	195	48	97	39	6
	60	890	757	422	881	752	420	88	44	5
	100	988	946	763	990	959	766	101	45	8
	300	1000	1000	1000	1000	1000	1000	87	50	11
(2,3)	30	282	153	41	322	186	55	68	32	8
	60	812	665	329	886	754	393	95	40	10
	100	929	858	588	985	952	776	90	44	8
	300	996	986	945	1000	1000	1000	90	45	7
(3,10)	30	187	112	26	339	203	50	104	55	8
	60	529	381	155	864	728	399	102	54	7
	100	623	477	238	979	952	771	97	57	15
	300	818	727	532	1000	1000	1000	118	61	18

In der **Gruppe (3,10)** sind die Werte des Tests so schlecht, dass er nicht zuverlässig zwischen kointegrierten und integrierten Systemen unterscheiden kann. Wie in den anderen beiden Gruppen reagiert er kaum auf die Verzerrung bei den stationären Systemen. Die Einflüsse bei den integrierten und kointegrierten Systemen sind jedoch deutlich größer als in den anderen Gruppen. Der ADF-Test erreicht hier gerade einmal 818 richtig klassifizierte Kointegrationsbeziehungen und ist bei den anderen Längen deutlich schlechter. Bei Länge 100 wird auf dem 0.05 Niveau nicht mal die Hälfte aller stationären Linearkombinationen erkannt. Bei den integrierten Systemen ist die Abweichung so groß, dass die 950-Marke auf dem 0.05 Niveau nicht mehr erreicht wird. Zwar erzielt er bei den integrierten Systemen auf dem 0.01 Niveau Werte über 980 und auch auf dem 0.05 mehr als 940. Aber die Ergebnisse für die kointegrierten Systeme sind so schlecht, dass er in dieser Gruppe nicht eingesetzt werden kann.

Johansentest

Kointegrierte Systeme Der Johansentest konnte für alle kointegrierten Systeme bis auf 14 der Länge 30 in der Gruppe (3,10) ausgeführt werden (siehe Tabelle 3.77, S.251). Auch auf dieser Skala waren die Werte der Systeme mit 30 Datenpunkten nicht ausreichend gut, um sie in der Beschreibung der Ergebnisse zu erwähnen.

In den Gruppen (2,3) und (3,10) steigt die Anzahl der richtig klassifizierten Systeme mit der Menge der Beobachtungspunkte. Dieser Effekt ist in der Gruppe (1,1) nicht zu erkennen. Dafür ist die Klassifikation in dieser Gruppe auf dem 0.01 Niveau besser als auf 0.05 Niveau und noch besser als auf dem 0.1 Niveau. Diese Systematik ist in den anderen beiden Gruppen

Tabelle 3.77: Klassifikation 1000 kointegrierter Systeme auf der Ordinalskala 4

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	268	633	99	426	533	41	722	271	7	1000
	60	2	876	122	9	919	72	63	920	17	1000
	100	0	863	137	0	931	69	0	979	21	1000
	300	0	863	137	0	923	77	0	988	12	1000
(2,3)	30	277	633	90	446	518	36	726	268	6	1000
	60	11	899	90	33	921	46	150	839	11	1000
	100	1	910	89	5	955	40	18	972	10	1000
	300	0	912	88	0	965	35	0	997	3	1000
(3,10)	30	460	449	77	580	374	32	807	177	2	986
	60	213	707	80	312	647	41	511	482	7	1000
	100	99	821	80	156	807	37	301	689	10	1000
	300	15	903	82	30	934	36	68	927	5	1000

nicht bei allen Längen zu sehen. Beide Effekte waren in allen Gruppen auf allen Niveaus im unverzerrten Fall vorhanden (vgl. Tabellen 3.12, S.170, 3.13, S.171 und 3.14, S.171).

In der Gruppe (1,1) gibt es allgemein kaum Abweichungen zum Klassifikationsverhalten im unverzerrten Fall (vgl. Tabelle 3.12, S.170). Nur bei Länge 30 auf dem 0.01 Niveau gibt es mehr als 11 anders klassifizierte Systeme. In der Gruppe (2,3) sind es auf dem 0.01 mehr als 80, auf den anderen Niveaus bis 41 bzw. 27 bei Länge 30 und immer noch 1-27 bei den anderen Längen. Die Abweichungen in der Gruppe (3,10) betragen zum Teil mehrere hundert Systeme. Sie sind bei Länge 300 auf dem 0.1 Niveau mit 15 Systemen am geringsten und bei Länge 60 auf dem 0.01 Niveau mit 440 am größten. Hier werden auch mehr Systeme als integriert eingestuft als als kointegriert. Die Werte sind insgesamt besser als auf der dritten Ordinalskala (vgl. Tabelle 3.77, S.251).

In den Gruppen (1,1) und (2,3) werden ab Länge 100 auf dem 0.01 und in Gruppe (2,3) auch auf dem 0.05 Niveau sehr gute Klassifikationswerte erreicht.

Die Schätzung für β (siehe Tabelle 3.78, S.252) haben sich in der Gruppe (1,1) auf zwei Nachkommastellen nicht verändert. Dies gilt auch für die Boxplot-Statistiken⁴². In den anderen beiden Gruppen hat sich der Median an der zweiten Nachkommastelle bei allen Längen konstant verschoben. In Gruppe (2,3) wird er nun auf -0,68, in (3,10) auf -0,31 geschätzt. Die Boxplot-Statistiken haben sich dementsprechend verschoben.

Stationäre Systeme Der Test wurde für alle stationären Systeme durchgeführt. Auf dem 0.1 Niveau wurden in allen Gruppen alle Systeme ab einer Länge von 60 als stationär erkannt.

⁴²Ausnahme: Oberer Whisker hat sich bei Länge 60 um -0,013 verschoben.

Tabelle 3.78: Boxplot-Statistiken für die Schätzung von β bei kointegrierten Systemen auf der Ordinalskala 4

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-10,280	-1,497	-1,139	-1,004	-0,900	-0,545	8,788	78
	60	-1,688	-1,205	-1,054	-1,004	-0,954	-0,804	-0,406	53
	100	-1,361	-1,122	-1,034	-1,002	-0,975	-0,888	-0,628	42
	300	-1,091	-1,039	-1,009	-0,999	-0,989	-0,960	-0,922	34
(2,3)	30	-303,681	-0,871	-0,740	-0,696	-0,651	-0,528	429,537	66
	60	-1,136	-0,793	-0,714	-0,685	-0,661	-0,582	-0,465	23
	100	-0,855	-0,764	-0,705	-0,683	-0,663	-0,607	-0,574	11
	300	-0,774	-0,738	-0,692	-0,674	-0,661	-0,615	-0,612	9
(3,10)	30	-3,697	-0,482	-0,360	-0,321	-0,277	-0,155	3,407	75
	60	-1,519	-0,439	-0,344	-0,314	-0,280	-0,186	0,361	67
	100	-2,723	-0,438	-0,346	-0,314	-0,281	-0,187	-0,132	36
	300	-0,491	-0,432	-0,340	-0,314	-0,278	-0,189	-0,127	19

Tabelle 3.79: Klassifikation 1000 stationärer Systeme auf der Ordinalskala 4

(a,b)	T	Niveau 0.1			Niveau 0.05			Niveau 0.01			num
		I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	
(1,1)	30	89	84	827	204	194	602	530	271	199	1000
	60	0	0	1000	0	0	1000	4	9	987	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	84	97	819	209	197	594	542	277	181	1000
	60	0	0	1000	0	1	999	7	10	983	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	88	97	815	204	200	596	548	255	197	1000
	60	0	0	1000	0	0	1000	8	9	983	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

In den Gruppen (1,1) und (3,10) trifft dies auch auf dem 0.05 Niveau zu. In der Gruppe (2,3) wird hier ein System als kointegriert klassifiziert. Auf dem 0.01 Niveau kommt es in allen Gruppen bei Länge 60 zu einigen (9-11) Fehlklassifikationen. Auf den Längen 100 und 300 werden alle Systeme als stationär klassifiziert. Dies Ergebnis ist minimal besser als auf der dritten Ordinalskala (vgl. Tabelle 3.79, S.252). β wird um Null herum geschätzt (vgl. Tabelle D.20, S.377).

Tabelle 3.80: Klassifikation 1000 integrierter Systeme auf der Ordinalskala 4

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	788	187	24	866	124	9	961	38	0	999
	60	745	242	13	855	141	4	960	39	1	1000
	100	704	271	25	812	180	8	929	71	0	1000
	300	626	335	39	729	260	11	879	119	2	1000
(2,3)	30	777	192	15	858	120	6	953	31	0	984
	60	839	146	9	909	81	4	969	25	0	994
	100	794	192	11	872	123	2	960	37	0	997
	300	755	227	16	822	174	2	929	69	0	998
(3,10)	30	642	152	11	705	97	3	775	29	1	805
	60	680	175	16	744	118	9	821	50	0	871
	100	626	229	18	703	165	5	800	72	1	873
	300	519	370	34	621	294	8	734	188	1	923

Integrierte Systeme In der Gruppe (1,1) konnte nur bei einem System der Länge 30 der Johansentest nicht ausgeführt werden. In den anderen Gruppen konnten 984-998 bzw. 805-923 Systeme verwendet werden (siehe Tabelle 3.80, S.253).

Auf dem 0.01 Niveau werden nur vereinzelt 1-2 Systeme als stationär fehlklassifiziert; auf dem 0.05 sind es nicht mehr als 11 und auch auf dem 0.1 Niveau werden nie mehr als 39, meist weniger als 20 Systeme für stationär gehalten. Entscheidend ist jedoch, ob der Test zwischen integrierten und kointegrierten Systemen unterscheiden kann. Für kointegrierte Systeme waren die Ergebnisse in den ersten beiden Gruppen dafür geeignet.

Wenn man die Systeme der Länge 30 nicht berücksichtigt, werden stets immer weniger Systeme als integriert klassifiziert, je länger die Beobachtungszeit wird. So wird die 95% Marke in der Gruppe (1,1) nur bei der Länge 60, in der Gruppe (2,3) bei 60 und 100 auf dem 0.01 erreicht und in der Gruppe (3,10) bei Länge 60 mit 94,26% knapp verfehlt. Auf den anderen Niveaus sind die Klassifikationsraten stets unter 90%⁴³.

Die Unterscheidung zwischen integrierten und kointegrierten Systemen liegt auch auf dem 0.01 Niveau nicht bei der verlangten Qualität. Auch hier streuen die Werte der Schätzungen

⁴³Ausnahme: 91,45% in Gruppe (2,3) bei Länge 60 auf 0.05 Niveau.

von β um Null. Nur in der Gruppe (1,1) liegt der Median bei einer Länge von 100 bei 0.13 (siehe Tabelle D.20, S.377).

Ergebnis: Messung auf Ordinalskala 4

Augmented Dickey-Fuller-Test In der Gruppe (1,1) kann der ADF-Test auf dieser Ordinalskala bei Systemen ab 100 Datenpunkten eingesetzt werden. In der Gruppe (2,3) sind die Ergebnisse bei 300 Beobachtungen ebenfalls noch ausreichend gut. Die stationäre Linearkombination der kointegrierten Systeme in Gruppe (3,10) gelingt jedoch nicht mehr.

Johansen-Test Der Johansen-Test erkennt auch auf dieser Ordinalskala die stationären Systeme jeder Gruppe ab einer Länge von 60. Für sie kann er auch stets ausgeführt werden. Dies gilt auch für fast alle kointegrierten Systeme. Hier ist die Klassifikation jedoch nur in den Gruppen (1,1) und (2,3) und nur auf dem 0.01 Niveau ausreichend gut. Für integrierte Systeme wird die 950-Marke jedoch nur in der Gruppe (2,3) bei Länge 100 ebenfalls überschritten, so dass der Test in der Gruppe (1,1) das gesetzte Erfolgskriterium knapp verfehlt.

3.5.5 Ergebnis: Messung auf Ordinalskalenniveau

Augmented Dickey-Fuller-Test Die Verzerrung durch die konstruierten Ordinalskalen war in allen Fällen so groß, dass der Test in keiner Gruppe bedenkenlos zur Unterscheidung von kointegrierten und integrierten Systemen eingesetzt werden kann. Die Auswirkungen waren stets in der Gruppe (3,10) am stärksten. Auch in der Gruppe (1,1) wurde die 950-Marke nicht bei allen Systemgruppen gleichzeitig erreicht. Dennoch lag der ADF-Test meist nur knapp unter diesem Richtwert. Nur bei Reihen mit gleicher Varianz könnte der Test überhaupt eingesetzt werden, aber auch hier waren die Simulationsergebnisse meist unter den gesetzten Mindestanforderungen.

Johansen-Test Für den Johansen-Test zeigt sich ein deutlicher Unterschied zwischen den ersten beiden Ordinalskalen, die die Randbereich nicht ausreichend gut differenzieren und den anderen beiden 10-stufigen Ordinalskalen. Durch die Decken- und Bodeneffekte verliert man zu viele Informationen, um integrierte und kointegrierte Systeme unterscheiden zu können. Durch die Beschränkung des Wertebereichs gelingt die Klassifikation der Systeme mit 300 Datenpunkten schlechter als für kürzere Systeme.

Durch die Simulationsstudien mit diesen Ordinalskalen bestätigen sich vor allem die Ergebnisse der vorigen Studien zu Stufenanzahl und Ceilingeffekten. Die Ergebnisse auf den letzten beiden Ordinalskalen lassen vermuten, dass ein Einsatz von Ordinalskalen dann möglich ist, wenn die Randbereiche ausreichend gut abgedeckt werden.

3.6 Missing Values

Da man bei der Erhebung der Daten in der Psychologie meist auf das Mitwirken der Versuchspersonen angewiesen ist, hat man häufig das Problem, dass nicht für alle Messzeitpunkte Daten vorliegen. Da die Schätzverfahren fehlende Werte jedoch nicht verarbeiten können und in der Implementierung in Softwarepaketen kein Standardvorgehen mit dem Umgang solcher *missing values* enthalten ist, müssen wir die Frage klären, ob Kointegrationstests auch noch eingesetzt werden können, wenn nicht über den gesamten Messzeitraum Daten vorliegen.

Im Kapitel 3.2 (ab S.160) haben wir gesehen, dass zuverlässige Ergebnisse der Kointegrationsverfahren erst bei einer Länge von ca. 100 Messzeitpunkten gegeben sind. Bei täglicher Messung ist also über einen Messzeitraum von etwa drei Monaten zu erheben. Auf wie viele Messwerte in diesem Zeitraum könnte man eventuell verzichten, wenn sie nicht vorliegen, ohne die Erhebung von Neuem beginnen zu müssen. Kann Kointegrationsanalyse auch dann eingesetzt werden, wenn es nicht möglich ist wirklich täglich zu messen? Kann beispielsweise eine Messung wöchentlich (z.B. sonntags) ausgelassen werden, ohne die Auswertungsmöglichkeiten einzuschränken?

Wir werden anhand von Simulationsstudien diese Fragen klären, indem wir aus den vorliegenden Systemen in einer ersten Simulationsserie in regelmäßigen Abständen den jeweils zehnten, siebten, fünften oder jeden dritten Messwert entfernen und die Ergebnisse der Kointegrationstests mit denen an den vollständigen Datensätzen aus Kapitel 3.2 vergleichen. Dabei werden aus beiden Zeitreihen jeweils zu denselben Zeitpunkten Messwerte entfernt, um eine zeitliche Verschiebung der beiden Reihen gegeneinander zu vermeiden. Wenn in der Praxis nur in einer Zeitreihe ein Messpunkt fehlt, muss aus diesem Grund der Messwert desselben Tages aus der zweiten Reihe ebenfalls entfernt werden.

In einer zweiten Serie werden dann an zufälligen Zeitpunkten Messwerte aus den Systemen entfernt. Um Vergleichbarkeit mit der ersten Serie zu erzielen, wird jeweils dieselbe Anzahl von Messzeitpunkten entfernt, also zunächst ein Zehntel, dann ein Siebtel, ein Fünftel, bis in der letzten Studie nur noch zwei Drittel der Messzeitpunkte berücksichtigt werden.

Implizit wird durch diese Vorgehen auch die Frage der Äquidistanz zwischen Messungen beantwortet, da sich durch das regelmäßige oder zufällige Entfernen einiger Messwerte verschieden große zeitliche Abstände zwischen den vorhandenen Messwerten ergeben. Da durch das Entfernen einiger Werte der Wertebereich selbst nicht verändert wird, kann der Johansen-Test in jeder Gruppe für alle 1000 Systeme durchgeführt werden.

3.6.1 Regelmäßiges Entfernen jeder 10. Messung

In der ersten Simulationsstudie wird regelmäßig jeder 10. Beobachtungspunkt entfernt, was in Abbildung 3.21 auf Seite 257 dargestellt ist. Dadurch haben die Reihen nicht mehr die

Längen 30, 60, 100 bzw. 300, sondern nun noch 27, 54, 90 bzw. 270. Der Wertebereich wird dabei nicht verändert. Wir messen also auf einer Verhältnisskala mit Wertebereich \mathbb{R} .

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Die Ergebnisse des Augmented Dickey-Fuller-Test unterscheiden sich in den drei Gruppen kaum (siehe Tabelle 3.81 auf Seite 256).

Tabelle 3.81: Ergebnisse des ADF-Tests nach Entfernung jeder 10. Messung: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	428	248	73	417	263	86	104	45	10
	60	812	654	311	812	653	295	72	34	7
	100	966	901	629	968	912	637	84	36	2
	300	1000	1000	1000	1000	1000	999	69	41	6
(2,3)	30	435	272	72	406	251	60	87	45	10
	60	806	650	313	804	630	292	74	32	4
	100	961	890	631	971	899	628	84	45	6
	300	1000	1000	1000	1000	1000	1000	72	36	7
(3,10)	30	423	262	82	402	251	73	88	46	9
	60	796	641	298	799	634	305	77	35	6
	100	961	894	640	971	908	628	81	40	6
	300	1000	1000	1000	1000	1000	1000	76	37	8

Bei den kointegrierten und stationären Systemen werden alle bis auf eine Linearkombination in der Gruppe (1,1) auf dem 0.01 Niveau bei einer Länge von ursprünglich 300 Datenpunkten erkannt. Bei nun 90 Beobachtungen werden auf dem 0.1 Niveau mehr als 960, auf dem 0.05 Niveau ca. 900 und auf dem 0.01 Niveau etwa 630 stationäre Linearkombinationen richtig eingeschätzt. Bei den Systemen der Länge 60 sind es um die 800, 640 bzw. 300 Linearkombinationen. Bei den kürzesten betrachteten Systemen werden auf keinem Niveau mehr als die Hälfte richtig klassifiziert und auf dem 0.01 Niveau sogar weniger als 100.

Die Erkennungsrate bei den integrierten Systemen ist stärker vom Signifikanzniveau als der betrachteten Länge abhängig. Auf dem 0.01 Niveau werden maximal 10 Linearkombinationen für stationär gehalten. Auf dem 0.05 Niveau sind es etwa 4% und auf dem 0.1 Niveau 7-9%.

Johansenstest

Kointegrierte Systeme Die Klassifikation der kointegrierten Systeme (siehe Tabelle 3.82, S.258) ist bei einer Länge von nun 27 Datenpunkten zu schlecht, um den Test sinnvoll einset-

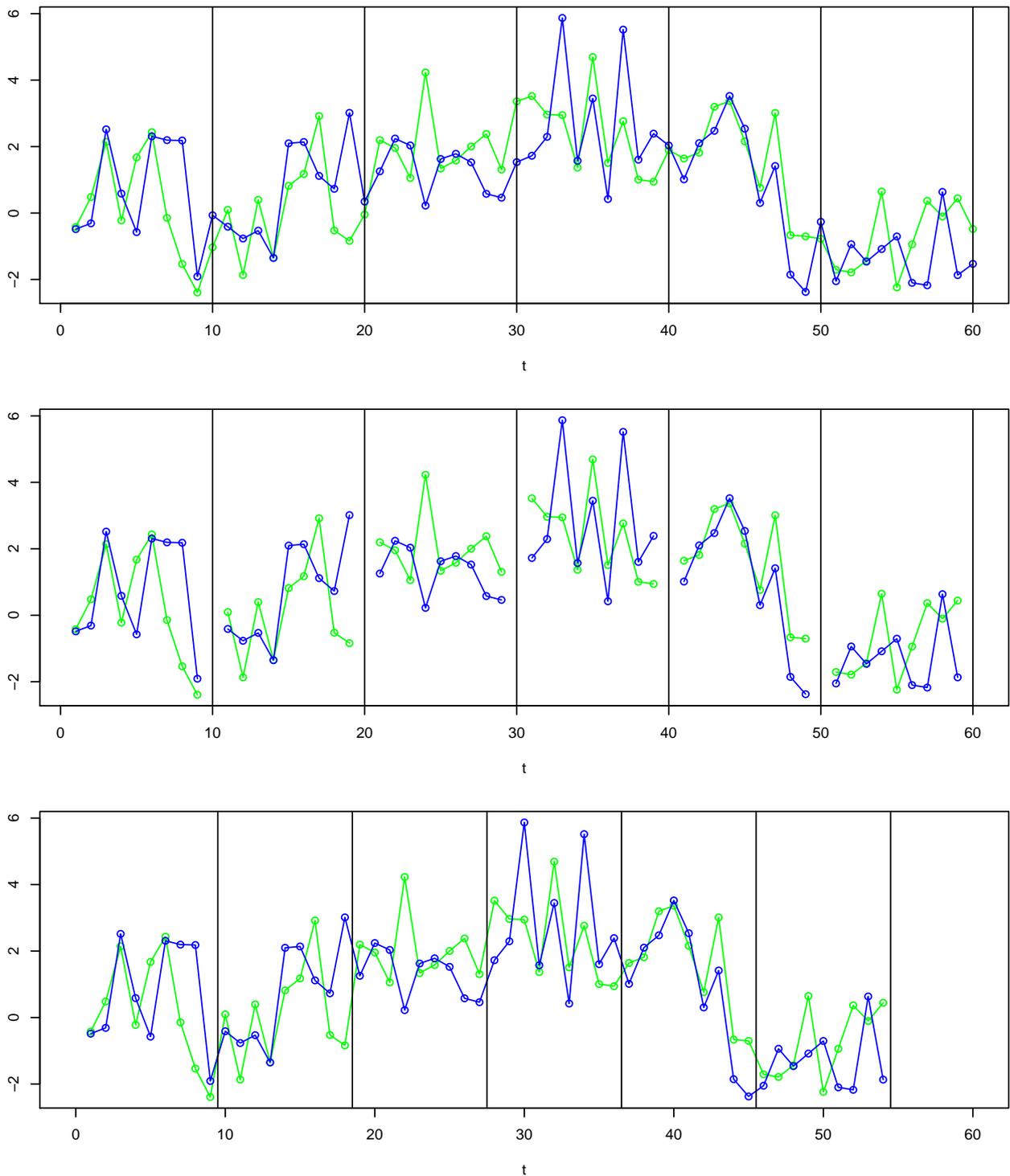


Abbildung 3.21: Entfernen jedes 10. Messwertes, hier am Beispiel eines kointegrierten Systems der Länge 60 aus der Gruppe (1,1). Oben: In der vollständigen Zeitreihe werden die zu entfernenden Messpunkte markiert. Mitte: Die Messwerte wurden entfernt. Unten: Zeitreihe nach Entfernen der Messwerte, markiert sind die Stellen fehlender Werte. Dadurch verkürzt sich die Zeitreihe (hier auf 54 Punkte).

Tabelle 3.82: Klassifikation 1000 kointegrierter Systeme nach Entfernen jeder 10. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	357	562	81	493	472	35	797	201	2	1000
	60	4	874	122	20	914	66	140	846	14	1000
	100	0	877	123	0	931	69	1	984	15	1000
	300	0	885	115	0	936	64	0	990	10	1000
(2,3)	30	342	582	76	491	482	27	783	214	3	1000
	60	5	892	103	24	923	53	149	837	14	1000
	100	0	904	96	0	945	55	0	986	14	1000
	300	0	917	83	0	960	40	0	994	6	1000
(3,10)	30	310	609	81	499	478	23	785	214	1	1000
	60	5	895	100	28	921	51	136	848	16	1000
	100	0	903	97	0	948	52	0	983	17	1000
	300	0	914	86	0	959	41	0	994	6	1000

zen zu können. Während auf dem 0.1 Niveau noch die Mehrheit der Systeme als kointegriert erkannt wird, werden auf dem 0.05 Niveau schon mehr Systeme als integriert eingestuft. Auf dem 0.01 Niveau ist der gemeinsame Trend nur bei ca. 200 Systemen zu erkennen.

Ab einer ursprünglichen Länge von 60 Beobachtungen unterscheiden sich die Ergebnisse kaum von den Originaldaten und sind zum Teil sogar etwas besser (vgl. Tabellen 3.12, S.170, 3.13, S.171 und 3.14, S.171). Bei einer Länge von 60 werden in jeder Gruppe auf allen Niveaus deutlich mehr als 800 Systeme als kointegriert erkannt. Auf dem 0.01 Niveau werden bei 300 Datenpunkten in allen drei Gruppen 99% richtig eingestuft. Auf diesem Niveau sind es bei Länge 100 (nun 90) noch mehr als 98%. Auf dem 0.05 werden bei beiden Längen in allen Gruppen mehr als 930 Systeme, auf dem 0.1 Niveau nur in der Gruppe (1,1) weniger als 900 korrekt klassifiziert.

Die Schätzungen von β (Tabelle 3.83) sind in allen Gruppen bis auf die erste Nachkommastelle (oft sogar bis auf die zweite) nicht von den Statistiken im Fall mit allen Daten zu unterscheiden. Es wird also nicht nur erkannt, dass die Systeme kointegriert sind, sondern die Kointegrationsbeziehung wird auch sehr genau geschätzt.

Stationäre Systeme Bei den stationären Systemen (siehe Tabelle 3.84, S.259) gibt es nur bei Länge 30 einen deutlichen Unterschied zu den Klassifikationsergebnissen mit vollem Datenumfang (vgl. Tabellen 3.16, S.174, 3.17, S.175 und 3.18, S.175). Hier werden auf allen Niveaus zu wenig Systeme richtig klassifiziert. Für die Längen 90 und 270 ist die Klassifikation auf allen Niveaus perfekt. Bei 54 Datenpunkten werden nur auf dem 0.01 Niveau weniger als 99%, nämlich 941-961 richtig eingestuft.

Tabelle 3.83: Schätzung für β bei kointegrierten Systemen nach Entfernung jeder 10. Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-57,809	-1,527	-1,142	-0,995	-0,882	-0,494	3,958	79
	60	-1,968	-1,211	-1,059	-1,007	-0,952	-0,796	-0,468	53
	100	-1,430	-1,124	-1,034	-1,003	-0,974	-0,885	-0,679	36
	300	-1,071	-1,037	-1,009	-0,999	-0,990	-0,963	-0,932	34
(2,3)	30	-147,920	-0,809	-0,703	-0,666	-0,632	-0,530	136,306	93
	60	-8,342	-0,723	-0,680	-0,667	-0,651	-0,608	-0,525	61
	100	-0,751	-0,698	-0,675	-0,666	-0,658	-0,635	-0,578	54
	300	-0,702	-0,677	-0,670	-0,667	-0,664	-0,656	-0,648	29
(3,10)	30	-1,829	-0,338	-0,310	-0,300	-0,292	-0,265	0,099	85
	60	-0,347	-0,315	-0,304	-0,300	-0,296	-0,284	-0,255	43
	100	-0,316	-0,309	-0,302	-0,300	-0,298	-0,291	-0,283	41
	300	-0,306	-0,303	-0,301	-0,300	-0,299	-0,297	-0,295	45

Tabelle 3.84: Klassifikation 1000 stationärer Systeme nach Entfernen jeder 10. Messung

(a,b)	T	Niveau 0.1			Niveau 0.05			Niveau 0.01			num
		I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	
(1,1)	30	148	167	685	300	258	442	640	247	113	1000
	60	0	0	1000	2	1	997	16	23	961	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	150	147	703	317	245	438	647	249	104	1000
	60	1	1	998	1	4	995	29	29	942	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	168	137	695	305	251	444	663	237	100	1000
	60	0	0	1000	0	3	997	23	36	941	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Tabelle 3.85: Klassifikation 1000 integrierter Systeme nach Entfernen jeder 10. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	802	180	18	884	115	1	969	31	0	1000
	60	822	169	9	891	105	4	971	29	0	1000
	100	795	188	17	889	108	3	968	32	0	1000
	300	829	158	13	898	98	4	971	29	0	1000
(2,3)	30	831	158	11	914	83	3	980	19	1	1000
	60	874	120	6	940	58	2	987	13	0	1000
	100	866	125	9	931	69	0	991	9	0	1000
	300	903	88	9	959	40	1	992	8	0	1000
(3,10)	30	835	154	11	895	98	7	971	29	0	1000
	60	883	111	6	942	56	2	984	16	0	1000
	100	877	113	10	936	63	1	992	8	0	1000
	300	907	86	7	954	44	2	992	8	0	1000

Integrierte Systeme Auch die Klassifikation der integrierten Systeme (siehe Tabelle 3.85, S.260) hat sich durch den Datenverlust kaum verändert (vgl. Tabellen 3.19, S.177, 3.20, S.178 und 3.21, S.179). Manche Ergebnisse sind nach dem Entfernen von 10% der Daten sogar etwas besser (z.B. auf Länge 100 und 300 in Gruppe (1,1)). In allen Gruppen werden auf dem 0.01 Niveau bei allen Längen mehr als 969, in den Gruppen (2,3) und (3,10) eher 98% der Systeme richtig als integriert erkannt. Auf dem 0.05 Niveau sind es in der ersten Gruppe ca. 89%, in den anderen beiden Gruppen ca. 93%. Auf dem 0.1 Niveau werden in allen Gruppen bei allen Längen noch 800-900 Systeme richtig erkannt.

Ergebnis: Regelmäßiges Entfernen jeder 10. Messung

Augmented Dickey-Fuller-Test Die Ergebnisse des ADF-Tests haben sich durch den Datenverlust kaum verändern, so dass der Test bei einer Länge von ursprünglich 300 Beobachtungen auf dem 0.05 und 0.01 Niveau sehr gut eingesetzt werden kann. 90 Datenpunkte waren jedoch schon bei den vollständigen Daten zu wenige, um den Test auch hier einzusetzen. Für stationäre Linearkombinationen sind nur die Ergebnisse auf dem 0.1 Niveau ausreichend gut. Auf diesem Niveau werden jedoch zu wenig integrierte Linearkombinationen erkannt.

Johansen-Test Der Johansen-Test klassifiziert bei Fehlen jedes zehnten Wertes ab Länge 100 (nun 90) auf dem 0.01 Niveau ausreichend viele kointegrierte und stationäre Systeme. Auf diesem Niveau werden bei jeder Länge fast alle integrierten Systeme richtig eingeschätzt.

3.6.2 Regelmäßiges Entfernen jeder 7. Messung

In dieser Simulationsstudie wurde regelmäßig jeder 7. Wert aus den beiden Reihen entfernt. Dies könnte in der Anwendung der Fall sein, wenn einmal pro Woche keine Daten erhoben werden können. So haben die Reihen jetzt nur noch 26, 52, 86 bzw. 258 Datenpunkte. Die Abbildung 3.22 zeigt am Beispiel eines kointegrierten Systems der Länge 60 aus der Gruppe (1,1), wie sich das Entfernen auswirkt.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Die Auswirkungen auf die Ergebnisse des Augmented Dickey-Fuller-Tests sind in Tabelle 3.86 zusammengestellt.

Tabelle 3.86: Ergebnisse des ADF-Tests nach Entfernung jeder 7. Messung: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	399	239	62	381	246	67	94	45	12
	60	793	627	264	794	637	280	75	29	6
	100	949	875	578	938	871	584	84	43	3
	300	1000	1000	1000	1000	1000	1000	66	38	3
(2,3)	30	388	247	66	376	228	58	76	42	10
	60	762	594	261	789	610	279	81	34	8
	100	946	866	551	958	861	568	79	48	8
	300	1000	1000	1000	1000	1000	1000	69	35	6
(3,10)	30	398	234	65	385	234	64	73	35	10
	60	783	617	278	771	624	282	79	39	6
	100	943	876	573	953	872	556	79	41	6
	300	1000	1000	1000	1000	1000	1000	74	34	6

Gruppe (1,1) Der ADF-Test (siehe Tabelle 3.81, S.256) zeigt bei allen Arten von Systemen Abweichungen zum vollständigen Fall. Nur bei Länge 300 sind sie kaum zu bemerken. Auf allen Niveaus erkennt er alle stationären Linearkombinationen der kointegrierten und stationären Systeme. Bei den integrierten Systemen stuft er nur 3-9 weitere Linearkombinationen als stationär ein und erkennt auf dem 0.1 Niveau sogar bei Länge 258 und 52 einige stationäre Linearkombinationen mehr als im Fall ohne *missing values*. So werden auf dem 0.05 und 0.01 Niveau weiterhin die meisten Einheitswurzeln in den Linearkombinationen erkannt. Die Ergebnisse sind mit denen der Länge 90 bei den Originaldaten vergleichbar (vgl. Tabelle 3.7, S.164).

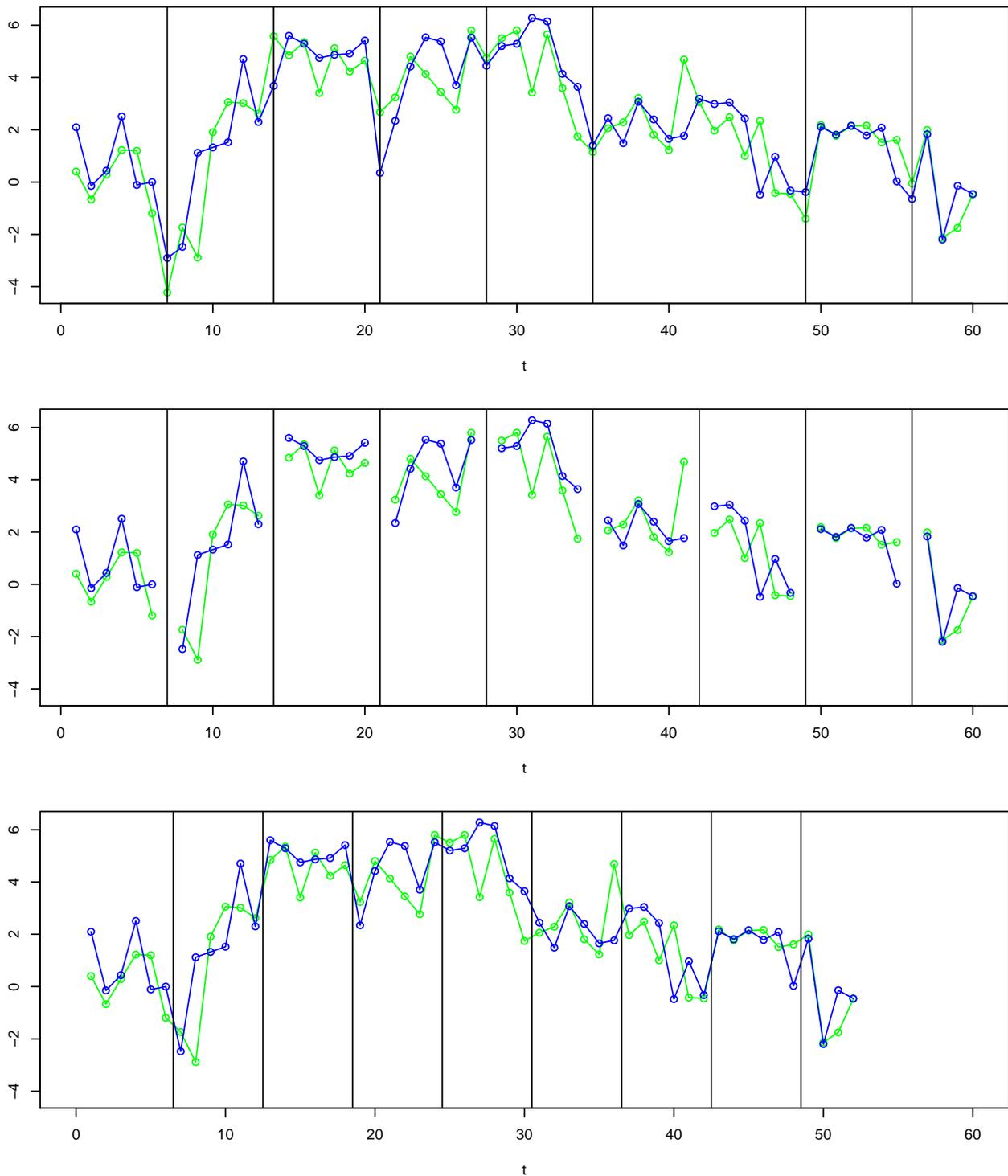


Abbildung 3.22: Entfernen jedes 7. Messwertes, hier am Beispiel eines kointegrierten Systems der Länge 60 aus der Gruppe (1,1). Oben: In der vollständigen Zeitreihe werden die zu entfernenden Messpunkte markiert. Mitte: Die Messwerte wurden entfernt. Unten: Zeitreihe nach Entfernen der Messwerte, markiert sind die Stellen fehlender Werte. Dadurch verkürzt sich die Zeitreihe (hier auf 52 Punkte).

Gruppe (2,3) Der ADF-Test reagiert bei den integrierten Systemen am geringsten auf den Informationsverlust. Bei einer Ursprungslänge von 300 ist jedoch auch bei den kointegrierten und stationären Systemen kein Effekt zu bemerken. Hier wird auf keinem Niveau eine Linearkombination missklassifiziert. Für integrierte Systeme weicht er nur um wenige Systeme ab, so dass er auf dem 0.05 und 0.01 Niveau sehr gute Klassifikationsergebnisse erzielt und verwendet werden könnte.

Bei den längsten Systemen (nun 258 Datenpunkte) werden alle stationären Linearkombinationen der kointegrierten und stationären Systeme erkannt. Die Ergebnisse für die Systeme der Länge 100 (nun 86) sind auf dem 0.1 Niveau ebenfalls noch gut, jedoch nur bei den stationären Systemen über 950. Auf dem 0.01 Niveau werden bei dieser Länge etwas mehr als 550 stationäre Linearkombinationen erkannt.

Gruppe (3,10) In der Gruppe (3,10) gibt es ähnliche Ergebnisse wie in den anderen beiden Gruppen. Der Test erkennt bei Länge 300 alle stationären Linearkombinationen auf allen Niveaus und auf dem 0.05 und 0.01 Niveau ausreichend viele integrierte. Für die integrierten Linearkombinationen ist auch die Erkennungsrate bei den anderen Längen sehr gut. Hier sorgt der Informationsverlust jedoch dafür, dass die stationären Linearkombinationen nur auf dem 0.1 Niveau noch zu 94-95% erkannt werden. Bei 26 bzw. 52 Datenpunkten werden weniger als 800, auf dem 0.01 Niveau bei Länge 30 nur ca. 6% richtig eingeschätzt.

Johansen test

Kointegrierte Systeme Die größten Abweichungen von der ursprünglichen Klassifikation gibt es bei einer Länge von 30 in allen Gruppen (siehe Tabelle 3.87, S.264). Da selbst auf dem 0.1 Niveau maximal 57% aller kointegrierten Systeme richtig klassifiziert wird, vernachlässigen wir diese Länge bei der weiteren Betrachtung. Mit nun 26 Beobachtungen sind die Systeme eindeutig zu kurz für eine zuverlässige Klassifikation.

Bei den Längen 100 und 300 werden auf dem 0.01 Niveau in allen Gruppen ca. 98% aller kointegrierten Systeme richtig erkannt. Die Abweichung von der ursprünglichen Klassifikation beträgt hier maximal 10 Systeme. Auch bei Länge 60 sind es maximal 32 Systeme, die anders eingestuft werden (vgl. Tabellen 3.12, S.170, 3.13, S.171 und 3.14, S.171). Die Werte sind hier zwar schlechter als bei Länge 60 ohne Informationsverlust, jedoch immer noch besser als bei Länge 50⁴⁴.

Auf dem 0.05 und 0.1 Niveau konnten schon beim vollen Datensatz nicht mehr als 950 kointegrierte Systeme richtig klassifiziert werden, so dass es auch mit weniger Datenpunkten nicht gelingt. Dennoch liegen die Werte mit etwa 930-950 auf dem 0.05 Niveau in ähnlicher

⁴⁴Den Zeitreihen der Länge 50 fehlen die letzten 10 Werte, während die zehn Messwerte nicht vom Ende der Messung, sondern aus dem mittleren Verlauf entfernt wurden. So enthalten die Prozesse mehr Informationen über die Entwicklung des Verlaufs. Mathematisch hat der Trend am Zeitpunkt 60 bereits 60 Summanden, während es bis zum Zeitpunkt 50 nur eben 50 sind.

Größenordnung. Die Ergebnisse sind auch etwas schlechter als in der ersten Missing Values Studie, wo jeder zehnte Messwert entfernt wurde. Die Schätzungen von β haben sich nicht verändert (siehe Tabelle D.22, S.379).

Tabelle 3.87: Klassifikation 1000 kointegrierter Systeme nach Entfernen jeder 7. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	388	543	69	569	399	32	841	157	2	1000
	60	10	876	114	36	902	62	205	781	14	1000
	100	0	870	130	0	934	66	0	983	17	1000
	300	0	880	120	0	934	66	0	991	9	1000
(3,10)	30	381	544	75	518	449	33	776	221	3	1000
	60	10	888	102	39	911	50	177	812	11	1000
	100	0	901	99	0	946	54	1	987	12	1000
	300	0	914	86	0	955	45	0	995	5	1000
(3,10)	30	349	571	80	528	442	30	804	192	4	1000
	60	13	883	104	44	902	54	177	809	14	1000
	100	0	901	99	0	952	48	1	981	18	1000
	300	0	911	89	0	954	46	0	995	5	1000

Stationäre Systeme Für die stationären Systeme (siehe Tabelle 3.88, S.265) hat der Datenverlust erst bei Systemen der ursprünglichen Länge 60 Auswirkungen. Die längeren Systeme werden in allen drei Gruppen auf allen Niveaus perfekt klassifiziert. Auf dem 0.1 Niveau gelingt dies auch noch bei nur 52 Datenpunkten. Auf dem 0.05 Niveau sind es noch mehr als 99% und auf dem 0.01 Niveau noch mehr als 900 richtig eingestufte Systeme. Die Ergebnisse sind besser als bei den Originaldaten bei Länge 50, jedoch schlechter als bei Länge 60 (vgl. Tabellen 3.16, S.174, 3.17, S.175 und 175, S.175).

Integrierte Systeme Die Klassifikation der integrierten Systeme (siehe Tabelle 3.89, S.265) hat sich durch den Datenverlust sogar etwas verbessert (vgl. Tabellen 3.19, S.177, 3.20, S.178 und 3.21, S.179). Auf dem 0.01 Niveau werden in allen Gruppen 97-99% der Systeme als integriert eingestuft. In der Gruppe (1,1) sind es auf dem 0.05 Niveau ca. 900, in den anderen Gruppen 92-95% richtig klassifizierter Systeme. Auch auf dem 0.1 Niveau werden in allen Gruppen bei jeder Länge mehr als 800 Systeme als integriert erkannt.

Ergebnis: Regelmäßiges Entfernen jeder 7. Messung

Augmented Dickey-Fuller-Test Auch nach Entfernen jedes siebten Wertes kann der ADF-Test bei den Systemen der Länge 300 (nun 258) noch auf dem 0.05 und 0.01 Niveau zuver-

Tabelle 3.88: Klassifikation 1000 stationärer Systeme nach Entfernen jeder 7. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	176	158	666	325	279	396	654	248	98	1000
	60	0	0	1000	1	5	994	32	39	929	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	170	186	644	325	271	404	675	248	77	1000
	60	0	0	1000	1	2	997	40	39	921	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	179	173	648	343	252	405	668	237	95	1000
	60	0	0	1000	2	1	997	36	52	912	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Tabelle 3.89: Klassifikation 1000 integrierter Systeme nach Entfernen jeder 7. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	812	173	15	897	95	8	980	20	0	1000
	60	816	176	8	910	87	3	980	19	1	1000
	100	816	167	17	894	102	4	967	33	0	1000
	300	831	158	11	905	92	3	973	27	0	1000
(2,3)	30	847	145	8	920	78	2	978	22	0	1000
	60	876	119	5	936	63	1	991	9	0	1000
	100	875	118	7	934	65	1	990	10	0	1000
	300	908	84	8	955	44	1	989	11	0	1000
(3,10)	30	842	146	12	911	88	1	975	25	0	1000
	60	880	114	6	942	57	1	988	12	0	1000
	100	876	115	9	939	61	0	988	12	0	1000
	300	908	84	8	957	41	2	991	9	0	1000

lässig eingesetzt werden. Bei den kürzeren Zeitreihen ist der Informationsverlust stärker zu bemerken.

Johansentest Der Verlust jedes siebten Messwerts hat auf die Klassifikation der Systeme ab einer Länge von 100 (nun 86) keinen Einfluss. Auf dem 0.01 Niveau werden alle stationären, mehr als 99% der integrierten und über 98% der kointegrierten Systeme richtig eingestuft. Auch auf die Schätzung der Kointegrationsbeziehung hat der Datenverlust keine Auswirkung.

3.6.3 Regelmäßiges Entfernen jeder 5. Messung

Nach regelmäßiger Entfernung jeder fünften Messungen haben die Systeme noch eine verbleibende Länge von 24, 48, 80 bzw. 240 Datenpunkten. In der Abbildung 3.23 ist am Beispiel eines kointegrierten Systems der Länge 60 aus der Gruppe (1,1) zu sehen, wie stark sich der Verlauf dadurch verändert.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Die Ergebnisse des Augmented Dickey-Fuller-Tests für diese Simulationsstudie sind in Tabelle 3.90 zusammengefasst. Der Datenverlust von 20% aller Messwerte wirkt sich in allen Gruppen

Tabelle 3.90: Ergebnisse des ADF-Tests nach Entfernung jeder 5. Messung: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	350	196	45	327	192	61	93	42	10
	60	706	533	199	734	552	227	85	39	8
	100	927	805	488	924	814	497	79	42	6
	300	1000	1000	1000	1000	999	999	64	37	7
(2,3)	30	330	207	58	340	189	58	89	48	10
	60	690	529	215	706	523	192	82	41	9
	100	917	800	472	940	817	480	78	39	6
	300	1000	1000	999	1000	1000	1000	75	35	6
(3,10)	30	312	182	51	346	197	49	85	54	7
	60	693	524	201	681	517	207	82	39	6
	100	906	784	461	926	837	468	71	37	5
	300	1000	1000	1000	1000	1000	1000	75	36	7

ähnlich aus (siehe Tabelle 3.90, S.266). In der Gruppe (1,1) wird bei Länge 240 auf dem 0.05 und 0.01 Niveau eine Linearkombination der stationären Systeme nicht mehr als stationär

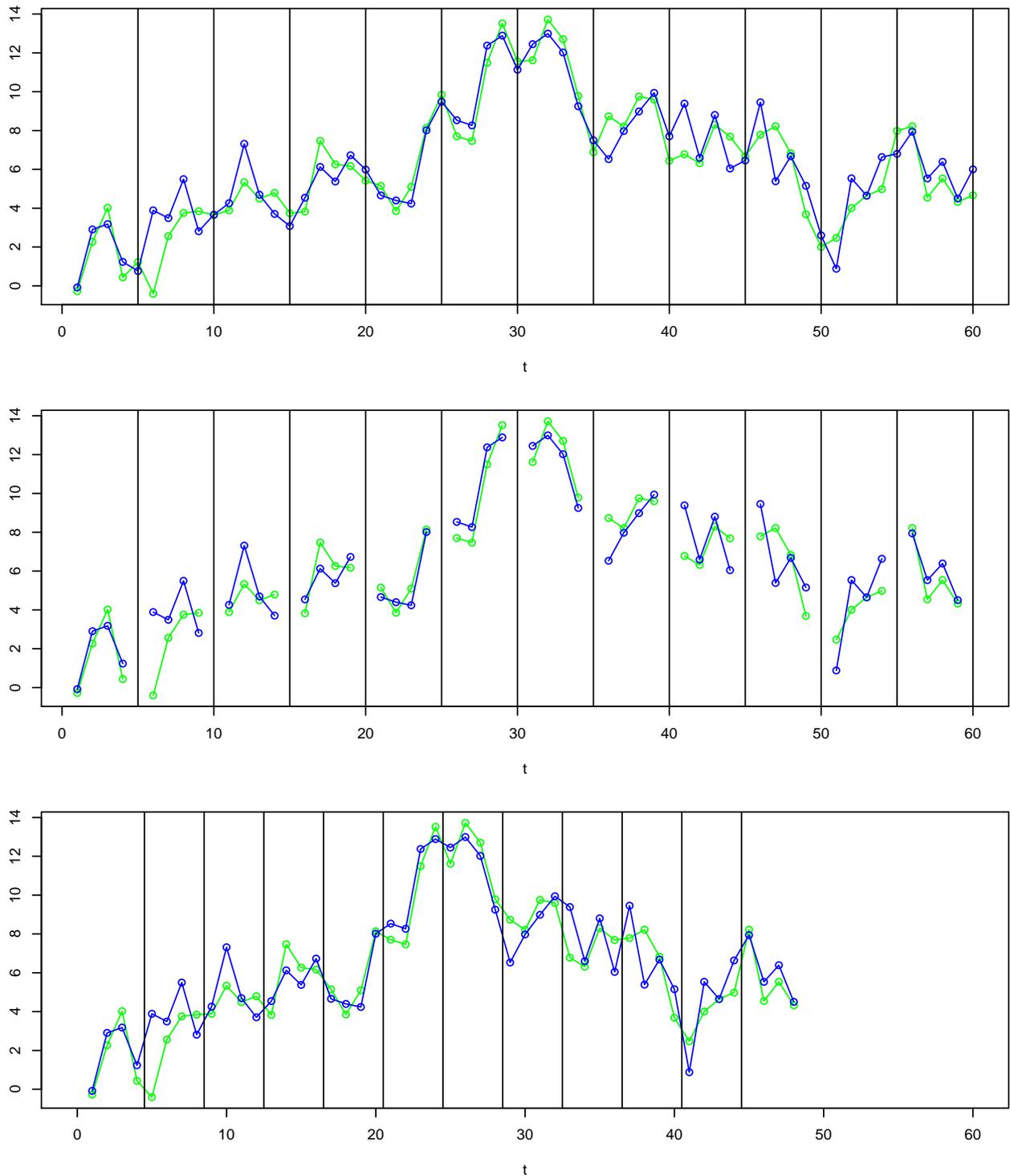


Abbildung 3.23: Entfernen jedes 5. Messwertes, hier am Beispiel eines kointegrierten Systems der Länge 60 aus der Gruppe (1,1). Oben: In der vollständigen Zeitreihe werden die zu entfernenden Messpunkte markiert. Mitte: Die Messwerte wurden entfernt. Unten: Zeitreihe nach Entfernen der Messwerte, markiert sind die Stellen fehlender Werte. Dadurch verkürzt sich die Zeitreihe (hier auf 48 Punkte).

erkannt. Dies gilt auch für eine Kointegrationsbeziehung in der Gruppe (2,3) dieser Länge auf dem 0.01 Niveau. Auf dem 0.1 Niveau werden bei 80 verbleibenden Messpunkten noch mehr als 900, aber weniger als 950 Linearkombinationen als stationär erkannt. Auf dem 0.01 Niveau sind es bei dieser Länge bereits weniger als die Hälfte. Die Ergebnisse bei kürzeren Reihen sind in allen Gruppen deutlich schlechter. Dieser Effekt ist auch stärker als bei Entfernen jeder siebten Messung (vgl. Tabelle 3.86, S.261). Die Auswirkung bei den integrierten Systemen ist nicht so deutlich. In allen Gruppen werden passend zum jeweiligen Niveau ausreichend viele Einheitswurzeln in den Systemen erkannt.

Johansenstest

Kointegrierte Systeme Für die Ergebnisse des Johansenstest nach Entfernen jedes fünften Messwertes (siehe Tabelle 3.91, S.268) betrachten wir nur die Werte ab einer Originallänge von 60. Bei kürzeren Systemen ist die Unterscheidbarkeit von integrierten und kointegrierten Systemen nicht gegeben.

Der Großteil der Systeme wird in allen Gruppen auf allen Niveaus richtig als kointegriert klassifiziert. Erst ab Länge 100 (jetzt 80) werden jedoch Werte über 900 Systemen erreicht. Auf dem 0.01 Niveau sind es in allen Gruppen mehr als 980 richtig klassifizierte Systeme. Hier werden bei Länge 100 vereinzelt (0-3) Systeme als integriert missklassifiziert. Die Werte auf dem 0.05 Niveau liegen bei den Längen 100 und 300 deutlich über 900, die 950-Marke kann jedoch nur in den Gruppen (2,3) und (3,10) bei Länge 300 überschritten werden. Auf dem 0.1 Niveau werden noch über 80% der Systeme als kointegriert klassifiziert.

Tabelle 3.91: Klassifikation 1000 kointegrierter Systeme nach Entfernen jeder 5. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	435	496	69	609	361	30	856	142	2	1000
	60	14	869	117	50	892	58	247	740	13	1000
	100	0	877	123	1	927	72	2	984	14	1000
	300	0	889	111	0	938	62	0	989	11	1000
(2,3)	30	417	516	67	583	396	21	827	169	4	1000
	60	18	877	105	54	893	53	249	736	15	1000
	100	0	903	97	0	944	56	0	985	15	1000
	300	0	911	89	0	959	41	0	994	6	1000
(3,10)	30	405	528	67	572	409	19	826	172	2	1000
	60	18	880	102	60	890	50	255	732	13	1000
	100	0	899	101	0	949	51	3	980	17	1000
	300	0	912	88	0	961	39	0	994	6	1000

Die Boxplot-Statistiken für die Schätzung von β haben sich nur auf der zweiten Nachkom-

mastelle, der Median gar erst auf der dritten verändert (siehe Tabelle D.23, S.380). Je länger die Systeme werden und je größer der Varianzunterschied der Reihen, desto geringer ist die Abweichung. Insbesondere bei der Länge 300 liegen der minimale und maximale Schätzwert sehr nah an den beiden Whisker-Statistiken.

Stationäre Systeme Der Johansentest klassifiziert ab einer Länge von 100 (jetzt 80) auf allen Niveaus alle 1000 stationären Systeme in allen Gruppen richtig (siehe Tabelle 3.92, S.269). Auch bei Länge 60 liefert er noch eine sehr gute Klassifikationsrate, die nur auf dem 0.01 Niveau die 900-Marke nicht erreicht. Auf dem 0.01 Niveau werden bei dieser Länge 82, 75 bzw. 84 Systeme als integriert eingestuft. Dies passiert auf dem 0.05 und 0.1 Niveau nur in Einzelfällen.

Tabelle 3.92: Klassifikation 1000 stationärer Systeme nach Entfernen jeder 5. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	218	205	577	411	274	315	744	203	53	1000
	60	1	0	999	9	1	990	82	62	856	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	241	216	543	403	307	290	730	216	54	1000
	60	0	3	997	8	10	982	75	87	838	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	235	216	549	392	295	313	740	217	43	1000
	60	1	2	997	3	6	991	84	81	835	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Integrierte Systeme Auch nach einem Datenverlust von 20% ist das Klassifikationsverhalten des Johansentest bei den integrierten Systemen nicht verändert (siehe Tabelle 3.93, S.270). Teilweise werden sogar mehr Systeme als integriert eingestuft als bei den Systemen mit allen Messwerten (vgl. Tabellen 3.19, S.177, 3.20, S.178 und 3.21, S.179). Auf dem 0.01 Niveau werden bei allen Längen in der ersten Gruppe mehr als 97%, in den anderen beiden Gruppen 98-99% der Systeme richtig eingestuft. Die Klassifikationsrate liegt auf dem 0.05 Niveau bei mehr als 89% und erreicht bei den Systemen der Länge 240 in den Gruppen (2,3) und (3,10) die 950-Marke. Auch auf dem 0.1 Niveau werden mehr als 800 Systeme richtig klassifiziert.

Tabelle 3.93: Klassifikation 1000 integrierter Systeme nach Entfernen jeder 5. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	796	184	20	890	106	4	977	23	0	1000
	60	822	169	9	900	96	4	976	24	0	1000
	100	820	167	13	888	110	2	974	26	0	1000
	300	844	146	10	912	85	3	979	21	0	1000
(2,3)	30	827	160	13	917	78	5	982	18	0	1000
	60	874	119	7	938	60	2	990	10	0	1000
	100	863	126	11	929	71	0	992	8	0	1000
	300	909	81	10	957	41	2	993	7	0	1000
(3,10)	30	824	161	15	895	98	7	969	31	0	1000
	60	875	118	7	942	57	1	986	14	0	1000
	100	883	113	4	934	65	1	988	12	0	1000
	300	911	80	9	956	42	2	993	7	0	1000

Ergebnis: Regelmäßiges Entfernen jeder 5. Messung

Augmented Dickey-Fuller-Test Mir 240 verbleibenden Datenpunkten sind die längsten betrachteten Systeme auch nach Entfernen jedes fünften Messwertes noch lang genug, um den ADF-Test für die Unterscheidung zwischen stationärer und integrierter Linearkombination einzusetzen.

Johansen test Da die Zeitreihen mit ursprünglich 100 Beobachtungen durch die Entfernung jedes fünftes Messwertes auf eine Länge von 80 Punkten reduziert wurden, sind sie immer noch lang genug, dass der Johansen test auf dem 0.01 Niveau alle Systemgruppen sehr gut einschätzen kann. Er erkennt in jeder Gruppe mehr als 98% der kointegrierten, alle stationären und mehr als 97% bzw. 99% der integrierten Systeme bei den Längen 80 und 240.

3.6.4 Regelmäßiges Entfernen jeder 3. Messung

Durch das regelmäßige Entfernen jedes dritten Beobachtungspunktes werden nur zwei Drittel der ursprünglichen Messwerte berücksichtigt. Dadurch verkürzen sich die Längen der Systeme auf 20, 40, 67 bzw. 200 Messpunkte. In der Abbildung 3.24 ist dieser Vorgang beispielhaft dargestellt.

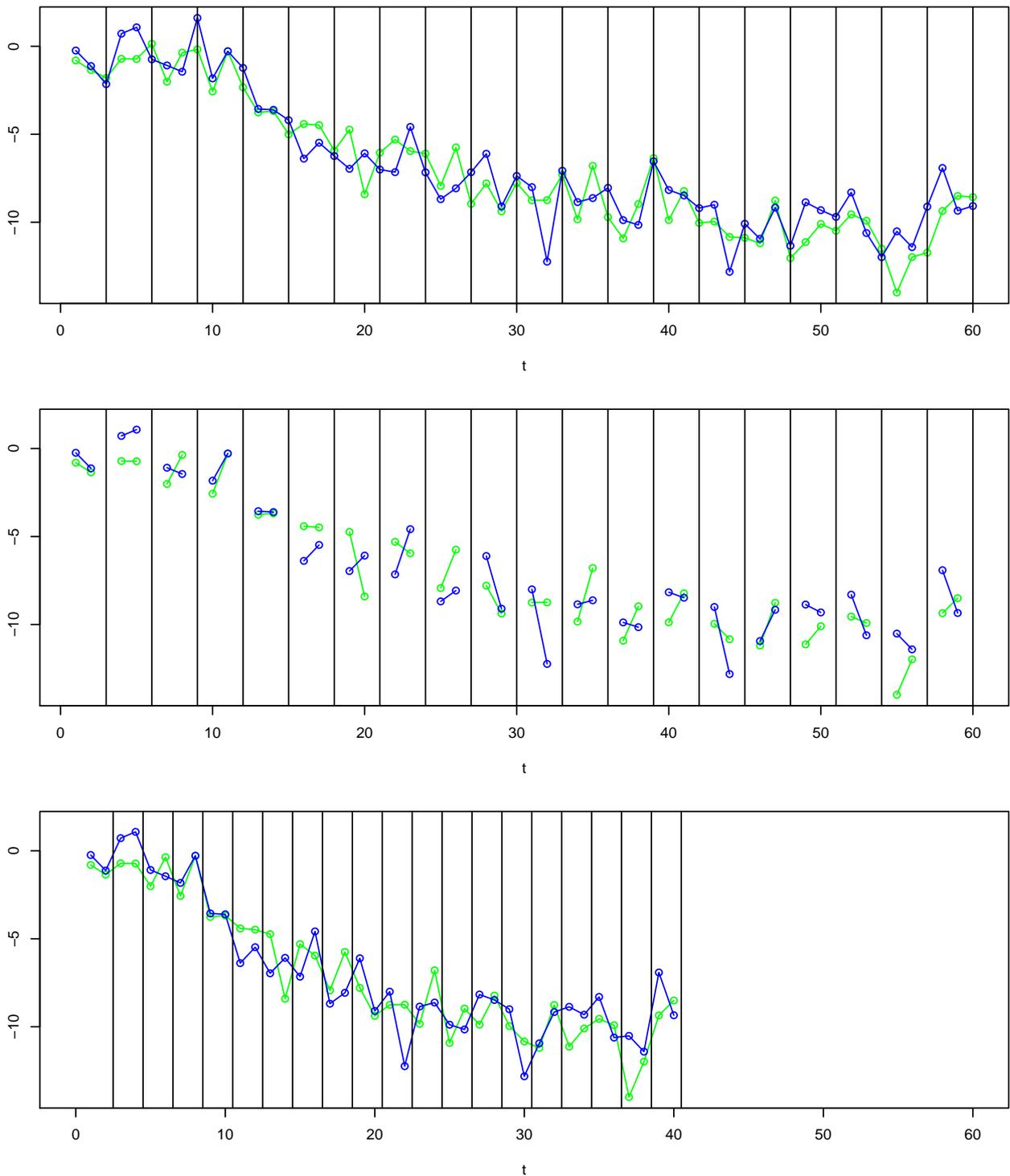


Abbildung 3.24: Entfernen jedes 3. Messwertes, hier am Beispiel eines kointegrierten Systems der Länge 60 aus der Gruppe (1,1). Oben: In der vollständigen Zeitreihe werden die zu entfernenden Messpunkte markiert. Mitte: Die Messwerte wurden entfernt. Unten: Zeitreihe nach Entfernen der Messwerte, markiert sind die Stellen fehlender Werte. Dadurch verkürzt sich die Zeitreihe (hier auf 40 Punkte).

Tabelle 3.94: Ergebnisse des ADF-Tests nach Entfernung jeder 10. Messung: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	251	146	40	257	151	48	93	40	11
	60	511	343	112	528	366	129	85	40	8
	100	779	633	297	792	626	306	83	41	3
	300	1000	1000	999	1000	1000	1000	68	38	5
(2,3)	30	227	126	30	242	143	39	69	42	6
	60	529	365	117	518	342	94	89	32	9
	100	795	629	272	786	624	277	74	34	2
	300	1000	1000	1000	1000	1000	999	71	36	7
(3,10)	30	251	131	34	253	156	45	68	28	4
	60	521	330	102	528	369	93	85	47	12
	100	774	624	281	801	608	271	68	34	3
	300	1000	1000	998	1000	1000	999	74	35	9

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

In der Tabelle 3.94 auf Seite 272 sind die Ergebnisse des Augmented Dickey-Fuller-Tests für alle drei Gruppen nach Entfernen jedes dritten Messwertes zusammengefasst.

Wenn nun nur noch zwei von drei Messwerten berücksichtigt werden, ist die Erkennung stationärer Linearkombinationen bei Reihen, die nun kürzer als 100 Beobachtungen sind, zu gering. Sie ist außerdem schlechter als dies bei Systemen ähnlicher Länge ohne Datenverlust der Fall war (vgl. Tabellen 3.7, 3.9 und 3.11, S.164), 3.9 bzw. 3.11). Da der Test selbst auf dem 0.1 Niveau maximal 801 stationäre Linearkombinationen erkennt, kann er auch bei 67 Beobachtungspunkten in keiner Gruppe mehr eingesetzt werden. Die Ergebnisse für stationäre Linearkombinationen der ursprünglichen Länge 300, die nun noch 200 Datenpunkte haben, und für integrierte Linearkombinationen jeder Länge haben sich kaum verändert. So kann der Test bei Systemen der Länge 300 auch dann auf dem 0.05 und 0.01 noch zuverlässig eingesetzt werden, wenn jeder dritte Messwert entfernt wird.

Johansen test

Kointegrierte Systeme Bei einer Länge von 200 (ehemals 300) werden maximal 15 kointegrierte Systeme (siehe Tabelle 3.95, S.273) anders klassifiziert als bei voller Länge (vgl. Tabellen 3.12, S.170, 3.13, S.171 und 3.14, S.171). Hier wird auf dem 0.01 Niveau die 950-Marke deutlich überschritten (etwa 99%) und auch auf dem 0.05 Niveau liegen die Werte in den Gruppen (2,3) und (3,10) über dieser Marke, in der Gruppe (1,1) knapp darunter. Auf dem 0.1 Niveau wurden schon nicht genügend ungekürzte Systeme richtig klassifiziert.

Jedoch wird in keiner Gruppe auf keinem Niveau ein System der Länge 200 nun als integriert eingeschätzt.

Die Werte bei einer Länge von 100 (nun 67) sind auf dem 0.01 und 0.05 Niveau auch noch immer gut. Denn sie liegen auf dem 0.05 Niveau deutlich über 900, auf dem 0.01 Niveau sogar über 950. Bei einer Länge von 60 (nun 40) ist die Differenzierung zwischen integrierten und kointegrierten Systemen nicht mehr groß genug. In der Gruppe (3,10) werden auf dem 0.01 Niveau sogar mehr als die Hälfte der Systeme als integriert eingestuft.

Tabelle 3.95: Klassifikation 1000 kointegrierter Systeme nach Entfernen jeder 3. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	529	401	70	671	300	29	885	111	4	1000
	60	95	802	103	192	755	53	476	512	12	1000
	100	0	882	118	1	929	70	26	958	16	1000
	300	0	895	105	0	946	54	0	991	9	1000
(2,3)	30	480	449	71	630	341	29	839	156	5	1000
	60	70	834	96	150	798	52	428	555	17	1000
	100	1	895	104	4	935	61	20	965	15	1000
	300	0	909	91	0	957	43	0	995	5	1000
(3,10)	30	510	412	78	658	310	32	867	132	1	1000
	60	85	820	95	180	768	52	483	507	10	1000
	100	0	893	107	2	943	55	17	968	15	1000
	300	0	914	86	0	954	46	0	993	7	1000

Die Werte für beta liegen immer noch extrem nah an den wahren Werten.

Stationäre Systeme Auf den Längen 100 (nun 67) und 300 (nun 200) werden alle stationären Systeme auf dem 0.1 und 0.05 Niveau richtig eingestuft (siehe Tabelle 3.96, S.274). Auf dem 0.01 Niveau kommt es zu 3-4 Fehlklassifikationen. Hier werden bei einer Länge von 60 (nun 40) mehr als 200 Systeme als integriert eingeschätzt. Ein Effekt, der auf den anderen Niveaus nur geringfügig auftritt. Auf dem 0.1 Niveau werden auch bei dieser Länge mehr als 97% aller Systeme richtig klassifiziert, auf 0.05 sind es ca. 95%.

Integrierte Systeme Von den integrierten Systemen können bei Länge 300 (nun 200) auf dem 0.01 Niveau mehr als 950 als integriert erkannt werden (siehe Tabelle 3.97, S.274). In den Gruppen (2,3) und (3,10) ist dies auch auf dem 0.05 Niveau der Fall. Auf diesem Niveau werden in jeder Gruppe ab einer Länge von nun 40 Datenpunkten (vorher 60) mehr als 900 Systeme richtig klassifiziert. Die restlichen werden vor allem für kointegriert, nur im Einzelfall für stationär gehalten. Auf dem 0.1 Niveau sind diese Werte höher, so dass der

Tabelle 3.96: Klassifikation 1000 stationärer Systeme nach Entfernen jeder 3. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	375	257	368	536	276	188	802	168	30	1000
	60	12	12	976	41	37	922	227	177	596	1000
	100	0	0	1000	0	0	1000	1	3	996	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	341	301	358	529	307	164	817	161	22	1000
	60	14	16	970	44	39	917	254	183	563	1000
	100	0	0	1000	0	0	1000	0	3	997	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	367	251	382	532	284	184	806	163	31	1000
	60	14	14	972	47	45	908	226	200	574	1000
	100	0	0	1000	0	0	1000	2	2	996	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Test auf diesem Niveau bei keiner Länge mehr ausreichend gute Ergebnisse bringt. Es ist ein Effekt der Länge bei der Klassifikation zu bemerken.

Tabelle 3.97: Klassifikation 1000 integrierter Systeme nach Entfernen jeder 3. Messung

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	756	220	24	841	151	8	945	55	0	1000
	60	814	180	6	900	96	4	977	23	0	1000
	100	824	165	11	904	95	1	975	25	0	1000
	300	857	134	9	922	74	4	982	18	0	1000
(2,3)	30	782	190	28	858	135	7	961	39	0	1000
	60	852	138	10	923	75	2	984	16	0	1000
	100	856	134	10	935	64	1	982	18	0	1000
	300	905	86	9	956	41	3	992	8	0	1000
(3,10)	30	746	227	27	832	159	9	939	60	1	1000
	60	844	150	6	920	78	2	982	18	0	1000
	100	858	131	11	930	69	1	986	14	0	1000
	300	902	89	9	953	44	3	992	8	0	1000

Ergebnis: Regelmäßiges Entfernen jeder 3. Messung

Augmented Dickey-Fuller-Test Wenn nur zwei Drittel aller Daten berücksichtigt werden können, sind die Systeme der ursprünglichen Länge 100 zu kurz für eine Auswertung mit

dem ADF-Test. Bei den Systemen der Länge 200 (zuvor 300) werden auf dem 0.01 Niveau nicht mehr alle stationären Linearkombinationen erkannt. Die wenigen Einzelfälle verändern das Verhalten des Tests jedoch kaum, so dass er auch bei so großem Informationsverlust noch eingesetzt werden kann.

Johansentest Und auch der Johansentest kann bei allen Systemen der ursprünglichen Länge 100 bzw. 300 noch eingesetzt werden. Die stationären Systeme der Länge 200 werden weiterhin perfekt klassifiziert. Bei dieser Länge werden auch mehr als 99% der kointegrierten und 97%-99% der integrierten Systeme erkannt. Für die Systeme der Länge 100 ist bei Entfernen jeder dritten Messung ein Effekt zu bemerken. Auf dem 0.01 Niveau wurden 3-4 stationäre Systeme nicht mehr erkannt. Bei den kürzeren Zeitreihen kann der Test jedoch nicht mehr eingesetzt werden.

3.6.5 Zufälliges Entfernen eines Zehntels der Messungen

In einer weiteren Serie von Simulationsstudien wird nun jeweils dieselbe Anzahl von Beobachtungen wie in den eben beschriebenen Simulationsstudien aus den Daten entfernt. Nun werden jedoch zufällig Messpunkte ausgewählt, deren Messwert aus den Systemen gelöscht wird. Um Vergleichbarkeit zwischen den einzelnen Systemen und den Systemgruppen zu gewährleisten, wurden für jede Länge stets dieselben Zeitpunkte entfernt. Dazu wurde eine Zufallsstichprobe entsprechender Länge aus der Gesamtheit aller Messpunkte der entsprechenden Länge gezogen. Die geschah mit Hilfe der R-Funktion `sample`, deren Einsatz im Anhang D.6.4 (S.396) beschrieben ist. Um Regelmäßigkeiten auszuschließen, wurde für jede Länge die Zufallsstichprobe neu gezogen. Die bei der Länge 30 entfernten Messpunkte können bei den anderen Längen also zufällig wieder enthalten sein.

In einer ersten Studie wurden 10% aller Messwerte entfernt. Die dafür ausgeschlossenen Messpunkte sind in Abbildung 3.25 auf Seite 276 zu sehen und im Anhang D.6.4 aufgelistet. Dadurch verkürzen sich die Zeitreihen um jeweils ein Zehntel auf die Längen 27, 54, 90 und 270 Datenpunkte.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Die Ergebnisse des Augmented Dickey-Fuller-Tests für alle Gruppen sind in Tabelle 3.98 dargestellt.

Gruppe (1,1) Der ADF-Test zeigt in der ersten Gruppe bei allen Systemtypen eine Verschlechterung im Vergleich zu den vollen Datensätzen (vgl. Tabelle 3.7, S.164). Die Werte sind vergleichbar, wenn auch minimal schlechter, als in der Studie mit regelmäßigem Entfernen jedes 10. Messwerts (vgl. Tabelle 3.81, S.256). Bei Länge 300 (jetzt 270) werden alle

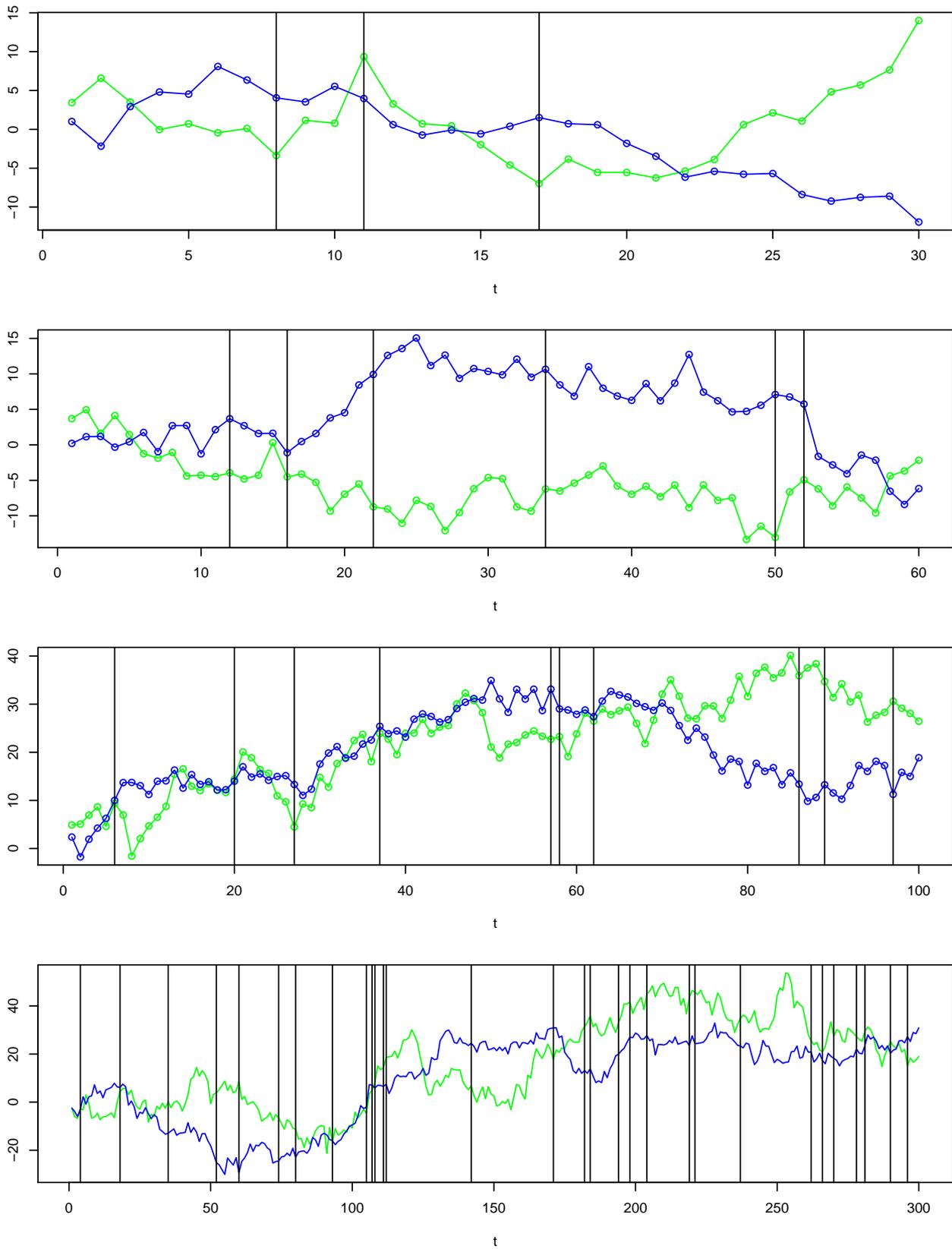


Abbildung 3.25: Zufällig entfernte Messpunkte bei den Längen 30, 60, 100 und 300 bei Entfernung eines Zehntels aller Daten (Für genau Wertangaben siehe Anhang D.6.4.).

Tabelle 3.98: Ergebnisse des ADF-Tests nach Entfernung von einem Zehntel der Messungen: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	435	268	73	400	242	67	96	47	8
	60	819	642	314	816	650	301	68	30	5
	100	958	892	645	965	906	607	85	41	11
	300	1000	1000	1000	1000	1000	1000	63	35	4
(2,3)	30	423	266	70	398	239	78	85	43	6
	60	824	643	311	813	653	291	73	34	3
	100	969	919	649	958	901	621	83	49	10
	300	1000	1000	1000	1000	1000	1000	68	33	8
(3,10)	30	437	277	63	404	250	66	66	32	6
	60	798	631	277	821	655	279	74	35	2
	100	962	899	634	970	902	630	81	41	12
	300	1000	1000	1000	1000	1000	1000	74	33	6

stationären Linearkombinationen und ein Großteil (937-996) der integrierten erkannt. Bei den kürzeren kointegrierten und stationären Systemen kommen nur die Werte auf dem 0.1 Niveau bei Länge 100 (jetzt 90) über die 950-Marke. Auf dem 0.05 Niveau bewegen sich die Werte nahe 900, was auf dem 0.01 Niveau und bei den Systemen der Länge 60 (jetzt 54) und 30 (jetzt 27) nicht mehr der Fall ist. Bei den integrierten Systemen sind die Ergebnisse auf dem 0.01 Niveau bei jeder Länge sehr gut (989-996 richtig erkannte Einheitswurzeln). Auf dem 0.05 Niveau werden noch mehr als 950 Linearkombinationen richtig erkannt. Insgesamt über alle Systemgruppen ist der Test jedoch nur bei Länge 300 auf dem 0.05 oder 0.01 Niveau zuverlässig einsetzbar.

Gruppe (2,3) In der zweiten Gruppe erkennt der ADF-Test bei einer Länge von 300 (jetzt 270) auf allen Niveaus die 1000 stationären Linearkombinationen der kointegrierten und stationären Systeme und auf dem 0.05 und 0.01 Niveau mehr als 950 integrierte (992 auf 0.01 und 967 auf 0.05, bei 0.1 sind es noch 932). Bei Länge 100 (jetzt 90) übersteigen die Werte bei stationären Linearkombinationen auf dem 0.1 Niveau die 950-Marke und auch auf dem 0.05 Niveau werden mehr als 90% erkannt. Auch für die integrierten Systeme wird auf dem 0.05 und 0.01 Niveau bei dieser Länge noch in ausreichend vielen Linearkombinationen die Einheitswurzel erfasst. Dies gilt ebenfalls für alle anderen Längen. Auch in dieser Gruppe ist der Einsatz des ADF-Test nur bei einer Länge über 100 sinnvoll, da bei kürzeren Zeitreihen zwar die integrierten, jedoch nicht die stationären Linearkombinationen zuverlässig klassifiziert werden.

Gruppe (3,10) Die Abweichungen in der dritten Gruppe sind bei den integrierten Systemen im Vergleich zu den vollständigen Datensätzen (vgl. Tabelle 3.11, S.168) am geringsten, obwohl er bei Länge 300 für kointegrierte und stationäre Systeme auch weiterhin alle Linearkombinationen richtig erkennt und somit überhaupt nicht vom Ursprungsergebnis abweicht. Bei den integrierten Systemen sind es bei dieser Länge jedoch auch nur 1 Linearkombination, die nun als stationär eingeschätzt wird. Bei dieser Länge kann der Test also auch bei 10% fehlender Messdaten auf dem 0.01 und 0.05 Niveau eingesetzt werden. Bei einer Länge von unter 100 liegen auf dem 0.05 Niveau die Werte in allen Systemgruppen nahe der 900-Marke. Jedoch werden nur bei den integrierten Systemen gleichzeitig mehr als 950 Linearkombinationen richtig eingeschätzt. Für stationäre Linearkombinationen der kürzeren Systeme sind die Erkennungsraten auf allen Niveaus zu schlecht, um den Test sinnvoll zur Unterscheidung von stationären und nicht-stationären Linearkombinationen einzusetzen.

Johansentest

Kointegrierte Systeme Abgesehen von den Systemen der Länge 30 (nun noch 27) wird der Großteil der Systeme auf jedem Niveau als kointegriert eingestuft (siehe Tabelle 3.99, S.279). In allen Gruppen nimmt die Klassifikationsrate mit der Länge zu. Ein Unterschied zwischen den Niveaus ist ebenfalls in allen Gruppen zu erkennen. So ist Anzahl als kointegriert erkannter Systeme auf dem 0.01 Niveau stets am höchsten. Die Ergebnisse der Gruppe (3,10) sind etwas besser als in der Gruppe (2,3), die Gruppe (1,1) hat mit geringer Differenz die niedrigsten Klassifikationsergebnisse. Bei einer Beobachtungslänge von 300 Datenpunkten werden hier auf dem 0.01 Niveau jedoch 992 Systeme als kointegriert kategorisiert und auch bei 100 Punkten wird die 950-Marke in dieser Gruppe wie in den beiden anderen klar überschritten. Auf dem 0.05 Niveau werden jedoch nur in den Gruppen (2,3) und (3,10) ca. 95-96% der Systeme der Länge 100 bzw. 300 als kointegriert klassifiziert. In der Gruppe (1,1) wird mit 930 bzw. 933 Systemen die gesetzte Marke nicht erreicht. Die Ergebnisse auf dem 0.1 Niveau sind ebenfalls gut, sie liegen jedoch mit Werten um 900 auch unter dem kritischen Wert.

Die Schätzungen von β sind in allen Gruppen sehr gut (siehe Tabelle 3.100, S.279). Nur in der ersten Gruppe zeigen sich bei Länge 30 größere Ausreißer.

Stationäre Systeme Für alle verkürzten stationären Systeme der Längen 100 und 300 (jetzt 90 bzw. 270) werden in allen Gruppen alle Systeme auf allen Niveaus als stationär eingeschätzt (siehe Tabelle 3.101, S.280). Die Systeme der ursprünglichen Länge 60 werden auf dem 0.1 und 0.05 Niveau auch noch zu mehr 99% erkannt. Auf dem 0.01 Niveau liegt die Klassifikationsrate bei dieser Länge bei ca. 94%. Für die Systeme der Länge 30 kann der Test nicht mehr eingesetzt werden, obwohl auf dem 0.1 Niveau noch die meisten Systeme (692-705) richtig eingestuft werden.

Tabelle 3.99: Klassifikation 1000 kointegrierter Systeme nach zufälligem Entfernen eines Zehntels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	348	570	82	523	439	38	806	191	3	1000
	60	9	879	112	24	915	61	155	828	17	1000
	100	0	869	131	0	930	70	1	982	17	1000
	300	0	885	115	0	933	67	0	992	8	1000
(2,3)	30	305	620	75	471	501	28	777	223	0	1000
	60	4	901	95	16	935	49	146	844	10	1000
	100	0	895	105	0	949	51	0	985	15	1000
	300	0	911	89	0	960	40	0	994	6	1000
(3,10)	30	339	589	72	501	470	29	778	220	2	1000
	60	5	891	104	28	926	46	150	839	11	1000
	100	0	896	104	0	951	49	2	984	14	1000
	300	0	915	85	0	960	40	0	995	5	1000

Tabelle 3.100: Schätzung für β bei kointegrierten Systemen nach zufälligem Entfernen eines Zehntels der Messungen

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
(1,1)	30	-20,705	-1,513	-1,141	-1,007	-0,889	-0,516	30,311	80
	60	-2,354	-1,205	-1,055	-1,003	-0,953	-0,805	-0,539	59
	100	-1,480	-1,114	-1,030	-1,005	-0,974	-0,893	-0,722	43
	300	-1,079	-1,036	-1,009	-0,999	-0,991	-0,963	-0,944	36
(2,3)	30	-3,968	-0,801	-0,699	-0,665	-0,631	-0,528	1,603	91
	60	-3,313	-0,725	-0,680	-0,666	-0,650	-0,606	-0,548	56
	100	-0,852	-0,699	-0,674	-0,666	-0,658	-0,633	-0,587	40
	300	-0,701	-0,677	-0,669	-0,667	-0,664	-0,657	-0,648	34
(3,10)	30	-1,959	-0,340	-0,311	-0,300	-0,291	-0,262	1,739	80
	60	-0,346	-0,316	-0,304	-0,300	-0,296	-0,284	-0,236	40
	100	-0,317	-0,309	-0,302	-0,300	-0,298	-0,292	-0,282	51
	300	-0,305	-0,303	-0,301	-0,300	-0,299	-0,297	-0,294	38

Tabelle 3.101: Klassifikation 1000 stationärer Systeme nach zufälligem Entfernen eines Zehntels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	150	145	705	289	257	454	639	244	117	1000
	60	0	1	999	0	5	995	23	42	935	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	141	167	692	308	239	453	653	244	103	1000
	60	0	0	1000	0	4	996	22	36	942	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	145	156	699	300	250	450	668	222	110	1000
	60	0	0	1000	0	0	1000	22	42	936	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Integrierte Systeme Wie die Tabelle 3.102 (S.281) zeigt, werden bei allen Längen auf allen Niveaus in allen Gruppen mehr als 800 Systeme als integriert klassifiziert. Die Kategorisierung ist umso besser je strenger das Niveau gewählt wird. Zwischen den Gruppen gibt es eine leichte Verbesserung mit steigendem Varianzunterschied. Ein Effekt der Länge ist kaum auszumachen. Auf dem 0.01 Niveau werden stets mehr als 96%, in den Gruppen (2,3) und (3,10) bei Länge 300 sogar 99% richtig klassifiziert. Die Werte auf dem 0.05 Niveau liegen in der Gruppe (1,1) bei ca. 88-89% und in den anderen beiden Gruppen bei 90-95%.

Ergebnis: Zufälliges Entfernen eines Zehntels der Messungen

Augmented Dickey-Fuller-Test Auch wenn 10% der Messwerte zufällig entfernt werden, kann der ADF-Test bei den Systemen der ursprünglichen Länge 300 noch eingesetzt werden. Bei Länge 100 sind die Ergebnisse bei stationären Linearkombinationen auf dem 0.1 Niveau noch ausreichend gut, für integrierte Linearkombinationen trifft dies allerdings nur auf dem 0.05 und 0.01 Niveau zu, so dass die Zeitreihen für eine zuverlässige Unterscheidung zu kurz sind.

Johansentest Auch die Ergebnisse des Johansentest sind sehr ähnlich zu denen bei regelmäßiger Entfernung jeder zehnten Messung, so dass der Test ab einer ursprünglichen Länge von 100 Beobachtungspunkten eingesetzt werden kann. Für kürzere Zeitreihen ist sein Einsatz nicht zu empfehlen.

Tabelle 3.102: Klassifikation 1000 integrierter Systeme nach zufälligem Entfernen eines Zehntels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	807	179	14	882	114	4	961	39	0	1000
	60	833	160	7	890	107	3	972	28	0	1000
	100	794	190	16	884	114	2	968	32	0	1000
	300	826	164	10	895	99	6	968	32	0	1000
(2,3)	30	841	147	12	911	83	6	974	25	1	1000
	60	877	117	6	932	66	2	988	12	0	1000
	100	865	128	7	932	67	1	988	12	0	1000
	300	902	89	9	952	46	2	990	10	0	1000
(3,10)	30	838	152	10	906	88	6	974	26	0	1000
	60	881	112	7	947	51	2	987	13	0	1000
	100	872	116	12	932	67	1	988	12	0	1000
	300	908	83	9	952	46	2	992	8	0	1000

3.6.6 Zufälliges Entfernen eines Siebtels der Messungen

In dieser Simulationsstudie werden zufällig ein Siebtel (ca. 14,3%) aller Messdaten vor der Auswertung entfernt. Die genauen Stellen der Punkte, die entfernt wurden, sind in Abbildung 3.26 auf Seite 282 zu sehen und in Anhang D.6.4 (S.396) aufgelistet. Die Systeme haben nun noch die Längen 26, 52, 86 bzw. 258.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Da sich die Ergebnisse des Augmented Dickey-Fuller-Tests nach zufälliger Entfernung eines Siebtels aller Messwerte in den drei Gruppen kaum unterscheiden, wie man in der Tabelle 3.103 auf Seite 283 sehen kann, werden sie im Folgenden zusammengefasst dargestellt.

Der ADF-Test erkennt in allen drei Gruppen auf allen Niveaus alle 1000 stationären Linearkombinationen bei den kointegrierten und stationären Systemen der Länge 300. Auch die Abweichungen bei den integrierten Systemen sind bei dieser Länge sehr gering. So werden selbst auf dem 0.1 Niveau mindestens 926 integrierte Linearkombinationen erkannt; auf dem 0.05 Niveau sind es mindestens 963 und bei 0.01 sogar 992-996.

Bei einer Länge von nun noch 86 (ursprünglich 100) Messpunkten wird die 950-Marke bei den stationären Linearkombinationen jedoch nur noch auf dem 0.1 Niveau erreicht oder knapp verfehlt. Schon auf dem 0.05 Niveau werden bei dieser Länge deutlich unter 900 und auf dem 0.01 Niveau nur noch knapp mehr als die Hälfte aller Linearkombinationen als stationär erkannt. Bei den kürzeren Systemen werden nicht mehr als 800 bei Länge 60 oder 400 bei Länge 30 als stationär eingeschätzt.

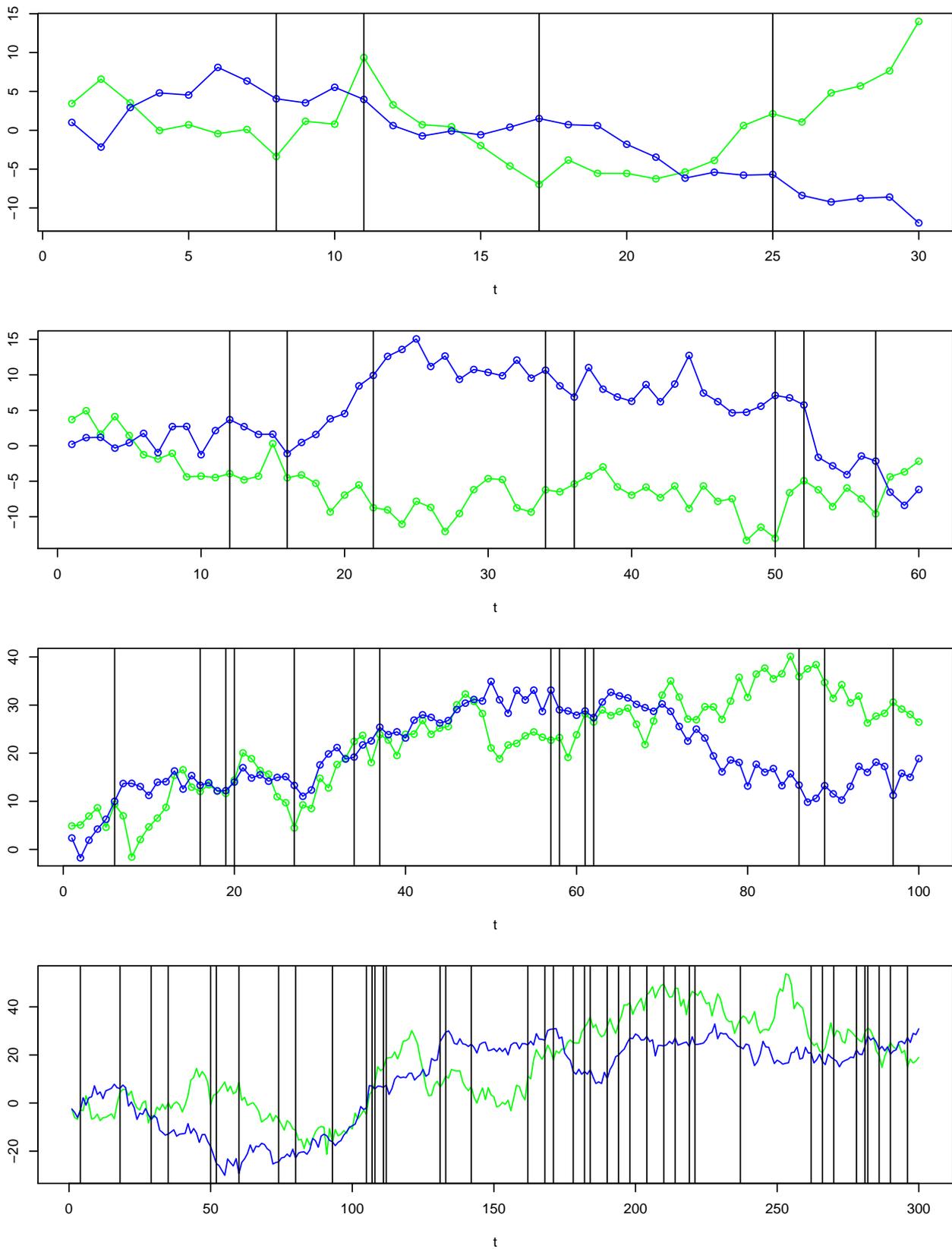


Abbildung 3.26: Zufällig entfernte Messpunkte bei den Längen 30, 60, 100 und 300 bei Entfernung eines Siebtels aller Daten.

Tabelle 3.103: Ergebnisse des ADF-Tests nach Entfernung eines Siebtels der Messungen: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	382	233	65	380	213	55	92	45	8
	60	797	631	296	788	604	294	72	29	4
	100	952	873	586	947	875	569	78	38	10
	300	1000	1000	1000	1000	1000	1000	66	37	6
(2,3)	30	402	241	54	379	223	63	88	38	6
	60	786	605	270	792	617	252	62	29	4
	100	961	874	583	947	868	583	78	43	9
	300	1000	1000	1000	1000	1000	1000	68	30	8
(3,10)	30	382	237	64	382	239	48	65	30	6
	60	760	582	252	775	618	247	71	29	1
	100	941	864	550	957	864	562	78	42	12
	300	1000	1000	1000	1000	1000	1000	74	35	7

Auf die Klassifikation der integrierten Linearkombinationen hat der Informationsverlust keine Auswirkung. Hier werden weiterhin zuverlässig in allen Gruppen die drei Niveaustufen eingehalten, so dass bei jeder Länge auf dem 0.05 und 0.01 Niveau ausreichend viele Einheitswurzeln identifiziert werden.

Johansentest

Kointegrierte Systeme Die Tabelle 3.104 auf Seite 284 zeigt, dass in allen Gruppen ab einer Länge von nun 52 (vorher 60) Datenpunkten auf allen Niveaus mehr als 800 Systeme als kointegriert kategorisiert werden (Ausnahme: Gruppe (1,1), Länge 60, 0.01 Niveau: 799). Ab einer Länge von 86 (zuvor 100) Beobachtungen werden auf dem 0.01 Niveau mehr als 98%, ab Länge 300 mehr als 99% der Systeme richtig eingestuft. Auf dem 0.05 Niveau liegen die Werte bei diesen Längen zwar über 900, jedoch nur bei den längsten Systemen in den Gruppen (2,3) und (3,10) auch über der 950-Marke. Auf dem 0.1 Niveau werden zwar noch ca. 88-92% der Systeme richtig eingeschätzt, diese Raten sind jedoch nach dem gesetzten Maßstab zu gering. Die Kointegrationsbeziehungen werden bei allen Längen in allen drei Gruppen richtig erkannt (siehe Tabelle D.26 im Anhang auf Seite D.26.)

Stationäre Systeme Wie in Tabelle 3.105 auf Seite 284 zu sehen ist, werden in allen Gruppen auf allen Niveaus alle 1000 Systeme der Längen 100 und 300 (nun 86 bzw. 258) als stationär erkannt. Bei 60 (nun 52) Datenpunkten gilt dies auf dem 0.1 Niveau in den Gruppen (2,3) und (3,10) ebenfalls. In der Gruppe (1,1) und in allen Gruppen auf dem 0.05 Niveau werden bei dieser Länge noch mehr als 99% richtig eingeschätzt. Auf dem 0.01 Niveau

Tabelle 3.104: Klassifikation 1000 kointegrierter Systeme nach zufälligem Entfernen eines Siebtels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	370	557	73	538	427	35	816	181	3	1000
	60	13	880	107	37	907	56	187	799	14	1000
	100	0	872	128	0	928	72	2	981	17	1000
	300	0	888	112	0	942	58	0	992	8	1000
(2,3)	30	340	582	78	509	466	25	801	198	1	1000
	60	7	899	94	23	930	47	188	801	11	1000
	100	0	896	104	0	947	53	1	985	14	1000
	300	0	920	80	0	961	39	0	995	5	1000
(3,10)	30	365	560	75	531	444	25	804	196	0	1000
	60	9	895	96	38	919	43	185	804	11	1000
	100	0	901	99	0	948	52	5	982	13	1000
	300	0	917	83	0	963	37	0	995	5	1000

Tabelle 3.105: Klassifikation 1000 stationärer Systeme nach zufälligem Entfernen eines Siebtels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	170	164	666	319	259	422	669	236	95	1000
	60	0	1	999	0	6	994	31	61	908	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	172	167	661	350	256	394	688	225	87	1000
	60	0	0	1000	1	5	994	28	46	926	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	171	181	648	360	248	392	693	219	88	1000
	60	0	0	1000	1	2	997	36	60	904	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

sind es noch mehr als 900 stationäre Linearkombinationen.

Tabelle 3.106: Klassifikation 1000 integrierter Systeme nach zufälligem Entfernen eines Siebtels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	803	183	14	900	98	2	961	39	0	1000
	60	828	166	6	896	103	1	984	16	0	1000
	100	803	185	12	892	106	2	968	32	0	1000
	300	827	162	11	905	91	4	972	28	0	1000
(2,3)	30	836	151	13	911	83	6	975	24	1	1000
	60	878	115	7	936	63	1	989	11	0	1000
	100	867	126	7	932	65	3	986	14	0	1000
	300	906	87	7	957	42	1	991	9	0	1000
(3,10)	30	832	159	9	905	91	4	978	22	0	1000
	60	878	116	6	947	52	1	989	11	0	1000
	100	873	116	11	936	63	1	988	12	0	1000
	300	909	83	8	953	45	2	993	7	0	1000

Integrierte Systeme Unabhängig von der Länge werden auf dem 0.01 Niveau in allen Gruppen mehr als 96% der Systeme als integriert klassifiziert (siehe Tabelle 3.106, S.285). Der Johantest erreicht in den Gruppen (2,3) und (3,10) sogar Klassifikationsraten um 98-99%. Auf dem 0.05 Niveau liegen die Werte immer noch um 900 in der ersten und bei ca. 93% in den anderen beiden Gruppen. Die 950-Marke wird nur in den letzten beiden Gruppen bei einer Länge von ursprünglich 300 Messpunkten erreicht. Die Klassifikationsraten auf dem 0.1 Niveau liegen bei allen Längen über 80% und in den beiden Gruppen (2,3) und (3,10) noch etwas besser, übersteigen jedoch in diesen Gruppen nur bei den längsten Systemen auch die 900.

Ergebnis: Zufälliges Entfernen eines Siebtels der Messwerte

Augmented Dickey-Fuller-Test Der ADF-Test zeigt ähnliche Ergebnisse wie bei regelmäßiger Entfernung jedes siebten Wertes, erkennt bei den längsten Systemen alle stationären und hinreichend viele integrierte und kann deshalb zur Unterscheidung eingesetzt werden.

Johantest Der Johantest klassifiziert auf dem 0.01 Niveau ab einer Länge von nun 86 Punkten über 98% der kointegrierten, alle stationären und mindestens 96% der integrierten Systeme korrekt.

3.6.7 Zufälliges Entfernen eines Fünftels der Messungen

Wenn zufällig 20% der Messwerte entfernt werden (siehe Abbildung 3.27, S.287), dann verkürzen sich die betrachteten Zeitreihensysteme auf 24, 48, 80 bzw. 240 Datenpunkte.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

Tabelle 3.107: Ergebnisse des ADF-Tests nach Entfernung eines Fünftels der Messungen: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	343	212	51	345	197	54	88	47	5
	60	708	539	218	714	545	208	79	30	5
	100	917	805	455	919	821	473	77	34	8
	300	1000	1000	999	1000	1000	1000	67	31	3
(2,3)	30	317	185	51	327	215	54	75	38	3
	60	701	533	197	701	520	197	71	33	3
	100	918	809	457	922	834	495	76	38	9
	300	1000	1000	999	1000	1000	1000	68	34	9
(3,10)	30	324	189	51	335	207	47	68	31	7
	60	690	501	177	712	523	186	80	36	5
	100	915	805	467	932	832	439	75	37	8
	300	1000	1000	1000	1000	1000	999	65	35	9

Auch für diese Simulationsstudie werden die Ergebnisse der drei Gruppen aus Tabelle 3.107 (S.286) gemeinsam beschrieben. Der ADF-Test zeigt erstmals auch bei den stationären Linearkombinationen der Länge 300 eine Abweichung. Ähnlich wie in der Studie, in der jeder 5. Datenpunkt entfernt wurde, gibt es in jeder Gruppe eine stationäre Linearkombination, die nun für integriert gehalten wird (vgl. Tabelle 3.90, S.266). In den ersten beiden Gruppen ist dies für ein kointegriertes System der Fall, in der Gruppe (3,10) bei einem stationären. Bei Länge 100 wird nun in keiner Gruppe selbst auf dem 0.1 Niveau die 950-Marke erreicht und auf dem 0.05 Niveau selbst 900 nicht überschritten. Auf die Ergebnisse bei den integrierten Daten hat das Entfernen der Messpunkte keine Auswirkung. Auf dem 0.05 Niveau werden in jeder Gruppe bei allen Längen bei mehr als 950 (meist sogar 960) Systemen die Einheitswurzel identifiziert; auf dem 0.01 Niveau gelingt dies zuverlässig bei mehr als 99%. Der ADF-Test kann bei einem Datenverlust von 20% bei kürzeren Reihen nicht mehr eingesetzt werden, wie dies auch bei den Originaldaten, die kürzer als 100 sind, der Fall war (vgl. Tabellen 3.7, S.164), 3.9, S.166) und 3.11, S.168)

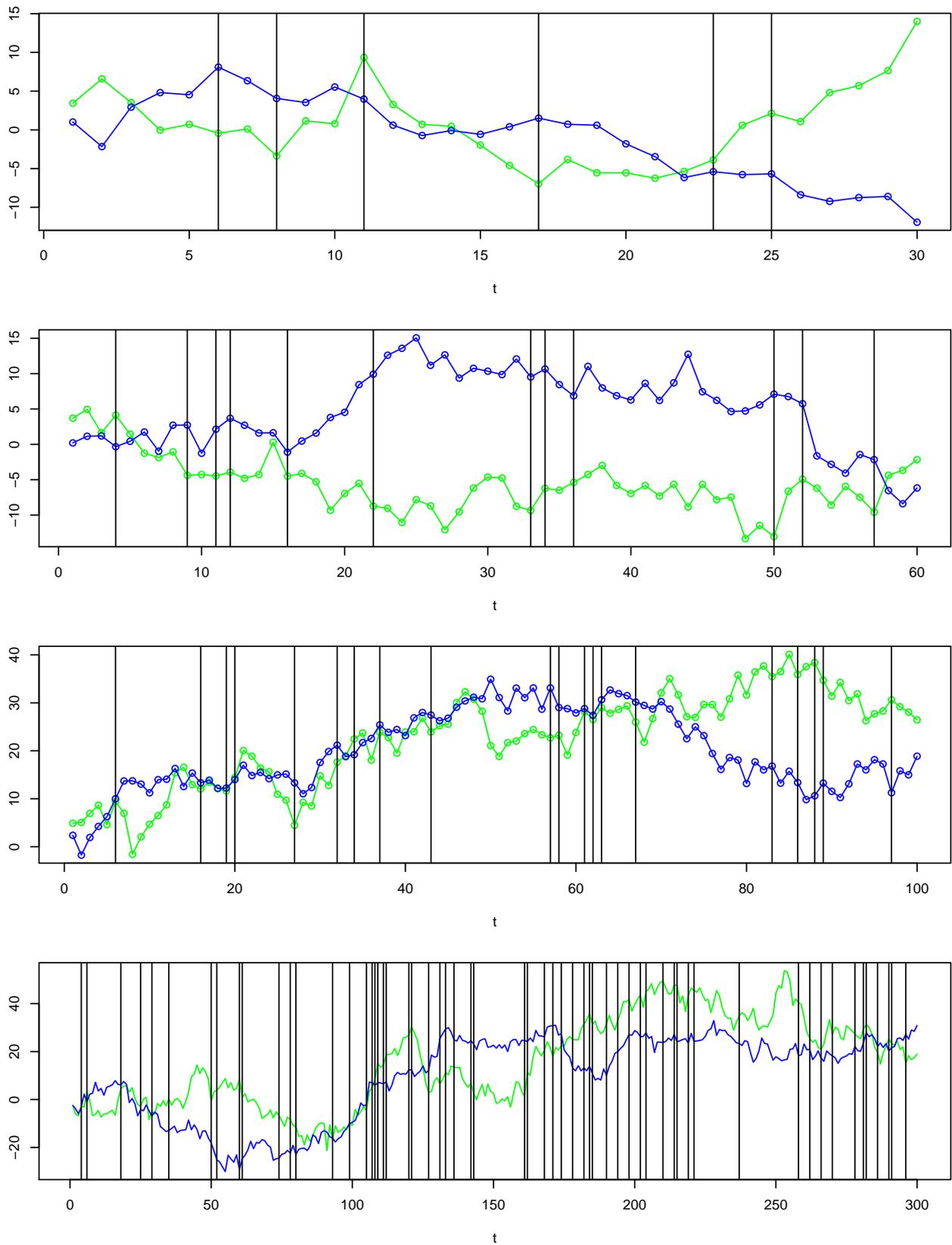


Abbildung 3.27: Zufällig entfernte Messpunkte bei den Längen 30, 60, 100 und 300 bei Entfernung eines Fünftels aller Daten.

Tabelle 3.108: Klassifikation 1000 kointegrierter Systeme nach zufälligem Entfernen eines Fünftels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	428	501	71	586	383	31	838	158	4	1000
	60	30	854	116	63	875	62	250	737	13	1000
	100	0	875	125	0	938	62	1	985	14	1000
	300	0	888	112	0	941	59	0	991	9	1000
(2,3)	30	379	551	70	556	424	20	832	168	0	1000
	60	23	875	102	64	887	49	260	727	13	1000
	100	0	901	99	0	954	46	4	982	14	1000
	300	0	919	81	0	960	40	0	995	5	1000
(3,10)	30	409	522	69	594	389	17	840	159	1	1000
	60	22	873	105	61	891	48	268	721	11	1000
	100	0	906	94	1	952	47	6	980	14	1000
	300	0	916	84	0	962	38	0	994	6	1000

Johansenstest

Kointegrierte Systeme In keiner der Gruppen werden mehr als 900 kointegrierte Systeme der Länge 60 als solche erkannt (siehe Tabelle 3.108, S.288). Ab einer Länge von ursprünglich 100 (jetzt 80) Beobachtungen klassifiziert der Johansenstest auf dem 0.01 Niveau in allen Gruppen mehr als 98%, bei Länge 300 mehr als 99% korrekt. In den Gruppen (2,3) und (3,10) werden mehr als 950 Systeme der Länge 100 bzw. 300 richtig eingeschätzt, in der Gruppe (1,1) wird dieser kritische Wert mit 938 bzw. 941 richtig kategorisierter Systeme knapp verfehlt. Auf dem 0.1 Niveau werden trotz des hohen Datenverlusts noch ca. 88-92% der Systeme richtig eingestuft. Auf die Schätzung des Kointegrationsparameters hat dies jedoch kaum einen Einfluss (siehe Tabelle D.27, S.384).

Stationäre Systeme Ähnlich wie in der vorangegangenen Simulationsstudie mit einem Siebtel entfernter Messpunkte werden weiterhin alle 1000 Systeme der Länge 300 und über 99% der Systeme mit 100 (jetzt noch 80) Datenpunkten als stationär erkannt. Die Ergebnisse in der Tabelle 3.109 auf Seite 289 zeigen außerdem, dass auch bei nun 48 Punkten auf dem 0.05 und 0.1 Niveau noch ausreichend viele stationäre System korrekt klassifiziert werden. Einzig auf dem 0.01 Niveau wird die 950-Marke bei dieser Länge nicht erreicht. Mit 24 Datenpunkten sind die verkürzten Systeme der ursprünglichen Länge 30 zu kurz, um sie richtig zu kategorisieren.

Integrierte Systeme Weiterhin erzielt die Klassifikation der integrierten Systeme die besten Erkennungsraten (siehe Tabelle 3.110, S.290). In keiner Gruppe wird mehr als ein System

Tabelle 3.109: Klassifikation 1000 stationärer Systeme nach zufälligem Entfernen eines Fünftels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	12	399	589	25	596	379	108	801	91	1000
	60	0	13	987	8	33	959	81	160	759	1000
	100	0	0	1000	0	0	1000	0	3	997	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	3	335	662	12	502	486	89	716	195	1000
	60	1	15	984	2	44	954	35	243	722	1000
	100	0	0	1000	0	0	1000	0	1	999	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	1	262	737	2	459	539	30	716	254	1000
	60	1	10	989	2	27	971	53	167	780	1000
	100	0	0	1000	0	0	1000	0	2	998	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

auf dem 0.01 Niveau für stationär gehalten. Bei allen Längen werden 97-98% der Systeme, bei Länge 300 in den Gruppen (2,3) und (3,10) sogar mehr als 99% richtig eingestuft. Nur wenige Ausnahmen werden für kointegriert gehalten. Auf dem 0.05 Niveau sind es jedoch schon über 50, in der Gruppe (1,1) sogar über 100 Systeme, bei denen nun ein gemeinsamer Trend ausgemacht wird. In den letzten beiden Gruppen werden bei Länge 300 jedoch noch 951 Systeme richtig erkannt. Die Klassifikationsraten auf dem 0.1 Niveau liegen bei über 80% und nehmen mit der Länge zu, so dass in den Gruppen (2,3) und (3,10) mehr als 900 Systeme der Länge 300 als integriert klassifiziert werden. Insgesamt sind sie jedoch zu gering, um den Test auf diesem Niveau einzusetzen.

Ergebnis: Zufälliges Entfernen eines Fünftels der Messungen

Augmented Dickey-Fuller-Test Durch das zufällige Entfernen von 20% aller Daten werden die Systeme der Länge 100 zu kurz, um den ADF-Test zuverlässig anzuwenden. Bei den Systemen der Länge 300 gibt es erste Schwierigkeiten auf dem 0.01 Niveau alle stationären Linearkombinationen zu erkennen. Außer bei diesen Einzelfällen ist der Test bei dieser Länge jedoch nahezu perfekt und erkennt auch auf dem 0.05 und 0.01 Niveau mehr als 96% der integrierten Linearkombinationen.

Johansentest Die Klassifikationsrate des Johansentest auf dem 0.01 Niveau bleibt bei den kointegrierten und integrierten Systemen der Länge 100 bzw. 300 trotz des großen Datenverlusts bei 98-99%. Für die stationären Systeme der Länge 100 kann auf diesem Niveau die perfekte Klassifikationsrate nicht mehr gehalten werden, sie liegt jedoch immer noch bei

Tabelle 3.110: Klassifikation 1000 integrierter Systeme nach zufälligem Entfernen eines Fünftels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	819	169	12	895	102	3	971	29	0	1000
	60	811	180	9	887	112	1	977	22	1	1000
	100	820	169	11	898	101	1	981	19	0	1000
	300	835	156	9	909	88	3	981	19	0	1000
(2,3)	30	835	152	13	915	79	6	978	21	1	1000
	60	858	133	9	920	79	1	989	11	0	1000
	100	877	113	10	932	67	1	988	12	0	1000
	300	902	91	7	951	47	2	990	10	0	1000
(3,10)	30	850	136	14	909	88	3	977	22	1	1000
	60	856	136	8	926	72	2	987	13	0	1000
	100	883	110	7	937	62	1	989	11	0	1000
	300	906	88	6	951	47	2	992	8	0	1000

mindestens 99,7%.

3.6.8 Zufälliges Entfernen eines Drittels der Messungen

In dieser abschließenden Simulationsstudie zu *missing values* werden nun nur noch zwei Drittel der Ausgangsdaten für die Auswertung verwendet. In der Abbildung 3.28 auf Seite 291 und im Anhang D.6.4 ist zu sehen, welche Messpunkte aus den Systemen entfernt wurden. Es bleiben Systeme der Längen 20, 40, 67 und 200.

Test mit bekannter Kointegrationsbeziehung - Augmented Dickey-Fuller-Test

In Tabelle 3.111 auf Seite 292 sind die Ergebnisse des Augmented Dickey-Fuller-Tests bei Berücksichtigung von zufällig ausgewählten zwei Drittel der Daten dargestellt. Schon bei den Originaldaten hatte der ADF-Test bei einer Länge von 200 auf dem 0.01 Niveau nicht alle stationären Linearkombinationen bei den kointegrierten und stationären Systemen erkannt (vgl. Tabelle 3.7, S.164). Dies tritt nun auch hier bei den Systemen der ursprünglichen Länge 300 (jetzt 200) auf. Und ebenfalls wie bei den Originaldaten werden bei einer Länge von jetzt 67 (vorher 100) deutlich weniger als 900 (meist sogar weniger als 800) stationäre Linearkombinationen erkannt. Bei den integrierten Systemen bewirkt auch ein Informationsverlust von einem Drittel der Daten keine Veränderung bei der Identifikation der Einheitswurzeln.

Der ADF-Test kann also bei einer Länge von 300 ursprünglichen Datenpunkten auch dann verwendet werden, wenn 100 Messwerte nicht berücksichtigt werden können. Für kürzere

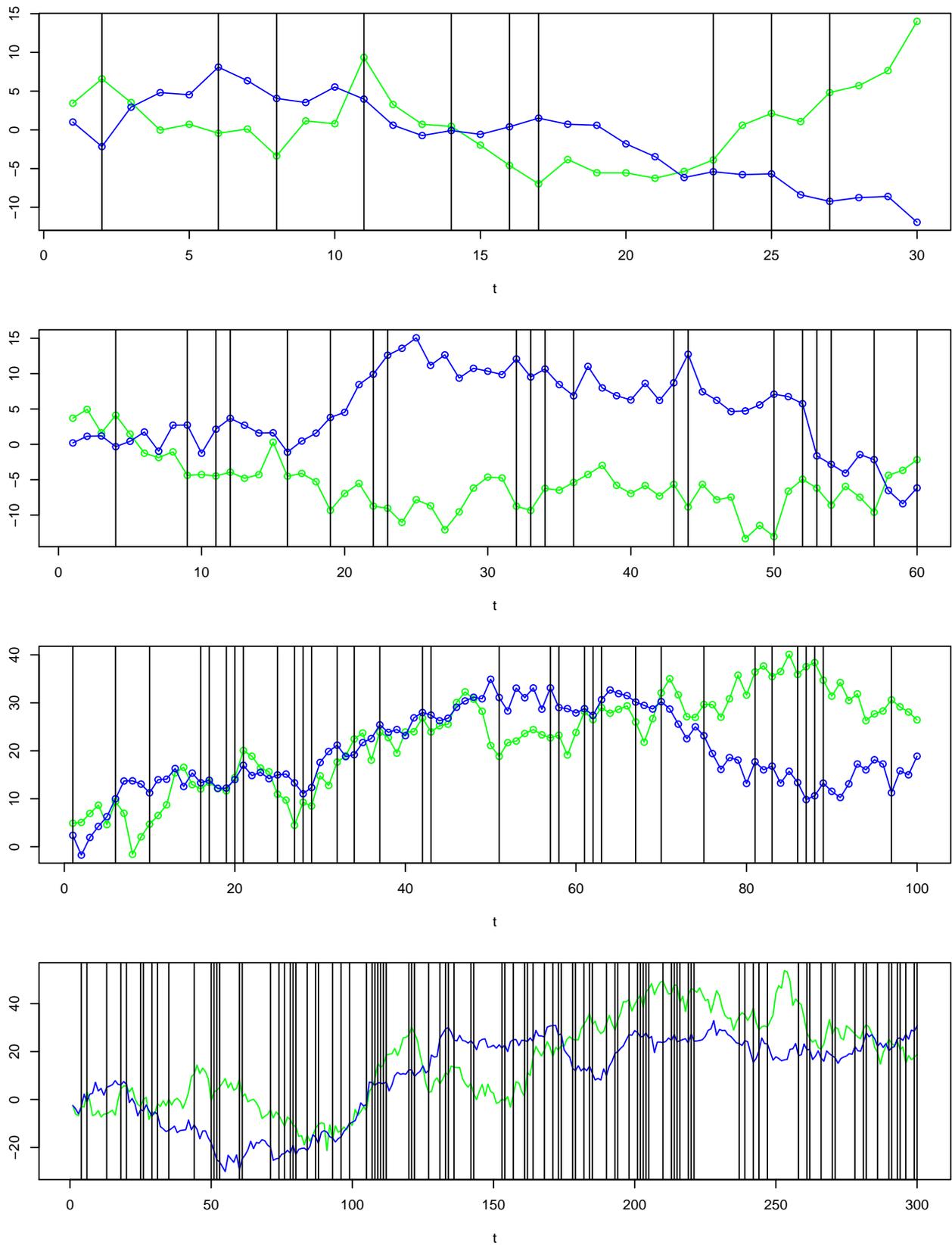


Abbildung 3.28: Zufällig entfernte Messpunkte bei den Längen 30, 60, 100 und 300 bei Entfernung eines Drittels aller Daten.

Tabelle 3.111: Ergebnisse des ADF-Tests nach Entfernung eines Drittels der Messungen: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	242	140	47	263	162	47	79	45	10
	60	551	365	114	535	353	119	87	33	5
	100	787	625	303	792	637	285	80	42	7
	300	1000	1000	998	1000	1000	996	67	37	5
(2,3)	30	242	145	39	229	140	48	78	43	11
	60	541	362	108	499	343	116	84	38	8
	100	807	653	271	802	648	282	85	48	4
	300	1000	1000	1000	1000	1000	1000	77	31	9
(3,10)	30	258	156	57	257	156	36	77	33	5
	60	528	344	93	504	344	105	78	36	11
	100	783	626	270	798	636	294	80	39	10
	300	1000	1000	998	1000	1000	999	70	30	8

Zeitreihen ist er jedoch zur Unterscheidung von stationären und integrierten Linearkombinationen nicht zuverlässig einsetzbar.

Johansenstest

Kointegrierte Systeme Der Johansenstest klassifiziert auf dem 0.01 Niveau in allen Gruppen ausreichend viele Systeme der Längen 100 (nun 67) und 300 (nun 200) als kointegriert (siehe Tabelle 3.112, S.293). Für die längsten Systeme sind es sogar mehr als 99%. Auf dem 0.05 Niveau wird die 950-Marke bei Systemen dieser Länge in der ersten Gruppe und bei Länge 100 auch in der zweiten Gruppe knapp verfehlt und in der Gruppe (3,10) sowie bei Länge 300 in Gruppe (2,3) knapp überschritten. Die Werte auf dem 0.1 Niveau liegen um 900.

Stationäre Systeme Auch wenn nur zwei Drittel der Daten verwendet werden, gelingt es immer noch alle stationären Systeme der Länge 300 und mehr als 99% der Länge 100 auf in allen Gruppen auf allen Niveaus richtig zu klassifizieren (siehe Tabelle 3.113, S.293). Die Werte für Systeme der ursprünglichen Länge 60, bei denen nur noch 40 Beobachtungen berücksichtigt werden, übertreffen auf dem 0.1 Niveau noch die 950-Marke, sind insbesondere auf dem 0.01 Niveau jedoch zu gering, um die Klassifikation zuverlässig zu verwenden.

Integrierte Systeme Die Ergebnisse des Johansenstest für integrierte Systeme haben sich durch den Datenverlust sogar noch leicht verbessert. In der Tabelle 3.114 auf Seite 294 ist zu sehen, dass sogar bei den kürzesten Systemen, die nun nur noch 20 Datenpunkte

Tabelle 3.112: Klassifikation 1000 kointegrierter Systeme nach zufälligem Entfernen eines Drittels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	522	400	78	663	313	24	865	131	4	1000
	60	71	834	95	153	807	40	465	520	15	1000
	100	1	889	110	3	941	56	30	957	13	1000
	300	0	893	107	0	949	51	0	994	6	1000
(2,3)	30	498	429	73	653	314	33	870	127	3	1000
	60	73	841	86	163	797	40	467	525	8	1000
	100	0	902	98	2	948	50	28	963	9	1000
	300	0	920	80	0	962	38	0	994	6	1000
(3,10)	30	485	437	78	642	333	25	852	143	5	1000
	60	91	830	79	173	788	39	458	535	7	1000
	100	1	904	95	2	953	45	16	973	11	1000
	300	0	923	77	0	966	34	0	993	7	1000

Tabelle 3.113: Klassifikation 1000 stationärer Systeme nach zufälligem Entfernen eines Drittels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	363	261	376	530	268	202	819	156	25	1000
	60	4	15	981	31	55	914	239	183	578	1000
	100	0	0	1000	0	0	1000	1	1	998	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	363	264	373	533	283	184	815	163	22	1000
	60	11	16	973	39	49	912	216	212	572	1000
	100	0	0	1000	0	0	1000	1	7	992	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	381	276	343	536	282	182	813	163	24	1000
	60	11	12	977	37	56	907	235	206	559	1000
	100	0	0	1000	0	0	1000	1	3	996	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000

Tabelle 3.114: Klassifikation 1000 integrierter Systeme nach zufälligem Entfernen eines Drittels der Messungen

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
(1,1)	30	763	200	37	848	139	13	950	48	2	1000
	60	857	138	5	912	86	2	981	18	1	1000
	100	841	147	12	908	87	5	981	19	0	1000
	300	867	122	11	931	65	4	986	14	0	1000
(2,3)	30	798	174	28	880	112	8	966	34	0	1000
	60	884	110	6	946	51	3	990	10	0	1000
	100	890	100	10	943	54	3	987	13	0	1000
	300	907	84	9	955	44	1	993	7	0	1000
(3,10)	30	775	196	29	864	125	11	957	42	1	1000
	60	874	120	6	935	64	1	992	8	0	1000
	100	889	102	9	934	63	3	990	10	0	1000
	300	916	77	7	957	42	1	994	6	0	1000

haben, mehr als 950 Systeme richtig als integriert klassifiziert werden. Bei den längeren Systemen aller Gruppen liegen die Werte über 98% bzw. 99%. Auf den anderen Niveaus werden ebenfalls noch die meisten Systeme als integriert erkannt. Jedoch übersteigen nur noch die Klassifikationsraten der längsten Systeme in den Gruppen (2,3) und (3,10) die 950-Marke.

Ergebnis: Zufälliges Entfernen eines Drittels der Messungen

Augmented Dickey-Fuller-Test Auch wenn ein Drittel der Daten zufällig wegfällt, kann der ADF-Test für die Systeme der Länge 300 (nun 200) stationäre von integrierten Linearkombinationen auf dem 0.05 und 0.01 Niveau zuverlässig unterscheiden.

Johansen-Test Die Klassifikation des Johansen-Tests für die Systeme der Länge 100 und 300 ändert sich auch bei diesem enormen Datenverlust nur unmerklich. Auf dem 0.01 Niveau werden beinahe alle Systeme der Länge 300 richtig klassifiziert. Bei Länge 100 (nun 67) liegt die Rate der richtig erkannten kointegrierten Systeme bei 95,7-97,3%. Es werden weiterhin mehr als 990 stationäre und mehr als 980 integrierte Systeme erkannt.

3.6.9 Ergebnis: Messung mit Missing Values

Durch das Entfernen von Messwerten verkürzen sich die Zeitreihen. Dadurch verliert man Informationen. Die Simulationsstudien zeigten jedoch, dass es weder für den Augmented

Dickey-Fuller-Test noch für den Johansen-Test einen Unterschied macht, in welchen Abständen die Daten entfernt wurden.

Die Ergebnisse beider Tests waren nur von der verbleibenden Anzahl der Datenpunkte abhängig. Daraus lässt sich schließen, dass man den zufälligen Verlust von Daten oder eingeschränkte Erhebungsmöglichkeiten dadurch ausgleichen kann, dass man den Erhebungszeitraum verlängert, bis man eine ausreichende Anzahl von Messzeitpunkten erreicht hat.

3.7 Ergebnisse der Simulationsstudien

Die in diesem Kapitel vorgestellten Simulationsstudien haben viele interessante Ergebnisse geliefert, aus denen man wichtige Erkenntnisse für die Anwendbarkeit der Kointegrationsanalyse bei psychologischen Daten gewinnen kann.

Eine entscheidende Größe bei der Versuchsplanung einer Längsschnittstudie oder einer Verlaufsanalyse ist die Zeit. Allgemein findet man häufig eine Empfehlung von mindestens 30 Datenpunkten für zeitreihenanalytische Auswertungen. Die eingesetzten Schätzverfahren basieren auf der Korrelation der Daten. Dabei wird im Standardfall jedoch beispielsweise die Unabhängigkeit vom Fehlerterm vorausgesetzt. Diese Forderung ist in Zeitreihen nicht sinnvoll, da man insbesondere an den inneren Abhängigkeiten (Autokorrelationen, Kap. 1.9) interessiert ist. Diese Korrelationsstruktur führt zu einer Verzerrung der Schätzer (vgl. S.36), so dass sie nicht mehr erwartungstreu sind. Durch die asymptotische Normalverteilung bleibt jedoch die Eigenschaft der Konsistenz erhalten. Längeneempfehlungen wie 30 oder 50 Datenpunkte erinnern an die Mindestanzahl von Beobachtungen im Gesetz der großen Zahlen oder dem Zentralen Grenzwertsatz.

In Kapitel 3.2 erwies sich der Phillips-Perron-Test schon ab 30 Datenpunkten als sehr zuverlässig, um eine falsche Nullhypothese „Prozess enthält Einheitswurzel“ abzulehnen. Für die Akzeptanz einer richtigen Nullhypothese konnte er jedoch erst ab 200 Beobachtungen und nur in den beiden Gruppen mit Varianzunterschied zwischen den beiden Zeitreihen sein Niveau halten. In der folgenden Studie mit simulierter Messung auf Intervallskalen verschiedener Stufenanzahl erkannte er zwar weiterhin selbst bei den Systemen mit Länge 30 mehr als 99% der stationären Linearkombinationen, stuft jedoch auf der 10-stufigen Intervallskala zum Teil mehr als die Hälfte, auf der 5-stufigen teilweise über 80% aller integrierten Linearkombinationen als stationär ein. Der Phillips-Perron-Test kann also nicht dazu verwendet werden, ein System auf eine bekannte Kointegrationsbeziehung zu überprüfen. Insbesondere dann nicht, wenn man eine Skala mit eingeschränkter Stufenzahl verwendet.

Dazu ist jedoch der Augmented Dickey-Fuller-Test sehr gut geeignet. Unabhängig von der Länge der Zeitreihe hielt er sein Niveau auch den nach simulierten Messungen stets ein. Im Gegensatz zum Phillips-Perron-Test nimmt er die Nullhypothese jedoch bei Systemen mit einer Länge kleiner als 100 Datenpunkten zu häufig fälschlicherweise an. Für einen Test auf

eine bekannte Kointegrationsbeziehung mit dem ADF-Test benötigt man folglich mindestens 100 Messungen. Für die beiden vorgestellten Unit-Root-Tests gibt es jedoch auch zahlreiche Abwandlungen (z.B. Maddala & Kim, 2007). Hier ist zu prüfen, ob einige dieser Verfahren bereits bei kürzeren Systemen bessere Ergebnisse liefern.

Der Johansen-Test kann zur Klassifikation von bivariaten integrierten, kointegrierten und stationären Systemen verwendet werden, da er zunächst den Kointegrationsrang eines Systems bestimmt und anschließend die Kointegrations- und Ladungsmatrizen schätzt (vgl. Kap.2.7). Die Simulationsstudien haben gezeigt, dass er ab einer Länge von 60-70 Beobachtungen sehr zuverlässig eingesetzt werden kann. Bei täglicher Messung entspricht dies einem Erhebungszeitraum von etwa zwei Monaten. Diese Ergebnisse werden im Zusammenhang multivariater Zeitreihenmodellierung (Stadnytska & Gruber, 2011) veröffentlicht werden. Da er die Kointegrationsbeziehung zwischen den beiden Zeitreihen sehr genau schätzt (vgl. Tabellen der Schätzung von β in den entsprechenden Unterkapiteln), kann er auch zur Überprüfung einer bestehenden Kointegrationsbeziehung verwendet werden, indem man mit Hilfe der Verteilungsfunktion von β die Schätzungen verschiedener Systeme vergleichen kann. Der ADF-Test benötigt für eine ähnliche Analyse ca. zusätzliche 40 Datenpunkte für eine ähnliche Analyse.

Da die Voraussetzungen an die Skalenqualität der Messdaten nicht mit mathematisch analytische Methoden geprüft werden können (Bortz & Döring, 1995) und die meisten Erkenntnisse zur Kointegrationsanalyse in der wirtschaftswissenschaftlichen Forschung entstanden sind, gibt es bislang keine Aussagen über die Anzahl der benötigten Stufen auf einer Skala oder deren Skalenniveau. Aus den Ergebnissen der Simulationen in Kapitel 3.3 kann man schließen, dass eine Messung auf einer 10-stufigen Intervallskala die Klassifikationsergebnisse des Johansen-Test oder das Erkennen der stationären Linearkombinationen durch den ADF-Test kaum verändert. Auch auf einer 7- oder 5-stufigen Intervallskala sind die Ergebnisse noch ausreichend gut, um die Tests einsetzen. Bei nur 3 Stufen gibt es jedoch Probleme. Zwar werden die meisten Systeme noch richtig klassifiziert. Die Konstruktion der Skala führt jedoch dazu, dass der Johansen-Test bei vielen Systemen mit unterschiedlicher Varianz nicht mehr ausgeführt werden kann, da einer Zeitreihe durch die Messung ein (fast) konstanter Verlauf zugeordnet wird. Stationäre Systeme werden noch sehr zuverlässig erkannt. Die Klassifikation von integrierten und kointegrierten Systemen gelingt jedoch nicht mehr in ausreichender Anzahl.

Der Effekt, dass der Johansen-Test nicht mehr ausgeführt werden kann, da eine Zeitreihe nach der simulierten Messung als konstant dargestellt wird, tritt bei Ceiling- und Flooreffekten besonders stark auf. Die Ergebnisse aus Kapitel 3.4 zeigen deutlich, dass eine Abdeckung des gesamten Wertebereichs durch die Skalenwerte entscheidend ist. Durch das Abschneiden der Randbereiche kann die Varianzentwicklung nicht mehr richtig abgebildet werden, so dass stochastische Trends nicht mehr erkannt bzw. verschiedene integrierte Prozesse nicht mehr unterschieden werden können. Dadurch wird das Verhalten der Tests so stark verändert, dass sie nicht mehr effektiv eingesetzt werden können. Die Ergebnisse der simulierten Messung

auf Ordinalskalenniveau lassen sich ebenfalls vor allem auf Ceiling- und Flooreffekte zurückführen. Bei der Unterscheidung zwischen integrierten und kointegrierten Systemen gab es auf diesen Skalen ähnliche Problemen wie bei Randeffekten. Da die eingesetzten Statistiken vor allem auf Autokorrelationen basieren und diese auf Ordinalskalenniveau sinnvoll interpretiert werden können, spricht theoretisch nichts gegen den Einsatz von einer Ordinalskala, die Randeffekte vermeidet.

Besonders bei den integrierten Systemen gab es häufiger die Beobachtung, dass sich die Klassifikationsergebnisse bei Systemen der Länge 300 durch die Messung auf Skalen weniger Stufen verschlechterten. Durch die Einschränkung des Wertebereichs auf eine diskrete Anzahl von Merkmalsausprägungen kann der für integrierte Prozesse typische Verlauf nicht mehr korrekt abgebildet werden. Diese Prozesse zeichnen sich dadurch aus, dass die Varianz mit fortschreitender Zeit immer stärker anwächst und sich der Prozess immer weiter von seinem Ausgangspunkt entfernt. Da dieses Verhalten auf einer Skala mit wenigen Stufen ausbleibt, werten die Tests dies eher als einen regelmäßigen, also stationären Verlauf. Bei Messungen mit wenigen Stufen sollten also weitere Untersuchungen über eine maximale Länge folgen.

Falls man Zeitreihendaten für eine Kointegrationsanalyse verwendet möchte, sollte man bei der Skalenkonstruktion einige Kriterien beachten. Die Skalenwerte sollten so gewählt sein, dass sie den gesamten Bereich aller möglichen Merkmalsausprägungen abdecken, um Ceiling- und Flooreffekte zu vermeiden. In den Simulationsstudien wurde jeweils eine Skala für ein bivariates System verwendet, auch wenn die Verläufe starke Varianzunterschiede aufwiesen. Vermutlich wird eine Versuchsperson diesen Varianzunterschied bei ihrer Bewertung ausgleichen, so dass Systeme mit ähnlicher Varianz entstehen, wie wir sie in Gruppe (1,1) betrachtet haben. Jedes Merkmal sollte jedoch auf einer eigenen passenden Skala gemessen werden, die alle Ausprägungen berücksichtigt. Durch die Konstruktion der Testdaten für die Simulationsstudien (siehe Anhang D.4, S.386) werden zwei kointegrierte Reihen auf einen identischen Verlauf abgebildet, wenn man jede Zeitreihe auf einer eigenen Skala misst (siehe Abbildung 3.29). Interessant wäre jedoch das Klassifikationsergebnis, wenn Skalen mit unterschiedlicher Stufenzahl verwendet werden.

Als Ratingskalen werden häufig sogenannte *Likert-Skalen* eingesetzt, auf denen die einzelnen Skalenstufen verbal repräsentiert werden, z.B. „trifft zu“ - „trifft eher zu“ - „weder noch“ - „trifft eher nicht zu“ - „trifft nicht zu“. Solche Skalen haben typischerweise Ordinalskalenniveau. Ein Fragebogen besteht jedoch meist aus mehreren solcher Likert-Skalen und man nimmt an, dass die Summe solcher ordinalskalierten Items Intervallskalenniveau erhält. In jedem Fall wird dadurch die Variabilität und im Sinne der Simulationen die Stufenanzahl größer.

Ein weiteres entscheidendes Ergebnis der Simulationsstudien ist, dass das Auslassen von Messwerten die Klassifikation kaum beeinflusst. Entscheidender als die Äquidistanz der Zeitpunkte der Messungen ist die absolute Anzahl an verbleibenden Werten. Selbst bei zufälligem Ausfall von einem Drittel aller Messungen konnten der ADF- und Johansenstest noch

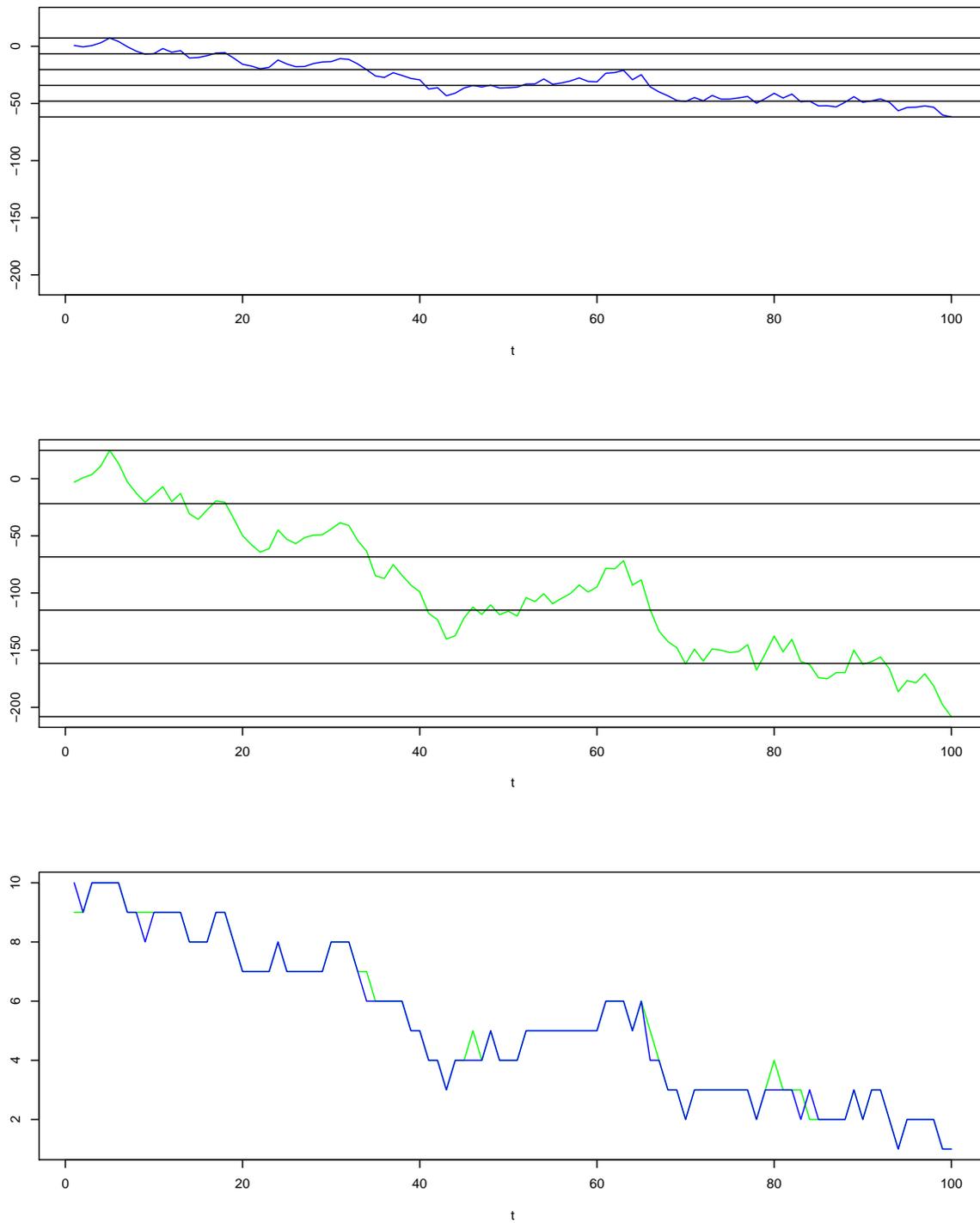


Abbildung 3.29: Kointegriertes System der Länge 100 aus der Gruppe (3,10): Wenn jede Zeitreihe auf einer eigenen Skala gemessen wird, werden beide auf einen identischen Verlauf abgebildet.

zuverlässig eingesetzt werden. Der ADF-Test benötigt jedoch weiterhin mindestens 100 Datenpunkte und der Johantest etwa 70. Um dies zu erreichen, kann der Erhebungszeitraum um die entsprechende Anzahl von Messungen erweitert werden. Wichtig ist jedoch, dass bei Fehlen eines Messwertes die Daten zum selben Zeitpunkt aus dem gesamten System entfernt werden müssen, um zu gewährleisten, dass man zeitgleiche Messungen in einem System hat.

Durch Messausfälle kann es vorkommen, dass in einer Panelstudie die Systeme verschiedener Versuchspersonen unterschiedlich lang sind. Da wir gesehen haben, dass die Schätzung des Kointegrationsparameters jedoch von der Länge des Systems kaum beeinflusst wird, ist ein Vergleich der Kointegrationsbeziehungen auch dann noch möglich, wenn sich die Systeme in ihrer Länge unterscheiden.

Kointegrationsanalyse kann auch eingesetzt werden, wenn in regelmäßigen oder unregelmäßigen Abständen keine Messdaten erhoben werden können. Entscheidend ist die absolute Zahl von Messpunkten, wobei für den Johantest mindestens 60-70 Beobachtungen vorausgesetzt werden. Außerdem sollte die Merkmalsausprägung jeder Variablen durch die Skala abgedeckt werden, da die größten Effekte auf das Klassifikationsverhalten durch Decken- und Bodeneffekte bewirkt wurden. Je größer die Variabilität der Messungen ist, desto besser können stochastische Trends dargestellt werden. Bei Systemen der Länge 100 ist dies auf 5 Stufen jedoch noch ausreichend gut möglich.

Als Gesamtergebnis kann man festhalten, dass die Voraussetzungen an die Qualität der Daten für Kointegrationsanalyse ähnlich sind wie bei anderen Standardverfahren der Psychologie.

Ausblick

Ziel der vorliegenden Dissertation war es dem sozialwissenschaftlichen Leser den Zugang zur Kointegrationsanalyse zu ermöglichen und zu prüfen, ob die dargestellten Verfahren auch bei einer Datenqualität eingesetzt werden können, wie sie für psychologische Daten typisch sind.

Durch die auf das für Kointegrationsanalyse Wesentliche beschränkte Einführung in die allgemeine Zeitreihenanalyse ist ein Einstieg in die Thematik der Kointegration schnell möglich. Basierend auf der Unterscheidung von stationären, trendstationären und integrierten Prozessen durch das Vorhandensein eines deterministischen oder stochastischen Trends wurden kointegrierte Systeme, die gemeinsame stochastische Trends enthalten, eingeführt. Im dreidimensionalen Modell wurden verschiedene Konstellationen gemeinsamer Trends veranschaulicht, bevor in den verschiedenen Darstellungsmöglichkeiten eines multivariaten kointegrierten Systems mathematische Grundlagen der Kointegration beleuchtet wurden. Nachdem durch das konstruierte Beispiel einer Betrunkenen und ihres Hundes die Fehlerkorrekturmechanismen anschaulich gemacht wurden, folgte eine ausführliche Betrachtung des Fehlerkorrekturmodells, das sowohl langfristige Regelmäßigkeiten als auch kurzfristige Abweichungen darstellen kann und somit zur Überprüfung psychologischer Theorien und Modelle wie der Kollusionstheorie (Willi, 1984) gut geeignet ist. Außerdem wurde erläutert, wie zusätzliche deterministische Terme in das Fehlerkorrekturmodell aufgenommen werden können, und gezeigt, welche Auswirkung dies bei der Interpretation der Kointegrationsbeziehung haben kann. Damit der Anwender Vertrauen in die Testmethoden auf Kointegration fassen kann, wurde insbesondere der Johantest mit vielen analytischen Details dargestellt, so dass der Ablauf des Verfahrens und die Interpretation der Ergebnisausgabe in Softwarelösungen verständlich wurde. An einem System von psychologischen Variablen und körperlichen Symptomen wurde der Einsatz des Johantest und des Fehlerkorrekturmodells veranschaulicht.

Durch komplexe Simulationsstudien wurden anschließend entscheidende Fragen für die Anwendung der Methoden bei psychologischen Daten beantwortet. Da die Verfahren auch bei nicht äquidistanten Zeitreihen, auf Skalen mit wenigen Stufen und sogar bei einem Drittel Datenverlust noch sehr gute Ergebnisse erbrachten, können sie in der psychologischen Forschung sinnvoll eingesetzt werden, ohne die üblichen Anforderungen an die Datenqualität zu verschärfen.

In den Simulationsstudien wurden stets bivariate Systeme reiner stationäre $I(0)$ - oder differenzstationärer $I(1)$ -Modelle betrachtet und auf Kointegration untersucht. In Kapitel 2.6 (S.114) wurde jedoch gezeigt, dass man auch deterministische Anteile in die Kointegrationsbeziehung aufnehmen kann. Eine Zeitreihe kann folglich außer einem konstanten Mittelwert und eines integrierten Anteils zusätzlich einen deterministischen Trend beinhalten. Die Un-

terscheidung von trendstationären und differenzstationären Anteilen ist entscheidend, da sie die Interpretation des Modells in der Anwendung verändert (Kirchgässner & Wolters, 2007). Folgt die Zeitreihe einem linearen Trend, wird sie nie zu ihrer Ausgangsposition zurückkehren. Bei einem differenzstationären Prozess gibt es jedoch eine positive Wahrscheinlichkeit für diese Rückkehr (Klenke, 2006, Kap.17).

Es ist möglich mit dem Johansentest Kointegration auch dann noch zu erkennen, wenn es zusätzliche deterministische Anteile im System gibt. Zusätzlich ist es möglich, auch deterministische saisonale Einflüsse zu berücksichtigen⁴⁵. Darüber hinaus kann es jedoch stochastische saisonale Aspekte geben, die der Johansentest nicht mehr aufdecken kann.

Saisonale Kointegration kann dann vorliegen, wenn es im System saisonale Einheitswurzeln gibt. Dann ist ein Differenzieren der Zeitreihen nicht mehr ausreichend, um sie in stationäre Prozesse zu transformieren, vielmehr muss ein saisonaler Operator $S(L) = (1 - L^S)(1 - L)^{-1}$ angewandt werden (Engle, 1991, Kap.14). Hylleberg, Engle, Granger und Yoo (1990) haben das Prinzip des Augmented Dickey-Fuller-Tests so erweitert, dass saisonale Unit Roots identifiziert werden können. Darüber hinaus gibt es weitere Testverfahren anderer Autoren (Dickey, Hasza & Fuller, 1984; Canova & Hansen, 1995). Die Gruppe entwickelte außerdem ein Testverfahren für saisonale Kointegration (Engle, Granger, Hylleberg & Lee, 1993) und Lee (1992) ergänzte eine Abwandlung des Johansentests bei saisonalen Einheitswurzeln. Bei Maddala und Kim (2007, 12. Seasonal Unit Roots and Seasonal cointegration) ist eine ausführliche vergleichende Darstellung dieser Verfahren zu finden.

Auch in psychologischen Daten lassen sich Zyklen und Perioden vermuten. Die hohe Autokorrelation auf dem siebten Lag in der Zeitreihe zur Aggressivität der Ehefrau (siehe Kapitel 2.8, S.138) könnte auf einen wöchentlichen Zyklus hinweisen. Um Saisonalität jedoch genau analysieren zu können, reicht der hier dargestellte Ansatz der Zeitreihenanalyse nicht aus. Vielmehr wird der Einsatz von spektralanalytischen Verfahren wie der Fourieranalyse notwendig (siehe Haßler, 2007; Rinne & Specht, 2002 für eine Einführung).

Das dynamische Verhalten vieler psychologischer Phänomene lässt sich nicht hinreichend genau über stochastische Trends ganzzahliger Integrationsordnung erklären. Zeitreihen, die durch Anwendung des Differenzenoperators mit einem Integrationsgrad von $0 < |d| < 0.5$ in einen stationären Prozess überführbar sind, nennt man *fraktional* integriert (z.B. Beran, 1998). Lässt sich der stationäre Prozess in einem ARMA-Modell darstellen, spricht man auch von einem *ARFIMA*-Prozess (z.B. Fan & Yao, 2003, S.66, Def.2.10). Solche Prozesse weisen ein langes Gedächtnis auf. Diese Eigenschaft findet man in vielen psychologischen Prozessen (siehe Wagenmakers et al. (2004) für einen Überblick). Messungen der Hirnaktivität via EEG zeigen solche Muster, ebenso wie die Ergebnisse diverser kognitiver Experimente zu Wahlentscheidungen oder Schätzverfahren und visueller Reizverarbeitung (Gilden, 1997, 2001; Farrell, Wagenmakers & Ratcliff, 2006). In klinischen Studien konnten Peterson und Leckman (1998) zeigen, dass die Ticks beim Tourette-Syndrom einem dynamischen Verlauf

⁴⁵In der Funktion `ca.jo` durch den Parameter `season`.

folgen, der sich durch fraktionale Integration erklären lässt. Ähnliches gilt für die Stimmungsschwankungen bei bipolarer Störung (Gottschalk, Bauer & Whybrow, 1995).

Wenn die Linearkombination zweier integrierter Zeitreihen zwar nicht mehr vom Integrationsgrad $I(1)$, aber auch noch nicht stationär $I(0)$ ist, also einen Integrationsrang $0 < d - b < 1$ besitzt, spricht man von *fraktionaler Kointegration* (Baillie & Bollerslev, 1994). Im Sinne von Dueker und Startz (1998) schließt der Begriff auch den allgemeinen Fall mit ein, in dem integrierte Prozesse der allgemeinen Integrationsordnung d eine Linearkombination besitzen, die einen nicht-ganzzahligen Integrationsgrad $d - b$ mit $b \geq 1$ besitzt. Aktuell werden Verfahren diskutiert, die fraktionale Kointegration mit linearen Trends verbinden (Hassler, Marmol & Velasco, 2008) oder auch saisonale fraktionale Kointegration identifizieren können (Gil-Alana, 2008).

Ähnlich wie bei der Regressionsanalyse kann die langfristige Beziehung der Zeitreihen eines Systems eine komplexere Gestalt haben, als dies über eine lineare Gleichung abzubilden ist. So kann es ein *nicht-lineares* Gleichgewicht, $Z_t = f(Y_t)$, zwischen verschiedenen Zeitreihen geben. Dies führt zu einer Verallgemeinerung des Konzepts der Kointegration (z.B. Karlsen, Myklebust & Tjøstheim, 2007). Im bivariaten Fall ist der Kointegrationsraum als Attraktor ist nun nicht länger eine Gerade, sondern eine beliebige Kurve im zweidimensionalen Raum. Wenn man überhaupt keine Voraussetzungen an den Daten generierenden Prozess stellt, müssen neue Kointegrationstests verwendet werden. Ähnlich wie bei der *nicht-parametrischen Regression* (Tsybakov, 2009) wird die Kointegrationsbeziehung als allgemeiner funktionaler Zusammenhang mittels Kernschätzern aus den Daten gewonnen (Bierens, 1997).

Alle diese Modelle haben gemeinsam, dass sie zur Erklärung des aktuellen Wertes einer Einzelreihe nur vergangene Werte des Gesamtsystems berücksichtigen. Zwar wird der Lag-Einfluss der übrigen Komponentenprozesse miteinbezogen, gleichzeitige Wechselwirkungen werden jedoch nicht erfasst. Um diese kontemporären Faktoren analysieren zu können, müssen die Modelle erweitert werden. Die *strukturellen vektorautoregressiven und Fehlerkorrekturmodelle* (*sVAR bzw. sVECM*) basieren auf der Impulsantwortanalyse (siehe Kapitel 1.15, S.52) und setzen eine Wold-Kausalstruktur voraus (Lütkepohl, 2005, Kap.9). Für die Schätzung der Modelle gibt es verschiedene Ansätze, die sogenannten *A- und B-Modelle*. Die Restriktionen an das Modell müssen so gewählt werden, dass eine eindeutige Lösung existiert. Dies ist nur dann möglich, wenn man die Einzelprozesse eindeutig untereinander darstellen kann. Diese Verfahren sind bei Stadnytska (2010) sehr anschaulich dargestellt. Hier werden auch die Zeitreihendaten aus Kapitel 2.8 mit Hilfe dieser Methode auf gleichzeitige Einflüsse untersucht.

Die Simulationsstudien zur Skalenqualität konnten erste wichtige Antworten für den psychologischen Anwender geben. So konnte ein Orientierungswert für die Mindestlänge eines Systems für die Kointegrationsanalyse gegeben werden. Durch die Beschränkung des Wertebereichs auf wenige Stufen trat bei längeren Zeitreihen jedoch auch der Effekt auf, dass sich die Klassifikation verschlechterte, da stochastische Trends nicht ausreichend gut abgebil-

det werden konnten. Daraus ergibt sich die Frage einer maximalen Länge, die durch weitere Untersuchungen zu klären ist.

Außerdem wurde deutlich, dass es notwendig ist, jede Zeitreihe auf einer geeigneten Skala zu messen, die den gesamten Wertebereich abdeckt. Durch die Konstruktion der Skalen anhand von Minimal- und Maximalwerten beider Reihen, wurde die Zeitreihe mit der geringeren Varianz häufig zu stark geglättet, so dass ihr Verlauf durch die simulierte Messung als (beinahe) konstant erschien. Bei ähnlicher Skalenkonstruktion für jeden Einzelprozess erhält man durch das Verfahren jedoch bei den kointegrierten Systemen einen fast identischen Verlauf, so dass der Einsatz eines Kointegrationsverfahren nicht notwendig ist, da schon die grafische Betrachtung den identischen Verlauf zeigt (siehe Abbildung 3.29, S.298). Interessant wäre es nun zu prüfen, welche Ergebnisse sich bei der Klassifikation ergeben, wenn die Zeitreihen auf Skalen mit unterschiedlicher Stufenzahl gemessen werden.

Eine wichtige Erkenntnis ist zudem, dass die Äquidistanz nicht so entscheidend ist wie die Länge einer Zeitreihe. Auch bei einem Drittel fehlender Daten gab es bei ausreichender Länge kaum Abweichungen in der Klassifikation. Allerdings wurde bei dieser Simulationsstudie der Wertebereich nicht zusätzlich eingeschränkt. Da der Anwender bei psychologischen Daten jedoch auch mit *missing values* bei Daten auf einer wenige Stufen umfassenden Skala umgehen muss, könnte eine weitere Simulationsstudie einen wichtigen Erkenntnisgewinn und eine Anleitung zum Umgang mit diesem Problem bieten. Im Rahmen einer weiteren eigener Veröffentlichung sollen diese Fragen geklärt werden (Gruber & Stadnytska, 2011).

Nicht allein durch moderne Erhebungsverfahren wie EEG- und fMRT-Messungen, digitalisierte Tagebücher oder physiologischer Messungen gibt es eine Vielzahl von Anwendungsbereichen für zeitreihenanalytische Verfahren im kognitiven, psychosomatischen und Therapie begleitenden Bereich. Die Kointegrationsanalyse bietet die Möglichkeit, den Systemcharakter mehrerer Zeitreihen zu identifizieren und zudem komplexe Wirkungsweisen in ihrem gemeinsamen dynamischem Verhalten zu beschreiben und zu erklären und darüber hinaus Fehlklassifikationen durch Scheinregression zu vermeiden. Die vielen Erweiterungen bieten eine hohe Flexibilität, um den dynamischen Charakter eines Zeitreihensystem adäquat analysieren zu können. Die vorliegende Arbeit hat gezeigt, wie die Kointegrationsanalyse die psychologische Forschung bereichern kann.

A Mathematischer Anhang

A.1 Algebraische Strukturen

In der vorliegenden Arbeit werden häufig Ergebnisse verwendet, die in jeder mathematischen Grundvorlesung vermittelt werden, in der Methodenausbildung von Psychologen jedoch nicht immer in der für diese Arbeit notwendigen Tiefe vermittelt werden. Eine (abstrakte) Einführung von Gruppen, Ringen und Körpern ist für das Verständnis der vorliegenden Arbeit sehr hilfreich, würde im Hauptteil der Arbeit jedoch zu weit vom behandelten Thema abweichen. Um das Rechnen mit Polynomen und Matrizen auf eine formale Grundbasis zu stellen, werden in diesem Anhang deshalb die verwendeten Formeln im Kontext der algebraischen Strukturen dargestellt.

A.1.1 Gruppe

Die Grundlage aller algebraischen Strukturen bildet die sogenannte Halbgruppe. Die aus ihr entstehende Gruppe ist für das Verständnis der Körpertheorie unerlässlich (z.B. Bosch, 2009).

Definition A.1.1 (Halbgruppe, Monoid, Gruppe) *Sei G eine Menge, $G = \{g_1, g_2, \dots\}$ mit einer Verknüpfung $*$, die in der Menge abgeschlossen ist (Abgeschlossenheit).*

$$\begin{aligned} * : G \times G &\rightarrow G \\ (g_1, g_2) &\mapsto g_1 * g_2 \end{aligned}$$

G heißt eine Halbgruppe, wenn das Assoziativgesetz gilt:

$$\forall g_1, g_2, g_3 \in G : g_1 * (g_2 * g_3) = (g_1 * g_2) * g_3$$

Eine Halbgruppe heißt Monoid, wenn in G ein neutrales Element für diese Verknüpfung existiert.

$$\exists i \in G \forall g \in G : i * g = g$$

Ein Monoid heißt Gruppe, wenn es zu jedem Element ein inverses Element existiert.

$$\forall g \in G \exists \tilde{g} : g * \tilde{g} = i$$

Bemerkung A.1.2 (Eindeutigkeit) *Jeder Monoid besitzt genau ein Einselement.*

Zu jedem Element g einer Gruppe G gibt es genau ein Inverses $\tilde{g} =: g^{-1}$ und es gilt

$$(g_1 * g_2)^{-1} = g_2^{-1} * g_1^{-1}.$$

Definition A.1.3 *Eine Halbgruppe, ein Monoid oder eine Gruppe G heißt kommutativ (oder abelsch), wenn das Kommutativgesetz gilt:*

$$\forall g_1, g_2 \in G : g_1 * g_2 = g_2 * g_1$$

Bemerkung A.1.4 *Jede Gruppe ist auch ein Monoid und jeder Monoid ist auch eine Halbgruppe.*

Bemerkung A.1.5 *Damit eine Menge G mit einer Verknüpfung $*$ eine Gruppe bildet, müssen also folgende Axiome gelten:*

- *Abgeschlossenheit: Die Verknüpfung bildet zwei Elemente aus der Gruppe immer auf ein weiteres Element der Gruppe ab. Es gibt keine zwei Gruppenelemente, deren Verknüpfung außerhalb der Gruppe liegt.*
- *Assoziativgesetz: Bei der Verknüpfung von mehreren Elementen ist die Reihenfolge der Verknüpfung für das Ergebnis egal.*
- *Neutrales Element: Es gibt in der Gruppe ein Element, dessen Verknüpfung mit jedem anderen Gruppenelement nichts verändert. Es gibt genau ein solches Element.*
- *Inverse Elemente: Zu jedem Gruppenelement gibt es ein weiteres Element der Gruppe, so dass ihre Verknüpfung gleich dem neutralen Element ist.*

Für eine kommutative Gruppe gilt außerdem:

- *Kommutativgesetz: Bei der Verknüpfung von zwei Gruppenelementen kann ich die beiden Elemente vertauschen, ohne dass sich das Ergebnis ändert.*

Beispiel A.1.6 (Halbgruppe, Monoid, Gruppe) *Hier einige Beispiele, die deutlich machen, dass die Eigenschaft Gruppe, Monoid und Halbgruppe nicht nur von der Verknüpfung, sondern auch von der Menge abhängig sind.*

1. $(\mathbb{N} = \{0, 1, 2, 3, \dots\}, +)$: *Die Menge der natürlichen Zahlen bildet zusammen mit der gewöhnlichen Addition als Verknüpfung einen Monoid. Das natürliche Element ist die Null. Da die negativen Zahlen $\{-1, -2, -3, \dots\}$ nicht in der Menge enthalten sind, gibt es keine inversen Elemente. Der Monoid ist kommutativ.*
2. $(\mathbb{N}_0 = \{1, 2, 3, \dots\}, +)$: *Betrachtete man die Menge der natürlichen Zahlen ohne die Null, bleibt eine Halbgruppe, da es kein neutrales Element in der Menge gibt.*
3. $(\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}, +)$: *Die Menge der ganzen Zahlen bildet zusammen mit der gewöhnlichen Addition eine kommutative Gruppe.*

4. (\mathbb{Q}, \cdot) : Die Menge der rationalen Zahlen bildet zusammen mit der gewöhnlichen Multiplikation eine kommutative Gruppe. Das neutrale Element der Multiplikation ist Eins, das inverse Element zu a ist $\frac{1}{a}$. Die Multiplikation ist abgeschlossen, assoziativ und kommutativ.
5. (\mathbb{Z}, \cdot) : Der Menge der ganzen Zahlen fehlen die multiplikativ inversen Elemente. Deshalb bildet sie zusammen mit der Multiplikation nur einen kommutativen Monoid.

A.1.2 Ring & Körper

Definition A.1.7 (Ring) R sei eine Menge mit zwei Verknüpfungen „+“ und „ \cdot “. Das Tripel $(R, +, \cdot)$ heißt ein Halbring, falls gilt:

1. $(R, +)$ ist eine kommutative Gruppe mit neutralem Element (Nullelement) 0_R .
2. (R, \cdot) ist eine Halbgruppe.
3. Es gelten die Distributivgesetze

$$\forall a, b, c \in R : (a + b) \cdot c = ac + bc$$

$$\text{und } a \cdot (b + c) = ab + ac$$

$(R, +, \cdot)$ heißt ein Ring, wenn auch die Multiplikation ein neutrales Element (Einselement) besitzt, wenn also $(R \setminus 0, \cdot)$ ein Monoid ist.

Ein Ring (Halbring) heißt kommutativ, wenn auch die Multiplikation kommutativ ist.

$$\forall a, b \in R : a \cdot b = b \cdot a$$

Bemerkung A.1.8 In der Literatur wird ein Halbring manchmal als Ring und ein Ring als Ring mit Eins bezeichnet.

Definition A.1.9 (Körper) Eine Menge K mit zwei Verknüpfungen „+“ und „ \cdot “ heißt ein Körper, falls

1. $(K, +)$ ist eine kommutative Gruppe mit neutralem Element 0_K .
2. (K, \cdot) ist eine kommutative Gruppe mit neutralem Element 1_K .
3. Es gilt das Distributivgesetz:

$$\forall a, b, c \in K : (a + b) \cdot c = ac + bc$$

Da die Multiplikation kommutativ ist, braucht man nur ein Distributivgesetz.

Bemerkung A.1.10 *Der größte Unterschied zwischen einem (kommutativen) Ring (mit Eins) und einem Körper ist die Existenz von multiplikativ inversen Elementen. Doch dieser „kleine“ Unterschied ist manchmal sehr entscheidend.*

Beispiel A.1.11 (Ringe, Körper) *Hier nun einige Beispiel, die den Unterschied zwischen Ring und Körper veranschaulichen sollen.*

1. $(\mathbb{Z}, +, \cdot)$: *Die Menge der ganzen Zahlen zusammen mit der gewöhnlichen Addition und Multiplikation bildet einen kommutativen Ring (mit Eins).*
2. $(\mathbb{Q}, +, \cdot)$: *In der Menge der rationalen Zahlen sind die multiplikativ inversen Elemente enthalten. Sie bildet zusammen mit der gewöhnlichen Addition und Multiplikation einen Körper.*
3. $(\mathbb{R}, +, \cdot)$: *Die reellen Zahlen bilden einen Körper. Da jede rationale Zahl auch eine reelle Zahl ist, die reellen Zahlen die rationalen also enthalten, spricht man von \mathbb{R} auch als Oberkörper zu \mathbb{Q} .*
4. $(\mathbb{C}, +, \cdot)$: *Die komplexen Zahlen sind ein Oberkörper von \mathbb{R} .*

A.1.3 Matrizenringe

Definition A.1.12 *Seien m und n natürliche Zahlen. Eine $m \times n$ -Matrix über einem Körper K ist ein rechteckiges Schema aus m Zeilen und n Spalten, deren Einträge Elemente aus K sind. Also sieht eine $m \times n$ -Matrix A mit den Elementen $a_{ij} \in K (i = 1, \dots, m, j = 1, \dots, n)$ folgendermaßen aus:*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Man schreibt für eine Matrix auch

$$A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \quad (a_{ij} \in K).$$

Die Menge aller $m \times n$ -Matrizen über K wird mit $K^{m \times n}$ bezeichnet.

Definition A.1.13 (Matrizenaddition) *Auf $K^{m \times n}$ kann man eine Addition komponentenweise erklären. Die Summe der Matrizen*

$$A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \quad \text{und} \quad B = (b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$$

ist definiert als

$$A + B = (a_{ij} + b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$$

In anderer Schreibweise für 2×2 -Matrizen also als

$$A + B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{pmatrix}.$$

Definition A.1.14 (Matrizenmultiplikation) Wir definieren für zwei allgemeine Matrizen ihr Produkt (z.B. (Meyberg, 1999, §6.2 Die Matrizenmultiplikation) oder (Beutelspacher, 2003, S. 93ff)) . A sei eine $m \times n$ -Matrix und B eine $r \times s$ -Matrix, wobei m, n, r, s natürliche Zahlen sind. Seien also $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ und $B = (b_{kl})_{1 \leq k \leq r, 1 \leq l \leq s}$ zwei Matrizen. Damit ein Produkt von A und B überhaupt erklärt werden kann, muss die Anzahl der Spalten von A gleich der Anzahl der Zeilen von B sein. Sei also $n = r$.

In diesem Fall ergibt sich als Produkt eine $m \times s$ -Matrix $C = (c_{il})_{1 \leq i \leq m, 1 \leq l \leq s}$, deren Elemente wie folgt definiert sind:

$$c_{il} = \sum_{j=1}^n a_{ij} b_{jl}$$

$$c_{il} = (a_{i1}, a_{i2}, \dots, a_{in}) \cdot \begin{pmatrix} b_{1l} \\ b_{2l} \\ \vdots \\ b_{rl} \end{pmatrix} = a_{i1} \cdot b_{1l} + \dots + a_{in} \cdot b_{nl}, \text{ da } r = n.$$

Beispiel A.1.15 (2×2 Matrix)

$$A \cdot B = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

$$\begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix} \cdot \begin{pmatrix} 3 & 5 \\ 7 & 9 \end{pmatrix} = \begin{pmatrix} 2 \cdot 3 + 4 \cdot 7 & 2 \cdot 5 + 4 \cdot 9 \\ 6 \cdot 3 + 8 \cdot 7 & 6 \cdot 5 + 8 \cdot 9 \end{pmatrix} = \begin{pmatrix} 6 + 28 & 10 + 36 \\ 18 + 56 & 30 + 72 \end{pmatrix} = \begin{pmatrix} 34 & 46 \\ 74 & 102 \end{pmatrix}$$

Theorem A.1.16 Die Menge der Matrizen mit Einträgen aus einem Körper K ist zusammen mit der Matrizenaddition und der Matrizenmultiplikation ein Ring, der mit $K^{n \times n}$ bezeichnet wird; dieser hat ein Einselement (die Einheitsmatrix), ist aber für $n \geq 2$ nicht kommutativ.

Beispiel A.1.17 Ein Vergleich mit Beispiel A.1.15 zeigt die Nicht-Kommutativität der Ma-

trizenmultiplikation.

$$\begin{pmatrix} 3 & 5 \\ 7 & 9 \end{pmatrix} \cdot \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix} = \begin{pmatrix} 3 \cdot 2 + 5 \cdot 6 & 3 \cdot 4 + 5 \cdot 8 \\ 7 \cdot 2 + 9 \cdot 6 & 7 \cdot 4 + 9 \cdot 8 \end{pmatrix} = \begin{pmatrix} 36 & 52 \\ 68 & 100 \end{pmatrix}$$

Transponierte Matrix

Die transponierte Matrix entsteht durch „Stürzen“ der Matrix (Beiglböck, 1983, S. 128).

Definition A.1.18 (Transponierte Matrix) *Man erhält die transponierte Matrix A' einer $m \times n$ -Matrix $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$, indem man die Zeilen als Spalten bzw. die Spalten als Zeilen schreibt*

$$A' = (a_{ji})_{1 \leq j \leq n, 1 \leq i \leq m}.$$

Dies wird an einem Beispiel schnell klar.

Beispiel A.1.19 (2×3 -Matrix)

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

$$A' = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Da wir die transponierten Matrix in einigen Gleichungen verwenden, führen wir hier die wichtigsten Rechenregeln (Fischer, 2002, S.96, S.147) mit transponierten Matrizen ein.

Bemerkung A.1.20 (Rechenregeln für die transponierte Matrix) *Seien A, B zwei Matrizen und λ eine reelle Zahl. Dann gelten folgende Rechenregeln.*

$$\begin{aligned} (A + B)' &= A' + B' \\ (\lambda \cdot A)' &= \lambda \cdot A' \\ (A')' &= A \\ (A \cdot B)' &= B' \cdot A' \end{aligned}$$

Man beachte, dass sich die Reihenfolge der Multiplikation durch die Transposition umkehrt.

Invertieren von Matrizen

Allgemein bestimmt die Inverse einer Matrix über den Gauß-Jordan-Algorithmus (Beiglböck, 1983, 2.3 Der Gauß-Jordan-Algorithmus). Für uns sind eigentlich nur zwei Beispiele entschei-

dend.

Beispiel A.1.21 (Diagonalmatrix) *Die Inverse einer Diagonalmatrix*

$$D = \begin{pmatrix} d_1 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & & \ddots & \\ 0 & \dots & d_n \end{pmatrix}$$

ist die Diagonalmatrix

$$D^{-1} = \begin{pmatrix} \frac{1}{d_1} & \dots & 0 \\ 0 & \frac{1}{d_2} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & \dots & \frac{1}{d_n} \end{pmatrix}.$$

Beispiel A.1.22 (2×2 -Matrix) *Für eine 2×2 -Matrix*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

ist die Inverse

$$A^{-1} = \frac{1}{ad - bc} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

In der Inversen einer 2×2 -Matrix taucht ihre Determinante auf.

Bemerkung A.1.23 (Determinante einer 2×2 -Matrix) *Die Determinante einer 2×2 -Matrix ist (z.B. Meyberg, 1999)*

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Matrizenmultiplikation in Schreibweise $[\beta'; \tau']$

Im Abschnitt 2.6 wird ein Vektor τ der transponierten $r \times K$ -Matrix β angehängt und dabei folgende Schreibweise verwendet

$$[\beta'; \tau'] = \begin{pmatrix} \beta_{11} & \beta_{21} & \dots & \beta_{K1} & \tau_1 \\ \vdots & & & & \\ \beta_{1r} & \beta_{2r} & \dots & \beta_{Kr} & \tau_r \end{pmatrix}.$$

Für das Produkt dieser Matrix mit einer $K \times r$ -Matrix α kann man dann schreiben

$$= \alpha [\beta'; \tau'] = [\alpha\beta'; \alpha\tau'],$$

da die Multiplikation für Matrizen spaltenweise definiert ist (siehe Definition A.1.14)

$$\begin{aligned} \alpha [\beta'; \tau'] &= \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1r} \\ & \vdots & \\ \alpha_{K1} & \dots & \alpha_{Kr} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} & \dots & \beta_{K1} & \tau_1 \\ \vdots & & & & \\ \beta_{1r} & \beta_{2r} & \dots & \beta_{Kr} & \tau_r \end{pmatrix} \\ &= \begin{pmatrix} \alpha_{11}\beta_{11} \dots \alpha_{1r}\beta_{1r} & \alpha_{11}\beta_{21} \dots \alpha_{1r}\beta_{2r} & \dots & \alpha_{11}\beta_{K1} \dots \alpha_{1r}\beta_{Kr} & \alpha_{11}\tau_1 \dots \alpha_{1r}\tau_r \\ \alpha_{21}\beta_{11} \dots \alpha_{2r}\beta_{1r} & \alpha_{21}\beta_{21} \dots \alpha_{2r}\beta_{2r} & \dots & \alpha_{21}\beta_{K1} \dots \alpha_{2r}\beta_{Kr} & \alpha_{21}\tau_1 \dots \alpha_{2r}\tau_r \\ & \vdots & & & \\ \alpha_{K1}\beta_{11} \dots \alpha_{Kr}\beta_{1r} & \alpha_{K1}\beta_{21} \dots \alpha_{Kr}\beta_{2r} & \dots & \alpha_{K1}\beta_{K1} \dots \alpha_{Kr}\beta_{Kr} & \alpha_{K1}\tau_1 \dots \alpha_{Kr}\tau_r \end{pmatrix} \\ &= \left[\begin{array}{cccc|c} \alpha_{11}\beta_{11} \dots \alpha_{1r}\beta_{1r} & \alpha_{11}\beta_{21} \dots \alpha_{1r}\beta_{2r} & \dots & \alpha_{11}\beta_{K1} \dots \alpha_{1r}\beta_{Kr} & \alpha_{11}\tau_1 \dots \alpha_{1r}\tau_r \\ \alpha_{21}\beta_{11} \dots \alpha_{2r}\beta_{1r} & \alpha_{21}\beta_{21} \dots \alpha_{2r}\beta_{2r} & \dots & \alpha_{21}\beta_{K1} \dots \alpha_{2r}\beta_{Kr} & \alpha_{21}\tau_1 \dots \alpha_{2r}\tau_r \\ & \vdots & & & \\ \alpha_{K1}\beta_{11} \dots \alpha_{Kr}\beta_{1r} & \alpha_{K1}\beta_{21} \dots \alpha_{Kr}\beta_{2r} & \dots & \alpha_{K1}\beta_{K1} \dots \alpha_{Kr}\beta_{Kr} & \alpha_{K1}\tau_1 \dots \alpha_{Kr}\tau_r \end{array} \right] \\ &= [\alpha\beta'; \alpha\tau']. \end{aligned}$$

Matrizenmultiplikation mit Elementarmatrizen

Es gibt verschiedene Formen von Elementarmatrizen (Beutelspacher, 2003, S.195).

- Einheitsmatrix
- Diagonalmatrizen mit einem von Eins verschiedenen Eintrag auf der Diagonalen
- Matrizen, die sich von der Einheitsmatrix durch einen Eintrag außerhalb der Diagonalen unterscheiden

Die Multiplikation mit Matrizen des ersten Typs bewirkt die Multiplikation einer Zeile/Spalte mit einer Konstanten. Die Multiplikation mit Matrizen des zweiten Typs bewirkt die Addition des Vielfachen einer Zeile/Spalte zu einer anderen.

Multiplikation von rechts bewirkt Spaltenoperationen, während Multiplikation von links in den Zeilen operiert.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Spaltenoperationen

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} &= \begin{pmatrix} a & 2b \\ c & 2d \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 2a & b \\ 2c & d \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} a & b+2a \\ c & d+2c \end{pmatrix} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} &= \begin{pmatrix} a+2b & b \\ c+2d & d \end{pmatrix} \end{aligned}$$

Zeilenoperationen

$$\begin{aligned} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} a & b \\ 2c & 2d \end{pmatrix} \\ \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} 2a & 2b \\ c & d \end{pmatrix} \\ \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} a+2c & b+2d \\ c & d \end{pmatrix} \\ \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} a & b \\ c+2a & d+2b \end{pmatrix} \end{aligned}$$

Die Multiplikation mit mehreren Elementarmatrizen bewirkt mehrere Spalten- bzw. Zeilenumformungen nacheinander.

Bemerkung A.1.24 (Hinweis Smith Mc Millan Zerlegung) *Im zweiten Schritt der Smith Mc Millan Zerlegung führt man also so lange Umformungen durch, bis man eine Diagonalmatrix erhält. Die verwendeten Zeilenumformungen werden in der Matrix $U(L)$, die Spaltenumformungen in $V(L)$ zusammengefasst.*

Die Matrizen werden für die Darstellung von $A(L)$ gebraucht.

A.1.4 Polynomring

Die Polynome in einer Unbekannten mit Koeffizienten aus einem Körper bilden einen Ring (Bosch, 2009; Meyberg, 1999). Beutelspacher (2003, 6. Polynomringe) erklärt sehr anschaulich, was man unter einer „Unbekannten“ verstehen muss, und beschäftigt sich ausführlich mit den Besonderheiten von Polynomringen.

Definition A.1.25 (Polynomring) *Unter einem Polynomring $K[z]$ verstehen wir die Menge*

aller Polynome

$$P(z) := a_0z^0 + a_1z^1 + a_2z^2 + \dots + a_pz^p$$

mit Koeffizienten $a_0, a_1, \dots, a_p \in K$, wobei K ein Körper ist, zusammen mit der Polynomaddition und Polynommultiplikation. Man nennt a_0 auch absolutes Glied und p mit $a_p \neq 0$, aber $a_n = 0 \forall n > p$ Grad des Polynoms.

Wir werden meist Polynome mit reellen ($K = \mathbb{R}$) oder auch komplexen ($K = \mathbb{C}$) Koeffizienten betrachten. Wir definieren nun eine Addition und eine Multiplikation auf dieser Menge und zeigen, dass $K[z]$ mit ihnen einen Ring, aber keinen Körper bildet.

Definition A.1.26 (Polynomaddition) *Man addiert zwei Polynome $P(z), Q(z) \in K[z]$ mit $m = \max(p, q)$, indem man die Koeffizienten für jede z -Potenz addiert*

$$\begin{aligned} P(z) &= p_0 + p_1z + p_2z^2 + \dots + p_pz^p \\ Q(z) &= q_0 + q_1z + q_2z^2 + \dots + q_qz^q \\ P(z) + Q(z) &= (p_0 + q_0) + (p_1 + q_1)z + (p_2 + q_2)z^2 + \dots + (p_m + q_m)z^m \\ &= \sum_{i=1}^m (p_i + q_i)z^i \end{aligned}$$

Alle Koeffizienten p_i, q_i , für die i größer als der Grad von P bzw. Q ist, sind gleich Null.

p bezeichne hier den Grad des Polynoms P , sowie q den von Q . Dass auch die Koeffizienten der Polynome mit diesen Buchstaben gekennzeichnet sind, dient der leichteren Lesart und führt nicht zu Konflikten, da sie stets indiziert sind. So ist p_p der p -te Koeffizient des Polynoms P .

Bemerkung A.1.27 *Die Menge aller Polynome $K[z]$ bildet mit der komponentenweisen Addition eine kommutative Gruppe. Die Abgeschlossenheit, das Assoziativ- und Kommutativgesetz ergeben sich direkt aus den Eigenschaften des zugrundeliegenden Körpers K . Das neutrale Element der Addition ist das Nullpolynom $N(z) = 0$, die additiv Inversen bildet man, indem man das Polynom mit -1 multipliziert: $P(z) + (-1)P(z) = 0$.*

Definition A.1.28 (Polynommultiplikation) *Die Multiplikation von Polynomen ist durch eine sogenannte Faltung definiert.*

$$P(z) \cdot Q(z) = \sum_{k=0}^{p+q} \left(\sum_{j=0}^k p_j q_{k-j} \right) z^k, \tag{A.1}$$

wobei

$$\begin{aligned} P(z) &= p_0 + p_1z + p_2z^2 + \dots + p_pz^p \\ Q(z) &= q_0 + q_1z + q_2z^2 + \dots + q_qz^q \end{aligned}$$

Beispiel A.1.29 Betrachten wir folgendes Beispiel.

$$\begin{aligned} P(z) &= 5 + 7z \\ Q(z) &= 1 + 2z + 3z^2 \end{aligned}$$

Das Polynom P hat den Grad $p = 1$, Q hat den Grad $q = 2$. Das Produkt $P \cdot Q(z)$ hat also den Grad $p + q = 3$.

$$P(z) \cdot Q(z) = (5 + 7z) \cdot (1 + 2z + 3z^2) \tag{A.2}$$

$$= 5 \cdot (1 + 2z + 3z^2) + 7z \cdot (1 + 2z + 3z^2) \tag{A.3}$$

$$= (5 + 10z + 15z^2) + (7z + 14z^2 + 21z^3) \tag{A.4}$$

$$= 5 + (10 + 7)z + (15 + 14)z^2 + (21)z^3 \tag{A.5}$$

$$= 5 + 17z + 29z^2 + 21z^3 \tag{A.6}$$

Dabei ist die Gleichung (A.5) genau in der Gestalt, wie wir die Multiplikation definiert hatten.

$$\begin{aligned} \sum_{k=0}^{p+q} \left(\sum_{j=0}^k p_j q_{k-j} \right) z^k &= \sum_{k=0}^3 \left(\sum_{j=0}^k p_j q_{k-j} \right) z^k \\ &= \left(\sum_{j=0}^0 p_j q_{k-j} \right) + \left(\sum_{j=0}^1 p_j q_{k-j} \right) z + \left(\sum_{j=0}^2 p_j q_{k-j} \right) z^2 + \left(\sum_{j=0}^3 p_j q_{k-j} \right) z^3 \\ &= (p_0 q_0) + (p_0 q_1 + p_1 q_0) z + (p_0 q_2 + p_1 q_1 + p_2 q_0) z^2 \\ &\quad + (p_0 q_3 + p_1 q_2 + p_2 q_1 + p_3 q_0) z^3 \end{aligned}$$

Mit $p_0 = 5, p_1 = 7, p_2 = p_3 = 0$ und $q_0 = 1, q_1 = 2, q_2 = 3, q_3 = 0$ ergibt sich

$$\begin{aligned} \sum_{k=0}^3 \left(\sum_{j=0}^k p_j q_{k-j} \right) z^k &= (5 \cdot 1) + (5 \cdot 2 + 7 \cdot 1) z + (5 \cdot 3 + 7 \cdot 2 + 0 \cdot 1) z^2 \\ &\quad + (5 \cdot 0 + 7 \cdot 3 + 0 \cdot 2 + 0 \cdot 1) z^3 \\ &= 5 + 17z + 29z^2 + 21z^3 \end{aligned}$$

Bemerkung A.1.30 Die Polynommultiplikation ist kommutativ. Das Einselement ist $E(z) = 1$. Es gibt jedoch im Allgemeinen keine Inversen. Nur die konstanten Polynome $A(z) = a_0$ lassen sich durch $B(z) = \frac{1}{a_0}$ multiplikativ invertieren: $A(z) \cdot B(z) = a_0 \cdot \frac{1}{a_0} = \frac{a_0}{a_0} = 1$. Man nennt die Elemente eines Rings, die multiplikativ Inverse besitzen Einheiten des Rings.

Einen Ersatz für echte multiplikativ Inverse kann man mit folgendem Verfahren berechnen.

Polynominvertierung durch Koeffizientenvergleich

Bei der Polynominvertierung gibt es folgendes Problem. Die Polynome mit reellen Koeffizienten bilden ähnlich wie die ganzen Zahlen \mathbb{Z} keinen Körper, sondern nur einen Ring mit Einselement. Das heißt insbesondere, dass es im Allgemeinen keine Inversen gibt (Ausnahme: konstante Polynome). Durch Koeffizientenvergleich kann man ein inverses Polynom bekommen, das aber unendlichen Grad hat und somit nicht im Polynomring liegt.

Ist Q das inverse Polynom von P , dann soll gelten:

$$P(z) \cdot Q(z) = 1$$

P habe den Grad p , der Grad von Q muss vorerst als theoretisch ∞ angenommen werden.

$$P(z) = p_0 + p_1z + p_2z^2 + \dots + p_pz^p$$

Durch Koeffizientenvergleich ergibt sich (vgl. Faltung von Polynomen):

$$\begin{aligned} 1 &= Q(z) \cdot P(z) \\ &= (q_0 + q_1z + q_2z^2 + \dots) \cdot (p_0 + p_1z + p_2z^2 + \dots + p_pz^p) \\ &= \begin{array}{cccccccc} q_0 \cdot p_0 & + & q_0 \cdot p_1z & + & q_0 \cdot p_2z^2 & + & q_0 \cdot p_3z^3 & + & \dots & + & q_0 \cdot p_pz^p & + \\ & & q_1 \cdot p_0z & + & q_1 \cdot p_1z^2 & + & q_1 \cdot p_2z^3 & + & \dots & + & q_1 \cdot p_{p-1}z^p & + & q_1 \cdot p_pz^{p+1} \\ & & & & q_2 \cdot p_0z^2 & + & q_2 \cdot p_1z^3 & + & \dots & + & q_2 \cdot p_{p-2}z^p & + & \dots & + & q_2p_pz^{p+2} \end{array} \\ &= \dots \\ &= q_0 \cdot p_0 \\ &+ (q_0 \cdot p_1 + q_1 \cdot p_0) z \\ &+ (q_0 \cdot p_2 + q_1 \cdot p_1 + q_2 \cdot p_0) z^2 \\ &+ (q_0 \cdot p_3 + q_1 \cdot p_2 + q_2 \cdot p_1 + q_3 \cdot p_0) z^3 \\ &+ \dots \\ &+ (q_0 \cdot p_p + q_1 \cdot p_{p-1} + q_2 \cdot p_{p-2} + \dots + q_p \cdot p_0) z^p \\ &+ \dots \\ &+ \left(q_0 \cdot \underbrace{p_k}_{=0} + q_1 \cdot p_{k-1} + \dots + q_k \cdot p_0 \right) z^k \\ &+ \dots \end{aligned}$$

mit $p_k = 0$ für $k > p$.

Daraus folgt ein Gleichungssystem, aus dem man die Koeffizienten q_k bestimmen kann.

$$\begin{aligned}
 p_0 q_0 &= 1 \\
 p_1 q_0 + p_0 q_1 &= 0 \\
 q_0 \cdot p_2 + q_1 \cdot p_1 + q_2 \cdot p_0 &= 0 \\
 \dots & \\
 q_0 &= \frac{1}{p_0} \\
 \Rightarrow q_1 &= (-q_0 \cdot p_1) \cdot \left(\frac{1}{p_0}\right) \\
 q_2 &= (-q_0 \cdot p_2 - q_1 \cdot p_1) \cdot \left(\frac{1}{p_0}\right) \\
 \dots &
 \end{aligned}$$

Daraus ergeben sich die Koeffizienten für Q durch :

$$\begin{aligned}
 q_0 &= \frac{1}{p_0} \\
 q_k &= -\frac{1}{p_0} \sum_{j=0}^{k-1} q_j p_{k-j}
 \end{aligned}$$

Ein Beispiel findet sich im Anhang zu Lagpolynomen B.2.

A.2 Einführung in die komplexen Zahlen

Die komplexen Zahlen bilden einen Körper, der alle reellen Zahlen enthält und in dem jedes Polynom Nullstellen hat. Nach dem Fundamentalsatz der Algebra besitzt jedes (nichtkonstante) Polynom sogar so viele komplexe Nullstellen wie sein Grad angibt. Die folgende Darstellung lehnt sich an Freitag und Busam (2006) an. Weitere Einführungen in die komplexen Zahlen finden sich bei Wolter und Dahn (2000) und Meyberg (1999).

A.2.1 Die komplexen Zahlen als Oberkörper von \mathbb{R}

Es gibt Polynome mit reellen Koeffizienten, die jedoch keine reelle Nullstelle besitzen.

$$P(x) = x^2 + 1$$

ist ein einfaches Beispiel. Mit den komplexen Zahlen findet man einen Zahlenbereich, der die reellen Zahlen umfasst und eine Nullstelle dieses Polynoms enthält. Als erste Idee betrachtet man nicht länger einzelne reelle Zahlen, sondern Paare (x, y) reeller Zahlen $x, y \in \mathbb{R}$.

Der Körper der komplexen Zahlen

Definition A.2.1 (Körper der komplexen Zahlen) *Die Menge $\mathbb{R} \times \mathbb{R} := \{(x, y), x, y \in \mathbb{R}\}$ der Paare reeller Zahlen bildet zusammen mit folgender Addition und Multiplikation einen Körper.*

$$\begin{aligned} + : (x_1, y_1) + (x_2, y_2) &= (x_1 + x_2, y_1 + y_2) \\ \cdot : (x_1, y_1) \cdot (x_2, y_2) &= (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1) \end{aligned}$$

Wir nennen ihn Körper der komplexen Zahlen und bezeichnen ihn mit \mathbb{C} .

Um diese Aussage zu prüfen, muss man nachweisen, dass die Zahlenmenge mit den angegebenen Verknüpfungen jeweils eine kommutative Gruppe bildet und das Distributivgesetz gilt. Dabei ist

- $(0, 0)$ das neutrale Element der Addition (Nullelement) und
- $(1, 0)$ das neutrale Element der Multiplikation (Einselement).

Die inversen Elemente zu (x, y) sind

- $-(x, y) = (-x, -y)$ bezüglich der Addition und
- $\left(\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2}\right)$ in der Multiplikation.

Oberkörper der reellen Zahlen

Wir betrachten zunächst besondere komplexe Zahlen, nämlich solche der Gestalt $(x, 0)$. Es ergibt sich

$$\begin{aligned} (x_1, 0) + (x_2, 0) &= (x_1 + x_2, 0) \\ (x_1, 0) \cdot (x_2, 0) &= (x_1 \cdot x_2, 0). \end{aligned}$$

Die Summe und das Produkt solcher Zahlen haben also wieder dieselbe Gestalt $(x_3, 0)$. Sie werden addiert und multipliziert wie reelle Zahlen. Deshalb kann man $(x, 0) \in \mathbb{C}$ mit $x \in \mathbb{R}$ identifizieren. Der Körper der komplexen Zahlen $\mathbb{C} := \{(x, y), x, y \in \mathbb{R}\}$ enthält also einen Unterkörper $\{(x, 0), x \in \mathbb{R}\}$, der isomorph (= strukturgleich) zum Körper der reellen Zahlen ist, und deshalb mit diesem identifiziert werden kann. Somit ist \mathbb{C} ein Oberkörper von \mathbb{R} .

Definition A.2.2 (Oberkörper) *Ein Oberkörper C ist ein Körper, der einen anderen (Unter“=)Körper R enthält. Das heißt R ist eine Teilmenge von C und Addition und Multiplikation in R entstehen durch Einschränkung der Addition und Multiplikation in C auf Elemente aus R .*

Die imaginäre Einheit

Nun betrachten wir die zweite Komponente einer komplexen Zahl. Dazu untersuchen wir die Zahl $(0, 1)$ und berechnen zunächst ihr Quadrat, was durch die Definition der Multiplikation möglich ist

$$(0, 1)^2 = (0, 1) \cdot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 0 \cdot 1) = (-1, 0).$$

Da wir $(-1, 0)$ mit -1 identifizieren können, ist $(0, 1)$ eine Lösung von

$$x^2 + 1 = 0.$$

Diese „Wurzel aus -1 “ nennt man *imaginäre Einheit* und bezeichnet sie mit i .

Mit diesem Wissen kann man eine komplexe Zahl $z = (x, y)$ wie folgt in ihre Komponenten zerlegen, wobei x und y weiterhin reelle Zahlen sind.

$$\begin{aligned} z = (x, y) &= x \cdot (1, 0) + y \cdot (0, 1) \\ &= x + iy \end{aligned}$$

Die komplexe Zahl $z = x + iy$ ist so eindeutig festgelegt. Man nennt $\operatorname{Re}(z) = x$ den Realteil und $\operatorname{Im}(z) = y$ den Imaginärteil.

In dieser Schreibweise mit imaginärer Einheit wird auch die merkwürdig anmutende Multiplikationsvorschrift verständlich.

$$\begin{aligned} z_1 \cdot z_2 &= (x_1 + iy_1) \cdot (x_2 + iy_2) = x_1x_2 + iy_1x_2 + ix_1y_2 + i^2y_1y_2 \\ &= (x_1x_2 - y_1y_2) + i(x_2y_1 + x_1y_2) \end{aligned}$$

Anders als der Körper der reellen Zahlen ist \mathbb{C} nicht geordnet. Das heißt, dass es keine größer/kleiner-Relation gibt. Der Grund dafür ist, dass das Produkt aus zwei komplexen Zahlen ohne Vorzeichen nicht mehr stets wieder ohne Vorzeichen ist (vgl. i^2).

Der Betrag einer komplexen Zahl

Ändert man das Vorzeichen zwischen der Verknüpfung der beiden Komponenten erhält man das *konjugiert Komplexe* einer komplexen Zahl.

$$\bar{z} := x - iy$$

Die Zahl $z\bar{z} = (x + iy) \cdot (x - iy) = x^2 + y^2 \in \mathbb{R}$ ist stets eine nicht negative Zahl. Die Wurzel

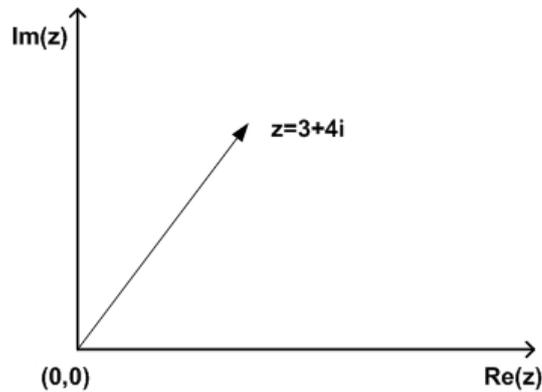


Abbildung A.1: Darstellung einer komplexen Zahl

aus dieser Zahl definiert den Betrag von z .

$$|z| := \sqrt{z\bar{z}}$$

Mit Hilfe des Betrags erhält man eine einfache Darstellung des multiplikativ Inversen.

$$\begin{aligned} z^{-1} &= \frac{1}{z} = \frac{\bar{z}}{z \cdot \bar{z}} = \frac{\bar{z}}{|z|^2} \\ &= \frac{x - iy}{x^2 + y^2} \end{aligned}$$

A.2.2 Geometrische Veranschaulichung und Polarkoordinaten

Die Menge $\mathbb{R} \times \mathbb{R}$ bildet einen Vektorraum. Man kann jedes Element als Linearkombination der Basiselemente $e_1 = (1, 0)$ und $e_2 = (0, 1)$ schreiben und als Punkt in der Zeichenebene darstellen

$$(3, 4) = 3 \cdot (1, 0) + 4 \cdot (0, 1) = 3e_1 + 4e_2.$$

Da wir \mathbb{C} mit den Paaren reeller Zahlen identifizieren, kann man jede komplexe Zahl als Punkt im $\mathbb{R} \times \mathbb{R}$ darstellen. Man bedenke, dass wir $(1, 0)$ mit 1 und $(0, 1)$ mit i gleichsetzen.

Die Addition von komplexen Zahlen kann man deshalb einfach als vektorielle Addition darstellen.

Der Betrag ist der euklidische Abstand zum Nullpunkt. Das konjugiert Komplexe erhält man durch Spiegelung an der reellen Achse.

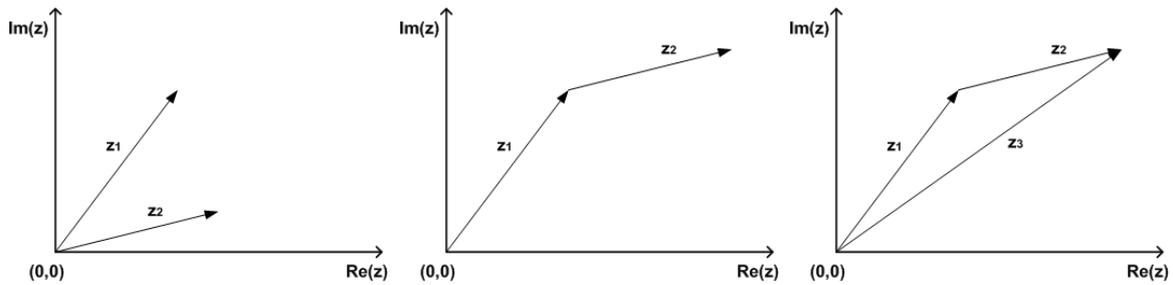


Abbildung A.2: Addition zweier komplexer Zahlen

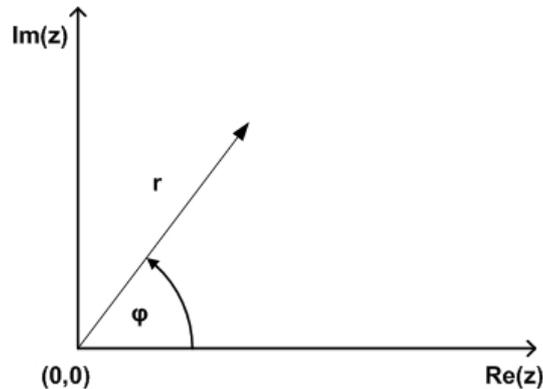


Abbildung A.3: Darstellung einer komplexen Zahl in Polarkoordinaten

Polarkoordinaten

Eine geometrische Deutung der Multiplikation erhält man mit Hilfe von Polarkoordinaten. Die Idee dabei ist, dass man jeden Punkt nicht nur durch seine Standardkoordinaten (kartesische Koordinaten), sondern auch durch die Länge (Betrag) und Lage (Winkel) zur waagerechten (reellen) Achse seines Ortsvektors beschreiben kann. Der Betrag ist dabei der euklidische Abstand zum Koordinatenursprung $(0,0)$.

Jeder Punkt aus $\mathbb{R} \times \mathbb{R}$ und damit auch aus \mathbb{C} lässt sich als

$$(x, y) = r(\cos(\phi), \sin(\phi)), \quad r > 0,$$

schreiben. Dabei ist der Betrag $r = \sqrt{x^2 + y^2}$ eindeutig bestimmt, der Winkel ϕ (der im Bogenmaß angegeben wird) jedoch nur bis auf Addition eines ganzzahligen Vielfaches von 2π (was 360° , also einer vollen Umdrehung entspricht). Jedes ϕ , für das obige Gleichung gilt, heißt ein Argument von z . Man nennt $\phi \in]-\pi, \pi]$ den Hauptwert des Arguments. Durch Angabe von Betrag und *Haupt*Argument ist der Punkt eindeutig bestimmt.

Multiplikation

In dieser Darstellung kann man die Multiplikation geometrisch veranschaulichen. Dazu brauchen wir noch

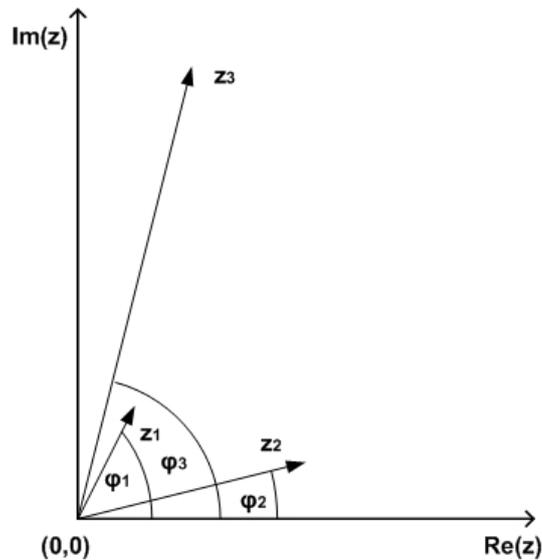


Abbildung A.4: Multiplikation zweier komplexer Zahlen

Theorem A.2.3 (Additionstheoreme der Winkelfunktionen) *Es gilt:*

$$\begin{aligned} &(\cos(\phi) + i \sin(\phi)) \cdot (\cos(\phi') + i \sin(\phi')) \\ &= (\cos(\phi + \phi') + i \sin(\phi + \phi')). \end{aligned}$$

Multipliziert man (A.7) aus, erhält man

$$= [\cos(\phi) \cdot \cos(\phi') - \sin(\phi) \sin(\phi')] + i[\cos(\phi) \sin(\phi') + \cos(\phi') \sin(\phi)].$$

Durch Vergleich der Real- und Imaginärteile in (A.7) und (A.7) erhält man die Additionstheoreme.

$$\begin{aligned} \cos(\phi + \phi') &= \cos(\phi) \cdot \cos(\phi') - \sin(\phi) \cdot \sin(\phi') \\ \sin(\phi + \phi') &= \sin(\phi) \cdot \cos(\phi') + \cos(\phi) \cdot \sin(\phi') \end{aligned}$$

Für die Multiplikation zweier komplexer Zahlen $z = r(\cos(\phi) + i \sin(\phi))$ und $z' = r'(\cos(\phi') + i \sin(\phi'))$ folgt also:

$$\begin{aligned} zz' &= r(\cos(\phi) + i \sin(\phi)) \cdot r'(\cos(\phi') + i \sin(\phi')) \\ &= rr'(\cos(\phi + \phi') + i \sin(\phi + \phi')) \end{aligned}$$

Komplexe Zahlen werden multipliziert, indem man die Beträge multipliziert und die Argumente addiert (Drehstreckung).

Einheitswurzeln

Durch mehrfache Anwendung der Multiplikation definiert man Potenzen mit komplexer Basis und ganzzahligem Exponenten ($n \in \mathbb{Z}$) durch

$$\begin{aligned} z^n &= \underbrace{z \cdot \dots \cdot z}_{n\text{-mal}} \\ z^0 &= 1 \\ z^{-n} &= \left(\frac{1}{z}\right)^n. \end{aligned}$$

Es gelten dabei die bekannten Rechenregeln

$$\begin{aligned} z^n \cdot z^m &= z^{n+m} \\ (z^n)^m &= z^{n \cdot m}. \end{aligned}$$

Definition A.2.4 (Einheitswurzel) *Eine komplexe Zahl ζ heißt n -te Einheitswurzel ($n \in \mathbb{N}$), falls $\zeta^n = 1$ gilt.*

Theorem A.2.5 (Gestalt der Einheitswurzeln) *Es gibt zu jedem $n \in \mathbb{N}$ genau n verschiedene n -te Einheitswurzeln, nämlich*

$$\zeta_{k,n} := \cos\left(\frac{2\pi}{n}k\right) + i \sin\left(\frac{2\pi}{n}k\right), \quad 0 \leq k \leq n-1.$$

Bemerkung A.2.6 $\zeta_{0,n} = 1$ ist für jedes n eine n -te Einheitswurzel.

Bemerkung A.2.7 Mit $\zeta_1 = \zeta_{1,n} = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$ gilt für $k = 0, 1, \dots, n-1$

$$\zeta_k = \zeta_1^k.$$

Beispiel A.2.8 *Einige Einheitswurzeln sind*

$$\begin{aligned} n = 1 & \quad \{1\} \\ n = 2 & \quad \{1, -1\} \\ n = 4 & \quad \{1, i, -1, -i\} \end{aligned}$$

Sämtliche Einheitswurzeln haben den Betrag 1 und liegen somit auf dem Rand des Einheitskreises, der Einheitskreislinie

$$S^1 := \{z \in \mathbb{C}; |z| = 1\}.$$

Sie bilden die Eckpunkte eines gleichseitigen n -Ecks, das S^1 eingeschrieben ist. Ein Eckpunkt ist stets $(1, 0) = 1$.

Aus diesem Grund nennt man die Gleichung $z^n = 1$ auch Kreisteilungsgleichung. Denn ihre Lösungen, die Einheitswurzeln, „teilen“ den Einheitskreis in n gleich große Abschnitte. Es

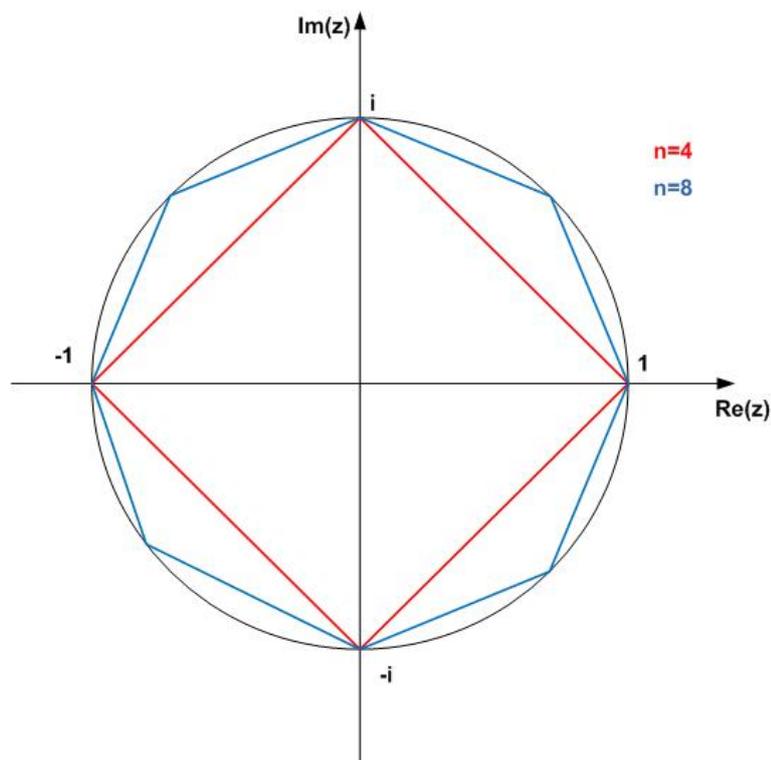


Abbildung A.5: Die n -ten Einheitswurzeln bilden ein regelmäßiges n -Eck auf dem Einheitskreis: hier dargestellt für $n = 4$ (rot) und $n = 8$ (blau)

gilt

$$z^n - 1 = (z - \zeta_0) \cdot (z - \zeta_1) \cdot \dots \cdot (z - \zeta_{n-1}).$$

Die $\zeta_{k,n}$ sind die Nullstellen des Polynoms

$$P(z) := z^n - 1.$$

Das Polynom hat also n verschiedene Nullstellen. Dies ist ein Spezialfall des

Theorem A.2.9 (Fundamentalsatz der Algebra) *Jedes nichtkonstante komplexe Polynom besitzt so viele Nullstellen, wie sein Grad angibt.*

Dabei müssen die Nullstellen mit ihrer Vielfachheit gerechnet werden.

A.3 Verallgemeinerte Binomialkoeffizienten

A.3.1 Binomialkoeffizienten

Wir beginnen mit der bekannten Definition der *einfachen* Binomialkoeffizienten (z.B. Forster, 2001), um diese später erweitern zu können.

Definition A.3.1 (Fakultät) *Unter der Fakultät einer natürlichen Zahl $n \in \mathbb{N}$ versteht man das Produkt aller natürlichen Zahlen von 1 bis n .*

$$\begin{aligned} n! &:= 1 \cdot 2 \cdot \dots \cdot n \\ 1! &= 1 \\ 2! &= 2 \\ 3! &= 6 \\ &\dots \end{aligned}$$

Für Null ist die Fakultät als Eins definiert.

$$0! := 1.$$

Mit dieser Schreibweise kann man die Binomialkoeffizienten definieren (z.B. Bronstein, Semendjaev, Musiol & Mühlig, 2008; Meyberg, 1999).

Definition A.3.2 (Binomialkoeffizienten) *Seien $n, k \in \mathbb{N}$. Dann ist der Binomialkoeffizient $\binom{n}{k}$ (sprich: „ n über k “ oder „ k aus n “) definiert durch:*

$$\binom{n}{k} := \prod_{j=1}^k \frac{n-j+1}{j} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{1 \cdot 2 \cdot \dots \cdot k} = \frac{n!}{k!(n-k)!}$$

Anschaulich zieht man aus einer Menge von n nummerierten Kugeln k Stück ohne Beachten der Reihenfolge ohne Zurücklegen.

Beispiel A.3.3 (Lotto) *Beim Lotto werden 6 Kugeln aus 49 nummerierten gezogen. Eine gezogene Kugel wird nicht wieder zurückgelegt und die Reihenfolge der gezogenen Kugeln ist für einen Gewinn nicht entscheidend.*

$$\binom{49}{6} := \prod_{j=1}^6 \frac{49-j+1}{j} = \frac{49 \cdot 48 \cdot \dots \cdot 44}{1 \cdot 2 \cdot \dots \cdot 6} = \frac{49 \cdot 48 \cdot \dots \cdot 44 \cdot 43 \cdot \dots \cdot 1}{1 \cdot 2 \cdot \dots \cdot 6 \cdot 43 \cdot \dots \cdot 1} = \frac{49!}{6!43!} \approx 13.983.816$$

Bemerkung A.3.4 (Eigenschaften der Binomialkoeffizienten) *Zunächst gilt*

$$\begin{aligned} \binom{n}{1} &= n \quad \forall n \geq 0 \\ \binom{n}{0} &= 1 \quad \forall n \geq 0 \end{aligned}$$

Man kann genau n verschiedene Einzelkugeln aus einer Menge von n verschiedenen Kugeln ziehen. „Man kann nur einmal nicht ziehen.“

Falls k größer als n oder negativ ist, wird der Binomialkoeffizient gleich Null gesetzt.

$$\begin{aligned} \binom{n}{k} &= 0, \text{ falls } k > n \\ \binom{n}{k} &= 0 \quad \forall k < 0 \end{aligned} \tag{A.7}$$

Binomialkoeffizienten sind symmetrisch.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k} \quad \forall n \in \mathbb{N}, k \in \mathbb{Z}$$

Diese wichtige Eigenschaft spiegelt sich auch im Binomischen Lehrsatz wieder, durch den auch das Pascalsche Dreieck (Meyberg, 1999, S.9) seine Bedeutung erhält.

A.3.2 Binomischer Lehrsatz

Die Binomialkoeffizienten sind wichtiger Bestandteil des Binomischen Lehrsatzes (z.B. Meyberg, 1999, Satz 2.2).

Theorem A.3.5 (Binomischer Lehrsatz) *Seien $x, y \in \mathbb{R}, k \in \mathbb{N}$. Dann gilt:*

$$\begin{aligned} (x+y)^n &= \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \\ &= x^n + \binom{n}{1} x^{n-1} y^1 + \binom{n}{2} x^{n-2} y^2 + \dots + \binom{n}{n-1} x^1 y^{n-1} + y^n \\ &= \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} \end{aligned}$$

Der Binomische Lehrsatz ist eine Erweiterung der 1. Binomischen Formel. Oder anders gesprochen: Die erste Binomische Formel $(a+b)^2 = a^2 + 2ab + b^2$ ist für $n = 2$ Teil des Binomischen Lehrsatzes.

Beispiel A.3.6 ($n=4$)

$$\begin{aligned} (x+y)^4 &= \sum_{k=0}^4 \binom{4}{k} x^{4-k} y^k \\ &= \binom{4}{0} x^4 y^0 + \binom{4}{1} x^3 y^1 + \binom{4}{2} x^2 y^2 + \binom{4}{3} x^1 y^3 + \binom{4}{4} x^0 y^4 \\ &= 1x^4 + 4x^3 y^1 + 10x^2 y^2 + 4x^1 y^3 + 1y^4 \\ &= 1y^4 + 4y^3 x^1 + 10y^2 x^2 + 4y^1 x^3 + 1x^4 \\ &= \sum_{k=0}^4 \binom{4}{k} x^k y^{4-k} \end{aligned}$$

A.3.3 Gammafunktion

Die Gammafunktion spielt eine wichtige Rolle in der Funktionentheorie (z.B. Freitag & Busam, 2006). Hier wird sie nur zur Konstruktion von verallgemeinerten Binomialkoeffizienten gebraucht. Dennoch werden hier ihre wichtigsten Eigenschaften und einige Werte vorgestellt. Wir beschränken uns dabei auf die reellen Zahlen.

A.3.4 Definition und Funktionalgleichung

Definition A.3.7 (Gammafunktion) *e* bezeichne die Eulersche Zahl. Dann ist die Gammafunktion für reelle x definiert durch das uneigentliche Integral (z.B. Beichelt, 2003, S.49):

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$$

Dies ist (absolut) konvergent für $x > 0$. Für komplexe $z := x + iy$ mit $\operatorname{Re}(z) > 0$ ist die Gammafunktion analog definiert, wobei $t^{z-1} := e^{(z-1) \cdot \log(t)}$. Dann ist es für $\operatorname{Re}(z) > 0$ (also auf der rechten Halbebene) absolut konvergent.

Bemerkung A.3.8 (Funktionalgleichung der Gammafunktion) Für reelle $x \notin \{0, -1, -2, \dots\}$ gilt folgende Relation.

$$\Gamma(x) = \frac{\Gamma(x+1)}{x} \text{ bzw. } \Gamma(x) = (x-1)\Gamma(x-1) \quad (\text{A.8})$$

Beweis. Durch partielle Integration mit $u(t) = t^x$ und $v'(t) = e^{-t}$.

$$\begin{aligned} \Gamma(x+1) &= \int_0^{\infty} t^x \cdot e^{-t} dt \\ &= [-t^x \cdot e^{-t}]_0^{\infty} + x \int_0^{\infty} e^{-t} \cdot t^{x-1} dt \\ &= x \cdot \Gamma(x) \end{aligned}$$

□

Bemerkung A.3.9 (Zusammenhang zur Fakultät) Insbesondere gilt für $n \in \mathbb{N}$:

$$\Gamma(n+1) = n! \quad (\text{A.9})$$

Bemerkung A.3.10 (Fortsetzung der Gammafunktion) *Iterierte Anwendung der Funktio-*

nalgleichung liefert

$$\begin{aligned}\Gamma(x) &= \frac{\Gamma(x+1)}{x} \\ &= \frac{\Gamma(x+n+1)}{x \cdot (x+1) \cdot \dots \cdot (x+n)}\end{aligned}$$

Man kann also mit der rechten Seite der Gleichung auch Werte für $x < 0$ berechnen. Wir verwenden diese Gleichung als Definition der Gammafunktion für diese Werte und verstehen unter der Gammafunktion im Folgenden diese Fortsetzung.

Theorem A.3.11 (Eigenschaften Gammafunktion) Die Γ -Funktion ist in die ganze komplexe Ebene mit Ausnahme der Stellen

$$S := \{0, -1, -2, -3, \dots\}$$

(eindeutig) analytisch fortsetzbar und genügt dort der Funktionalgleichung. Die Ausnahmestellen sind Pole erster Ordnung.

Das heißt insbesondere, dass sie für alle reellen $x \notin S$ definiert und beliebig oft differenzierbar ist.

Definition A.3.12 (Reelle Fakultät) Für reelle x wird die Fakultät analog zu (A.3.9) über die Γ -Funktion definiert.

$$x! := \Gamma(x+1) \text{ für } x \in \mathbb{R}$$

An den Polstellen (Definitionslücken) wird die Fakultät gleich ∞ gesetzt.

$$x! := \pm\infty \text{ für } x \in \mathbb{Z}, x < 0$$

A.3.5 Werte der Gammafunktion

Für $x > 0$ (bzw. $\operatorname{Re}(z) > 0$) kann man $\Gamma(x)$ über das uneigentliche Integral berechnen. Für weitere Werte bedient man sich der Funktionalgleichung.

Beispiel A.3.13

$$\Gamma(1) = \int_0^{\infty} e^{-t} dt = [-e^{-t}]_0^{\infty} = 0 - (-1) = 1$$

Für $1 \leq x < 2$ sind die Werte der Gammafunktion sogar tabellarisiert. Weitere Werte für

$x > 2$ und $x < 1, x \notin \{0, -1, -2, \dots\}$ lassen sich mit der Funktionalgleichung berechnen.

x	$\Gamma(x)$	x	$\Gamma(x)$
1,0	1,00000	1,5	0,88623
1,1	0,95135	1,6	0,89352
1,2	0,91817	1,7	0,90864
1,3	0,89747	1,8	0,93138
1,4	0,88726	1,9	0,96177

Bemerkung A.3.14 Die Gammafunktion hat keine Nullstellen.

A.3.6 Betafunktion

Definition A.3.15 (Beta-Funktion) Für positive $x, y \in \mathbb{R}$ (bzw. komplexe $z, w \in \mathbb{C}$ mit positivem Realteil) sei (z.B. Beichelt, 2003, S.51)

$$B(x, y) := \int_0^1 t^{x-1}(1-t)^{y-1} dt$$

Bemerkung A.3.16 (Eigenschaften der Betafunktion) Es gilt

$$B(x+1, y) = \frac{x}{x+y} \cdot B(x, y)$$

Man kann die Betafunktion auf die Gammafunktion zurückführen.

$$B(x, y) := \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{A.10}$$

A.3.7 Verallgemeinerte Binomialkoeffizienten

Definition A.3.17 (Verallgemeinerte Binomialkoeffizienten) Für $d \in \mathbb{R} \setminus \{-1, -2, -3, \dots\}$ und $k \in \mathbb{N}$ definiert man den verallgemeinerten Binomialkoeffizienten $\binom{d}{k}$ mit Hilfe der Gamma- bzw. Betafunktion.

$$\binom{d}{k} := \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} \tag{A.11}$$

$$\binom{d}{k} := \frac{1}{(d+1)B(k+1, d-k+1)} \tag{A.12}$$

In dieser Definition ist $\binom{d}{k} = 0$, falls $d \in \{-1, -2, -3, \dots\}$, da Γ dort Polstellen hat.

Definition A.3.18 (Verallgemeinerte Binomialkoeffizienten) Für $d \in \mathbb{R}$ und $k \in \mathbb{N}$ definiert

man alternativ:

$$\binom{d}{k} := \prod_{j=1}^k \frac{d-j+1}{j} = \frac{d(d-1) \cdot \dots \cdot (d-k+1)}{k!} \quad (\text{A.13})$$

Hier sind die Binomialkoeffizienten auch für $d \in \{-1, -2, \dots\}$ definiert und nicht wie in Gleichung (A.7) gleich Null. (siehe Bsp. A.3.20, S.330)

Theorem A.3.19 Die Definitionen (A.11), (A.12) und (A.13) sind äquivalent.

Beweis. Einfache Rechnungen und die Verwendung obiger Definitionen und Eigenschaften zeigen:

$$\begin{aligned} \binom{d}{k} &\stackrel{(\text{A.12})}{=} \frac{1}{(d+1)B(k+1, d-k+1)} \\ &\stackrel{(\text{A.10})}{=} \frac{1 \cdot \Gamma(k+1+d-k+1)}{(d+1)\Gamma(k+1)\Gamma(d-k+1)} \\ &\stackrel{(\text{A.8})}{=} \frac{\Gamma(d+2)}{(d+1)\Gamma(k+1)\Gamma(d-k+1)} \\ &= \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} = (\text{A.11}) \\ \text{und außerdem:} &= \frac{\Gamma(d+1)}{k!\Gamma(d-k+1)} \stackrel{(\text{A.8})}{=} \frac{d\Gamma(d)}{k!\Gamma(d-k+1)} = \frac{d(d-1)\Gamma(d-1)}{k!\Gamma(d-k+1)} \\ &= \frac{d(d-1) \cdot \dots \cdot (d-k+1)\Gamma(d-k+1)}{k!\Gamma(d-k+1)} \\ &= \frac{d(d-1) \cdot \dots \cdot (d-k+1)}{k!} = (\text{A.13}) \end{aligned}$$

□

Beispiel A.3.20

$$\begin{aligned} \binom{2,5}{2} &= \frac{2,5 \cdot 1,5}{2} = \frac{3,75}{2} = 1,875 \\ \binom{-1}{k} &= \frac{(-1)(-2) \cdot \dots \cdot (-k)}{k!} = (-1)^k \end{aligned}$$

B Zeitreihenanalyse

B.1 Stationarität

Die folgenden Ergänzungen zum Abschnitt *Stationarität* sind bei Fahrmeir et al. (1981); Rozanov (1975) oder Bhattacharya und Waymire (1990) zu finden.

Definition B.1.1 (Endlich-dimensionale Verteilungen) *Für einen stochastischen Prozess $(Y_t)_{t \in \mathbb{T}}$ heißen die Verteilungen der Zufallsvektoren $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n})$ mit $n \geq 1, t_1, \dots, t_n \in \mathbb{T}$ endlich-dimensionale Verteilungen.*

Wenn man die endlich-dimensionalen Verteilungen eines Prozesses betrachtet, dann beobachtet man Folgendes:

Lemma B.1.2 (Konsistenzbedingung) *Die endlich-dimensionalen Verteilungen eines (existierenden) stochastischen Prozesses erfüllen folgende Konsistenzbedingung.*

Die Verteilungsfunktionen in einer kleineren Dimension müssen als Randverteilungen der höherdimensionalen Verteilungsfunktionen ableitbar sein.

Durch eine höher dimensionale Verteilungsfunktion werden implizit alle Verteilungsfunktionen in kleineren Dimensionen als Randverteilungen vorgegeben. Ein System von Verteilungsfunktionen ist *konsistent*, wenn die explizite Vorgabe dieser Verteilungsfunktionen nicht von der impliziten Definition abweicht.

Gibt man umgekehrt eine Familie von konsistenten Wahrscheinlichkeitsmaßen vor, so existiert ein stochastischer Prozess, der diese als endlich-dimensionale Verteilungen hat. Über die endlich-dimensionalen Verteilungen kann man einen stochastischen Prozess festlegen.

Sind die endlich-dimensionalen Verteilungen zeitinvariant, dann nennt man den Prozess *stark stationär* (Fahrmeir et al., 1981; Brockwell & Davis, 1996).

Definition B.1.3 (Stark stationär) *Ein stochastischer Prozess $(Y_t)_{t \in \mathbb{T}}$ heißt stationär, falls $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}) \stackrel{d}{=} (Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h})$ gilt für alle $n \geq 1, t_1, \dots, t_n \in \mathbb{T}$ und h so, dass $(t_i + h) \in \mathbb{T} \forall t_i$.*

B.2 Lag Operationen

Der Lag-Operator

Hier werden weitere Eigenschaften des in Kapitel 1.5 eingeführten Lag-Operators zusammengefasst.

Bemerkung B.2.1 (Eigenschaften des Lag-Operators) *Der Lag-Operator hat folgende Eigenschaften:*

1. Angewendet auf einen konstanten Prozess $X_t = c$ ergibt sich:

$$Lc = c$$

2. Die Umkehrung des Lag-Operators ist der Lead- oder Forward-Operator. Dieser schiebt den Zeitindex um eine Periode in die Zukunft. Er wird mit L^{-1} bezeichnet.

$$L^{-1}X_t = X_{t+1}$$

3. Da $L^{-1}LX_t = X_t$ wird L^0 gleich der Identität gesetzt.

$$L^0 = 1$$

4. Mehrfache Anwendung des Operators bewirkt eine Verschiebung um mehrere Zeiteinheiten.

$$L^2X_t = L(LX_t) = LX_{t-1} = X_{t-2}$$

Die k -fache Anwendung schreibt man als k -te Potenz des Operators. Für $k \in \mathbb{Z}$ gilt:

$$L^k X_t = X_{t-k}$$

5. Es gelten die üblichen Rechenregeln für Potenzen.

$$L^m X_{t-n} = L^m(L^n X_t) = L^{m+n} X_t = X_{t-(m+n)}$$

Bemerkung B.2.2 (Linearität) *Der Lag-Operator ist ein linearer Operator.*

$$(aL^m + bL^n)(X_t + Y_t) = aX_{t-m} + bX_{t-n} + aY_{t-m} + bY_{t-n}$$

Deshalb ist es auch möglich, Lagpolynome zu definieren.

Invertierbarkeit von Lagpolynomen

Wir haben gesehen, dass die Invertierung von Polynomen nicht ohne Schwierigkeiten ist. Deshalb betrachten wir hier eine Bedingung, wann ein Lagpolynom invertierbar ist. Dazu führen wir zunächst den Begriff des kausal invertierbaren Lagpolynoms (Haßler, 2007) ein.

Definition B.2.3 (Kausal invertierbares Lagpolynom) *Ein Lagpolynom $P(L)$ ist kausal invertierbar, wenn eine Potenzreihenentwicklung mit nicht-negativen Potenzen und absolut summierbaren Koeffizienten existiert:*

$$(P(L))^{-1} = \sum_{j=0}^{\infty} \alpha_j L^j$$

mit

$$\sum_{j=0}^{\infty} |\alpha_j| < \infty. \quad (\text{B.1})$$

Eine Bedingung für die Existenz einer kausalen Inversen gewinnt man mit Hilfe der z -Transformierten, die wir bereits in Definition 1.5.4 auf Seite 28 in Kapitel 1.5 definiert hatten.

Theorem B.2.4 (Bedingung für kausale Invertierbarkeit) *Ein Lagpolynom $P(L)$ ist genau dann kausal invertierbar, wenn alle Nullstellen der z -Transformierten dem Betrage nach größer als 1 sind:*

$$P(z) = 0 \Rightarrow |z| > 1.$$

Dieses Theorem erinnert sehr an die Stationaritätsbedingung für autoregressive Prozesse. Deshalb betrachten wir nun das Beispiel des AR(1)-Prozesses, um die Herleitung noch einmal zu verfolgen.

Beispiel B.2.5 (AR(1)-Polynom) *Betrachten wir den Spezialfall eines Polynoms erster Ordnung*

$$A_1(L) = 1 - aL.$$

Durch Koeffizientenvergleich erhält man folgende Reihenentwicklung:

$$A_1(L)^{-1} = (1 - aL)^{-1} = \sum_{j=0}^{\infty} a^j L^j.$$

Dies ist die bekannte geometrische Reihe (siehe auch Anhang B.3, S.334), die nur für $|a| < 1$

konvergiert mit

$$\sum_{j=0}^{\infty} a^j = \frac{1}{1-a} \Leftrightarrow |a| < 1.$$

Eine absolut summierbare Reihenentwicklung des inversen Polynoms $A_1(L)^{-1}$ existiert also nur für bestimmte Werte von a , nämlich für $-1 < a < 1$. Die Invertierbarkeit von $A_1(L)$ ist also abhängig vom Parameter a .

Für die Nullstellen der z -Transformierten $A_1(z) = 1 - az$ gilt

$$\begin{aligned} A_1(z) = 1 - az &= 0 \\ \Leftrightarrow 1 &= az \\ \Leftrightarrow z &= \frac{1}{a} \end{aligned}$$

Im Fall $|a| < 1$ liegen sie alle außerhalb des Einheitskreises, sind im Betrag also größer als 1,

$$|a| < 1 \Leftrightarrow \frac{1}{|a|} > 1.$$

B.3 Wolddarstellung

Geometrische Reihe

Unter einer geometrischen Reihe versteht man die Summe der Potenzen einer reellen Zahl. Für jede Basis (außer 1) ergibt sich ein leicht zu berechnender Ausdruck. Für solche, die vom Betrag echt kleiner als Eins sind, konvergiert sogar die unendliche Summe (=Reihe) (Wolter & Dahn, 2000; Meyberg, 1999).

Theorem B.3.1 (Geometrische Reihe) Für $x \neq 1$ und jede natürliche Zahl n gilt

$$\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x}.$$

Theorem B.3.2 (Unendliche geometrische Reihe) Die Reihe $\sum_{k=0}^{\infty} x^k$ konvergiert für alle $|x| < 1$ mit dem Grenzwert

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

B.4 Impulsantwortfunktion

B.4.1 partielle Ableitung

Eine Funktion von mehreren Variablen hat nicht eine einzige Ableitung, sondern mehrere partielle Ableitungen. Je nachdem, ob die Funktion die Variablen auf einen (reellen) Wert oder auf einen Vektor abbildet, gibt es für jede Variable eine oder mehrere partielle Ableitungen (z.B. Wolter & Dahn, 2000).

Betrachtet man eine Funktion, die zwei Veränderliche x, y auf einen reellen Wert abbildet, dann hat die Funktion zwei partielle Ableitungen, eine „nach x “ und eine „nach y “. Eine partielle Ableitung berechnet man, indem man die Funktion „ganz normal“ nach einer Variablen ableitet und die anderen Variablen dabei als konstant betrachtet.

Beispiel B.4.1 (\mathbb{R}^2 nach \mathbb{R}) *Wir betrachten zunächst die Funktion f , die die beiden Variablen x, y auf den reellen Wert $x^2 + 5y$ abbildet*

$$f(x, y) = x^2 + 5y.$$

Hält man y und damit auch $5y$ fest und leitet f nach x ab, ergibt sich

$$\frac{\partial f}{\partial x} = 2x,$$

da die Ableitung von $f(x) = x^2 + c$ gleich $2x$ ist. Ebenso ergibt sich die partielle Ableitung nach y

$$\frac{\partial f}{\partial y} = 5,$$

da die Ableitung von $f(y) = 5y$ gleich 5 ist.

Komplizierter wird es wenn die beiden Variablen nicht nur additiv, sondern multiplikativ verknüpft sind

$$f(x, y) = x^2 y^3.$$

Die partiellen Ableitungen von f ergeben sich, in dem man f nach jeweils einer Variablen ableitet und die andere dabei als Konstante betrachtet

$$\frac{\partial f}{\partial x} = 2xy^3,$$

da die Ableitung von $f(x) = x^2 \cdot c$ gleich $2x \cdot c$ ist. Die partielle Ableitung nach y ist

$$\frac{\partial f}{\partial y} = 3x^2 y^2.$$

Bislang sind wir davon ausgegangen, dass die Funktion f die beiden Variablen x, y auf eine reelle Zahl abbildet. Dies kann man auch so betrachten, dass die Funktionen einen zweidimensionalen Vektor auf eine (eindimensionale) Zahl abbildet: $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Nun werden wir Funktionen betrachten, die einen zweidimensionalen Vektor wieder auf einen zwei- oder auch drei- (mehr-)dimensionalen Vektor abbilden.

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

Für die partiellen Ableitungen betrachtet man die rechte Seite nicht als Ergebnisvektor einer Funktion, sondern als Vektor von Funktionsergebnissen

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \\ f_3(x, y) \end{pmatrix}.$$

Jede der Funktionen f_1, f_2, f_3 hängt nun wieder von zwei Variablen x, y ab und hat somit zwei partielle Ableitungen. Für die Gesamtfunktion f kann man also insgesamt 6 partielle Ableitungen berechnen. Es ist üblich diese in der sogenannten Hessematrix zusammenzufassen

$$\Psi = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} \end{pmatrix}.$$

Beispiel B.4.2 (\mathbb{R}^2 nach \mathbb{R}^3) Für die Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 + 5y \\ x^2 y^3 \\ x \end{pmatrix}$$

sind die partiellen Ableitungen

$$\begin{array}{lll} \frac{\partial f_1}{\partial x} = 2x & \frac{\partial f_1}{\partial y} = 5 & \frac{\partial f_2}{\partial x} = 2xy^3 \\ \frac{\partial f_2}{\partial y} = 3x^2y & \frac{\partial f_3}{\partial x} = 1 & \frac{\partial f_3}{\partial y} = 0 \end{array}$$

C Kointegrationsanalyse

C.1 Integrierte Prozesse

C.1.1 Differenzenoperator

Wir hatten in Kapitel 2.1 den Differenzenoperator Δ verwendet und auch darauf hingewiesen, dass die Darstellung als Lagpolynom besonders bei mehrfacher Anwendung sehr nützlich ist. Hier sollen nun die mathematische Definition (Schröder & Buscher, 2002, S.141) und die Darstellung des allgemeinen Differenzenoperators $(1 - L)^d$ als Lagpolynom nachgereicht werden.

Definition C.1.1 (Differenzen-Operator) *Mit Hilfe des Lag-Operators ergibt sich für die Bildung von ersten Differenzen folgende Schreibweise:*

$$\begin{aligned}\Delta Y_t &:= (1 - L)Y_t = Y_t - Y_{t-1} \\ \Delta_k Y_t &:= (1 - L^k)Y_t = Y_t - Y_{t-k}.\end{aligned}$$

Interessant ist vor allem die mehrfache Anwendung des Operators.

$$\Delta^d := (1 - L)^d = \underbrace{(1 - L) \cdot \dots \cdot (1 - L)}_{d \text{ mal}} \quad (\text{C.1})$$

Um die Darstellung von $(1 - L)^d$ als Lagpolynom herzuleiten, gibt es zwei verschiedene Möglichkeiten. Zum einen kann man die Gleichung (C.1) einfach ausmultiplizieren und das entstandene Lagpolynom auf den Prozess anwenden, zum anderen kann man den Operator $(1 - L)$ auch direkt mehrfach auf den Prozess anwenden. Dass beide Herangehensweisen zum selben Ergebnis führen, wollen wir am Fall $d = 2$ illustrieren.

Beispiel C.1.2 ($d = 2$) *Für das Ausmultiplizieren von $(1 - L)^2$ kann man die 2. Binomische Formel $(a - b)^2 = a^2 - 2ab + b^2$ verwenden.*

$$\begin{aligned}\Delta^2 Y_t &= (1 - L)^2 Y_t \\ &\stackrel{\text{Bin. Formel}}{=} (1 - 2L + L^2) Y_t \\ &= Y_t - 2Y_{t-1} + Y_{t-2}\end{aligned}$$

Zweifaches Anwenden von $(1 - L)$ bedeutet, dass man auf den Differenzenprozess, der durch

das erste Anwenden entsteht, erneut den Differenzen-Operator anwendet. Dies ergibt:

$$\begin{aligned}
 \Delta^2 Y_t &= \Delta(\Delta Y_t) \\
 &= \Delta(Y_t - Y_{t-1}) \\
 &= (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \\
 &= Y_t - 2Y_{t-1} + Y_{t-2}
 \end{aligned}$$

Die beiden Herangehensweisen führen also zum selben Ergebnis.

Mehrfaches Differenzieren entspricht der mehrfachen Anwendung des Differenzen-Operators, was wir mit einer Potenz notieren. Um die Darstellung des Operators $(1 - L)^d$ als Lagpolynom zu erhalten, kann man den Binomischen Lehrsatz (siehe Theorem A.3.5, S.326) verwenden.

Theorem C.1.3 (*d*-fache Differenzierung) *Wendet man den Binomischen Lehrsatz auf den Differenzenoperator $(1 - L)^d$ an, erhält man folgende Darstellung des Operators als Lagpolynom.*

$$(1 - L)^d = \sum_{k=0}^d \binom{d}{k} 1^{d-k} (-L)^k \quad (\text{C.2})$$

$$= \sum_{k=0}^d \binom{d}{k} 1^k (-L)^{d-k} \quad (\text{C.3})$$

Für unsere Zwecke ist die Formel (C.2) sinnvoller, da das d nicht weiter als Exponent auftaucht:

$$(1 - L)^d = \sum_{k=0}^d \binom{d}{k} (-1)^k L^k$$

Bemerkung C.1.4 (Differenzenoperator als Lagpolynom) *Dabei wird der Differenzenoperator $(1 - L)^d$ wie folgt als Lagpolynom verstanden:*

$$\begin{aligned}
 (1 - L)^d &= \sum_{k=0}^d \binom{d}{k} (-1)^k L^k \\
 &= 1 - dL + \binom{d}{2} L^2 - \binom{d}{3} L^3 + \dots
 \end{aligned}$$

C.1.2 Integrationsoperator

Die Darstellung (C.2) ist auch bei negativen Exponenten gültig, wenn man die sogenannten verallgemeinerten Binomialkoeffizienten benutzt, wie sie in Anhang A.3 (S.324ff) definiert sind.

Man erhält dieselbe Darstellung jedoch auch, wenn man $(1 - L)^{-d}$ als inversen Operator von $(1 - L)^d$ versteht

$$(1 - L)^{-d} = \left((1 - L)^d \right)^{-1}.$$

Bei ganzzahligem d ist $(1 - L)^d$ als endliches Lagpolynom darstellbar. Dass das Invertieren bei Polynomen nicht problemlos möglich ist, wurde in Anhang A.1 (S.305ff) beschrieben. Dies führt dazu, dass der Operator $(1 - L)^{-d}$ als Lagpolynom eine unendliche Ordnung hat, wie in der Reihendarstellung des Operators sichtbar wird.

Dass man durch Invertieren des Lagpolynoms dieselben Koeffizienten wie bei Verwendung von $-d$ über den Binomischen Lehrsatz erhält, soll folgendes Beispiel veranschaulichen.

Beispiel C.1.5 $((1 - L)^{-2})$ Wir versuchen durch Umstellen Y_t über ϵ_t darzustellen und so eine Darstellung für den inversen Operator zu erhalten. Zur Vereinfachung sei $X_t = 0$ für alle negativen Zeitpunkte $t < 0$. Die Betrachtung beginnt also bei Y_0 .

Zunächst kann man $(1 - L)^2$ als Lagpolynom darstellen und die Summanden auf die andere Seite der Gleichung bringen.

$$\begin{aligned} (1 - L)^2 Y_t &= \epsilon_t \\ Y_t - 2Y_{t-1} + Y_{t-2} &= \epsilon_t \\ Y_t &= \epsilon_t + 2Y_{t-1} - Y_{t-2} \end{aligned}$$

Diese allgemeine Darstellung zu einem Zeitpunkt t gilt nun auch am ersten Beobachtungszeitpunkt Y_0 . Der Einfachheit hatten wir angenommen, dass alle vorangegangenen Zustände gleich Null waren.

$$\begin{aligned} Y_0 &= \epsilon_0 + 2Y_{-1} - Y_{-2} \\ &= \epsilon_0 \end{aligned}$$

Da wir Y_t über die letzten zwei Zeitpunkte dargestellt haben, können wir die Darstellung von Y_0 verwenden, um Y_1 nur durch Lags von ϵ_t darzustellen.

$$\begin{aligned} Y_1 &= \epsilon_1 + 2Y_0 - Y_{-1} \\ &= \epsilon_1 + 2\epsilon_0 \end{aligned}$$

Dies kann man für die nachfolgenden Beobachtungspunkte immer weiter fortsetzen.

$$\begin{aligned} Y_2 &= \epsilon_2 + 2Y_1 - Y_0 \\ &= \epsilon_2 + 2(\epsilon_1 + 2\epsilon_0) - \epsilon_0 \\ &= \epsilon_2 + 2\epsilon_1 + 3\epsilon_0 \\ Y_3 &= \epsilon_3 + 2Y_2 - Y_1 \\ &= \epsilon_3 + 2(\epsilon_2 + 2\epsilon_1 + 3\epsilon_0) - (\epsilon_1 + 2\epsilon_0) \\ &= \epsilon_3 + 2\epsilon_2 + 3\epsilon_1 + 4\epsilon_0 \end{aligned}$$

So gelangt man zu einer allgemeinen Darstellung von Y_t über die Lags von ϵ_t

$$\begin{aligned} Y_t &= \epsilon_t + 2\epsilon_{t-1} + 3\epsilon_{t-2} + 4\epsilon_{t-3} + \dots \\ &= \left(\sum_{j=1}^{\infty} j \cdot L^{j-1} \right) \epsilon_t \end{aligned}$$

Daraus folgt also die Darstellung des inversen Operators mit theoretisch unendlicher Ordnung.

$$(1 - L)^{-2} = \left(\sum_{j=1}^{\infty} j \cdot L^{j-1} \right)$$

Zum Vergleich einige verallgemeinerte Binomialkoeffizienten:

$$\begin{aligned} \binom{-2}{0} &= 1 \\ \binom{-2}{1} &= \frac{-2}{1} = -2 \\ \binom{-2}{2} &= \frac{(-2)(-3)}{2!} = 3 \\ \binom{-2}{3} &= \frac{(-2)(-3)(-4)}{3!} = -4 \end{aligned}$$

Berücksichtigt man das alternierende Vorzeichen im Binomialsatz erhält man dieselben Koeffizienten wie oben.

Da die Koeffizienten des Operators $(1 - L)^{-2}$ nicht absolut summierbar sind, ist die Bedingung (B.1) aus der Definition B.2.3 von S.333 verletzt. Deshalb ist der Prozess $(1 - L)^{-2} \epsilon_t$ nach Theorem B.2.4 (S.333) nicht stationär.

C.1.3 Logarithmus

Der Logarithmus von x zur Basis b ist definiert als derjenige Exponent a für den gilt, dass seine Potenz zur Basis b gleich x ist, also $b^a = x$.

$$\begin{aligned}\log_b(x) &= a, \text{ falls} \\ b^a &= x\end{aligned}$$

Der Logarithmus zur Basis b ist also die Umkehrfunktion des Potenzierens zur Basis b .

$$\log_b(b^a) = a$$

Der Logarithmus von Eins ist deshalb gleich Null für alle Basen b , $b^0 = 1$ für beliebiges b gilt.

$$\log_b(1) = 0 \text{ für alle } b$$

Der Logarithmus, der Potenzieren zur Basis e (Eulersche Zahl $e \approx 2,71$) heißt natürlicher Logarithmus. Er wird mit \ln oder auch als \log (ohne Basis) geschrieben⁴⁶.

Der Logarithmus macht aus Produkten Summen.

$$\begin{aligned}\log(a \cdot b) &= \log(a) + \log(b) \\ \log\left(\prod_{i=1}^n a_i\right) &= \sum_{i=1}^n \log(a_i)\end{aligned}$$

C.2 Darstellungsformen

C.2.1 Umrechnung Moving Average zu Common Trend Darstellung

Ausgangspunkt unserer Überlegungen ist die MA-Darstellungen eines bivariaten kointegrierten Prozesses

$$\begin{pmatrix} \Delta Y_{1,t} \\ \Delta Y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}.$$

Ausgeschrieben ergibt sich folgendes System mit den Einzeldarstellungen der Differenzprozesse

$$\begin{aligned}\Delta Y_{1,t} &= \mu_1 + \psi_{11}(L)\epsilon_{1,t} + \psi_{12}(L)\epsilon_{2,t} \\ \Delta Y_{2,t} &= \mu_2 + \psi_{21}(L)\epsilon_{1,t} + \psi_{22}(L)\epsilon_{2,t}.\end{aligned}$$

⁴⁶Die Schreibweis \log wird häufig auch für den Logarithmus zur Basis 10 verwendet.

Wir betrachten die Darstellung der ersten Prozesskomponente nun genauer. Analoges gilt auch für die zweite Komponente und im mehrdimensionalen Fall auch für jede weitere. Zunächst lösen wir den Differenzenoperator $\Delta Y_t = Y_t - Y_{t-1}$ auf und stellen nach Y_t um

$$\begin{aligned}\Delta Y_{1t} = Y_{1,t} - Y_{1,t-1} &= \mu_1 + \psi_{11}(L)\epsilon_{1,t} + \psi_{12}(L)\epsilon_{2,t} \\ Y_{1,t} &= Y_{1,t-1} + \mu_1 + \psi_{11}(L)\epsilon_{1,t} + \psi_{12}(L)\epsilon_{2,t}.\end{aligned}\tag{C.4}$$

Nun haben wir Y_{1t} durch den Wert des Prozesses zum vorigen Zeitpunkt $Y_{1,t-1}$ dargestellt. Dieser lässt sich nun wiederum über den Wert zum Zeitpunkt $t - 2$ darstellen und so fort.

$$Y_{1,t-1} = Y_{1,t-2} + \mu_1 + \psi_{11}(L)\epsilon_{1,t-1} + \psi_{12}(L)\epsilon_{2,t-1}$$

Setzt man diese Darstellung für Y_{t-1} in die Gleichung (C.4) ein und sortiert die Summanden etwas um

$$\begin{aligned}Y_{1,t} &= \underbrace{Y_{1,t-2} + \mu_1 + \psi_{11}(L)\epsilon_{1,t-1} + \psi_{12}(L)\epsilon_{2,t-1}}_{Y_{1,t-1}} + \mu_1 + \psi_{11}(L)\epsilon_{1,t} + \psi_{12}(L)\epsilon_{2,t} \\ &= Y_{1,t-2} + 2\mu_1 + \psi_{11}(L)[\epsilon_{1,t} + \epsilon_{1,t-1}] + \psi_{12}(L)[\epsilon_{2,t} + \epsilon_{2,t-1}],\end{aligned}$$

erhält man eine Darstellung von $Y_{1,t}$ über $Y_{1,t-2}$. Dabei summieren sich die Werte der beiden Fehlerprozesse „rückwärts“ auf.

Rekursiv kann man nun auch $Y_{1,t-2}$ über seinen Vorgänger darstellen und somit in der Darstellung ersetzen

$$Y_{1,t} = Y_{1,t-3} + 3\mu_1 + \psi_{11}(L)[\epsilon_{1,t} + \epsilon_{1,t-1} + \epsilon_{1,t-2}] + \psi_{12}(L)[\epsilon_{2,t} + \epsilon_{2,t-1} + \epsilon_{2,t-2}].$$

Man kann hier schon sehen, dass die Darstellung vier Bestandteile hat: ein Lag von Y_t , ein Vielfaches des Mittelwertes μ_1 und jeweils die gewichtete Summe von Lags eines der beiden Fehlerprozesse.

Im i -ten Schritt der Rekursion hat $Y_{1,t}$ folgende Darstellung

$$Y_{1,t} = Y_{1,t-i} + i\mu_1 + \psi_{11}(L) \sum_{j=0}^{i-1} \epsilon_{1,t-j} + \psi_{12}(L) \sum_{j=0}^{i-1} \epsilon_{2,t-j}.$$

Führt man die Rekursion vollständig aus, gelangt man nach t Schritten zu einer Gestalt, in der $Y_{1,t}$ auf seinen Startwert $Y_{1,0}$ zurückgeführt wird

$$Y_{1,t} = Y_{1,0} + t\mu_1 + \psi_{11}(L) \sum_{j=0}^{t-1} \epsilon_{1,t-j} + \psi_{12}(L) \sum_{j=0}^{t-1} \epsilon_{2,t-j}.$$

Der letzte Summand in den beiden Summen ist $\epsilon_{1,t-(t-1)} = \epsilon_{1,1}$ bzw. $\epsilon_{2,1}$. Die ausgeschriebe-

nen Summen sind also beide von der Gestalt⁴⁷

$$\begin{aligned} \sum_{j=0}^{t-1} \epsilon_{t-j} &= \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \dots + \epsilon_2 + \epsilon_1 \\ &= \epsilon_1 + \epsilon_2 + \dots + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &= \sum_{k=1}^t \epsilon_k. \end{aligned}$$

Für die zweite Einzelreihe (und in höherdimensionalen Fällen für jede weitere Einzelreihe) gilt nach Rekursion eine analoge Darstellung

$$Y_{2,t} = Y_{2,0} + t\mu_2 + \psi_{21}(L) \sum_{k=1}^t \epsilon_{1,k} + \psi_{22}(L) \sum_{k=1}^t \epsilon_{2,k}.$$

Im bivariaten Fall also

$$\begin{pmatrix} Y_{1,t} \\ Y_{2,t} \end{pmatrix} = \begin{pmatrix} Y_{1,0} \\ Y_{2,0} \end{pmatrix} + t \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \psi_{11}(L) & \psi_{12}(L) \\ \psi_{21}(L) & \psi_{22}(L) \end{pmatrix} \begin{pmatrix} \sum_{k=1}^t \epsilon_{1,k} \\ \sum_{k=1}^t \epsilon_{2,k} \end{pmatrix}$$

Im bivariaten wie im allgemeinen multivariaten Fall gilt also folgende Vektorschreibweise⁴⁸

$$Y_t = Y_{t,0} + t\mu + \Psi(L) \sum_{k=1}^t \epsilon_t.$$

Verwendet man nun die Zerlegung der Matrix $\Psi(L)$ (Lemma 2.3.1, S.81), dann wird dies zu

$$Y_t = Y_{t,0} + t\mu + \Psi(1) \sum_{k=1}^t \epsilon_t + (1-L)\Psi^*(L) \sum_{k=1}^t \epsilon_t. \quad (\text{C.5})$$

Da $\Psi(1)$ den Rang $K-r$, im bivariaten Fall also gleich $r=1$, hat, wirken nicht alle in dem Vektor

$$\sum_{k=1}^t \epsilon_t = \begin{pmatrix} \sum_{k=1}^t \epsilon_{1,t} \\ \sum_{k=1}^t \epsilon_{2,t} \\ \vdots \\ \sum_{k=1}^t \epsilon_{K,t} \end{pmatrix}$$

enthaltenen Random Walks $\sum_{k=1}^t \epsilon_{i,t}$ im System ΔY_t . Da man eine $K \times K$ -dimensionale Matrix vom Rang $K-r$ als Matrix mit r Null-Spalten schreiben kann, werden r Random Walks durch die Multiplikation mit $\Psi(1)$ herausgekürzt, wie man am Beispiel einer bivariaten

⁴⁷Hier ist der Index für den ersten bzw. zweiten Fehlerterm ausgelassen. Man kann die Formel jedoch auch vektoriell verstehen. Mit ϵ_t als Vektor ist sie ebenfalls korrekt.

⁴⁸Hier ist ϵ_t ein K -dimensionaler Vektor. $\sum_{k=1}^t \epsilon_t$ ist damit der Vektor der Summen, also ein Vektor mit Random Walks.

Matrix sieht

$$\begin{aligned}\Psi(1) \sum_{k=1}^t \epsilon_t &= \begin{pmatrix} 0 & \psi_{12}(1) \\ 0 & \psi_{22}(1) \end{pmatrix} \begin{pmatrix} \sum_{k=1}^t \epsilon_{1,t} \\ \sum_{k=1}^t \epsilon_{2,t} \end{pmatrix} \\ &= \begin{pmatrix} \psi_{12}(1) \sum_{k=1}^t \epsilon_{2,t} \\ \psi_{22}(1) \sum_{k=1}^t \epsilon_{2,t} \end{pmatrix}.\end{aligned}$$

Hier wirkt nur noch $\sum_{k=1}^t \epsilon_{2,t}$. Der Rang $K - r$ der Matrix $\Psi(1)$ gibt also die Anzahl der gemeinsamen Trends (Random Walks) an.

Betrachtet man den hinteren Teil der Gleichung (C.5)

$$(1 - L)\Psi^*(L) \sum_{k=1}^t \epsilon_{t-k},$$

dann kann man den Faktor $(1 - L)$ in die Summe ziehen, wodurch sie sich stark vereinfacht

$$\begin{aligned}(1 - L)\Psi^*(L) \sum_{k=1}^t \epsilon_k &= \Psi^*(L) \sum_{k=1}^t (1 - L)\epsilon_k \\ &= \Psi^*(L) \sum_{k=1}^t (\epsilon_k - \epsilon_{k-1}) \\ &= \Psi^*(L) ((\epsilon_1 - \epsilon_0) + (\epsilon_2 - \epsilon_1) + (\epsilon_3 - \epsilon_2) + \dots + (\epsilon_{t-1} - \epsilon_{t-2}) + (\epsilon_t - \epsilon_{t-1}))\end{aligned}$$

Dies ist eine sogenannte *Teleskopsumme*. Das bedeutet, dass die Glieder der Summe (wie ein Teleskop) ineinander geschoben werden können, wodurch die Summe „kürzer wird“. Dies ist der Fall, weil jeder Summand einen Teil des vorigen wegkürzt. So ist $(\epsilon_1 - \epsilon_0) + (\epsilon_2 - \epsilon_1) = \epsilon_0 + \epsilon_2$ und so fort. Am Ende bleiben nur ein Teil des ersten und ein Teil des letzten Summanden stehen. In diesem Fall $-\epsilon_0 + \epsilon_t$, wodurch sich die Summe zu

$$(1 - L)\Psi^*(L) \sum_{k=1}^t \epsilon_{t-k} = \Psi^*(L)(-\epsilon_0 + \epsilon_t)$$

vereinfacht. Da ϵ_t als Weißes Rauschen per definitionem stationär ist, gilt dies auch für lineare (hier polynomielle) Umformungen. Zusammengefasst ergibt sich

$$Y_t = Y_0 + t\mu + \Psi(1) \sum_{k=1}^t \epsilon_t + \underbrace{\Psi^*(L)\epsilon_t - \Psi^*(L)\epsilon_0}_{V_t}$$

die Common-Trend-Darstellung des Prozesses Y_t .

C.2.2 Umrechnung VAR(p) zu Fehlerkorrekturdarstellung

Um aus der vektoriellen autoregressiven Darstellung in die vektorielle Fehlerkorrekturdarstellung eines Prozesses zu gelangen, sind nur wenige einfache Rechenschritte notwendig, die im Folgenden aufgezeigt werden.

Ausgangspunkt ist die VAR(p)-Darstellung des Prozesses

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t.$$

Der entscheidende Schritt, um auf die Fehlerkorrekturdarstellung zu kommen, ist die Subtraktion von Y_{t-1} auf beiden Seiten, dadurch gelangt man auf der linken Seite zum Differenzenprozess

$$\underbrace{Y_t - Y_{t-1}}_{=\Delta Y_t} = c - \underbrace{Y_{t-1}} + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t.$$

Im Folgenden addiert man auf der rechten Seite sukzessiv Terme, die in Summe Null ergeben, den Wert der Gleichung also nicht verändern, aber zu einer Umstrukturierung der Variablen führen. Im ersten Schritt ist dies der Term $+A_2 Y_{t-1} - A_2 Y_{t-1} (= 0)$

$$\Delta Y_t = c - Y_{t-1} + A_1 Y_{t-1} + \underbrace{+A_2 Y_{t-1} - A_2 Y_{t-1}} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \epsilon_t.$$

Die beiden Summanden, die gemeinsam Null ergeben, werden nun getrennt in die Darstellung eingebunden. Der positive Teil $+A_2 Y_{t-1}$ wird mit $-Y_{t-1}$ aus dem ersten Rechenschritt und $+A_1 Y_{t-1}$ aus der Modelldarstellung zusammengefasst zu $(-1 + A_1 + A_2) Y_{t-1}$. Der negative Summand $-A_2 Y_{t-1}$ wird mit $A_2 Y_{t-2}$ zur Differenz ΔY_{t-1} verbunden. $-A_2 Y_{t-1} + A_2 Y_{t-2} = -(A_2 Y_{t-1} - A_2 Y_{t-2}) = -A_2 (Y_{t-1} - Y_{t-2}) = -A_2 \Delta Y_{t-1}$.

$$\Delta Y_t = c + (-1 + A_1 + A_2) Y_{t-1} - A_2 \underbrace{(Y_{t-1} - Y_{t-2})}_{\Delta Y_{t-1}} + \dots + A_p Y_{t-p} + \epsilon_t$$

Dieses Prinzip wird sukzessiv fortgesetzt. Dazu müssen in jedem Schritt mehr Summanden aufgenommen werden. Im nächsten Schritt kommen mit der Matrix A_3 gewichtete Summanden hinzu: $+A_3 Y_{t-1} - A_3 Y_{t-1} + A_3 Y_{t-2} - A_3 Y_{t-2} (= 0)$

$$\begin{aligned} \Delta Y_t &= c + (-1 + A_1 + A_2) Y_{t-1} + \underbrace{+A_3 Y_{t-1} - A_3 Y_{t-1} + A_3 Y_{t-2} - A_3 Y_{t-2}} \\ &\quad - A_2 \Delta Y_{t-1} + A_3 Y_{t-2} \dots + A_p Y_{t-p} + \epsilon_t \\ &= c + (-1 + A_1 + A_2 + A_3) Y_{t-1} + (-A_2 - A_3) \Delta Y_{t-1} - A_3 \underbrace{(Y_{t-2} - Y_{t-3})}_{\Delta Y_{t-2}} \dots + A_p Y_{t-p} + \epsilon_t \end{aligned}$$

Und so weiter, bis zum Schluss alle Lags von Y_t mit der Matrix A_p gewichtet noch addiert

werden: $+A_p Y_{t-1} - A_p Y_{t-1} + A_p Y_{t-2} - A_p Y_{t-2} + \dots + A_p Y_{t-p+1} - A_p Y_{t-p+1} (= 0)$

$$\begin{aligned} \Delta Y_t &= c + (-\mathbb{I} + A_1 + A_2 + \dots + A_p) Y_{t-1} \\ &\quad (-A_2 - \dots - A_p)(Y_{t-1} - Y_{t-2}) \\ &\quad (-A_3 - \dots - A_p)(Y_{t-2} - Y_{t-3}) \\ &\quad \vdots \\ &\quad (-A_p)(Y_{t-p+1} - Y_{t-p+1}) \\ &\quad + \epsilon_t \end{aligned}$$

Wie man erkennen kann, sind hier die Darstellung der Modellparameter Γ_i bzw. Π definiert (vgl. Abschnitt 2.3.3, S.92).

$$\begin{aligned} \Pi &= -\mathbb{I}_K + A_1 + \dots + A_p \\ \Gamma_i &= -A_{i+n} - \dots - A_p \\ \Gamma_{p-1} &= -A_p \end{aligned}$$

Aus diesen ergibt sich direkt, wie man bei geogener Fehlerkorrekturdarstellung die Parameter der autoregressiven Darstellung berechnet.

Schwieriger und deshalb eigentlich nur für theoretische Aspekte interessant, ist die Umformung zwischen den zwei Grundmodellen VAR- und MA-Darstellung eines kointegrierten Prozesses.

C.2.3 VAR(p) zu MA

Mit Hilfe einer speziellen Form der Smith McMillan-Zerlegung (siehe Anhang C.2.4, S.347) für Polynommatrizen kann man aus der Darstellung des kointegrierten Systems in vektorautoregressiver Form auf die Moving Average Darstellung schließen.

Dazu schreibt man den Prozess in seiner Form mit der Lagpolynommatrix $A(L)$ wie folgt

$$A(L)Y_t = c + \epsilon_t$$

und verwendet die Smith McMillan-Zerlegung (vgl. Lemma C.2.4, S.349)

$$A(L) = Q(L)M(L)R(L),$$

wobei $Q(L)$ und $R(L)$ Polynommatrizen sind, deren Determinante nur Nullstellen außerhalb des Einheitskreises besitzt, und $M(L)$ folgende Gestalt hat

$$M(L) = \begin{pmatrix} \mathbb{I}_{K-r} & 0 \\ 0 & (1-L)\mathbb{I}_r \end{pmatrix}.$$

Ersetzt man die VAR-Polynommatrix durch ihre Zerlegung und multipliziert mit $R^{-1}(L)$, erhält man

$$M(L)Q(L)Y_t = R^{-1}(L)(c + \epsilon_t).$$

Multipliziert man diese Gleichung mit einer Matrix von ähnlicher Gestalt wie $M(L)$, bei der an den Stellen, wo $M(L)$ den Eintrag Eins hat, $(1 - L)$ steht und umgekehrt

$$\tilde{M}(L) = \begin{pmatrix} (1 - L)\mathbb{I}_{K-r} & 0 \\ 0 & \mathbb{I}_r \end{pmatrix},$$

führt dies dazu, dass man auf den Differenzenprozess ΔY_t kommt ($\tilde{M}(L)M(L) = (1 - L)\mathbb{I}_K$)

$$Q(L)\Delta Y_t = \tilde{M}(L)R^{-1}(L)(c + \epsilon_t).$$

Im letzten Schritt

$$\Delta Y_t = \underbrace{Q^{-1}(L)\tilde{M}(L)R^{-1}(L)}_{\Psi(L)}(c + \epsilon_t).$$

Nun verwendet man noch, dass bei der Anwendung eines Lagpolynoms auf eine Konstante gilt $P(L)c = P(1)c$ (Bemerkung 1.5.3, S.28) und erhält

$$\Delta Y_t = \underbrace{\Psi(1)c}_{\mu} + \Psi(L)\epsilon_t.$$

Gesamt ergibt sich mit

$$\begin{aligned} \mu &= Q^{-1}(1)\tilde{M}(1)R^{-1}(1)c \\ \Psi(L) &= Q^{-1}(L)\tilde{M}(L)R^{-1}(L) \end{aligned}$$

die Moving Average Darstellung

$$\Delta Y_t = \mu + \Psi(L)\epsilon_t.$$

C.2.4 SmithMcMillan-Zerlegung

Der Vollständigkeit wegen wird hier die allgemeine Zerlegung einer Polynommatrix nach SmithMcMillan vorgestellt.

Da wir nur daran interessiert sind, die autogressive Lag-Polynommatrix $A(L)$ zu zerlegen, benötigen wir folgendes Lemma nicht in seiner vollen Tragweite, sondern nur für Polynommatrizen. Dennoch wird es hier in seiner Allgemeinheit nach Yoo (1987, p.25 Lemma 3.4)

wiedergegeben. Eine ausführliche Darstellung von Polynommatrizen und solchen mit gebrochenrationalen Funktionen als Einträge sowie der Smith- und SmithMcMillan-Form dieser Matrizen findet sich bei Kailath (1980, Kapitel 6)

Lemma C.2.1 (SmithMcMillan-Zerlegung) *Eine $K \times r$ -Matrix $C(L)$ mit gebrochenrationalen Funktionen als Einträgen, die den Rang r hat, kann wie folgt zerlegt werden*

$$C(L) = U^{-1}(L)N(L)V^{-1}(L),$$

wobei gilt

1. $U(L)$ und $V(L)$ sind unimodulare Polynommatrizen, d.h. dass auch ihre Inversen $U^{-1}(L)$ und $V^{-1}(L)$ Polynommatrizen sind (und keine gebrochenen Funktionen als Einträge haben) und dass insbesondere ihre Determinante eine von Null verschiedene Konstante, also ein Polynom vom Grad Null (und kein echtes Polynom) ist
2. $N(L)$ ist eine $K \times r$ -Matrix, die nur Einträge auf der Hauptdiagonalen hat.

Die Einträge der Matrix $N(L)$ sind wie folgt festgelegt

$$N(L) = \begin{pmatrix} N_1(L) & & 0 \\ & N_2(L) & \\ 0 & & N_r(L) \end{pmatrix}$$

mit $N_i(L) = \frac{S_i(L)}{W_i(L)}$, $i = 1, \dots, r$, wobei für die Lagpolynome $S_i(L)$ und $W_i(L)$ jeweils gilt dass

- $S_i(L)$ und $W_i(L)$ für jedes i teilerfremd sind, d.h. dass der Bruch so weit wie möglich gekürzt wurde.
- Dies muss nicht für Polynome $S_i(L)$ und $W_j(L)$ mit $i \neq j$ gelten.
- $S_i(L)$ teilt $S_{i+1}(L)$ für alle $i = 1, \dots, r$
- $W_i(L)$ teilt $W_{i+1}(L)$ für alle $i = 1, \dots, r$
- Der Koeffizient für den kleinsten Grad von $S_i(L)$ und $W_i(L)$ für alle $i = 1, \dots, r$ ist gleich Eins.
- $W_1(L)$ ist das kleinste gemeinsame Vielfache der Nenner-Polynome in der Matrix $C(L)$

$N(L)$ wird Smith McMillan-Form von $C(L)$ genannt, diese Matrix ist eindeutig, was für $U(L)$ und $V(L)$ nicht gilt.

Bemerkung C.2.2 (Verfahren zur Zerlegung in Smith McMillan Form) *Yoo (1987, Appendix B, p.35) erstellte ein schrittweises Verfahren, wie man eine Matrix mit gebrochenrationalen Einträgen schrittweise zerlegen kann.*

1. Zunächst bestimmt man das kleinste gemeinsame Vielfache $g(L)$ aller Nenner-Polynome der Matrix $C(L)$ und erhält eine Polynommatrix $P(L)$, indem man alle Einträge von

$C(L)$ mit $g(L)$ multipliziert $P(L) = g(L)C(L)$.

Ist die ursprüngliche Matrix $A(L)$ bereits eine Polynommatrix, wird dieser Schritt ausgelassen und man arbeitet mit $A(L) = P(L)$ weiter.

2. Nun multipliziert man solange Elementarmatrizen von beiden Seiten, bis man eine Diagonalmatrix erhält. $U(L)P(L)V(L) = D(L)$, wobei $U(L)$ und $V(L)$ Produkte von Elementarmatrizen sind.⁴⁹
3. Nun multipliziert man auf beiden Seiten mit $g^{-1}(L)$ und kürzt eventuell gemeinsame Faktoren. So erhält man die SmithMcMillan-Zerlegung $U(L)C(L)V(L) = N(L)$. Bei einer Polynommatrix $A(L)$ entfällt auch dieser dritte Schritt.

Lemma C.2.3 (Spezialform der Smith McMillan-Form) Die Smith McMillan-Matrix $M(L)$ kann weiter zerlegt werden

$$N(L) = W^{-1}(L)M(L)S(L).$$

Dabei sind $W(L)$ und $S(L)$ Diagonalmatrizen, die aus den obigen Polynomen bestehen.

- $W(L) = \text{diag}(W_1(L), \dots, W_r(L))$
- $S(L) = \text{diag}(S_1(L), \dots, S_r(L))$
- $M(L)$ ist eine Diagonalmatrix folgender Gestalt

$$M(L) = \begin{pmatrix} \mathbb{I}_{K-r} & 0 \\ 0 & (1-L)\mathbb{I}_r \end{pmatrix} \quad (\text{C.6})$$

In unserem Fall kann man zusätzlich annehmen, dass $S_i(L)$ nur Wurzeln außerhalb des Einheitskreises besitzt (Yoo, 1987, p.28).

Lemma C.2.4 (Smith McMillan Form einer Polynommatrix) Eine Polynommatrix (und auch eine mit gebrochenrationalen Einträgen) kann also wie folgt zerlegt werden

$$\begin{aligned} A(L) &= \underbrace{U^{-1}(L)W^{-1}(L)}_{Q(L)} M(L) \underbrace{S(L)V^{-1}(L)}_{R(L)} \\ A(L) &= Q(L)M(L)R(L), \end{aligned}$$

wobei die Polynommatrizen $Q(L)$ und $R(L)$ nur Nullstellen außerhalb des Einheitskreises haben und $M(L)$ die Darstellung (C.6) hat.

⁴⁹Die Multiplikation mit Elementarmatrizen bewirkt Spalten- bzw. Zeilenumformungen, wie im Anhang A.1.3, S.312 gezeigt wird.

Beispiel C.2.5 (bivariater Prozess) *Wir betrachten folgenden bivariaten Prozess Y_t*

$$\begin{aligned} Y_{1t} &= 3Y_{2,t-1} - 3Y_{2,t-2} + \epsilon_{1t} \\ Y_{2t} &= Y_{2,t-1} + \epsilon_{2t}, \end{aligned}$$

der sich in vektorautoregressiver Form so darstellen lässt

$$\begin{pmatrix} 1 & -3L + 3L^2 \\ 0 & 1 - L \end{pmatrix} \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

Da $A(L) = \begin{pmatrix} 1 & -3L + 3L^2 \\ 0 & 1 - L \end{pmatrix}$ bereits eine Polynommatrix ist ($g(L) = 1$), benötigen wir zu ihrer Zerlegung nur den zweiten Schritt aus Bemerkung C.2.2. Um $A(L)$ auf Diagonalgestalt zu bringen, reicht es aus, das $-3L$ -fache der zweiten Zeile zur ersten Zeile zu addieren. Dies geschieht mit der Matrix

$$Q(L) = \begin{pmatrix} 1 & -3L \\ 0 & 1 \end{pmatrix}$$

Denn $-3L + 3L^2 = -3L(1 - L)$, und so

$$\begin{pmatrix} 1 & -3L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -3L + 3L^2 \\ 0 & 1 - L \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 - L \end{pmatrix}$$

Da dies schon eine Diagonalmatrix ist

$$M(L) = \begin{pmatrix} 1 & 0 \\ 0 & 1 - L \end{pmatrix}$$

bedeutet dies, dass man keine Spaltenoperationen benötigt hat. Somit ist $R(L) = \mathbb{I}_2$ eine Einheitsmatrix. Daraus ergibt sich für die MA-Matrix

$$\begin{aligned} \Psi(L) &= Q^{-1}(L)\tilde{M}(L) \\ &= \begin{pmatrix} 1 & 3L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 - L & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 - L & -3L \\ 0 & 1 \end{pmatrix} \end{aligned}$$

C.3 Kointegrationstests

C.3.1 Multivariate Normalverteilung

Folgende Definition für die multivariate Normalverteilung ist Klenke (2006, Kap 15.52) entnommen.

Definition C.3.1 (Multivariate Normalverteilung) *Sei Σ eine (strikt) positiv definite symmetrische reelle $K \times K$ Matrix und $\mu \in \mathbb{R}^K$. Ein Zufallsvektor $X = (X_1, \dots, X_K)$ heißt K -dimensional normalverteilt mit Erwartungswert μ und Kovarianzmatrix Σ , falls X die Dichte*

$$f_{\mu, \Sigma}(x) = \frac{1}{\sqrt{(2\pi)^K \det(\Sigma)}} \exp\left(-\frac{1}{2}x'\Sigma^{-1}x\right)$$

für $x \in \mathbb{R}^K$ hat. Dabei ist $x'y = \sum_{i=1}^K x_i y_i$ das Standardskalarprodukt im \mathbb{R}^K . Man schreibt $X \sim \mathcal{N}_K(\mu, \Sigma)$.

C.3.2 Verteilung von ϵ_t

Da der Fehlerprozess $(\epsilon_t)_{t=1, \dots, T}$ unabhängig, identisch multivariat normalverteilt (siehe Abschnitt C.3.1) ist, ist die gemeinsame Verteilungsdichte $f(\epsilon_1, \dots, \epsilon_T)$ das Produkt der Einzelverteilungsdichten $f(\epsilon_1) \cdot \dots \cdot f(\epsilon_T)$ (Taylor, 2005, S.29). Wegen der Funktionalgleichung der Exponentialfunktion $\exp(x_1) \cdot \dots \cdot \exp(x_n) = \sum_{i=1}^n \exp(x_i)$ (Freitag & Busam, 2006) ergibt sich für die gemeinsame Verteilung des Fehlerprozesses

$$f(\epsilon_1, \dots, \epsilon_T) = (2\pi)^{-\frac{T}{2}} (\det(\Sigma))^{-\frac{T}{2}} \cdot \exp\left(-\frac{1}{2} \sum_{t=1}^T \epsilon_t' \Sigma^{-1} \epsilon_t\right).$$

Da man den Fehlerprozess zu jedem Zeitpunkt t über die Differenzen ΔY_{t-i} und $\alpha\beta'Y_{t-1}$ darstellen kann

$$\epsilon_t = \Delta Y_t - \alpha\beta'Y_{t-1} - \Gamma_1 \Delta Y_{t-1} - \dots - \Gamma_{p-1} \Delta Y_{t-(p-1)},$$

ist die gemeinsame Verteilungsdichte nicht nur von der Kovarianzmatrix Σ , sondern auch von den Parametern α, β und Γ_i abhängig. Die Likelihood-Funktion für diese Parameter lässt sich deshalb aus der gemeinsamen Verteilungsdichte der ϵ_t gewinnen.

$$L_{\epsilon_t}(\alpha, \beta, \Sigma, \Gamma_1, \dots, \Gamma_{p-1}) = f_{\alpha, \beta, \Sigma, \Gamma_1, \dots, \Gamma_{p-1}}(\epsilon_1, \dots, \epsilon_T)$$

C.3.3 Maximum-Likelihood-Ansatz

Will man dem Ergebnis eines Zufallsexperiments eine Wahrscheinlichkeit zuordnen, braucht man eine Zufallsvariable X , die die möglichen Ausgänge des Experimentes beschreibt. Die Verteilung $P(X)$ der Zufallsvariablen ist eine Funktion, die jedem Ergebnis x eine Wahrscheinlichkeit zuordnet.

Nun gibt es in der Praxis häufig den Fall, dass zwar die Verteilungsfamilie der Zufallsvariablen, aber nicht ihre genaue Verteilungsfunktion bekannt ist. Zum Beispiel nimmt man oft an, dass eine Zufallsvariable normalverteilt ist, ohne den Erwartungswert und die Varianz zu kennen.

$$P_{\vartheta}(X) = \mathcal{N}_{\mu, \sigma^2}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{X - \mu}{2}\right)$$

Um die genaue Verteilungsfunktion der Variablen zu bestimmen, muss man die Parameter der Verteilungsfamilie (hier $\vartheta_1 = \mu$, $\vartheta_2 = \sigma$) schätzen.

Im Maximum-Likelihood-Ansatz wählt man diejenigen Parameter ϑ_i , für die das Ergebnis x des Zufallsexperiments (Wert der Zufallsvariablen) am plausibelsten ist. Also die Parameterwerte, für die die Wahrscheinlichkeit $P_{\vartheta}(x)$ maximal wird.

Eine *Likelihood-Funktion* $L_X(\vartheta)$ ordnet also jeder Parameterkonstellation ϑ die Wahrscheinlichkeit zu, die der Ausgang x einer nach P_{ϑ} verteilten Zufallsvariablen hat.

$$L_X(\vartheta) = P_{\vartheta}(X)$$

Die Darstellung der Likelihoodfunktion entspricht also der der bekannten Verteilungsfunktion. Nur ist die Betrachtungsweise eine andere. Während in gewohnter Weise die Zuordnung $x \rightarrow P_{\vartheta}(X)$ bei bekannten Parametern ϑ geschieht, betrachtet man nun die Zuordnung $\vartheta \rightarrow L_X(\vartheta)$ bei bekanntem Ausgang $X = x$.

Als Maximum-Likelihood-Schätzerwert dient dann jene Parameterkonstellation, unter der $L_x(\vartheta)$ am größten ist.

C.3.4 Zwischenrechnung

Zum Beweis der Gleichung (2.34) von S.132 geben wir hier folgende Nebenrechnung von Neusser (2006, S.225) wieder.

$$\begin{aligned}
& \frac{1}{2} \sum_{t=1}^T (\Delta Y_t - \hat{\alpha} \beta' Y_{t-1})' \hat{\Sigma}^{-1} (\Delta Y_t - \hat{\alpha} \beta' Y_{t-1}) \\
&= \frac{1}{2} \text{Spur} \left(\sum_{t=1}^T (\Delta Y_t - \hat{\alpha} \beta' Y_{t-1}) (\Delta Y_t - \hat{\alpha} \beta' Y_{t-1})' \hat{\Sigma}^{-1} \right) \\
&= \frac{1}{2} \text{Spur} \left((T S_{00} - T \hat{\alpha} \beta' S_{01} \beta \hat{\alpha}' + T \hat{\alpha} S_{11} \beta \hat{\alpha}') \hat{\Sigma}^{-1} \right) \\
&= \frac{T}{2} \text{Spur} \left((S_{00} - \hat{\alpha} \beta' S_{10}) \hat{\Sigma}^{-1} \right) \\
&= \frac{TK}{2}
\end{aligned}$$

C.3.5 Beweis der Bemerkung „Minimum der Determinante $\hat{\Sigma}(\beta)$ “

Hier wird der Beweis der Bemerkung 2.7.4 von Seite 132 ausgeführt. Zur Erinnerung wiederholen wir noch einmal die Bemerkung.

Bemerkung C.3.2 *Das Minimum der Determinante ist gegeben durch*

$$\text{argmin det} \left(\hat{\Sigma}(\beta) \right) = (\det(S_{00})) \prod_{i=1}^K (1 - \lambda_i)$$

dabei sind λ_i die kanonischen Korrelationen zwischen $R_{0,t}$ und $R_{1,t}$.

Der folgende Beweis richtet sich nach dem Originalartikel (Johansen, 1988), verwendet jedoch erleichternde Zwischenschritte aus Neusser (2006) und wir ergänzen eigene Erläuterungen.

Beweis. Es ist zu zeigen, dass die Determinante $\det(\Sigma)$ tatsächlich durch das Produkt der λ_i minimiert wird.

Die Determinante von Σ ist bei einem gegebenem Prozess nur von der $(K \times r)$ -Matrix β abhängig (vgl. Gleichung (2.33), S.131)

$$\det(\Sigma) = |S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}|.$$

Wir betrachten zunächst die Determinante einer Hilfsmatrix S_β , die in vier Submatrizen geteilt ist. Ihre Determinante berechnet sich deshalb analog zu der Determinante einer 2×2 -

Matrix (siehe A.1.23, S.311)

$$\det(S_\beta) = \begin{vmatrix} S_{00} & S_{01}\beta \\ \beta'S_{10} & \beta'S_{11}\beta \end{vmatrix} = |S_{00}\beta'S_{11}\beta - \beta'S_{10}S_{01}\beta|.$$

Diese Determinante lässt sich durch „Ausklammern“ verschiedener Ausdrücke auf folgende Arten darstellen

$$\begin{aligned} \det(S_\beta) &= |S_{00}| |\beta'S_{11}\beta - \beta'S_{10}S_{00}^{-1}S_{01}\beta| \\ &= |\beta'S_{11}\beta| \underbrace{|S_{00} - \beta'S_{10}(\beta'S_{11}\beta)^{-1}S_{01}\beta|}_{\det(\hat{\Sigma}(\beta))}. \end{aligned}$$

Durch Umstellen dieser Gleichung erhält man

$$\frac{\det(\hat{\Sigma}(\beta))}{|S_{00}|} = \frac{|S_{00} - \beta'S_{10}(\beta'S_{11}\beta)^{-1}S_{01}\beta|}{|S_{00}|} = \frac{|\beta'S_{11}\beta - \beta'S_{10}S_{00}^{-1}S_{01}\beta|}{|\beta'S_{11}\beta|}. \quad (\text{C.7})$$

Die linke Seite der Gleichung und damit die Determinante von Σ ist minimal, wenn die rechte Seite minimiert wird.

Das Minimum dieser Funktion wird durch das folgende verallgemeinerte Eigenwertproblem gelöst

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0. \quad (\text{C.8})$$

Dieses Eigenwertproblem liefert K Eigenwerte

$$1 \geq \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K \geq 0$$

mit K dazu gehörigen Eigenvektoren e_1, \dots, e_K , die durch $E'S_{11}E = \mathbb{I}_K$ normiert sind.

Bekanntermaßen sind die Lösungen eines Eigenwertproblem $Ax = \lambda x = x\lambda$ die Nullstellen des charakteristischen Polynoms $\chi_A(\lambda) = |\lambda\mathbb{I} - A| = 0$. Ebenso erfüllen die Lösungen der Gleichung (C.8) das zugehörige verallgemeinerte Eigenwertproblem in Bezug auf S_{11}

$$S_{11}ED = S_{10}S_{00}^{-1}S_{01},$$

wobei E wie erwähnt normiert durch $E'S_{11}E = \mathbb{I}_K$ ist.

Man kann nun zeigen, dass der Ausdruck in (C.7) minimiert wird, wenn man für β die ersten r Eigenvektoren, also die ersten Spaltenvektoren der Matrix E wählt. Johansen selbst verweist hier auf weitere Literatur (Anderson, 1984).

Nun bleibt noch zu zeigen, dass sich der Ausdruck in (C.7) dann als Produkt der $(1 - \lambda_i)$

schreiben lässt. Dazu wählt man für $\beta = E\xi$, wobei ξ eine $K \times r$ -Matrix ist. Damit gilt

$$\beta' S_{kk} \beta = (E\xi)' S_{11} (E\xi) = \xi' \underbrace{E' S_{11} E}_{\mathbb{I}} \xi = \xi' \xi$$

und

$$\begin{aligned} \beta' S_{10} S_{00}^{-1} S_{01} \beta &= (E\xi)' S_{10} S_{00}^{-1} S_{01} (E\xi) \\ &= \xi' E' \underbrace{S_{10} S_{00}^{-1} S_{01} E}_{S_{11} E D} \xi \\ &= \xi' \underbrace{E' S_{11} E}_{\mathbb{I}} D \xi \\ &= \xi' D \xi \end{aligned}$$

Dadurch wird (C.7) zu

$$\frac{\det(\hat{\Sigma}(\beta))}{|S_{00}|} = \frac{|\beta' S_{11} \beta - \beta' S_{10} S_{00}^{-1} S_{01} \beta|}{|\beta' S_{11} \beta|} = \frac{|\xi' \xi - \xi' D' \xi|}{|\xi' \xi|}$$

Die Wahl von β als die ersten r Spaltenvektoren von E ist gleichbedeutend mit der Wahl von ξ als die ersten r Einheitsvektoren. Damit berechnet sich

$$\frac{|\xi' \xi - \xi' D' \xi|}{|\xi' \xi|} = \det(S_{00}) \prod_{i=1}^r (1 - \lambda_i) \quad (\text{C.9})$$

Dies ist die Darstellung des Minimums der Determinante, die in der Bemerkung gefordert war. Sie ist damit bewiesen. \square

Um die Rechnung zu illustrieren, betrachten wir folgendes Beispiel.

Beispiel C.3.3 ($K = 3, r = 2$) Für einem dreidimensionalen Prozess mit zwei Kointegrationsbeziehungen ist $\xi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$ und $D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$.

Damit kann man

$$\xi' \xi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow |\xi' \xi| = 1$$

und

$$\xi'D\xi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

berechnet, woraus direkt folgt:

$$\begin{aligned} \Rightarrow \xi'\xi - \xi'D\xi &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ \Rightarrow |\xi'\xi - \xi'D\xi| &= \begin{vmatrix} (1 - \lambda_1) & 0 \\ 0 & (1 - \lambda_2) \end{vmatrix} = (1 - \lambda_1)(1 - \lambda_2) \end{aligned}$$

Damit folgt also in diesem Beispiel für die Determinante folgende Darstellung

$$\det(\hat{\Sigma}(\beta)) = |S_{00}| \frac{|\xi'\xi - \xi'D\xi|}{|\xi'\xi|} = |S_{00}| \frac{(1 - \lambda_1)(1 - \lambda_2)}{1} = |S_{00}| (1 - \lambda_1)(1 - \lambda_2)$$

Die Determinante der 1×1 -Matrix ist gleich dem Betrag des einzigen Eintrages, das Produkt besteht aus den beiden Faktoren $(1 - \lambda_1)$ und $(1 - \lambda_2)$.

D Simulationsstudien

D.1 Die Länge einer Zeitreihe

D.1.1 Auswertung Ausreißer in Gruppe (2,3) der kointegrierten Systeme

```
#####  
# Johansen-Procedure #  
#####
```

Test type: maximal eigenvalue statistic (lambda max) , with linear trend

Eigenvalues (lambda):
[1] 0.2676743 0.2204788

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
r <= 1		16.94	6.50	8.18 11.65
r = 0		21.18	12.91	14.90 19.19

Eigenvectors, normalised to first column:
(These are the cointegration relations)

	X.12	Y.12
X.12	1.000	1.0000000
Y.12	619.744	-0.6222205

Weights W:
(This is the loading matrix)

	X.12	Y.12
X.d	-0.0004982508	-0.4993696
Y.d	-0.0008430563	0.2914538

Tabelle D.1: Boxplot-Statistiken für die Schätzung von β bei stationären Systeme

	Min	W1	Box1	Median	Box2	W2	Max	Out
T	Gruppe (1,1)							
10	-551,508	-4,149	-1,101	0,000	0,947	3,939	1179,790	155
20	-500,584	-3,895	-1,015	-0,038	0,927	3,783	8382,465	163
30	-1502,535	-3,859	-0,990	-0,037	0,959	3,784	289,583	164
40	-6240,623	-3,685	-0,818	0,092	1,098	3,896	561,226	157
50	-4708,296	-3,714	-0,827	0,054	1,107	3,993	805,503	161
60	-690,812	-3,653	-0,872	0,059	1,051	3,931	407,535	175
70	-2266,890	-3,760	-0,867	0,045	1,099	4,043	263,869	153
80	-1511,529	-3,934	-0,984	0,029	0,998	3,903	577,140	137
90	-892,745	-3,719	-0,902	0,041	1,003	3,819	826,908	147
100	-648,911	-3,793	-0,920	-0,003	1,000	3,869	271,285	156
150	-146,335	-3,939	-0,973	0,002	1,024	4,018	80,100	155
200	-386,131	-4,310	-1,144	-0,047	0,996	4,188	206,764	153
300	-359,770	-4,257	-1,102	-0,042	1,024	4,186	434,192	162
500	-13730,836	-3,876	-0,928	-0,019	1,087	4,090	466,628	144
	Gruppe (2,3)							
10	-295,674	-2,712	-0,706	-0,031	0,633	2,628	162,196	162
20	-256,811	-2,656	-0,683	-0,047	0,692	2,749	308,455	147
30	-119,872	-2,784	-0,722	-0,035	0,689	2,789	149,625	176
40	-130,097	-2,956	-0,715	0,016	0,786	2,958	2879,767	171
50	-141,822	-2,760	-0,632	0,075	0,793	2,922	1717,417	170
60	-10629,232	-2,918	-0,686	0,059	0,814	3,048	145,385	165
70	-120,481	-2,660	-0,635	0,040	0,719	2,742	10312,482	162
80	-210,088	-2,709	-0,688	-0,009	0,661	2,668	6855,670	151
90	-692,273	-2,738	-0,682	0,011	0,720	2,790	196,918	156
100	-194,540	-2,755	-0,654	0,015	0,805	2,988	7098,324	153
150	-201,536	-2,809	-0,726	-0,029	0,668	2,753	614,874	163
200	-215,998	-2,916	-0,747	-0,040	0,716	2,872	355,871	155
300	-281,863	-2,611	-0,645	-0,017	0,682	2,650	307,776	167
500	-3023,330	-2,533	-0,620	-0,025	0,673	2,459	256,691	162
	Gruppe (3,10)							
10	-683,388	-1,272	-0,306	0,020	0,359	1,329	3672,130	145
20	-81,372	-1,281	-0,311	-0,010	0,341	1,318	59,037	142
30	-31,428	-1,343	-0,346	-0,009	0,322	1,320	68,813	152
40	-275,190	-1,302	-0,325	0,003	0,335	1,302	88,630	173
50	-138,124	-1,260	-0,284	0,019	0,371	1,322	322,515	155
60	-78,570	-1,235	-0,296	0,010	0,341	1,293	32,775	172
70	-105,331	-1,208	-0,278	0,022	0,343	1,271	2963,534	166
80	-101,726	-1,258	-0,344	-0,017	0,280	1,161	267,015	170
90	-52,922	-1,138	-0,299	-0,008	0,290	1,142	5009,286	170
100	-202,255	-1,259	-0,316	0,001	0,315	1,210	1834,087	163
150	-158,570	-1,249	-0,306	0,016	0,324	1,256	267,404	166
200	-293,668	-1,250	-0,302	0,026	0,343	1,270	731,182	147
300	-204,154	-1,300	-0,319	0,023	0,340	1,306	42,361	150
500	-127,309	-1,228	-0,293	0,015	0,335	1,231	143,822	158

Tabelle D.2: Boxplot-Statistiken für die Schätzung von β bei integrierten Systeme

	Min	W1	Box1	Median	Box2	W2	Max	Out
T	Gruppe (1,1)							
10	-1106,042	-4,381	-1,180	-0,033	0,984	4,173	1899,801	148
20	-819,994	-3,646	-0,928	0,029	0,935	3,680	419,788	165
30	-58,533	-3,349	-0,809	0,044	1,027	3,683	626,024	144
40	-299,763	-3,654	-0,836	0,079	1,058	3,806	239,158	154
50	-1459,994	-3,510	-0,899	0,003	0,926	3,559	4517,169	156
60	-192,930	-3,481	-0,770	0,083	1,055	3,783	1444,446	165
70	-244,212	-3,641	-0,820	0,082	1,101	3,949	190,255	158
80	-787,580	-3,588	-0,802	0,097	1,067	3,773	422,041	163
90	-283,344	-3,397	-0,746	0,111	1,140	3,966	1400,993	155
100	-1337,176	-3,489	-0,790	0,118	1,019	3,656	303,489	156
150	-402,433	-3,411	-0,779	0,103	1,046	3,779	199,603	142
200	-269,252	-3,468	-0,821	0,110	0,951	3,587	364,700	151
300	-415,989	-3,804	-0,910	0,063	1,020	3,872	1766,194	136
500	-5558,524	-3,573	-0,876	0,086	0,927	3,606	12921,672	158
	Gruppe (2,3)							
10	-151,993	-2,881	-0,692	0,013	0,770	2,929	268,083	169
20	-65,292	-2,466	-0,645	0,003	0,603	2,449	61,408	159
30	-559,303	-2,448	-0,591	0,016	0,647	2,493	208,098	141
40	-403,774	-2,381	-0,593	0,004	0,613	2,380	422,434	154
50	-81,168	-2,365	-0,597	0,008	0,601	2,371	43,887	152
60	-4524,877	-2,314	-0,507	0,036	0,703	2,492	472,191	156
70	-194,633	-2,483	-0,565	0,032	0,718	2,638	76,104	164
80	-254,299	-2,373	-0,543	0,083	0,733	2,629	206,054	155
90	-375,348	-2,250	-0,504	0,057	0,699	2,482	56,437	154
100	-636,554	-2,287	-0,510	0,074	0,704	2,518	424,528	145
150	-170,140	-2,194	-0,493	0,100	0,695	2,442	225,121	142
200	-286,270	-2,230	-0,515	0,076	0,635	2,343	45,738	151
300	-1214,868	-2,473	-0,611	0,018	0,631	2,423	308,108	149
500	-357,802	-2,330	-0,574	0,038	0,601	2,319	60,084	151
	Gruppe (3,10)							
10	-158,273	-1,270	-0,323	0,009	0,316	1,266	4369,836	169
20	-48,515	-1,131	-0,289	0,000	0,274	1,099	189,738	161
30	-39,737	-1,048	-0,255	0,008	0,291	1,098	178,311	136
40	-216,187	-1,106	-0,275	0,004	0,280	1,107	51,880	146
50	-232,103	-1,084	-0,287	-0,008	0,267	1,094	260,155	163
60	-346,964	-1,069	-0,242	0,028	0,312	1,128	90,413	151
70	-421,234	-1,088	-0,255	0,014	0,315	1,170	440,943	160
80	-34,704	-1,050	-0,250	0,036	0,323	1,159	1953,316	150
90	-333,634	-1,074	-0,245	0,018	0,310	1,094	1481,335	150
100	-56,526	-1,034	-0,215	0,042	0,334	1,155	61,612	144
150	-1078,504	-0,990	-0,216	0,040	0,312	1,099	54,340	142
200	-180,552	-1,035	-0,238	0,034	0,296	1,089	38,085	149
300	-27,144	-1,088	-0,270	0,013	0,289	1,122	60,925	156
500	-267,722	-1,021	-0,256	0,028	0,282	1,086	349,668	151

D.2 Messung auf Intervallskalenniveau

D.2.1 Vorstudie

Test bei bekannter Kointegrationsbeziehung Der Phillips-Perron-Test zeigt nur bei $T = 30$ leichte Abweichungen (maximal 7 stationäre Linearkombinationen werden nicht erkannt) von den Ausgangswerten ohne Transformation der Daten. Bereits ab $T = 60$ wird bei allen 1000 Systeme auf allen Niveaus die stationäre Linearkombination erkannt. Der Augmented Dickey-Fuller-Test erkennt die stationäre Linearkombination bei $T = 300$ in allen Fällen. Die größte Abweichung zu den Ausgangsdaten gibt es in der Untergruppe mit $(a,b) = (2,3)$. Hier werden bei $T = 30$ 34 Linearkombinationen mehr auf dem 0,1 Niveau als stationär erkannt. Der Phillips-Perron-Test ist ab $T = 60$ auch bei den diskretisierten Werten perfekt. Minimale Abweichungen (maximal ± 3) gibt es bei $T = 30$. Der ADF-Test klassifiziert für $T = 300$ in allen Untergruppen perfekt. Für die übrigen Längen gibt es nie mehr als 17 (0.05 Niveau, $T = 30$) Abweichungen. Hier werden stationäre Systeme fälschlicherweise als kointegriert klassifiziert. Meist beträgt die Differenz jedoch nur weniger als 10 Systeme. Integrierte Systeme werden von den Unitroot-Tests dann richtig erkannt, wenn die vorgegebene Linearkombination als integriert erkannt wird. Werte unter 50 sind also gut. Der Phillips-Perron-Test ist nach der Diskretisierung bis 40 integrierte Linearkombinationen mehr. Dennoch ist seine Klassifikationsrate insbesondere für (1,1) ungenügend und nur für das 0.01-Niveau bei den anderen Untergruppen wirklich gut. Der ADF-Test zeigt kaum Abweichungen zu seinem Verhalten vor Diskretisierung. Die größte Abweichung beträgt 7.

Tabelle D.3: Ergebnisse des PP-Tests bei diskretisierten Daten: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	998	990	923	999	995	904	529	407	187
	60	1000	1000	1000	1000	1000	1000	586	488	301
	100	1000	1000	1000	1000	1000	1000	573	474	301
	300	1000	1000	1000	1000	1000	1000	410	314	196
(2,3)	30	998	987	908	998	996	909	201	127	29
	60	1000	1000	1000	1000	1000	1000	208	119	38
	100	1000	1000	1000	1000	1000	1000	215	137	45
	300	1000	1000	1000	1000	1000	1000	131	71	22
(3,10)	30	999	993	920	1000	997	923	144	74	19
	60	1000	1000	1000	1000	1000	1000	150	73	19
	100	1000	1000	1000	1000	1000	1000	146	88	17
	300	1000	1000	1000	1000	1000	1000	96	48	16

Tabelle D.4: Ergebnisse des ADF-Tests bei diskretisierten Daten: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	314	180	36	343	185	54	87	31	5
	60	891	749	399	885	742	416	76	37	6
	100	983	953	758	988	959	777	82	38	5
	300	1000	1000	1000	1000	1000	1000	75	34	4
(2,3)	30	339	197	50	325	185	48	80	28	9
	60	868	748	415	885	747	388	85	35	6
	100	988	946	767	986	960	768	85	51	8
	300	1000	1000	1000	1000	1000	1000	73	31	3
(3,10)	30	334	191	46	346	191	47	85	34	5
	60	872	740	374	863	722	394	87	36	1
	100	987	951	739	986	949	771	82	42	9
	300	1000	1000	1000	1000	1000	1000	73	35	6

Klassifikation durch Johantest Beim Johantest werden nur bei $T=30$ mehr als 10 Systeme anders eingestuft als in der Ausgangsdatenlage. Die Klassifizierung bei den anderen Längen beträgt auf allen Niveaus maximal drei Systeme Unterschied. Dabei werden zumeist kointegrierte Systeme als stationär eingestuft.

Die Klassifikation des Johantest der stationären Systeme ändert sich durch die Diskretisierung für die Längen ab $T=100$ überhaupt nicht. Bei $T=60$ werden maximal 3 Systeme anders eingestuft. Bei $T=30$ gibt es Abweichungen, die bis zu 23 betragen.

Für integrierte Systeme gibt es kaum Unterschiede in der Klassifikation durch den Johantest. In der (1,1)-Untergruppe werden bis zu 18 Systeme anders klassifiziert. In den anderen maximal 6. Meist wird fälschlicher Weise eine Kointegrationsbeziehung in den integrierten Systemen vermutet. Nur ganz wenige Systeme werden sogar als stationär eingestuft.

Ergebnis Durch die Diskretisierung der Daten wird nur in Einzelfällen die Klassifikation verändert. Die Information über die Eigenschaft kointegriert (oder nicht) bleibt in den Reihen enthalten. Weniger als 1% der Systeme wird anders klassifiziert, nachdem die Daten diskretisiert wurden.

Tabelle D.5: Klassifikation jeweils 1000 Systeme bei diskretisierten Daten

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
Kointegrierte Systeme											
(1,1)	30	267	606	127	435	502	63	756	233	11	1000
	60	3	863	134	7	921	72	67	915	18	1000
	100	0	860	140	0	922	78	0	979	21	1000
	300	0	867	133	0	929	71	0	992	8	1000
(2,3)	30	258	624	118	396	544	60	718	274	8	1000
	60	0	898	102	8	940	52	69	915	16	1000
	100	0	899	101	0	943	57	0	985	15	1000
	300	0	913	87	0	959	41	0	995	5	1000
(3,10)	30	243	644	113	412	536	52	727	264	9	1000
	60	2	892	106	8	941	51	70	915	15	1000
	100	0	902	98	0	947	53	0	982	18	1000
	300	0	914	86	0	956	44	0	994	6	1000
Stationäre Systeme											
(1,1)	30	86	40	874	207	71	722	550	150	300	1000
	60	0	0	1000	0	0	1000	2	6	992	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	85	28	887	213	76	711	548	160	292	1000
	60	0	0	1000	0	1	999	6	10	984	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	85	31	884	210	110	680	540	176	284	1000
	60	0	0	1000	0	1	999	7	8	985	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
Integrierte Systeme											
(1,1)	30	802	153	45	886	101	13	972	27	1	1000
	60	787	181	32	873	119	8	967	32	1	1000
	100	766	187	47	861	128	11	956	44	0	1000
	300	798	164	38	873	117	10	964	36	0	1000
(2,3)	30	842	131	27	902	92	6	978	21	1	1000
	60	870	111	19	926	68	6	985	15	0	1000
	100	861	117	22	928	68	4	983	17	0	1000
	300	902	85	13	954	44	2	991	9	0	1000
(3,10)	30	841	131	28	902	90	8	972	27	1	1000
	60	873	108	19	932	64	4	986	12	2	1000
	100	870	109	21	938	58	4	986	14	0	1000
	300	906	81	13	954	43	3	990	10	0	1000

Tabelle D.6: Boxplot-Statistiken für die Schätzung von β bei diskretisierten Daten

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-92,9481	-1,5038	-1,1375	-1,0073	-0,8929	-0,5405	11,3194	86
	60	-1,6799	-1,1969	-1,0522	-1,0056	-0,9556	-0,8114	-0,3602	59
	100	-1,3926	-1,1142	-1,0322	-1,0022	-0,9751	-0,8895	-0,6198	38
	300	-1,0715	-1,0359	-1,0089	-0,9996	-0,9901	-0,9636	-0,9322	30
(2,3)	30	-26,0491	-0,8033	-0,7008	-0,6663	-0,6324	-0,5304	0,7034	76
	60	-0,8486	-0,7202	-0,6801	-0,6661	-0,6508	-0,6077	0,4372	52
	100	-0,8041	-0,6995	-0,6746	-0,6663	-0,6577	-0,6324	-0,5999	43
	300	-0,6986	-0,6778	-0,6696	-0,6670	-0,6640	-0,6558	-0,6507	25
(3,10)	30	-1,0885	-0,3362	-0,3094	-0,3006	-0,2915	-0,2653	-0,1223	80
	60	-0,3373	-0,3155	-0,3035	-0,2997	-0,2954	-0,2832	-0,2409	36
	100	-0,3152	-0,3081	-0,3022	-0,3001	-0,2980	-0,2919	-0,2865	52
	300	-0,3050	-0,3025	-0,3006	-0,2999	-0,2993	-0,2973	-0,2939	46
Stationäre Systeme									
(1,1)	30	-203,1048	-3,8567	-0,9816	-0,0055	0,9419	3,8214	617,1582	150
	60	-279,1702	-3,7329	-0,8936	0,0341	1,0057	3,7522	75,0713	162
	100	-457,9025	-3,7673	-0,9738	-0,0183	0,9368	3,7383	360,8014	160
	300	-271,1225	-4,0293	-1,0637	-0,0751	0,9866	3,9845	260,8211	163
(2,3)	30	-683,9750	-2,8838	-0,7438	-0,0259	0,6978	2,8450	950,8403	173
	60	-303,9686	-2,7709	-0,6756	0,0544	0,7920	2,9631	377,6084	179
	100	-51296,7559	-2,8339	-0,7041	0,0211	0,7319	2,8434	1515,2769	150
	300	-452,2735	-2,5122	-0,5950	0,0176	0,7550	2,7059	735,8588	161
(3,10)	30	-96,1476	-1,3726	-0,3612	-0,0099	0,3154	1,2999	693,1896	151
	60	-177,2041	-1,2436	-0,3098	0,0069	0,3334	1,2932	144,2224	174
	100	-21,0446	-1,2366	-0,3125	0,0002	0,3098	1,2163	2855,3087	157
	300	-80,4640	-1,2661	-0,3182	0,0107	0,3223	1,2264	469,2287	157
Integrierte Systeme									
(1,1)	30	-726,1654	-3,4066	-0,8001	0,0310	0,9736	3,5770	177,4908	159
	60	-1634,6510	-3,3780	-0,7703	0,0856	1,0580	3,7405	130,6077	168
	100	-485,7161	-3,4054	-0,7650	0,1153	1,0356	3,7085	72,4833	158
	300	-239,0673	-3,5977	-0,9225	0,0629	1,0151	3,8957	644,6967	136
(2,3)	30	-16055,9621	-2,3963	-0,5654	0,0207	0,6658	2,4988	720,7988	141
	60	-69,7663	-2,3119	-0,5190	0,0367	0,6823	2,4708	594,3402	156
	100	-464,8941	-2,1758	-0,5005	0,0832	0,6827	2,4566	164,7943	146
	300	-1653,0607	-2,3997	-0,6175	0,0211	0,6286	2,4824	177,4566	149
(3,10)	30	-105,7454	-1,0195	-0,2555	0,0095	0,2938	1,1031	35,4771	135
	60	-39,9957	-1,0569	-0,2347	0,0244	0,3148	1,1207	529,8852	150
	100	-435,1303	-1,0346	-0,2152	0,0398	0,3339	1,1485	894,6390	143
	300	-675,8397	-1,0960	-0,2696	0,0134	0,2861	1,0999	418,1353	157

D.3 Boxplot-Statistiken für die Schätzung von β

D.3.1 Intervallskalen

Tabelle D.7: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf 10-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-920,734	-3,676	-0,899	-0,004	0,952	3,704	1167,465	167
	60	-831,125	-3,615	-0,867	0,026	0,988	3,701	658,916	183
	100	-389,605	-3,853	-0,975	-0,018	0,956	3,777	292,198	165
	300	-945,262	-4,223	-1,063	-0,063	1,080	4,290	868,527	158
(2,3)	30	-57,822	-2,932	-0,730	-0,016	0,759	2,991	288,421	164
	60	-621,294	-2,919	-0,630	0,105	0,908	3,183	103,449	160
	100	-346,451	-2,799	-0,685	-0,011	0,727	2,839	1002,861	162
	300	-158,231	-2,699	-0,638	0,004	0,766	2,837	2929,278	154
(3,10)	30	-82,482	-1,433	-0,357	-0,004	0,362	1,396	90,168	162
	60	-2747,415	-1,362	-0,363	-0,016	0,353	1,413	88,558	186
	100	-1101,381	-1,415	-0,353	0,002	0,373	1,443	84,501	140
	300	-71,347	-1,461	-0,355	0,022	0,396	1,445	875,945	142
Integrierte Systeme									
(1,1)	30	-2447,946	-3,378	-0,804	0,058	0,972	3,603	844,590	155
	60	-110,393	-3,579	-0,815	0,078	1,051	3,816	226,715	158
	100	-853,854	-3,562	-0,795	0,132	1,071	3,866	323,205	168
	300	-2017,781	-3,917	-0,935	0,077	1,058	4,040	515,147	141
(2,3)	30	-478,873	-2,376	-0,595	0,008	0,617	2,423	1779,465	139
	60	-72,757	-2,090	-0,466	0,029	0,626	2,258	63,046	174
	100	-193,371	-2,062	-0,477	0,084	0,642	2,264	999,420	150
	300	-260,512	-2,013	-0,471	0,045	0,577	2,099	1319,518	138
(3,10)	30	-85,009	-1,027	-0,252	0,018	0,276	1,067	56,269	132
	60	-38,397	-0,882	-0,205	0,015	0,256	0,941	184,159	148
	100	-37,680	-0,792	-0,182	0,029	0,245	0,871	462,898	136
	300	-184,332	-0,715	-0,174	0,018	0,193	0,736	66,463	105

Tabelle D.8: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf 7-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-3305,616	-3,798	-1,000	-0,016	0,905	3,736	84,880	161
	60	-689,840	-4,123	-0,966	0,064	1,140	4,276	885,076	151
	100	-1713,226	-3,930	-0,950	0,034	1,060	4,071	196,699	165
	300	-565,836	-4,221	-1,091	-0,032	1,072	4,309	76,043	161
(2,3)	30	-233,542	-3,062	-0,767	0,014	0,818	3,193	247,830	160
	60	-840,579	-2,945	-0,706	0,047	0,795	3,034	665,505	177
	100	-124,587	-2,883	-0,653	0,060	0,853	3,034	1636,653	146
	300	-326,236	-2,846	-0,764	-0,051	0,681	2,843	518,579	150
(3,10)	30	-1058,025	-1,581	-0,412	-0,014	0,372	1,542	292,873	154
	60	-193,444	-1,512	-0,389	0,004	0,374	1,488	123,061	166
	100	-554,215	-1,513	-0,398	-0,015	0,355	1,457	396,589	160
	300	-322,737	-1,409	-0,354	0,030	0,373	1,437	210,301	162
Integrierte Systeme									
(1,1)	30	-1760,322	-3,561	-0,850	0,021	0,963	3,658	601,735	150
	60	-519,013	-3,554	-0,779	0,086	1,071	3,806	206,259	158
	100	-65,426	-3,513	-0,762	0,137	1,077	3,824	1916,732	154
	300	-156,709	-3,840	-0,909	0,090	1,048	3,942	621,917	144
(2,3)	30	-175,381	-2,159	-0,536	0,010	0,573	2,229	159,945	148
	60	-85,620	-2,139	-0,484	0,019	0,621	2,229	463,483	163
	100	-76,789	-1,977	-0,419	0,106	0,646	2,234	5236,513	154
	300	-74566,405	-2,016	-0,484	0,049	0,543	2,056	144,720	131
(3,10)	30	-42,267	-1,053	-0,254	0,016	0,297	1,090	429,936	110
	60	-97,178	-0,959	-0,233	0,000	0,252	0,977	93,681	137
	100	-62,342	-0,871	-0,207	0,021	0,239	0,902	220,453	109
	300	-94,519	-0,808	-0,195	0,008	0,215	0,825	47,370	90

Tabelle D.9: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf 5-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-1211,149	-3,719	-0,968	-0,006	0,911	3,671	1242,353	171
	60	-588,636	-3,727	-0,927	0,062	1,081	4,028	1855,584	169
	100	-779,062	-3,874	-0,871	0,109	1,148	4,098	1772,709	158
	300	-5803,267	-4,652	-1,211	-0,050	1,085	4,508	4192,015	161
(2,3)	30	-246,080	-3,170	-0,766	0,010	0,839	3,127	226,928	153
	60	-81,790	-3,179	-0,735	0,049	0,896	3,338	394,919	153
	100	-524,369	-3,067	-0,766	-0,008	0,779	2,977	988,761	163
	300	-134,096	-2,873	-0,751	-0,031	0,737	2,969	132,886	156
(3,10)	30	-54,725	-1,899	-0,493	-0,058	0,454	1,863	27,544	115
	60	-164,930	-1,729	-0,502	-0,030	0,361	1,651	156,539	171
	100	-265,549	-1,597	-0,411	-0,014	0,383	1,567	68,150	156
	300	-135,903	-1,460	-0,361	0,004	0,395	1,520	692,233	147
Integrierte Systeme									
(1,1)	30	-335,999	-3,441	-0,802	0,048	0,968	3,593	126,836	137
	60	-628,783	-3,515	-0,830	0,043	0,965	3,608	1000,767	159
	100	-381,311	-3,548	-0,815	0,128	1,013	3,638	765,055	152
	300	-4908,870	-3,715	-0,907	0,058	1,007	3,836	305,082	144
(2,3)	30	-107,366	-2,275	-0,564	0,008	0,591	2,316	375,663	144
	60	-113,151	-2,091	-0,484	0,030	0,618	2,252	206,028	157
	100	-137,140	-2,043	-0,470	0,067	0,585	2,160	307,597	142
	300	-249,721	-2,011	-0,524	0,041	0,525	2,033	71,973	144
(3,10)	30	-113,669	-1,428	-0,363	-0,026	0,348	1,378	120,902	76
	60	-26,002	-1,035	-0,267	0,018	0,304	1,141	16,451	82
	100	-10,114	-1,068	-0,251	0,020	0,298	1,120	38,362	94
	300	-74,299	-0,931	-0,223	0,015	0,259	0,980	272,821	89

Tabelle D.10: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf 3-stufiger Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-5108,560	-4,045	-0,990	0,030	1,048	4,060	361,214	165
	60	-348,739	-4,071	-1,015	0,040	1,044	4,085	1666,172	152
	100	-471,491	-3,879	-0,974	-0,016	0,970	3,854	1300,877	150
	300	-411,828	-3,960	-1,000	-0,053	0,993	3,943	285,049	160
(2,3)	30	-748,615	-3,442	-0,815	-0,013	0,939	3,543	1190,646	152
	60	-58399,454	-3,308	-0,794	0,064	0,908	3,456	1044,486	161
	100	-413,272	-2,863	-0,664	0,064	0,894	3,230	198,735	133
	300	-198,377	-2,883	-0,760	0,003	0,685	2,799	465,088	153
(3,10)	30	-441,049	-2,480	-0,645	-0,020	0,621	2,498	20,492	35
	60	-30,048	-2,205	-0,587	-0,003	0,553	2,216	25,281	54
	100	-104,888	-1,675	-0,412	0,060	0,442	1,692	376,657	47
	300	-70,651	-1,342	-0,365	-0,010	0,288	1,263	20,573	45
Integrierte Systeme									
(1,1)	30	-1851,018	-3,459	-0,781	0,013	1,011	3,680	699,093	118
	60	-1851,018	-3,459	-0,781	0,013	1,011	3,680	699,093	118
	100	-117,535	-3,603	-0,838	0,063	1,033	3,741	108,106	133
	300	-180,786	-3,930	-1,009	0,091	1,036	3,976	998,888	139
(2,3)	30	-249,721	-2,011	-0,524	0,041	0,525	2,033	71,973	144
	60	-21,625	-2,662	-0,608	0,056	0,767	2,819	62,973	113
	100	-141,433	-2,537	-0,589	0,062	0,712	2,623	230,811	115
	300	-263,821	-2,577	-0,633	0,018	0,678	2,616	268,441	117
(3,10)	30	-24,963	-1,937	-0,542	-0,007	0,529	2,134	37,757	30
	60	-658,696	-2,106	-0,567	-0,039	0,477	2,040	146,976	46
	100	-164,715	-1,957	-0,562	-0,049	0,400	1,790	7,541	50
	300	-1405,563	-1,744	-0,487	-0,009	0,382	1,607	67,478	58

D.3.2 Beschränkte Intervallskalen

Tabelle D.11: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf von oben beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-207,046	-3,801	-1,010	-0,005	0,885	3,662	1953,559	158
	60	-545,994	-3,781	-0,827	0,119	1,177	4,070	487,990	167
	100	-1618,205	-3,836	-0,866	0,072	1,143	4,136	571,525	148
	300	-103,556	-4,387	-1,108	-0,040	1,095	4,334	431,176	160
(2,3)	30	-378,925	-3,065	-0,758	-0,011	0,836	3,206	160,785	160
	60	-3020,625	-2,990	-0,652	0,063	0,928	3,275	1713,563	152
	100	-556,408	-3,304	-0,855	-0,035	0,787	3,216	143,191	165
	300	-68,026	-3,070	-0,743	0,032	0,810	3,099	196,436	150
(3,10)	30	-1145,202	-1,667	-0,434	-0,030	0,391	1,615	410,090	164
	60	-209,536	-1,654	-0,406	0,009	0,436	1,690	74,326	164
	100	-217,589	-1,803	-0,457	0,006	0,451	1,805	906,987	146
	300	-78,480	-1,740	-0,429	0,018	0,468	1,806	133,055	158
Integrierte Systeme									
(1,1)	30	-147,747	-3,576	-0,944	0,000	0,875	3,451	96,266	154
	60	-325,896	-3,287	-0,813	0,048	0,959	3,563	1196,349	184
	100	-481,639	-3,634	-0,793	0,070	1,137	3,975	2499,676	184
	300	-8427,828	-3,907	-0,932	0,042	1,059	3,991	1159,639	221
(2,3)	30	-155,625	-2,473	-0,623	-0,001	0,701	2,580	180,187	147
	60	-186,806	-2,284	-0,538	0,015	0,677	2,450	727,717	177
	100	-148,009	-2,441	-0,525	0,041	0,774	2,659	198,859	163
	300	-6320,473	-2,177	-0,523	0,014	0,583	2,240	1660,365	188
(3,10)	30	-126,318	-1,588	-0,406	0,010	0,392	1,504	86,129	118
	60	-268,238	-1,198	-0,296	0,012	0,327	1,253	151,717	126
	100	-42,616	-1,040	-0,245	0,024	0,321	1,163	24,599	115
	300	-1330,611	-0,892	-0,218	0,011	0,235	0,912	326,819	101

Tabelle D.12: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf bis zur Hälfte von oben beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-499,0744	-3,5130	-0,7979	0,1301	1,0856	3,8318	1215,9102	135
	60	-463,5069	-3,8542	-0,7926	0,1521	1,2809	4,2386	201,2185	140
	100	-4334,9535	-3,8234	-0,8450	0,1210	1,1902	4,2024	2741,1545	151
	300	-217,2905	-4,2311	-1,0502	0,0383	1,0754	4,1972	3364,7415	154
(2,3)	30	-303,009	-3,078	-0,756	0,100	0,858	3,242	1403,713	133
	60	-6086,131	-2,461	-0,516	0,078	0,861	2,893	217,623	148
	100	-289,269	-2,442	-0,547	0,064	0,742	2,657	72,399	145
	300	-1397,058	-2,444	-0,595	0,053	0,670	2,526	340,883	161
(3,10)	30	-98,576	-1,776	-0,422	0,031	0,481	1,757	74,697	67
	60	-73,722	-1,591	-0,387	0,054	0,466	1,728	96,160	88
	100	-117,163	-1,357	-0,337	0,060	0,351	1,338	412,467	82
	300	-6085,669	-1,137	-0,261	0,051	0,337	1,234	45,134	101
Integrierte Systeme									
(1,1)	30	-232,294	-3,566	-1,035	-0,044	0,654	3,176	1831,988	104
	60	-2734,443	-3,454	-0,856	0,001	0,892	3,409	902,037	142
	100	-71,939	-3,641	-0,734	0,025	1,223	4,146	1165,256	152
	300	-4322,831	-4,179	-0,875	0,013	1,338	4,633	24763,694	193
(2,3)	30	-125,713	-1,792	-0,474	0,019	0,453	1,814	724,492	162
	60	-1533,414	-3,609	-0,902	0,004	1,045	3,962	613,861	131
	100	-1314,269	-3,556	-0,646	0,044	1,418	4,435	1364,857	147
	300	-2931,241	-3,659	-0,733	0,022	1,288	4,299	1285,229	176
(3,10)	30	-160,254	-3,296	-0,859	-0,014	0,780	3,187	521,122	81
	60	-273,369	-2,820	-0,591	0,072	0,969	3,279	360,353	95
	100	-2439,595	-2,059	-0,378	0,069	0,816	2,582	231,348	121
	300	-1777,704	-1,680	-0,354	0,030	0,555	1,811	230,865	134

Tabelle D.13: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf von beiden Seiten beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-542,905	-3,684	-0,905	0,016	0,953	3,704	190,405	172
	60	-518,779	-3,798	-0,857	0,023	1,111	4,005	697,192	163
	100	-657,840	-3,791	-0,973	-0,028	0,926	3,761	277,022	152
	300	-70,593	-4,425	-1,149	-0,075	1,041	4,318	316,500	154
(2,3)	30	-457,386	-3,409	-0,883	-0,002	0,831	3,334	827,361	160
	60	-174,748	-3,354	-0,734	0,130	1,022	3,632	1200,370	159
	100	-441,629	-3,321	-0,841	-0,032	0,831	3,273	290,204	165
	300	-560,870	-2,990	-0,762	-0,008	0,735	2,972	959,805	160
(3,10)	30	-95,206	-1,907	-0,466	-0,010	0,495	1,916	398,093	156
	60	-204,716	-1,767	-0,455	-0,001	0,447	1,799	827,089	173
	100	-2500,979	-1,775	-0,447	0,002	0,441	1,766	166,778	144
	300	-247,015	-1,678	-0,409	0,028	0,447	1,720	272,523	137
Integrierte Systeme									
(1,1)	30	-2146,015	-3,543	-0,916	0,013	0,946	3,736	396,377	172
	60	-174,756	-3,339	-0,777	0,030	0,962	3,561	429,595	192
	100	-292,130	-3,369	-0,804	0,043	0,932	3,384	189,642	188
	300	-818,444	-3,757	-0,922	0,057	1,032	3,938	3034,966	183
(2,3)	30	-204,435	-2,408	-0,648	-0,016	0,537	2,303	404,748	139
	60	-234,654	-2,058	-0,517	-0,003	0,530	2,099	95,001	170
	100	-123,571	-1,971	-0,466	0,042	0,540	1,981	100,907	159
	300	-125,713	-1,792	-0,474	0,019	0,453	1,814	724,492	162
(3,10)	30	-12,494	-1,309	-0,345	-0,005	0,305	1,257	38,371	71
	60	-116,488	-1,091	-0,281	-0,005	0,267	1,066	31,666	91
	100	-16,945	-0,877	-0,230	0,004	0,230	0,906	23,219	98
	300	-72,772	-0,737	-0,199	-0,001	0,172	0,709	51,279	80

D.3.3 Ergebnisse auf von unten beschränkter Intervallskala

Tabelle D.14: Ergebnisse des ADF-Tests auf von unten beschränkter Intervallskala: Anzahl der als stationär eingestuften Linearkombinationen

		kointegriert			stationär			integriert		
T		0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(1,1)	30	319	179	46	345	211	59	91	39	8
	60	866	752	414	877	748	422	85	41	4
	100	977	939	726	994	960	769	93	44	11
	300	1000	1000	1000	1000	1000	1000	87	36	9
(2,3)	30	227	131	39	344	193	55	89	43	8
	60	544	387	166	887	750	399	75	39	8
	100	654	531	289	983	956	778	92	51	14
	300	852	763	586	1000	1000	1000	93	44	11
(3,10)	30	186	103	26	324	173	39	82	39	10
	60	478	326	127	868	746	411	81	41	8
	100	626	513	266	985	949	777	109	61	12
	300	817	763	618	1000	1000	1000	120	75	20

Tabelle D.15: Boxplot-Statistiken für die Schätzung von β auf von unten beschränkter Intervallskala

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-28,173	-1,511	-1,138	-1,004	-0,889	-0,522	3057,326	93
	60	-1,898	-1,214	-1,056	-1,004	-0,950	-0,790	-0,298	62
	100	-1,535	-1,128	-1,032	-1,001	-0,968	-0,875	-0,618	43
	300	-1,074	-1,042	-1,010	-0,998	-0,988	-0,956	-0,931	30
(2,3)	30	-22,935	-1,319	-0,893	-0,752	-0,594	-0,145	99,730	66
	60	-7,838	-1,254	-0,860	-0,749	-0,595	-0,245	45,808	28
	100	-2,937	-1,219	-0,850	-0,756	-0,598	-0,317	1,880	16
	300	-1,874	-1,126	-0,833	-0,748	-0,596	-0,360	-0,360	3
(3,10)	30	-20,069	-0,638	-0,463	-0,397	-0,343	-0,165	23,232	87
	60	-4,059	-0,532	-0,419	-0,384	-0,342	-0,227	4,537	95
	100	-1,469	-0,509	-0,405	-0,375	-0,335	-0,235	1,037	121
	300	-0,545	-0,493	-0,399	-0,370	-0,333	-0,234	-0,016	132
Stationäre Systeme									
(1,1)	30	-379,338	-3,852	-0,962	-0,020	1,017	3,900	139,402	145
	60	-269,252	-3,593	-0,853	0,093	1,052	3,869	1322,165	171
	100	-1103,508	-3,987	-1,035	-0,034	0,935	3,819	2634,318	167
	300	-247,186	-4,438	-1,138	-0,059	1,064	4,358	436,208	155
(2,3)	30	-369,842	-3,001	-0,774	-0,019	0,729	2,968	76,303	164
	60	-1214,907	-2,933	-0,694	0,058	0,867	3,191	153,921	165
	100	-550,231	-3,051	-0,769	-0,014	0,764	3,060	441,372	175
	300	-196,544	-2,953	-0,786	-0,056	0,713	2,961	2648,531	163
(3,10)	30	-723,389	-1,689	-0,412	-0,012	0,453	1,700	202,957	153
	60	-429,413	-1,811	-0,473	-0,017	0,420	1,750	305,289	164
	100	-119,312	-1,680	-0,416	0,004	0,447	1,720	844,782	154
	300	-88,727	-1,674	-0,418	0,026	0,425	1,685	118,659	149
Integrierte Systeme									
(1,1)	30	-229,398	-3,354	-0,905	-0,011	0,799	3,303	250,426	172
	60	-525,993	-3,729	-0,843	0,043	1,144	4,041	175,338	184
	100	-732,622	-3,422	-0,827	0,047	0,988	3,655	527,980	204
	300	-2369,907	-3,395	-0,741	0,042	1,058	3,742	363,207	209
(2,3)	30	-250,513	-2,848	-0,753	0,001	0,784	3,015	318,345	164
	60	-445,249	-2,798	-0,640	0,036	0,806	2,921	935,315	157
	100	-286,362	-2,424	-0,564	0,035	0,680	2,487	1146,655	174
	300	-3090,062	-2,229	-0,543	0,018	0,583	2,268	116,647	164
(3,10)	30	-47,012	-1,665	-0,401	0,000	0,449	1,706	356,075	112
	60	-127,081	-1,398	-0,354	-0,007	0,347	1,352	144,331	116
	100	-20,100	-1,169	-0,281	0,002	0,316	1,209	58,975	114
	300	-22,991	-0,881	-0,227	0,007	0,239	0,899	293,927	95

Tabelle D.16: Klassifikation jeweils 1000 Systeme auf nach unten beschränkter Intervallskala

		Niveau 0.1			Niveau 0.05			Niveau 0.01			
(a,b)	T	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	I(1)	CI(1)	I(0)	num
Kointegrierte Systeme											
(1,1)	30	261	609	130	421	507	72	722	258	20	1000
	60	4	870	126	11	921	68	74	902	24	1000
	100	0	860	140	0	921	79	0	976	24	1000
	300	0	857	143	0	931	69	0	986	14	1000
(2,3)	30	425	478	97	562	390	48	800	186	14	1000
	60	95	804	101	155	798	47	353	638	9	1000
	100	15	898	87	36	917	47	115	871	14	1000
	300	0	913	87	0	956	44	2	992	6	1000
(3,10)	30	313	427	88	460	323	45	656	163	9	828
	60	122	665	95	186	653	43	378	493	11	882
	100	54	776	84	78	788	48	162	736	16	914
	300	19	851	81	24	886	41	40	902	9	951
Stationäre Systeme											
(1,1)	30	79	92	829	202	189	609	542	271	187	1000
	60	0	0	1000	0	0	1000	4	16	980	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(2,3)	30	71	109	820	187	203	610	542	271	187	1000
	60	0	0	1000	1	0	999	4	15	981	1000
	100	0	0	1000	0	0	1000	0	0	1000	1000
	300	0	0	1000	0	0	1000	0	0	1000	1000
(3,10)	30	80	104	794	203	198	577	522	267	189	978
	60	0	0	997	0	1	996	5	10	982	997
	100	0	0	999	0	0	999	0	0	999	999
	300	0	0	1000	0	0	1000	0	0	1000	1000
Integrierte Systeme											
(1,1)	30	602	226	35	684	161	18	789	73	1	863
	60	577	298	39	680	223	11	811	102	1	914
	100	526	353	58	636	279	22	777	158	2	937
	300	417	435	111	514	403	46	698	257	8	963
(2,3)	30	597	191	21	669	133	7	741	65	3	809
	60	633	200	19	705	141	6	784	67	1	852
	100	599	234	33	681	174	11	770	94	2	866
	300	516	334	46	610	266	20	745	149	2	896
(3,10)	30	485	147	14	539	100	7	597	49	0	646
	60	525	160	23	592	107	9	659	48	1	708
	100	465	195	29	541	132	16	630	56	3	689
	300	355	309	48	446	248	18	567	143	2	712

D.3.4 Ordinalskalen

Tabelle D.17: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf der Ordinalskala 1

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-2429,131	-3,797	-0,946	-0,022	1,020	3,830	187,184	167
	60	-355,705	-3,852	-0,887	0,085	1,145	4,139	440,878	158
	100	-2160,595	-3,738	-0,902	0,012	1,003	3,861	243,254	157
	300	-309,800	-4,117	-1,044	-0,014	1,011	3,946	17130,689	168
(2,3)	30	-314,155	-3,340	-0,824	-0,021	0,881	3,404	4518,053	163
	60	-481,563	-3,294	-0,713	0,086	1,026	3,596	653,059	152
	100	-171,421	-3,305	-0,821	-0,021	0,848	3,316	155,074	148
	300	-229,569	-3,159	-0,804	-0,020	0,767	3,123	904,977	151
(3,10)	30	-1098,190	-2,005	-0,519	-0,027	0,479	1,958	147,975	146
	60	-43,080	-1,888	-0,469	-0,013	0,496	1,936	328,256	162
	100	-151,495	-1,973	-0,522	-0,012	0,450	1,894	608,436	158
	300	-78,267	-2,007	-0,479	0,021	0,540	2,048	3135,407	139
Integrierte Systeme									
(1,1)	30	-281,843	-3,628	-0,855	0,018	0,997	3,763	3760,798	173
	60	-281,257	-3,290	-0,751	0,016	0,964	3,512	655,681	210
	100	-805,630	-3,234	-0,758	0,076	0,938	3,359	158,551	209
	300	-299,672	-3,787	-0,919	0,057	1,025	3,872	318,187	194
(2,3)	30	-268,244	-2,239	-0,579	0,016	0,541	2,150	463,318	165
	60	-1699,239	-1,719	-0,381	0,034	0,515	1,834	88,650	189
	100	-96,669	-1,735	-0,405	0,036	0,503	1,836	28,859	182
	300	-216,929	-1,728	-0,449	0,028	0,413	1,698	271,640	180
(3,10)	30	-52,715	-1,070	-0,283	-0,008	0,242	0,975	33,060	99
	60	-49,897	-0,825	-0,209	0,010	0,205	0,808	22,468	119
	100	-59,728	-0,702	-0,178	0,006	0,181	0,711	46,875	118
	300	-22,584	-0,638	-0,161	0,000	0,157	0,619	130,472	112

Tabelle D.18: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf der Ordinalskala 2

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-157,901	-3,887	-0,871	0,108	1,157	4,166	1030,882	169
	60	-132,874	-3,690	-0,887	-0,024	0,987	3,777	592,642	159
	100	-151,092	-3,800	-0,968	-0,049	0,952	3,817	3774,949	159
	300	-2440,248	-4,188	-1,166	-0,134	0,914	3,989	816,187	151
(2,3)	30	-107,315	-2,345	-0,561	0,025	0,646	2,450	45,915	160
	60	-1217,710	-2,340	-0,563	0,033	0,675	2,495	264,478	183
	100	-55,038	-2,396	-0,578	0,023	0,688	2,585	1088,410	144
	300	-535,946	-2,439	-0,604	0,008	0,628	2,465	158,121	163
(3,10)	30	-51,787	-1,283	-0,316	0,037	0,356	1,308	33,159	65
	60	-265,571	-1,246	-0,327	0,005	0,291	1,207	157,111	57
	100	-36,267	-1,130	-0,305	-0,003	0,276	1,140	11,811	61
	300	-95,607	-1,040	-0,245	0,024	0,297	1,079	9,589	70
Integrierte Systeme									
(1,1)	30	-475,068	-3,630	-0,846	0,004	1,016	3,794	1330,767	130
	60	-1615,279	-3,660	-0,878	0,072	1,011	3,811	12066,140	156
	100	-693,910	-3,616	-0,817	0,115	1,075	3,862	859,937	137
	300	-416,348	-3,843	-0,976	0,035	0,954	3,691	262,031	132
(2,3)	30	-5387,600	-2,548	-0,620	-0,005	0,699	2,638	12572,934	122
	60	-115,507	-2,493	-0,593	0,027	0,707	2,602	63,412	134
	100	-64,877	-2,460	-0,513	0,057	0,786	2,645	220,320	135
	300	-134,620	-2,555	-0,575	0,032	0,755	2,673	534,934	116
(3,10)	30	-29,301	-1,767	-0,384	0,001	0,558	1,942	37,878	99
	60	-128,115	-1,583	-0,387	-0,007	0,449	1,686	68,255	116
	100	-75,871	-1,612	-0,363	0,025	0,473	1,703	46,711	126
	300	-111,454	-1,478	-0,339	0,010	0,442	1,613	85,466	110

Tabelle D.19: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf der Ordinalskala 3

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-135,800	-4,011	-0,927	0,101	1,132	4,203	19208,966	160
	60	-999,574	-3,904	-0,901	0,059	1,127	4,161	106,124	176
	100	-125,538	-3,929	-0,953	0,057	1,052	3,861	356,787	161
	300	-129,313	-4,294	-1,117	-0,064	1,037	4,186	727213,383	148
(2,3)	30	-811,090	-2,566	-0,625	0,006	0,706	2,680	768,826	170
	60	-387,232	-2,715	-0,674	-0,018	0,712	2,769	104,408	169
	100	-43,099	-2,604	-0,619	0,006	0,734	2,750	146,348	142
	300	-648,114	-2,609	-0,602	0,061	0,760	2,791	1037,988	160
(3,10)	30	-1298,993	-1,269	-0,311	0,003	0,328	1,284	73,783	150
	60	-406,700	-1,279	-0,316	0,000	0,332	1,303	40,074	149
	100	-67,599	-1,285	-0,323	0,011	0,320	1,283	45,389	162
	300	-69,777	-1,505	-0,399	0,007	0,339	1,429	60,398	152
Integrierte Systeme									
(1,1)	30	-1134,792	-3,641	-0,882	0,012	0,959	3,718	175,309	168
	60	-149,775	-3,401	-0,808	0,028	0,971	3,597	1666,795	165
	100	-688,871	-3,537	-0,703	0,133	1,194	4,030	322,295	168
	300	-571,854	-3,886	-1,004	0,040	1,042	4,102	290,289	164
(2,3)	30	-460,518	-2,291	-0,524	0,024	0,660	2,400	278,622	170
	60	-209,511	-2,042	-0,452	0,047	0,613	2,192	299,334	185
	100	-130,454	-2,242	-0,488	0,049	0,690	2,449	49,512	158
	300	-177,881	-2,216	-0,508	0,038	0,633	2,341	1187,450	180
(3,10)	30	-318,054	-1,338	-0,306	0,002	0,396	1,441	511,643	132
	60	-1724,596	-0,991	-0,238	0,036	0,313	1,140	64,218	159
	100	-26,131	-1,001	-0,215	0,029	0,327	1,117	90,123	145
	300	-14451,302	-0,828	-0,194	0,015	0,248	0,886	34,777	151

Tabelle D.20: Boxplot-Statistiken für die Schätzung von β bei stationären und integrierten Systemen auf der Ordinalskala 4

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Stationäre Systeme									
(1,1)	30	-149,009	-3,854	-0,965	-0,007	0,986	3,878	3327,603	144
	60	-156,343	-3,799	-0,896	0,053	1,051	3,957	454,325	173
	100	-1032,779	-3,729	-0,916	-0,005	1,028	3,942	706,289	159
	300	-114,376	-4,146	-1,045	-0,040	1,024	4,066	145,199	147
(2,3)	30	-487,120	-2,829	-0,744	0,000	0,716	2,860	1164,104	180
	60	-176,241	-2,932	-0,671	0,083	0,880	3,170	100,511	169
	100	-661,567	-2,906	-0,700	-0,014	0,771	2,867	681,856	151
	300	-495,420	-2,643	-0,711	-0,034	0,635	2,575	283,748	164
(3,10)	30	-227,691	-1,452	-0,356	0,007	0,388	1,455	99,202	156
	60	-666,384	-1,380	-0,350	-0,003	0,343	1,379	1203,840	174
	100	-35,143	-1,436	-0,358	0,007	0,364	1,434	89,731	150
	300	-199,118	-1,423	-0,372	0,016	0,369	1,463	69,230	155
Integrierte Systeme									
(1,1)	30	-218,856	-3,427	-0,842	0,011	0,920	3,507	155,782	157
	60	-142,680	-3,533	-0,787	0,057	1,056	3,817	2731,230	169
	100	-4031,175	-3,454	-0,780	0,131	1,046	3,754	82,865	160
	300	-294,291	-3,766	-0,993	0,060	1,025	4,052	930,590	143
(2,3)	30	-396,667	-2,124	-0,530	0,010	0,582	2,241	390,788	157
	60	-70,269	-2,252	-0,517	0,023	0,643	2,358	776,958	162
	100	-132,065	-2,093	-0,461	0,082	0,650	2,313	830,998	148
	300	-376,536	-2,158	-0,535	0,036	0,549	2,169	108,309	140
(3,10)	30	-136,224	-1,156	-0,305	0,004	0,263	1,102	122,783	116
	60	-192,312	-0,862	-0,213	0,006	0,231	0,890	292,549	146
	100	-270,926	-0,836	-0,199	0,016	0,235	0,872	24,391	135
	300	-196,072	-0,680	-0,181	0,013	0,176	0,711	2395,503	127

Tabelle D.21: Schätzung für β nach regelmäßigem Entfernen jeder 10. Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-57,809	-1,527	-1,142	-0,995	-0,882	-0,494	3,958	79
	60	-1,968	-1,211	-1,059	-1,007	-0,952	-0,796	-0,468	53
	100	-1,430	-1,124	-1,034	-1,003	-0,974	-0,885	-0,679	36
	300	-1,071	-1,037	-1,009	-0,999	-0,990	-0,963	-0,932	34
(2,3)	30	-147,920	-0,809	-0,703	-0,666	-0,632	-0,530	136,306	93
	60	-8,342	-0,723	-0,680	-0,667	-0,651	-0,608	-0,525	61
	100	-0,751	-0,698	-0,675	-0,666	-0,658	-0,635	-0,578	54
	300	-0,702	-0,677	-0,670	-0,667	-0,664	-0,656	-0,648	29
(3,10)	30	-1,829	-0,338	-0,310	-0,300	-0,292	-0,265	0,099	85
	60	-0,347	-0,315	-0,304	-0,300	-0,296	-0,284	-0,255	43
	100	-0,316	-0,309	-0,302	-0,300	-0,298	-0,291	-0,283	41
	300	-0,306	-0,303	-0,301	-0,300	-0,299	-0,297	-0,295	45
Stationäre Systeme									
(1,1)	30	-462,640	-4,035	-0,957	0,006	1,103	4,175	458,858	156
	60	-418,291	-3,912	-1,039	-0,075	0,887	3,763	412,286	158
	100	-342,885	-3,831	-0,987	-0,018	0,915	3,656	1831,806	170
	300	-9212,466	-4,047	-1,090	-0,078	0,881	3,799	368,329	160
(2,3)	30	-1321,431	-2,820	-0,720	-0,008	0,709	2,824	295,533	172
	60	-190,663	-2,736	-0,679	0,025	0,711	2,783	247,684	163
	100	-467,932	-2,667	-0,663	-0,003	0,699	2,673	365,791	165
	300	-1075,250	-2,633	-0,622	-0,011	0,736	2,756	2129,597	149
(3,10)	30	-167,921	-1,207	-0,305	0,003	0,300	1,175	71,811	166
	60	-137,568	-1,198	-0,320	0,000	0,286	1,164	7225,776	181
	100	-84,879	-1,237	-0,312	-0,015	0,324	1,232	843,726	167
	300	-34,725	-1,288	-0,326	0,002	0,331	1,278	35,308	149
Integrierte Systeme									
(1,1)	30	-456,023	-3,428	-0,816	0,022	0,961	3,589	170,311	157
	60	-361,405	-3,419	-0,762	0,081	1,073	3,665	294,443	163
	100	-100,010	-3,528	-0,796	0,103	1,082	3,897	186,219	150
	300	-1221,954	-3,736	-0,924	0,047	1,002	3,775	601,140	139
(2,3)	30	-285,568	-2,391	-0,589	0,005	0,612	2,389	57,572	151
	60	-409,208	-2,260	-0,467	0,062	0,731	2,463	321,331	152
	100	-146,191	-2,288	-0,493	0,077	0,718	2,485	122,726	153
	300	-76,618	-2,410	-0,604	0,015	0,637	2,487	610,967	152
(3,10)	30	-34,330	-1,110	-0,268	0,005	0,293	1,133	68,756	161
	60	-200,542	-1,015	-0,224	0,029	0,331	1,161	349,030	147
	100	-42,936	-1,046	-0,232	0,035	0,335	1,175	29,910	144
	300	-114,382	-1,080	-0,273	0,011	0,282	1,076	32,234	154

Tabelle D.22: Schätzung für β nach regelmäßigem Entfernen jeder 7. Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-170,278	-1,513	-1,135	-0,992	-0,883	-0,509	7,331	97
	60	-2,803	-1,229	-1,059	-1,004	-0,945	-0,775	-0,273	46
	100	-1,522	-1,127	-1,033	-1,001	-0,971	-0,879	-0,560	52
	300	-1,093	-1,040	-1,010	-0,999	-0,990	-0,962	-0,947	28
(2,3)	30	-4,206	-0,813	-0,701	-0,664	-0,627	-0,516	2,784	82
	60	-0,881	-0,725	-0,679	-0,665	-0,649	-0,603	1,543	47
	100	-0,768	-0,699	-0,674	-0,665	-0,657	-0,632	-0,592	57
	300	-0,704	-0,677	-0,669	-0,667	-0,664	-0,656	-0,648	53
(3,10)	30	-0,898	-0,339	-0,311	-0,300	-0,292	-0,264	0,260	91
	60	-0,431	-0,316	-0,303	-0,300	-0,295	-0,284	-0,256	35
	100	-0,321	-0,309	-0,302	-0,300	-0,298	-0,291	-0,280	47
	300	-0,305	-0,303	-0,301	-0,300	-0,299	-0,297	-0,294	37
Stationäre Systeme									
(1,1)	30	-6114,433	-4,070	-1,003	0,039	1,063	4,126	261,779	158
	60	-13543,541	-3,786	-0,931	0,049	1,037	3,987	1131,866	149
	100	-16008,791	-3,735	-0,898	0,035	1,013	3,725	181,455	149
	300	-1222,857	-3,756	-0,963	0,035	0,952	3,715	2284,482	151
(2,3)	30	-736,276	-2,934	-0,725	-0,006	0,753	2,913	8103,434	158
	60	-61,663	-2,939	-0,770	-0,016	0,723	2,873	70,354	145
	100	-1248,195	-2,757	-0,657	-0,004	0,755	2,840	427,497	136
	300	-111,292	-2,598	-0,702	-0,071	0,577	2,463	214,502	162
(3,10)	30	-204,003	-1,334	-0,359	-0,021	0,302	1,292	115,978	166
	60	-82,023	-1,282	-0,323	0,009	0,319	1,277	30,587	168
	100	-77,235	-1,233	-0,300	0,004	0,347	1,313	1469,827	154
	300	-3772,137	-1,178	-0,299	-0,003	0,309	1,205	163,777	153
Integrierte Systeme									
(1,1)	30	-322,455	-3,533	-0,815	0,049	1,023	3,723	439,291	147
	60	-466,964	-3,530	-0,800	0,045	1,024	3,693	68,846	167
	100	-179,361	-3,450	-0,770	0,122	1,021	3,694	685,160	159
	300	-202,070	-3,737	-0,907	0,043	1,011	3,879	984,684	136
(2,3)	30	-180,710	-2,377	-0,566	0,026	0,650	2,474	376,171	150
	60	-161,053	-2,302	-0,504	0,059	0,706	2,486	203,648	161
	100	-1868,482	-2,283	-0,503	0,086	0,702	2,501	2394,941	150
	300	-733,576	-2,448	-0,607	0,023	0,626	2,421	249,297	152
(3,10)	30	-120,595	-1,036	-0,271	0,009	0,283	1,108	50,296	148
	60	-202,477	-1,071	-0,242	0,026	0,311	1,129	839,314	163
	100	-26,002	-1,049	-0,234	0,043	0,327	1,136	921,667	153
	300	-141,232	-1,078	-0,277	0,011	0,284	1,124	170,041	153

Tabelle D.23: Schätzung für β nach regelmäßigem Entfernen jeder 5. Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-7,435	-1,566	-1,153	-1,006	-0,874	-0,458	56,716	94
	60	-2,021	-1,235	-1,064	-1,004	-0,949	-0,783	-0,457	53
	100	-1,560	-1,127	-1,035	-1,002	-0,972	-0,879	-0,657	40
	300	-1,082	-1,040	-1,010	-1,000	-0,990	-0,961	-0,942	27
(2,3)	30	-17,313	-0,830	-0,709	-0,668	-0,626	-0,501	1,854	92
	60	-1,046	-0,729	-0,682	-0,667	-0,649	-0,602	0,651	61
	100	-0,767	-0,698	-0,674	-0,666	-0,658	-0,635	-0,584	72
	300	-0,709	-0,678	-0,670	-0,667	-0,664	-0,655	-0,644	30
(3,10)	30	-12,478	-0,341	-0,311	-0,300	-0,290	-0,259	2,459	83
	60	-0,351	-0,316	-0,304	-0,300	-0,296	-0,283	-0,096	56
	100	-0,318	-0,309	-0,302	-0,300	-0,298	-0,290	-0,275	47
	300	-0,306	-0,303	-0,301	-0,300	-0,299	-0,297	-0,294	48
Stationäre Systeme									
(1,1)	30	-10192,652	-3,920	-0,982	-0,066	0,998	3,952	205,018	167
	60	-626,775	-3,986	-1,120	-0,056	0,793	3,659	179,040	154
	100	-223,494	-3,783	-0,953	-0,034	0,935	3,689	1355,677	166
	300	-1113,527	-4,202	-1,159	-0,056	0,933	4,043	686,989	166
(2,3)	30	-1246,054	-3,152	-0,877	-0,028	0,647	2,878	15435,891	158
	60	-878,337	-2,546	-0,597	0,057	0,764	2,735	1127,217	181
	100	-671,209	-2,779	-0,701	0,021	0,701	2,742	103,000	164
	300	-3075,594	-3,040	-0,860	-0,059	0,635	2,844	610,926	149
(3,10)	30	-116,533	-1,317	-0,367	-0,018	0,271	1,206	73,289	155
	60	-233,255	-1,261	-0,340	0,003	0,306	1,262	132,817	162
	100	-394,540	-1,296	-0,316	0,002	0,345	1,297	256,539	152
	300	-45,179	-1,332	-0,345	-0,010	0,323	1,297	1945,536	143
Integrierte Systeme									
(1,1)	30	-170,289	-3,649	-0,867	0,029	0,989	3,767	119,051	140
	60	-132,978	-3,345	-0,759	0,067	1,014	3,653	498,209	157
	100	-45,096	-3,477	-0,772	0,109	1,051	3,775	632,215	156
	300	-392,011	-3,804	-0,942	0,021	0,997	3,782	1874,160	143
(2,3)	30	-61,980	-2,318	-0,555	0,007	0,628	2,400	475,041	158
	60	-2092,712	-2,356	-0,509	0,048	0,742	2,577	602,812	144
	100	-148,589	-2,205	-0,479	0,079	0,731	2,506	75,719	157
	300	-235,976	-2,397	-0,615	0,021	0,621	2,459	389,978	157
(3,10)	30	-32,524	-1,050	-0,259	-0,005	0,277	1,034	1157,524	163
	60	-4254,064	-1,059	-0,224	0,030	0,334	1,126	77,422	138
	100	-315,419	-1,064	-0,234	0,036	0,321	1,151	197,530	142
	300	-21,432	-1,095	-0,269	0,013	0,281	1,097	820,306	161

Tabelle D.24: Schätzung für β nach regelmäßigem Entfernen jeder 3. Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-172,617	-1,687	-1,178	-0,997	-0,839	-0,341	41,903	129
	60	-2,314	-1,271	-1,075	-1,007	-0,943	-0,746	1,706	54
	100	-3,823	-1,147	-1,041	-1,003	-0,970	-0,863	-0,563	48
	300	-1,080	-1,046	-1,012	-1,001	-0,989	-0,956	-0,890	26
(2,3)	30	-13,029	-0,857	-0,717	-0,669	-0,622	-0,480	365,713	117
	60	-11,909	-0,742	-0,686	-0,666	-0,647	-0,589	0,245	57
	100	-0,775	-0,705	-0,676	-0,666	-0,657	-0,628	-0,448	54
	300	-0,703	-0,679	-0,670	-0,667	-0,663	-0,654	-0,646	32
(3,10)	30	-2,408	-0,348	-0,312	-0,300	-0,287	-0,251	2,657	111
	60	-0,838	-0,319	-0,305	-0,300	-0,295	-0,282	0,536	70
	100	-0,335	-0,311	-0,303	-0,300	-0,297	-0,289	-0,271	55
	300	-0,305	-0,303	-0,301	-0,300	-0,299	-0,297	-0,294	35
Stationäre Systeme									
(1,1)	30	-292,090	-3,769	-0,943	-0,019	1,007	3,887	201,477	155
	60	-345,126	-3,846	-1,033	0,004	0,940	3,891	77,928	143
	100	-592,188	-3,643	-0,911	-0,011	0,977	3,782	286,525	167
	300	-149,060	-4,066	-1,112	-0,005	0,923	3,908	829,599	158
(2,3)	30	-27827,133	-2,728	-0,696	-0,055	0,666	2,617	295,189	170
	60	-152,446	-2,731	-0,730	-0,024	0,607	2,600	592,418	155
	100	-217,056	-2,624	-0,632	0,038	0,713	2,701	371,699	183
	300	-189,982	-2,735	-0,668	0,023	0,733	2,826	501,289	157
(3,10)	30	-89,067	-1,285	-0,362	-0,042	0,283	1,230	447,492	139
	60	-312,780	-1,217	-0,277	0,021	0,349	1,276	71,021	150
	100	-47,706	-1,154	-0,298	0,017	0,293	1,150	326,739	171
	300	-3426,071	-1,265	-0,316	-0,003	0,323	1,277	73,051	164
Integrierte Systeme									
(1,1)	30	-790,370	-3,378	-0,851	-0,021	0,897	3,506	1442,755	173
	60	-526,128	-3,514	-0,793	0,077	1,027	3,715	216,512	154
	100	-245,523	-3,507	-0,751	0,123	1,113	3,881	634,686	145
	300	-2167,224	-3,900	-0,982	0,029	0,983	3,899	200,432	145
(2,3)	30	-68,405	-2,402	-0,520	0,020	0,737	2,604	172,575	150
	60	-2760,792	-2,327	-0,522	0,050	0,696	2,501	530,214	157
	100	-36,813	-2,203	-0,482	0,106	0,697	2,456	568,161	141
	300	-195,472	-2,357	-0,599	0,026	0,630	2,370	126,352	158
(3,10)	30	-23,544	-1,077	-0,263	0,001	0,294	1,111	100,523	162
	60	-3988,676	-1,062	-0,243	0,015	0,314	1,144	77,523	150
	100	-475,884	-1,060	-0,237	0,033	0,312	1,082	35,341	138
	300	-253,308	-1,095	-0,269	0,010	0,282	1,078	133,637	165

Tabelle D.25: Schätzung für β nach zufälligem Entfernen eines Zehntels der Messungen

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-20,705	-1,513	-1,141	-1,007	-0,889	-0,516	30,311	80
	60	-2,354	-1,205	-1,055	-1,003	-0,953	-0,805	-0,539	59
	100	-1,480	-1,114	-1,030	-1,005	-0,974	-0,893	-0,722	43
	300	-1,079	-1,036	-1,009	-0,999	-0,991	-0,963	-0,944	36
(2,3)	30	-3,968	-0,801	-0,699	-0,665	-0,631	-0,528	1,603	91
	60	-3,313	-0,725	-0,680	-0,666	-0,650	-0,606	-0,548	56
	100	-0,852	-0,699	-0,674	-0,666	-0,658	-0,633	-0,587	40
	300	-0,701	-0,677	-0,669	-0,667	-0,664	-0,657	-0,648	34
(3,10)	30	-1,959	-0,340	-0,311	-0,300	-0,291	-0,262	1,739	80
	60	-0,346	-0,316	-0,304	-0,300	-0,296	-0,284	-0,236	40
	100	-0,317	-0,309	-0,302	-0,300	-0,298	-0,292	-0,282	51
	300	-0,305	-0,303	-0,301	-0,300	-0,299	-0,297	-0,294	38
Stationäre Systeme									
(1,1)	30	-4241,389	-3,809	-0,946	0,061	1,050	4,043	635,970	145
	60	-347,383	-3,780	-0,975	0,027	0,915	3,732	777,641	161
	100	-1566,223	-3,839	-0,967	0,058	0,954	3,794	3551,605	165
	300	-1739,488	-4,122	-1,105	-0,055	0,913	3,842	1342,890	151
(2,3)	30	-110,802	-3,061	-0,750	0,020	0,795	3,108	404,075	164
	60	-549,949	-2,803	-0,687	0,039	0,758	2,766	749,781	163
	100	-125,331	-2,651	-0,649	0,028	0,741	2,763	16726,423	153
	300	-75,655	-2,821	-0,681	0,003	0,751	2,896	1014,453	132
(3,10)	30	-47,230	-1,327	-0,324	0,010	0,348	1,351	3345,449	141
	60	-84,527	-1,179	-0,299	0,000	0,298	1,143	163,371	184
	100	-75,184	-1,202	-0,295	0,012	0,319	1,237	95,068	155
	300	-144,101	-1,254	-0,312	-0,004	0,319	1,213	102,002	147
Integrierte Systeme									
(1,1)	30	-69,852	-3,430	-0,796	0,036	0,973	3,471	4796,596	151
	60	-624,894	-3,474	-0,755	0,083	1,063	3,745	566,133	175
	100	-665,329	-3,507	-0,788	0,113	1,028	3,694	2348,872	157
	300	-540,813	-3,870	-0,954	0,028	0,991	3,764	203,597	134
(2,3)	30	-146,782	-2,327	-0,561	0,019	0,652	2,443	79,702	146
	60	-368,592	-2,332	-0,511	0,063	0,717	2,532	225,078	161
	100	-165,704	-2,295	-0,530	0,068	0,685	2,483	152,651	148
	300	-1302,041	-2,446	-0,606	0,028	0,626	2,452	2636,804	149
(3,10)	30	-37,118	-1,023	-0,244	0,003	0,285	1,072	785,171	142
	60	-123,347	-1,072	-0,238	0,031	0,336	1,187	265,370	147
	100	-241,195	-1,056	-0,233	0,038	0,321	1,131	119,619	148
	300	-78,746	-1,097	-0,268	0,016	0,298	1,111	1021,427	153

Tabelle D.26: Schätzung für β nach zufälligem Entfernen eines Siebtels der Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-24,078	-1,522	-1,141	-1,008	-0,885	-0,526	143,694	90
	60	-2,280	-1,225	-1,060	-1,004	-0,950	-0,793	0,005	51
	100	-1,512	-1,118	-1,033	-1,004	-0,972	-0,882	-0,749	38
	300	-1,077	-1,037	-1,009	-1,000	-0,990	-0,963	-0,947	40
(2,3)	30	-2,184	-0,812	-0,702	-0,664	-0,629	-0,523	14,610	87
	60	-1,933	-0,726	-0,681	-0,666	-0,651	-0,608	-0,548	60
	100	-1,034	-0,698	-0,674	-0,666	-0,658	-0,634	-0,582	52
	300	-0,697	-0,677	-0,670	-0,667	-0,664	-0,657	-0,647	40
(3,10)	30	-0,680	-0,340	-0,310	-0,300	-0,290	-0,261	1,047	79
	60	-0,346	-0,316	-0,304	-0,299	-0,296	-0,284	-0,231	41
	100	-0,318	-0,308	-0,302	-0,300	-0,298	-0,291	-0,280	55
	300	-0,305	-0,303	-0,301	-0,300	-0,299	-0,297	-0,293	44
Stationäre Systeme									
(1,1)	30	-1394,685	-3,986	-1,030	-0,019	0,954	3,897	304,954	144
	60	-597,321	-3,730	-0,957	-0,007	0,981	3,744	85,822	154
	100	-577,367	-3,980	-1,012	0,029	1,013	4,044	497,101	150
	300	-337,830	-4,026	-1,057	-0,053	0,963	3,856	576,878	157
(2,3)	30	-582,671	-3,131	-0,866	-0,039	0,649	2,856	562,790	158
	60	-134,996	-2,897	-0,740	0,027	0,718	2,880	436,850	157
	100	-144,010	-2,658	-0,636	0,032	0,728	2,672	1597,219	157
	300	-408,134	-2,637	-0,658	-0,008	0,697	2,676	733,634	142
(3,10)	30	-476,706	-1,384	-0,371	-0,008	0,312	1,331	270,086	152
	60	-62,520	-1,263	-0,318	0,003	0,315	1,247	115,568	173
	100	-562,259	-1,198	-0,295	0,004	0,318	1,182	6140,619	150
	300	-88,447	-1,169	-0,297	0,001	0,295	1,136	99,699	148
Integrierte Systeme									
(1,1)	30	-477,754	-3,442	-0,821	0,037	0,982	3,569	101,183	143
	60	-645,759	-3,505	-0,791	0,059	1,025	3,680	682,430	169
	100	-2583,068	-3,474	-0,778	0,121	1,038	3,727	124,575	151
	300	-829,922	-3,750	-0,922	0,047	1,002	3,885	415,512	135
(2,3)	30	-636,630	-2,378	-0,575	0,015	0,637	2,438	2294,825	143
	60	-240,434	-2,238	-0,504	0,059	0,707	2,508	368,126	163
	100	-27,762	-2,319	-0,515	0,084	0,692	2,500	1274,138	144
	300	-176,672	-2,430	-0,604	0,030	0,623	2,457	712,461	147
(3,10)	30	-86,120	-1,036	-0,257	0,003	0,285	1,091	96,950	147
	60	-61348,074	-1,082	-0,247	0,022	0,328	1,185	36,251	150
	100	-28,344	-1,016	-0,237	0,035	0,316	1,145	168,696	148
	300	-67,276	-1,098	-0,268	0,012	0,291	1,120	76,746	154

Tabelle D.27: Schätzung für β nach zufälligem Entfernen eines Fünftels der Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-12,151	-1,516	-1,149	-1,006	-0,883	-0,486	10,554	92
	60	-3,913	-1,236	-1,064	-1,003	-0,949	-0,777	-0,378	56
	100	-1,571	-1,126	-1,033	-1,002	-0,970	-0,877	-0,691	46
	300	-1,076	-1,040	-1,010	-1,000	-0,990	-0,961	-0,943	30
(2,3)	30	-2,172	-0,815	-0,704	-0,663	-0,628	-0,515	1,868	93
	60	-2,073	-0,731	-0,683	-0,667	-0,650	-0,604	-0,527	53
	100	-0,900	-0,699	-0,674	-0,666	-0,658	-0,633	-0,574	63
	300	-0,704	-0,678	-0,670	-0,667	-0,664	-0,656	-0,643	30
(3,10)	30	-5,312	-0,341	-0,310	-0,301	-0,290	-0,260	1,690	83
	60	-0,358	-0,317	-0,304	-0,300	-0,296	-0,284	-0,249	57
	100	-0,319	-0,309	-0,302	-0,300	-0,298	-0,291	-0,284	44
	300	-0,305	-0,303	-0,301	-0,300	-0,299	-0,297	-0,293	38
Stationäre Systeme									
(1,1)	30	-34,891	-0,481	0,764	1,151	1,666	3,003	464,919	54
	60	-94,865	-1,256	0,583	1,153	1,825	3,685	145,246	150
	100	-170,005	-0,144	1,047	1,405	1,845	3,011	75,665	87
	300	0,908	0,984	1,297	1,406	1,537	1,892	2,319	18
(2,3)	30	-34,991	-0,374	0,667	0,956	1,410	2,523	221,230	104
	60	-5,651	-0,099	0,750	1,016	1,330	2,172	36,686	87
	100	-12,341	0,069	0,782	1,001	1,283	2,033	17,838	55
	300	0,564	0,653	0,889	0,977	1,069	1,337	1,697	17
(3,10)	30	-27,147	-0,167	0,289	0,436	0,657	1,201	753,726	80
	60	-246,103	-0,179	0,294	0,423	0,621	1,095	118,795	112
	100	-12,777	0,053	0,359	0,461	0,600	0,960	29,820	70
	300	0,267	0,285	0,396	0,433	0,477	0,598	0,728	14
Integrierte Systeme									
(1,1)	30	-323,979	-3,473	-0,820	0,022	0,958	3,611	229,366	145
	60	-555,978	-3,479	-0,736	0,066	1,104	3,797	1479,553	169
	100	-335,988	-3,433	-0,775	0,110	1,008	3,520	382,601	154
	300	-534,629	-3,809	-0,961	0,038	0,956	3,782	625,852	134
(2,3)	30	-348,736	-2,323	-0,574	0,014	0,623	2,413	5251,467	150
	60	-637,005	-2,369	-0,525	0,063	0,708	2,547	139,202	163
	100	-128,456	-2,263	-0,531	0,086	0,669	2,454	241,198	146
	300	-6684,183	-2,478	-0,625	0,027	0,621	2,474	350,945	143
(3,10)	30	-90,771	-1,036	-0,261	-0,003	0,272	1,019	45,391	152
	60	-46,581	-1,077	-0,241	0,030	0,321	1,161	709,828	159
	100	-365,379	-1,060	-0,241	0,031	0,309	1,124	165,198	149
	300	-427,864	-1,079	-0,274	0,011	0,284	1,104	68,682	154

Tabelle D.28: Schätzung für β nach zufälligem Entfernen eines Drittels der Messung

(a,b)	T	Min	W1	Box1	Median	Box2	W2	Max	Out
Kointegrierte Systeme									
(1,1)	30	-12,386	-1,647	-1,167	-0,991	-0,847	-0,379	108,276	120
	60	-3,067	-1,247	-1,065	-1,000	-0,944	-0,765	30,159	72
	100	-1,596	-1,140	-1,039	-1,002	-0,967	-0,860	-0,611	42
	300	-1,068	-1,043	-1,010	-1,000	-0,989	-0,956	-0,917	29
(2,3)	30	-11,535	-0,852	-0,714	-0,664	-0,622	-0,484	16,984	109
	60	-3,630	-0,733	-0,682	-0,665	-0,648	-0,598	-0,404	57
	100	-0,772	-0,705	-0,676	-0,666	-0,656	-0,627	-0,523	54
	300	-0,700	-0,679	-0,670	-0,667	-0,664	-0,654	-0,644	31
(3,10)	30	-11,521	-0,349	-0,313	-0,301	-0,289	-0,253	1,531	115
	60	-0,356	-0,318	-0,305	-0,300	-0,295	-0,282	-0,178	60
	100	-0,326	-0,312	-0,303	-0,300	-0,297	-0,288	-0,276	38
	300	-0,306	-0,303	-0,301	-0,300	-0,299	-0,297	-0,291	47
Stationäre Systeme									
(1,1)	30	-193,417	-3,631	-0,907	0,060	0,991	3,799	928,500	164
	60	-255,390	-3,909	-0,937	0,053	1,058	4,018	366,647	159
	100	-494,098	-3,683	-0,891	0,084	1,032	3,890	268,440	162
	300	-707,163	-3,931	-0,958	-0,013	1,082	4,043	218,069	143
(2,3)	30	-526,798	-3,070	-0,789	0,024	0,732	2,998	64,589	156
	60	-209,533	-2,865	-0,700	0,081	0,758	2,903	129,401	167
	100	-133,103	-2,982	-0,769	-0,003	0,709	2,897	152,295	153
	300	-369,936	-2,777	-0,736	0,016	0,653	2,715	138,413	153
(3,10)	30	-45,982	-1,217	-0,318	-0,012	0,292	1,203	213,845	171
	60	-59,359	-1,196	-0,292	0,035	0,316	1,221	165,611	161
	100	-230,625	-1,322	-0,355	0,005	0,293	1,263	62,817	164
	300	-218,446	-1,181	-0,280	0,036	0,331	1,235	29,927	158
Integrierte Systeme									
(1,1)	30	-453,224	-3,665	-0,900	-0,010	0,985	3,654	271,942	147
	60	-161,431	-3,548	-0,780	0,065	1,080	3,829	136,490	167
	100	-255,062	-3,690	-0,813	0,127	1,169	4,087	79,317	139
	300	-351,032	-3,866	-0,950	0,060	1,001	3,921	159,799	150
(2,3)	30	-43,413	-2,448	-0,564	0,046	0,698	2,591	571,181	129
	60	-832,645	-2,365	-0,511	0,058	0,728	2,584	360,271	150
	100	-76,746	-2,333	-0,539	0,090	0,662	2,441	181,181	141
	300	-605,356	-2,512	-0,622	0,038	0,649	2,555	310,058	141
(3,10)	30	-85,934	-1,067	-0,264	0,012	0,294	1,129	251,964	142
	60	-77,519	-1,091	-0,235	0,025	0,340	1,192	42,700	140
	100	-150,303	-1,100	-0,257	0,038	0,307	1,149	65,714	143
	300	-231,328	-1,120	-0,270	0,020	0,306	1,165	154,268	151

D.4 Erstellung der Testsets

Diese Hilfsfunktion erzeugt einen Random Walk-Prozess zu einer vorgegebenen Länge n . Dazu wird zunächst ein Weißes Rauschen \mathbf{eps} der Länge $n+1$ erzeugt. Durch sukzessives Summieren entsteht daraus ein Random Walk Prozess. Da man mindestens ein Lag zurück greifen muss, braucht man als Startwert $W[1] = 0$. Um diese Null nun wieder auszuschließen, gibt man nur die Werte $2:(n+1)$ aus. Dieser Prozess hat die Länge n und keinen Startwert Null.

```
> Fkt.RW <- function(n) {
+   eps <- rnorm(n + 1)
+   W <- c(0)
+   for (i in 2:(n + 1)) {
+     W[i] <- W[i - 1] + eps[i]
+   }
+   W[2:(n + 1)]
+ }
```

Fkt.testset

Mit dieser Funktion kann man sich eine bestimmte Anzahl \mathbf{nsimul} von kointegrierten bivariaten Systemen einer vorgegebenen Länge n erzeugen. Das kointegrierte System entsteht aus einem Common Trend. Die Faktoren \mathbf{a}, \mathbf{b} können beliebig vorgegeben werden.

Zunächst wird mit Hilfe der Funktion `Fkt.RW` eine $(n \times \mathbf{nsimul})$ -Matrix mit \mathbf{nsimul} Random Walks der Länge n erstellt. Dabei werden nur solche Varianten verwendet, die beim Stationaritätstest `adf.test` einen p-Wert größer 0.1 aufweisen. Obwohl wir aus der Konstruktion des Random Walks (Funktion `Fkt.RW`) wissen, dass er durch Summation aus Weißem Rauschen entstanden ist und somit sein Differenzenprozess stationär, er selbst also integriert der Ordnung 1 ist, soll diese Bedingung zusätzlich dazu dienen, dass wir „fast stationäre“ Prozesse als Common Trend ausschließen.

Jede Spalte der Matrix `W.mx` enthält einen Random Walk der Länge n . Dieser Random Walk dient als Common Trend in dem bivariaten System, das aus den Einzelreihen X_t, Y_t besteht⁵⁰.

$$X_t = aW_t + \epsilon_{1t}$$

$$Y_t = bW_t + \epsilon_{2t}$$

Die so entstandenen integrierten Prozesse X_t, Y_t werden ebenfalls in Matrizen `X.mx` und `Y.mx` gespeichert.

Nach Konstruktion enthalten beide Prozesse einen gemeinsamen Trend W_t , so dass die Li-

⁵⁰Für die Simulation werden die Einzelreihen der bivariaten Systeme mit X_t, Y_t statt mit Y_{1t}, Y_{2t} bezeichnet.

nearkombination

$$\begin{aligned} Z_t &= bX_t - aY_t \\ &= b(aW_t + \epsilon_{1t}) - a(bW_t + \epsilon_{2t}) \\ &= b\epsilon_{1t} + a\epsilon_{2t} \sim I(0) \end{aligned}$$

stationär ist.

Da es in der vorliegenden Arbeit üblich ist, die Kointegrationsbeziehung normiert zu betrachten, wird dies auch im Coding verfolgt⁵¹.

$$Z_t = X_t - \frac{a}{b}Y_t$$

Dieser Abweichungsprozess wird ebenfalls in der Spalte einer Matrix `Z.mx` gespeichert. Außerdem werden die Ergebnisse der Stationaritätstest (ADF-Test, Phillips-Perron) in den Vektoren `p.pp` und `p.adf` gespeichert.

Zum späteren Verwenden der Testdaten sind folgende Informationen nötig, die ebenfalls gespeichert werden

- `a`
- `b`
- `n`
- `nsimul`
- `epsX` - der Fehlerprozess für X_t
- `epsY` - der Fehlerprozess für Y_t

```
> Fkt.testset <- function(nsimul, n, a, b) {
+   W.mx <- matrix(nrow = n, ncol = nsimul)
+   for (i in 1:nsimul) {
+     p <- 0.01
+     while (p < 0.1) {
+       W <- Fkt.RW(n)
+       p <- adf.test(W)\$p.value
+       if (p > 0.1) {
+         W.mx[, i] <- W
+       }
+     }
+   }
+ }
```

⁵¹Natürlich hätte man dies bereits bei der Definition von X_t und Y_t berücksichtigen können und stets $a = 1$ setzen. In dieser Form ist die Common Trend Idee jedoch anschaulicher.

```
+ X.mx <- matrix(nrow = n, ncol = nsimul)
+ Y.mx <- matrix(nrow = n, ncol = nsimul)
+ Z.mx <- matrix(nrow = n, ncol = nsimul)
+ epsX.mx <- matrix(nrow = n, ncol = nsimul)
+ epsY.mx <- matrix(nrow = n, ncol = nsimul)
+ p.adf <- c()
+ p.pp <- c()
+ ls <- list()
+ for (i in 1:nsimul) {
+   epsX.mx[, i] <- rnorm(n)
+   epsY.mx[, i] <- rnorm(n)
+   X.mx[, i] <- a * W.mx[, i] + epsX.mx[, i]
+   Y.mx[, i] <- b * W.mx[, i] + epsY.mx[, i]
+   Z.mx[, i] <- X.mx[, i] - (a/b) * Y.mx[, i]
+   p.adf[i] <- adf.test(Z.mx[, i])\$p.value
+   p.pp[i] <- pp.test(Z.mx[, i])\$p.value
+ }
+ ls <- list(a = a, b = b, n = n, nsimul = nsimul, W = W.mx,
+   X = X.mx, Y = Y.mx, Z = Z.mx, epsX = epsX.mx, epsY = epsY.mx,
+   p.adf = p.adf, p.pp = p.pp)
+ save(ls, file = "Testset.Rdata")
+ ls
+ }
```

Fkt.testset10

Mit dieser Funktion wird ein stationäres Testset erstellt. Die Paare X_t, Y_t sind stationäre Reihen, die aus zwei stationären Prozessen hervorgehen. Die Multiplikation mit a bzw. b bewirkt, dass die beiden Prozesse gegebenenfalls unterschiedliche Varianzen haben.

```
> Fkt.testset_10 <- function(nsimul, n, a, b) {
+   X.mx <- matrix(nrow = n, ncol = nsimul)
+   Y.mx <- matrix(nrow = n, ncol = nsimul)
+   Z.mx <- matrix(nrow = n, ncol = nsimul)
+   epsX.mx <- matrix(nrow = n, ncol = nsimul)
+   epsY.mx <- matrix(nrow = n, ncol = nsimul)
+   p.adf <- c()
+   p.pp <- c()
+   ls <- list()
+   for (i in 1:nsimul) {
+     epsX.mx[, i] <- rnorm(n)
```

```

+     epsY.mx[, i] <- rnorm(n)
+     X.mx[, i] <- a * rnorm(n) + epsX.mx[, i]
+     Y.mx[, i] <- b * rnorm(n) + epsY.mx[, i]
+     Z.mx[, i] <- X.mx[, i] - (a/b) * Y.mx[, i]
+     p.adf[i] <- adf.test(Z.mx[, i])\$p.value
+     p.pp[i] <- pp.test(Z.mx[, i])\$p.value
+   }
+   ls <- list(a = a, b = b, n = n, nsimul = nsimul, X = X.mx,
+     Y = Y.mx, Z = Z.mx, epsX = epsX.mx, epsY = epsY.mx, p.adf = p.adf,
+     p.pp = p.pp)
+   save(ls, file = "Testset_I0.Rdata")
+   ls
+ }

```

Fkt.testsetI1

Mit dieser Funktion wird ein integriertes Vergleichsset erstellt, das nach Konstruktion nicht kointegriert sein soll. Dazu werden zunächst zwei Random Walks erzeugt, deren Differenz nicht als stationär eingeschätzt wird. Anschließend werden die Prozesse X_t, Y_t aus diesen Random Walks erzeugt. Im Gegensatz zum kointegrierten Fall, dieses Mal jedoch nicht mit Hilfe desselben Random Walks, sondern mit zwei verschiedenen.

```

> Fkt.testset_I1 <- function(nsimul, n, a, b) {
+   Wx.mx <- matrix(nrow = n, ncol = nsimul)
+   Wy.mx <- matrix(nrow = n, ncol = nsimul)
+   X.mx <- matrix(nrow = n, ncol = nsimul)
+   Y.mx <- matrix(nrow = n, ncol = nsimul)
+   Z.mx <- matrix(nrow = n, ncol = nsimul)
+   epsX.mx <- matrix(nrow = n, ncol = nsimul)
+   epsY.mx <- matrix(nrow = n, ncol = nsimul)
+   p.adf <- c()
+   p.pp <- c()
+   ls <- list()
+   for (i in 1:nsimul) {
+     epsX.mx[, i] <- rnorm(n)
+     epsY.mx[, i] <- rnorm(n)
+     p.x <- 0.01
+     while (p.x < 0.1) {
+       W <- Fkt.RW(n)
+       p.x <- adf.test(W)\$p.value
+       if (p.x > 0.1) {

```

```
+           Wx.mx[, i] <- W
+         }
+       }
+       p.y <- 0.01
+       while (p.y < 0.1) {
+         W <- Fkt.RW(n)
+         p <- pp.test(W)\$p.value
+         if (p > 0.1) {
+           p.y <- pp.test(Wx.mx[, i] - W)\$p.value
+           if (p.y > 0.1) {
+             Wy.mx[, i] <- W
+           }
+         }
+       }
+       X.mx[, i] <- a * Wx.mx[, i] + epsX.mx[, i]
+       Y.mx[, i] <- b * Wy.mx[, i] + epsY.mx[, i]
+       Z.mx[, i] <- X.mx[, i] - (a/b) * Y.mx[, i]
+       p.adf[i] <- adf.test(Z.mx[, i])\$p.value
+       p.pp[i] <- pp.test(Z.mx[, i])\$p.value
+     }
+     ls <- list(a = a, b = b, n = n, nsimul = nsimul, Wx = Wx.mx,
+              Wy = Wy.mx, X = X.mx, Y = Y.mx, Z = Z.mx, epsX = epsX.mx,
+              epsY = epsY.mx, p.adf = p.adf, p.pp = p.pp)
+     save(ls, file = "Testset_I1.Rdata")
+   ls
+ }
```

D.5 Allgemeine Auswertung

Im Allgemeinen geht es darum zu prüfen, welche Transformationen die Information „kointegriert“ in den Daten lassen und welche sie auslöschen. Um dies zu prüfen, braucht man eine Funktion, die auf ein (kointegriertes) Prozesssystem eine Datentransformation anwendet und anschließend testet, ob die Kointegrationsbeziehung weiterhin besteht. Da wir verschiedene Datentransformationen testen wollen, ist die Funktion `Fkt.Simulation.Kointegrationstest` so allgemein geschrieben, dass man ihr die Transformationsfunktion `func` mitgeben kann. Außerdem erwartet sie ein Testset in der Form, wie es in `Fkt.testset` erstellt wird, also als Liste.

Zunächst wird die vorgegebene Datentransformation auf die Prozesse X_t und Y_t angewendet. Da R Funktionen auf Matrizen anwenden kann, ist dieses Statement sehr elegant

```
X.func <- func(ls\$X)
Y.func <- func(ls\$Y)
```

Da wir den Anteil des Common Trends W_t in beiden Prozessen kennen, können wir auf Ko-integration testen, indem wir einen Prozess $Z.func_t$ mit denselben Koeffizienten berechnen $Z.func_t = X.func_t - \frac{a}{b}Y.func_t$ und auf Stationarität prüfen. Dazu werden zusätzlich die Ergebnisse des ADF- und PP-Tests gespeichert bzw. ausgegeben. Auf die transformierten Zeitreihen wird der Johantest mit `ca.jo` angewendet. In `L.max` werden die Werte der Statistiken aus den beiden Schritten des Max-Tests gespeichert. Die Schätzung des Kointegrationsvektors behält man in `beta`. Alles diese Werte werden gemeinsam zur Weiterverarbeitung in einer Liste `ls.test` ausgegeben.

```
> Fkt.Simulation.Kointegrationstest <- function(ls, func) {
+   X.func <- func(ls\$X)
+   Y.func <- func(ls\$Y)
+   p.adf <- c()
+   p.pp <- c()
+   L.max <- matrix(ncol = 2, nrow = ls\$nsimul)
+   lambda <- matrix(ncol = 2, nrow = ls\$nsimul)
+   beta <- matrix(ncol = 2, nrow = ls\$nsimul)
+   Z.func <- X.func - (ls\$a/ls\$b) * Y.func
+   for (i in 1:ls\$nsimul) {
+     p.adf[i] <- adf.test(Z.func[, i])\$p.value
+     p.pp[i] <- pp.test(Z.func[, i])\$p.value
+     ca.jo.i <- ca.jo(matrix(c(X.func[, i], Y.func[, i]), ncol = 2,
+       dimnames = list(NULL, c("X", "Y"))))
+     L.max[i, ] <- attr(ca.jo.i, "teststat")
+     beta[i, ] <- attr(ca.jo.i, "V")[, 1]
+   }
+   ls.test <- list(ls = ls, func = func, X.func = X.func, Y.func = Y.func,
+     Z.func = Z.func, p.adf = p.adf, p.pp = p.pp, L.max = L.max, beta = beta)
+   ls.test
+ }
```

Bei Transformationen, die den Wertebereich einschränken, wurden fast identische Systeme und solche mit einem fast konstanten Komponentenprozess ausgeschlossen.

Auswertungsskript

Nachdem eines der neun Testsets geladen wurde und die Funktion für die Transformation `func` festgelegt wurde, werden die Zeitreihen zunächst bei der entsprechenden Länge abgeschnitten und anschließend der Auswertungsfunktion `Fkt.Simulation.Kointegrationstest`

übergeben.

Hier der erste Auswertungsschritt für untransformierte Daten der Länge 10 dargestellt.

```
Fkt.ident <- function(X){
X
}
func <- Fkt.ident

mx <- matrix(c(1:14), nrow = 14, ncol=17,
  dimnames = list(c("10", "20", "30", "40", "50", "60", "70", "80",
    "90", "100", "150", "200", "300", "500"),
    c("adf<=0.01", "adf<=0.05", "adf<=0.1",
      "pp<=0.01", "pp<=0.05", "pp<=0.1",
      "L.max1.10", "L.max1.5", "L.max1.1",
      "L.max2.10", "L.max2.5", "L.max2.1",
      "beta.min", "beta.mean", "beta.max", "numbeta", "num")))

X.part <- ls\X[1:10,]
Y.part <- ls\Y[1:10,]
ls.part <- list(X = X.part, Y = Y.part, nsimul = ls\nsimul, a = ls$a, b = ls$b)
ls.test <- Fkt.Simulation.Kointegrationstest(ls.part, func)

Die Ausgabewerte dieser Funktion können nun weiterverarbeitet werden. Die Informationen für die weitere Auswertung werden in einer Zeile der Matrix mx gespeichert. Dies sind die Anzahl der Zeitreihe, deren Statistik in den beiden Unit-Root-Tests den jeweiligen kritischen Wert überschritten hat, gleiches für die Max-Statistiken, wobei die Werte im zweiten Schritt nur dann berücksichtigt werden, wenn der erste Wert signifikant war, da der zweite Schritt sonst nicht ausgeführt werden muss, und schließlich die Box-Statistiken sowie minimaler und maximaler Wert der Schätzung für beta. So können sie in ein Tabellenkalkulationsprogramm übertragen und zur Darstellung aufbereitet werden.

adf <- c(length(ls.test\p.adf[ls.test\p.adf <= 0.01]),
  length(ls.test\p.adf[ls.test\p.adf <= 0.05]),
  length(ls.test\p.adf[ls.test\p.adf <= 0.1]))
pp <- c(length(ls.test\p.pp[ls.test\p.pp <= 0.01]),
  length(ls.test\p.pp[ls.test\p.pp <= 0.05]),
  length(ls.test\p.pp[ls.test\p.pp <= 0.1]))

na.L.max.1 <- na.omit(ls.test\L.max[,1])
na.L.max.2 <- na.omit(ls.test\L.max[,2])
NA.num <- length(na.L.max.1)
```

```

na.L.max <- na.omit(ls.test\ $L.max)

L.max.10 <- na.L.max
L.max.5 <- na.L.max
L.max.1 <- na.L.max

for(i in 1:NA.num){
  if(L.max.10[i,2] < 12.91 ) L.max.10[i,1] <- NA
  else{ L.max.10[i,2] <- NA}
}

for(i in 1:NA.num){
  if(L.max.5[i,2] < 14.90 ) L.max.5[i,1] <- NA
  else{ L.max.5[i,2] <- NA}
}

for(i in 1:NA.num){
  if(L.max.1[i,2] < 19.19 ) L.max.1[i,1] <- NA
  else{ L.max.1[i,2] <- NA}
}

na.L.max.1.10 <- na.omit(L.max.10[,1])
na.L.max.1.5 <- na.omit(L.max.5[,1])
na.L.max.1.1 <- na.omit(L.max.1[,1])

na.L.max.2.10 <- na.omit(L.max.10[,2])
na.L.max.2.5 <- na.omit(L.max.5[,2])
na.L.max.2.1 <- na.omit(L.max.1[,2])

L.max.1 <- c(length(na.L.max.1.10[na.L.max.1.10 > 6.50]),
  length(na.L.max.1.5[na.L.max.1.5 > 8.18]),
  length(na.L.max.1.1[na.L.max.1.1 > 11.65]))
L.max.2 <- c(1000 - length(na.L.max.2.10), 1000 - length(na.L.max.2.5),
  1000 - length(na.L.max.2.1))

na.beta <- na.omit(ls.test\ $beta)
beta.stats <- boxplot.stats(ls.test\ $beta[,2])\ $stats
beta.out <- length(boxplot.stats(ls.test\ $beta[,2])\ $out)
beta.sum <- c(mean(na.beta[,2]), min(na.beta[,2]), beta.stats,
  max(na.beta[,2]), beta.out)

```

```
mx[1,] <- c(adf, pp, L.max.1, L.max.2, beta.sum, NA.num)
```

D.6 Funktionen zur Datentransformation

D.6.1 Intervallskalen

Die Funktionen für die Intervallskala teilen das Intervall $D = [min, max]$ in 10, 7, 5 bzw. 3 gleichgroße Bereiche auf und projizieren alle Werte aus einem Teilintervall auf einen Punkt der Skala. Die Teilintervalle sind gleich groß, so dass verschiedene Intervallskalen entstehen. Die Funktionen für 7, 5 und 3 Skalenstufen sind analog angelegt wie `Fkt.Anz.Punkte.10`.

```
> Fkt.Anz.Punkte.10 <- function(X, min, max) {
+   n <- length(X)
+   D <- max - min
+   d1 <- min + 0.1 * D
+   d2 <- min + 0.2 * D
+   d3 <- min + 0.3 * D
+   d4 <- min + 0.4 * D
+   d5 <- min + 0.5 * D
+   d6 <- min + 0.6 * D
+   d7 <- min + 0.7 * D
+   d8 <- min + 0.8 * D
+   d9 <- min + 0.9 * D
+   d10 <- min + 1 * D
+   X.func <- c()
+   for (i in 1:n) {
+     if (X[i] < min) { X.func[i] <- 0}
+     else { if (X[i] < d1) { X.func[i] <- 1}
+           else { if (X[i] < d2) { X.func[i] <- 2}
+                 else { if (X[i] < d3) { X.func[i] <- 3}
+                       else { if (X[i] < d4) { X.func[i] <- 4}
+                             else { if (X[i] < d5) { X.func[i] <- 5}
+                                   else { if (X[i] < d6) { X.func[i] <- 6}
+                                         else { if (X[i] < d7) { X.func[i] <- 7}
+                                               else { if (X[i] < d8) { X.func[i] <- 8}
+                                                     else { if (X[i] < d9) { X.func[i] <- 9}
+                                                           else { if (X[i] <= max) { X.func[i] <- 10}
+                                                                 }}}}}}}}}}}
+     }}}}}}}}}}}
+   X.func
+ }
```

D.6.2 Beschränkte Intervallskalen

Die Funktionen für eine beschränkte Intervallskala sind nach ähnlichem Schema aufgebaut, fassen jedoch einen größeren Randbereich zu einer Skalenstufe zusammen. Zu ihrer Anwendung wird in der Auswertungsfunktion für das jeweilige Systems stets der minimale und maximale Wert mitgegeben.

```
Fkt.Schranke.max <- function(X, min, max){
X.func <- c()
for(i in 1:n){
  if(X[i] < min){X.func[i] <- 0}
  else{ if(X[i] <d1){X.func[i] <- 1}
    else{ if(X[i] <d2){X.func[i] <- 2}
      else{ if(X[i] <d3){X.func[i] <- 3}
        else{ if(X[i] <d4){X.func[i] <- 4}
          else{ if(X[i] <d5){X.func[i] <- 5}
            else{ if(X[i] <d6){X.func[i] <- 6}
              else{ if(X[i] >=d6){X.func[i] <- 7}
                }}}}]}
X.func}
}
```

Die bis zur Hälfte von oben beschränkte Skala fasst die Stufen 5, 6, 7 noch zusammen. Für die von unten beschränkte Skala werden alle Werte $X[i] < d4$ auf die Stufe 4 abgebildet, alle anderen in regelmäßigen Abständen auf die Stufen 5 - 10. Für die von beiden Seiten beschränkte Skala werden die $X[i] < d3$ auf die Stufe 3 und die $X[i] >= d7$ auf die Stufe 8 abgebildet.

D.6.3 Ordinalskalen

Da auch die Ordinalskalen nach demselben Prinzip konstruiert wurden, werden hier nur jeweils die Grenzwerte der Stufenbereiche angegeben (siehe Tabelle D.29).

Tabelle D.29: Werte für die Ordinalskalen

	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Skala 1	0.25	0.40	0.475	0.525	0.60	0.75	1.00			
Skala 2	0.05	0.15	0.35	0.65	0.85	0.95	1.00			
Skala 3	0.01	0.04	0.09	0.16	0.25	0.36	0.49	0.64	0.81	1.00
Skala 4	0.20	0.25	0.30	0.425	0.50	0.585	0.70	0.75	0.85	1.00

D.6.4 Missing Values

Um Messwerte regelmäßig aus den Zeitreihen zu entfernen, wurde abhängig von der Länge bestimmt, dass jeder n .te Wert durch NA ersetzt wird. Die Konstruktion ist für 10, 7, 5 und 3 Punkte gleich und wird deshalb allgemein mit n dargestellt.

```
Fkt.Testset.NA.n <- function(X){
  num <- floor( length(X) / n )
  pos <- n * c(1:num)
  is.na(X) <- pos
  X
}
```

Wenn zufällig Messwerte entfernt werden sollen, muss zunächst eine Stichprobe mit festgelegtem Prozentsatz aus der Gesamtheit aller Messpunkte gezogen werden. Durch die Angabe von `set.seed` kann man sicherstellen, dass immer dieselben Stellen entfernt werden. Auch hier ist das Prinzip für alle Funktionen gleich.

```
Fkt.Testset.NA.prob.n <- function(X){
  all <- c(1:length(X))
  num <- floor( length(X) / n )
  set.seed(1)
  pos <- sample(all, num)
  is.na(X) <- pos
  X
}
```

Tabelle D.30: Position der zufällig entfernten Datenpunkte bei Entfernen eines Zehntels der Messwerte

Länge 30	8	11	17							
Länge 60	12	16	22	34	50	52				
Länge 100	6	20	27	37	57	58	62	86	89	97
Länge 300	4	18	35	52	60	74	80	93	105	107
	108	111	112	142	171	182	184	194	198	204
	219	221	237	262	266	270	278	281	290	296

Tabelle D.31: Position der zufällig entfernten Datenpunkte

bei Entfernen eines Siebtels der Messungen										
Länge 30	8	11	17	25						
Länge 60	12	16	22	34	36	50	52	57		
Länge 100	6	16	19	20	27	34	37	57	58	61
	62	86	89	97						
Länge 300	4	18	29	35	50	52	60	74	80	93
	105	107	108	111	112	131	133	142	162	168
	171	178	182	184	190	194	198	204	210	214
	219	221	237	262	266	270	278	281	282	286
	290	296								
bei Entfernen eines Fünftels der Messungen										
Länge 30	6	8	11	17	23	25				
Länge 60	4	9	11	12	16	22	33	34	36	50
	52	57								
Länge 100	6	16	19	20	27	32	34	37	43	57
	58	61	62	63	67	83	86	88	89	97
Länge 300	4	6	18	25	29	35	50	52	60	61
	74	78	80	93	99	105	107	108	109	111
	112	120	121	127	131	133	136	142	143	161
	162	168	171	174	178	182	184	185	190	194
	198	202	204	210	214	215	219	221	237	258
	262	266	270	278	281	282	286	290	291	296
bei Entfernen eines Drittels der Messungen										
Länge 30	2	6	8	11	14	16	17	23	25	27
Länge 60	4	9	11	12	16	19	22	23	32	33
	34	36	43	44	50	52	53	54	57	60
Länge 100	1	6	10	16	17	19	20	21	25	27
	28	29	32	34	37	42	43	51	57	58
	61	62	63	67	70	75	81	83	86	87
	88	89	97							
Länge 300	4	6	13	18	20	25	26	29	31	35
	44	50	51	52	53	60	61	71	74	76
	78	79	80	84	87	88	93	96	99	105
	107	108	109	110	111	112	120	121	122	127
	131	133	134	136	142	143	153	154	157	161
	162	164	168	171	173	174	178	179	182	184
	185	190	193	194	198	201	202	203	204	205
	210	213	214	215	216	219	220	221	237	239
	242	244	247	258	261	262	266	270	271	278
	281	282	286	290	291	293	294	296	299	300

Danksagung

An dieser Stelle möchte ich Herrn Prof. Dr. Joachim Werner für seine Unterstützung während der Promotion besonders danken. Ohne seinen Einsatz wäre ein Quereinstieg in die Psychologie nicht möglich gewesen.

Ein besonderer Dank gilt Esther Stroe-Kunold und Tetiana Stadnytska für die produktive Zusammenarbeit und ihr hilfreiches Feedback. In den gemeinsamen Sitzungen der Arbeitsgruppe „Zeitreihenanalyse“ habe ich viele wertvolle Anregungen zur verständlichen Darstellung häufig komplexer mathematischer Zusammenhänge und für die Psychologie relevante Anwendungen erhalten. Für ihren besonderen Beitrag bei diesen Treffen danke ich deshalb Astrid Milde und Simone Braun.

Janina Reichelt danke ich dafür, dass sie mit ihrem mathematischen Fachwissen die Arbeit auf ihre Korrektheit hin überprüft hat. Dietmar Henkes hat vor allem zum theoretischen Kapitel wichtige Hinweise geben können und Sarah Schrick hat den Text nicht nur auf Verständlichkeit, sondern auch auf Rechtschreibung kontrolliert. Bei Fragen zur Formatierung und einer abschließenden Aufbereitung der Darstellung war Alexander Heinle eine große Hilfe. Allen danke ich für die vielen Stunden des Korrekturlesens, die aufgeworfenen Frage und die hilfreichen Bemerkungen, durch die die Arbeit an Qualität gewonnen hat.

Außerdem möchte ich Herrn Prof. Dr. Andreas Voß für die Bereitschaft danken, für diese umfangreiche Arbeit das zweite Gutachten zu übernehmen.

Erklärung

**Erklärung gemäß § 8 Abs. 1 Buchst. b)
der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens-
und Empirische Kulturwissenschaften:**

Hiermit versichere ich, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel verwendet und Zitate kenntlich gemacht habe.

**Erklärung gemäß § 8 Abs. 1 Buchst. c)
der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens-
und Empirische Kulturwissenschaften:**

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form keiner anderen Prüfungsbehörde vorgelegt habe.

Heidelberg, den 7. März 2011

Antje Gruber

Literaturverzeichnis

- Ahtola, J. & Tiao, G. (1987). Distribution of least squares estimators of autoregressive parameters for process with complex roots on the unit circle. *Journal of Time Series Analysis*, 8, 1-14.
- Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 27, 217-235.
- Alho, J. M. & Spencer, B. D. (2005). *Statistical demography and forecasting*. New York; [Heidelberg]: Springer.
- Amelang, M. & Zielinski, W. (2004). *Psychologische Diagnostik und Intervention* (3., korr., aktualisierte u. überarb. Aufl.). Berlin; Heidelberg: Springer.
- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2. Aufl.). New York: Wiley.
- Andrés, A. R. & Halicioglu, F. (2010). Determinants of suicides in denmark: Evidence from time series data. *Health Policy*, 98 (2-3), 263 - 269.
- Anglin, J. M. (1993). *Vocabulary development: A morphological analysis* (Bd. 58) (Nr. 10). Monographs of the Society for Research in Child Development.
- Aoki, M. (1988). Nonstationarity, cointegration, and error correction in economic modeling: Editor's introduction and overview. *Journal of Economic Dynamics and Control*, 12 (2-3), 199 - 201.
- Assenmacher, W. (2002). *Einführung in die Ökonometrie* (6., vollst. überarb. und erw. Aufl.). München; Wien: Oldenbourg.
- Assenmacher, W. (2003). *Deskriptive Statistik* (3., verb. Aufl.). Berlin; Heidelberg: Springer.
- Avedyan, E. D., Tsytkin, Y. Z. & Bittanti, S. (1998). Kalman estimation with Brownian disturbances. *Journal of the Franklin Institute. B. Engineering and Applied Mathematics*, 335 (7), 1195-1205.
- Baillie, R. T. & Bollerslev, T. (1994). Cointegration, fractional cointegration, and exchange rate dynamics. *The Journal of Finance*, 49 (2), pp. 737-745.
- Baker, B. O., Hardyck, C. D. & Petrinovich, L. F. (1966). Weak Measurements vs. Strong Statistics: An Empirical Critique of S. Stevens' Proscriptions in Statistics. *Educational and Psychological Measurement*, 26 (2), 291-309.
- Banerjee, A. (Hrsg.). (2000). *Co-integration, error correction, and the econometric analysis of non-stationary data* (2. Aufl.). New York: Oxford University Press.
- Beard, R., Pentikainen, T. & Pesonen, E. (1977). *Risk theory*. London: Chapman and Hall.
- Beichelt, F. H. (Hrsg.). (2003). *Teubner-Taschenbuch der Stochastik* (1. Aufl.). Stuttgart: Teubner.
- Beiglböck, W. D. (1983). *Lineare Algebra*. Berlin; Heidelberg: Springer.
- Beran, J. (1998). *Statistics for long-memory processes* (2. Aufl.) (Nr. 61). Boca Raton: Chapman & Hall.

- Berk, L. E. (2007). *Development through the lifespan* (4. Aufl.). Boston, Mass.; Munich: Pearson/Allyn and Bacon.
- Beutelspacher, A. (2003). *Lineare Algebra*. Wiesbaden: Vieweg.
- Bhattacharya, R. N. & Waymire, E. C. (1990). *Stochastic processes with applications*. New York: Wiley.
- Bierens, H. J. (1997). Nonparametric cointegration analysis. *Journal of Econometrics*, 77 (2), 379 - 404.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Heidelberg: Springer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2., vollst. überarb. u. aktualisierte Aufl.). Berlin; Heidelberg: Springer.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4., überarb. Aufl.). Heidelberg: Springer Medizin Verlag.
- Bosch, S. (Hrsg.). (2009). *Algebra*. Berlin, Heidelberg: Springer.
- Bowerman, B. L. & O'Connell, R. T. (1993). *Forecasting and time series* (3. Aufl.). Belmont, Calif; London: Duxbury Press; Chapman and Hall.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time series analysis* (überarb. Aufl.). San Francisco, California: Holden-Day.
- Brähler, E. & Scheer, J. W. (1983). *Der Giessener Beschwerdebogen*. Bern: Huber.
- Bremmer, D. & Kesselring, R. (2004). Divorce and female labor force participation: evidence from time-series data and cointegration. *Atlantic Economic Journal*, 32, 175-190.
- Brockwell, P. J. & Davis, R. A. (1996). *Introduction to time series and forecasting*. New York; Berlin; Heidelberg: Springer.
- Brockwell, P. J. & Davis, R. A. (2006). *Time series* (2. Aufl.). New York, Heidelberg: Springer.
- Bronstein, I. N., Semendjaev, K. A., Musiol, G. & Mühlig, H. (Hrsg.). (2008). *Taschenbuch der Mathematik* (7., vollst. überarb. und erg. Aufl.). Frankfurt am Main: Teubner.
- Bühlmann, H. (1970). *Mathematical methods in risk theory*. Berlin: Springer.
- Bullinger, M. (1989). Psychological effects of air pollution on healthy residents: A time-series approach. *Journal of Environmental Psychology*, 9 (2), 103 - 118.
- Bundesagentur für Arbeit. (2011, Februar). *Zeitreihen*. Verfügbar unter <http://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Themen/Zeitreihen/Zeitreihen-Nav.html>
- Bundeskriminalamt. (2011, Februar). *Kriminalstatistik*. Verfügbar unter <http://www.bka.de/pks/zeitreihen/>
- Bürgy, R. (1998). *Die zeitreihenanalytische Modellierung psychologischer Regulationsvorgänge* (Nr. 32). Hamburg: Kovac.
- Cai, Y., Chou, R. Y. & Li, D. (2009). Explaining international stock correlations with cpi fluctuations and market volatility. *Journal of Banking and Finance*, 33 (11), 2026 - 2035.

- Canova, F. & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics*, 13 (3), pp. 237-252.
- Cattell, R. B. (Hrsg.). (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand MacNally.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Amsterdam: North-Holland.
- Cooney, J. B., Clarke, J. F. & Morris, G. L. (1986). Analysis of the physiological stress profile: The interrupted time-series design. *Biofeedback & Self Regulation*, 11 (3), 231 - 245.
- Costa, P. T. & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The neo personality inventory. *Psychological Assessment*, 4, 5-13.
- Darrat, A. & Al-Yousif, Y. (1999). On the long-run relationship between population and economic growth: Some time series evidence for developing countries. *Eastern Economic Journal*, 25, 301-306.
- Davison, M. L. & Sharma, A. R. (1988). Parametric statistics and levels of measurement. *Psychological Bulletin*, 104 (1), 137 - 144.
- DeJong, R. (2003). Logarithmic spurious regressions. *Economics Letters*, 81, 13-21.
- Delignières, D., Fortes, M. & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics, Psychology & Life Sciences*, 8, 479-510.
- Deng, A. (2010). Local power of consistent tests for serial correlation against the nearly integrated, nearly white noise process. *Economics Letters*, 107 (1), 22 - 25.
- Deutsche Bundesbank. (2011, Februar). *Zeitreihen*. Verfügbar unter http://www.bundesbank.de/statistik/statistik_zeitreihen.php
- Deutscher Wetterdienst. (2011, Februar). *Klimadaten*. Verfügbar unter <http://www.dwd.de/>
- Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74 (366, part 1), 427-431.
- Dickey, D. A., Hasza, D. P. & Fuller, W. A. (1984). Testing for unit roots in seasonal time series. *Journal of the American Statistical Association*, 79 (386), pp. 355-367.
- Diebold, F. & Nerlove, M. (1990). Unit roots in economic time series: A selective survey. *Advances in Econometrics*, 8, 3-69.
- Diggle, P. J. (1990). *Time series*. Oxford: Clarendon Press.
- Dueker, M. & Startz, R. (1998). Maximum-likelihood estimation of fractional cointegration with an application to U.S. and canadian bond rates. *Review of Economics and Statistics*, 80 (3), 420-426.
- Dugas, M. J., Francis, K. & Bouchard, S. (2009). Cognitive behavioural therapy and applied relaxation for generalized anxiety disorder: A time series analysis of change in worry and somatic anxiety. *Cognitive Behaviour Therapy*, 38 (1), 29 - 41.
- Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 28, 233-244.

- Econometrics, J. of. (2010). *Twenty years of cointegration* (Bd. 158) (Nr. 1).
- Elder, J. & Kennedy, P. (2001). Testing for unit roots: What should students be taught? *Journal of Economic Education*, 32, 137-146.
- Elkins, S. R. & Moore, T. M. (2011). A time-series study of the treatment of panic disorder. *Clinical Case Studies*, 10 (1), 3 - 22.
- Engle, R. F. (Hrsg.). (1991). *Long-run economic relationships - readings in cointegration*. London: Oxford University Press.
- Engle, R. F. & Granger, C. W. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55, 251-276.
- Engle, R. F., Granger, C. W. J., Hylleberg, S. & Lee, H. S. (1993). The Japanese consumption function. *Journal of Econometrics*, 55 (1-2), 275 - 298.
- Engle, R. F. & Yoo, B. S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35, 143-159.
- Fahrenberg, J. (1983). An evaluation of trait, state, and reaction aspects of activation processes. *Psychophysiology*, 20 (2), 188 - 195.
- Fahrenberg, J., Myrtek, M., Pawlik, K. & Perrez, M. (2007). Ambulantes Assessment - Verhalten im Alltagskontext erfassen. *Psychologische Rundschau*, 58, 12-23.
- Fahrmeir, L., Kaufmann, H. L. & Ost, F. (1981). *Stochastische Prozesse - eine Einführung in Theorie und Anwendungen*. München; Wien: Hanser.
- Fahrmeir, L., Kneib, T. & Lang, S. (2007). *Regression*. Berlin; Heidelberg: Springer.
- Fan, J. & Yao, Q. (2003). *Nonlinear time series*. New York: Springer.
- Farrell, S., Wagenmakers, E.-J. & Ratcliff, R. (2006). 1/f noise in human cognition: Is it ubiquitous, and what does it mean?. *Psychonomic Bulletin & Review*, 13 (4), 737 - 741.
- Feller, W. (1970). *An introduction to probability theory and its applications* (3. Aufl.). New York, NY: Wiley.
- Ferstl, E. (2006). Die Relevanz psychophysiologischer Reaktionen und Prozesse für das Erbringen von Höchstleistungen am Beispiel des konzertierenden Musikers. *Musik-, Tanz- und Kunsttherapie*, 17 (2), 68 - 75.
- Fischer, G. (2002). *Lineare Algebra*. Braunschweig; Wiesbaden: Vieweg.
- Forster, O. (2001). *Differential- und Integralrechnung einer Veränderlichen*. Braunschweig: Vieweg.
- Frank, D. C., Esper, J., Raible, C. C., Büntgen, U., Trouet, V., Stocker, B. et al. (2010). Ensemble reconstruction constraints on the global carbon cycle sensitivity to climate. *Nature*, 463, 527-530.
- Freitag, E. & Busam, R. (2006). *Funktionentheorie 1*. Berlin, Heidelberg: Springer.
- Friz, P. K. & Victoir, N. B. (2010). *Multidimensional stochastic processes as rough paths* (1. Aufl.). Cambridge: Cambridge University Press.
- Fuller, W. A. (1996). *Introduction to statistical time series* (2. Aufl.). New York: Wiley.
- Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, 87 (3), 564 - 567.

- Gil-Alana, L. A. (2008). Testing the null hypothesis of no cointegration against seasonal fractional cointegration. *Applied Mathematical Sciences. Journal for Theory and Applications*, 2 (5-8), 363–379.
- Gilden, D. L. (1997). Fluctuations in the time required for elementary decisions. *Psychological Science*, 8, 296-301.
- Gilden, D. L. (2001). Cognitive emissions of $1/f$ noise. *Psychological Review*, 108, 33-56.
- Girlich, H.-J. (1973). *Diskrete stochastische Entscheidungsprozesse und ihre Anwendung in der Lagerhaltung*. Leipzig: Teubner.
- Glass, G. V., Willson, V. L. & Gottman, J. M. (1975). *Design and analysis of time-series experiments*. Boulder, Colorado: Colorado Associated University Press.
- Gottman, J. M. (1981). *Time-series analysis*. Cambridge: Cambridge University Press.
- Gottschalk, A., Bauer, M. S. & Whybrow, P. C. (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry*, 52, 947-959.
- Granger, C. W. (1969). Investigating casual relations by econometric and cross-spectral models. *Econometrica*, 37, 424-438.
- Granger, C. W. & Newbold, P. (1974). Spurious regression in econometrics. *Journal of Econometrics*, 2, 111-120.
- Granger, C. W. & Newbold, P. (1977). *Forecasting economic time series*. New York: Academic Pr.
- Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14 (2), 227 - 238.
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16 (1), 121 - 130.
- Granger, C. W. J. & Lee, T.-H. (1990). Multicointegration. In *Co-integration, spurious regressions, and unit roots* (S. 71 - 84). U CA, San Diego: Advances in Econometrics series, vol. 8.
- Gruber, A. & Stadnytska, T. (2011). Detecting cointegration in interrupted time series. (Manuscript submitted for publication)
- Gujarati, D. N. (2003). *Basic econometrics* (4. Aufl.). Boston: McGraw-Hill.
- Haig, B. (2003). What is a spurious correlation? *Understanding Statistic*, 2, 125-132,.
- Haldrup, N. (1994). The asymptotics of single-equation cointegration regressions with I(1) and I(2) variables. *Journal of Economics*, 63, 153-181.
- Haßler, U. (2007). *Stochastische Integration und Zeitreihenmodellierung*. Berlin, Heidelberg: Springer.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.
- Hannan, E. & Quinn, B. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, B* 41, 190-195,.
- Hassler, U., Marmol, F. & Velasco, C. (2008). Fractional cointegration in the presence of linear trends. *Journal of Time Series Analysis*, 29 (6), 1088–1103.
- Hatanaka, M. (1996). *Time series based econometrics*. Oxford: Oxford University Press.

- Hecheltjen, K. G. & Mertesdorf, F. (1973). Development of a multidimensional questionnaire on mood. *Gruppendynamik*, 2, 110 - 122.
- Henss, R. (1989). Zur Vergleichbarkeit von Ratingskalen mit unterschiedlicher Kategorienzahl. *Psychologische Beiträge*, 31 (3-4), 264 - 284.
- Herrmann, F. (2005). *Integration und Volatilität bei Emerging Markets* (1. Aufl.). Wiesbaden: Dt. University-Verl.
- Hochstädter, D. (1969). *Stochastische Lagerhaltungsmodelle*. Berlin: Springer.
- Hull, J. W., Clarkin, J. F. & Alexopoulos, G. S. (1993). Time series analysis of intervention effects: Fluoxetine therapy as a case illustration. *Journal of Nervous and Mental Disease*, 181 (1), 48 - 53.
- Hylleberg, S., Engle, R. F., Granger, C. W. J. & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44 (1-2), 215 - 238.
- Hylleberg, S. & Mizon, G. E. (1989). Cointegration and error correction mechanisms. *The Economic Journal*, 99, 113-125.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231-254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59 (6), pp. 1551-1580.
- Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. New York: Oxford University Press.
- Johansen, S. & Juselius, K. (1990). Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52 (2), 169 - 210.
- Kailath, T. (1980). *Linear systems*. Englewood Cliffs (New Jersey): Prentice-Hall.
- Karlsen, H. A., Myklebust, T. & Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *The Annals of Statistics*, 35 (1), 252-299.
- Karnath, H.-O. H. (Hrsg.). (2006). *Neuropsychologie* (2., aktualisierte und erw. Aufl.). Heidelberg: Springer Medizin.
- Kasa, K. (1997). Consumption-based versus production-based models of international equity markets. *Journal of International Money and Finance*, 16 (5), 653 - 680.
- Kirchgässner, G. & Wolters, J. (2007). *Introduction to modern time series analysis*. Berlin; Heidelberg: Springer.
- Klenke, A. (2006). *Wahrscheinlichkeitstheorie*. Berlin; Heidelberg: Springer.
- Kratochwill, T. R. H. (Hrsg.). (1978). *Single subject research*. New York: Academic Press.
- Krauth, J. (1995). *Testkonstruktion und Testtheorie*. Weinheim: Beltz, PsychologieVerlags-Union.
- Krengel, U. (2003). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Braunschweig; Wiesbaden: Vieweg.
- Kuo, H.-H. (2006). *Introduction to stochastic integration*. New York, NY; [Heidelberg]: Springer.

- Kupfer, J., Brosig, B. & Brähler, E. (2005). A multivariate time-series approach to marital interaction. *GMS Psycho-Social-Medicine*, 2, 1 - 10.
- Lasswell, H. (1937). Veränderungen an einer Versuchsperson während einer kurzen Folge von psychoanalytischen Interviews. *Imago*, 23, 375-380.
- Lee, H. S. (1992). Maximum likelihood inference on cointegration and seasonal cointegration. *Journal of Econometrics*, 54 (1-3), 1 - 47.
- Lester, B. M., Hoffman, J. & Brazelton, T. B. (1985). The rhythmic structure of mother-infant interaction in term and preterm infants. *Child Development*, 56 (1), 15 - 27.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Aufl.). Weinheim: Beltz, Psychologie-Verl.-Union.
- Lin, Z. & Brannigan, A. (2003). Advances in the analysis of non-stationary time series: An illustration of cointegration and error correction methods in research on crime and immigration. *Quality & Quantity*, 37, 151-168.
- Link, A., Endt, P., Oeff, M. & Trahms, L. (2000). Zeitreihen-Analyse von Herzsignalen zur Detektion abnormer Myokard-Aktivierung. *Herzschrittmachertherapie und Elektrophysiologie*, 11 (4), 219-224.
- Lissitz, R. W. & Green, S. B. (1975). Effect of the number of scale points on reliability: A monte carlo approach. *Journal of Applied Psychology*, 60 (1), 10 - 13.
- Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, 8 (12), 750 - 751.
- Luiz, J.-M. (2001). Temporal association, the dynamics of crime, and their economic determinants: A time series econometric model of south africa. *Social Indicators Research*, 53, 33-61.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin; Heidelberg: Springer.
- Lütkepohl, H. & Kräzig, M. (Hrsg.). (2004). *Applied time series econometrics*. Cambridge: Cambridge University Press.
- MacKinnon, J. G. (1981). Critical values for cointegration tests. In *Long-run economic relationships* (Kap. 13). Oxford University Press.
- Maddala, G. S. & Kim, I.-M. (2007). *Unit roots, cointegration and structural change*. Cambridge: Cambridge University Press.
- Matell, M. S. & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, 56 (6), 506 - 509.
- McCleary, R. & Hay, R. A. (1980). *Applied time series analysis for the social sciences* (1. Aufl.). Beverly Hills: Sage.
- McNown, R. (2003). A cointegration model of age-specific fertility and female labor supply in the united states. *Southern Economic Journal*, 70, 344-358.
- Meissner, W. (1978). The conceptualization of marriage and family dynamics from a psychoanalytical perspective. In T. J. H. Paolino (Hrsg.), *Marriage and marital therapy* (S. 71 - 84). Advances in Econometrics series, vol. 8.

- Meyberg, K. (1999). *Differential- und Integralrechnung, Vektor- und Matrizenrechnung* (5., korr. Aufl.). Berlin; Heidelberg: Springer.
- Mezer, A., Yovel, Y., Pasternak, O., Gorfine, T. & Assaf, Y. (2009). Cluster analysis of resting-state fmri time series. *NeuroImage*, 45 (4), 1117 - 1125.
- Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, 100 (3), 398 - 407.
- Mills, T. C. & Markellos, R. N. (2008). *The econometric modelling of financial time series* (3. Aufl.). Cambridge: Cambridge University Press.
- Moseley, E. C. (1963). Psychodiagnosis on the basis of the holtzman inkblot technique. *Journal of Projective Techniques*, 27 (1), 86 - 91.
- Murray, M. P. (1994). A drunk and her dog: an illustration of cointegration and error correction. *The American Statistician*, 48, 37-39.
- Neidhardt, P. (1964). *Informationstheorie und automatische Informationsverarbeitung* (2., überarb. und erw. Aufl.). Stuttgart: Berliner Union.
- Nelson, C. R. & Kang, H. (1984). Pitfalls in the use of time as an explanatory variable in regression. *Journal of Business & Economic Statistics*, 2, 73-82.
- Nelson, C. R. & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series : Some evidence and implications. *Journal of Monetary Economics*, 10 (2), 139 - 162.
- Neusser, K. (2006). *Zeitreihenanalyse in den Wirtschaftswissenschaften*. Wiesbaden: Teubner.
- Newstead, S. E. & Arnold, J. (1989). The Effect of Response Format on Ratings of Teaching. *Educational and Psychological Measurement*, 49 (1), 33-43.
- Nobelstiftung. (2003, Oktober). *Informationen der Nobelstiftung zur Preisverleihung 2003 an Sir Clive Granger*. Verfügbar unter [http : //nobelprize.org/nobel_prizes/economics/laureates/2003](http://nobelprize.org/nobel_prizes/economics/laureates/2003)
- O'Brien, R. (1999). Measuring the convergence/divergence of „serious crime“ arrest rates for males and females: 1960-1995. *Journal of Quantitative Criminology*, 15, 97-114.
- Oerter, R. H. (Hrsg.). (2008). *Entwicklungspsychologie* (6., vollst. überarb. Aufl.). Weinheim; Basel: BeltzPVU.
- Orth, B. (1974). *Einführung in die Theorie des Messens*. Stuttgart ; Berlin ; Köln ; Mainz: Kohlhammer.
- Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, 54 (3), 461 - 472.
- Pearson, K. (1896). Mathematical contributions to the theory of evolution: on a form of spurious correlation which may arise when indices are used in the measurement of organisms. *Proceedings of the Royal Society London*, 60, 489-498.
- Perron, P. (1988). Trends and random walks in macroeconomic time series : Further evidence from a new approach. *Journal of Economic Dynamics and Control*, 12 (2-3), 297 - 332.

- Petermann, F. (1978). *Veränderungsmessung*. Stuttgart; Berlin; Köln; Mainz: Kohlhammer.
- Petermann, F. (Hrsg.). (1989). *Einzelfallanalyse* (2., völlig überarb. Aufl.). München: Oldenbourg.
- Peterson, B. S. & Leckman, J. F. (1998). The temporal dynamics of tics in Gilles de la Tourette syndrome. *Biological Psychiatry*, 44 (12), 1337 - 1348.
- Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R* (2. Aufl.). New York: Springer. Verfügbar unter <http://www.pfaffikus.de>
- Phillips, P. (1986). Understanding spurious regression in econometrics. *Journal of Econometrics*, 33, 311-340.
- Phillips, P. (1991). Optimal inference in cointegrated systems. *Econometrica*, 59, 283-306.
- Phillips, P. & Durlauf, S. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, 53, 473-495.
- Phillips, P. C. B. & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75 (2), pp. 335-346.
- R Development Core Team. (2008). R: A language and environment for statistical computing [Software-Handbuch]. Vienna, Austria. Verfügbar unter <http://www.R-project.org>
- Rinne, H. & Specht, K. (2002). *Zeitreihen - statistische Modellierung, Schätzung und Prognose*. München: Vahlen.
- Rohrman, B. M. (Hrsg.). (1978). *Fluglärm und seine Wirkung auf den Menschen*. Bern: Huber.
- Rozanov, J. A. (1975). *Stochastische Prozesse* (Nr. 28). Berlin: Akademie-Verlag.
- Sachs, L. & Hedderich, J. (2009). *Angewandte Statistik - Methodensammlung mit R* (13., aktualisierte und erw. Aufl.). Berlin; Heidelberg: Springer.
- Said, S. E. & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71 (3), pp. 599-607.
- Schaie, K. W. (1996). Intellectual development in adulthood. In *Handbook of the psychology of aging* (4. Aufl., S. 266 - 286). Academic Press.
- Schaie, K. W. & Willis, S. L. (2011, Februar). *The Seattle Longitudinal Study*. Verfügbar unter <http://geron.psu.edu/sls/>
- Schlittgen, R. & Streitberg, B. (1999). *Zeitreihenanalyse* (8., überarb. Aufl.). München; Wien: Oldenbourg.
- Schmitz, B. & Otto, J. (1984). Die Analyse von Migräne mit allgemeinspsychologischen Stresskonzepten. Zeitreihenanalysen für einen Einzelfall. *Archiv für Psychologie*, 136 (3), 211 - 234.
- Schröder, M. H. & Buscher, H. S. (Hrsg.). (2002). *Finanzmarkt-Ökonometrie*. Stuttgart: Schäffer-Poeschel.
- Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J. et al. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25 (2), 167 - 177.
- Schwarz, G. (1978). Estimating the dimensions of a model. *Annual Statistics*, 6, 461-464.
- Seber, G. A. F. & Lee, A. J. (2003). *Linear regression analysis* (2. Aufl.). New York: Wiley.

- Sharpley, C. F. & Alavosius, M. P. (1988). Autocorrelation in behavioral data: An alternative perspective. *Behavioral Assessment*, 10 (3), 243 - 251.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48 (1), 1-48.
- Smith, A. & Harrison, R. (1995). *A drunk, her dog and a boyfriend: an illustration of multiple cointegration and error correction* (Bericht). University of Canterbury, Christchurch, New Zealand.
- Smith, J. D., Handler, L. & Nash, M. R. (2010). Therapeutic assessment for preadolescent boys with oppositional defiant disorder: A replicated single-case time-series design. *Psychological Assessment*, 22 (3), 593 - 602.
- Southern California Earthquake Data Center. (2011, Februar). *Earthquake data center*. Verfügbar unter <http://www.data.scec.org/>
- Stadnytska, T. (2010). Applied multivariate time series modeling. (Manuscript submitted for publication)
- Stadnytska, T. & Gruber, A. (2011). Multivariate time series modeling of nonstationary data. (Manuscript submitted for publication)
- Statistisches Bundesamt Deutschland. (2011, Februar). *Destatis*. Verfügbar unter <http://www.destatis.de/>
- Stier, W. (2001). *Methoden der Zeitreihenanalyse*. Berlin; Heidelberg: Springer.
- Stine, W. W. (1989). Meaningful inference: The role of measurement in statistics. *Psychological Bulletin*, 105 (1), 147 - 155.
- Stock, J. H. & Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, 83 (404), 1097-1107.
- Stroe-Kunold, E. (2005). *Multivariate Analyse instationärer Zeitreihen*. Heidelberg.
- Stroe-Kunold, E. (2010). *Analyzing dynamic process systems*. Unveröffentlichte Dissertation, Heidelberg, University, Diss., 2010. Verfügbar unter <http://archiv.ub.uni-heidelberg.de/volltextserver/volltexte/2010/10670>
- Stroe-Kunold, E., Gruber, A., Stadnytska, T., Werner, J. & Brosig, B. (2010). *Cointegration methodology for psychological researchers: An introduction to the analysis of dynamic process systems*. (Manuscript submitted for publication)
- Stroe-Kunold, E., Stadnytska, T., Werner, J. & Braun, S. (2009). Estimating long-range dependence in time series: An evaluation of estimators implemented in R. *Behavior Research Methods*, 41, 909-923.
- Stroe-Kunold, E. & Werner, J. (2007). Sind psychologische Prozesse kointegriert? Standortbestimmung und Perspektiven der Kointegrationsmethodologie in der psychologischen Forschung. *Psychologische Rundschau*, 58, 225-237.
- Stroe-Kunold, E. & Werner, J. (2008). Modeling human dynamics by means of cointegration methodology. *Methodology*, 4, 113-131.
- Stroe-Kunold, E. & Werner, J. (2009). A drunk and her dog: A spurious relation? cointegration tests as instruments to detect spurious correlations between integrated time series. *Quality & Quantity*, 43, 913-940.

- Taylor, S. J. (2005). *Asset price dynamics, volatility, and prediction*. Princeton, N.J.: Princeton University Press.
- Terman, L. M. H. (Hrsg.). (1926). *Genetic studies of genius*. Stanford, Calif.: University Press.
- Thompson, B. (1998). *Canonical correlation analysis* (12. Aufl.). Newbury Park, Calif.: Sage.
- Trapletti, A. & Hornik, K. (2009). tseries: Time series analysis and computational finance [Software-Handbuch]. Verfügbar unter <http://CRAN.R-project.org/package=tseries>
- Tränkle, U. (1987). Auswirkungen der Gestaltung der Antwortskala auf quantitative Urteile. *Zeitschrift für Sozialpsychologie*, 18 (2), 88 - 99.
- Tschacher, W. & Ramseyer, F. (2009). Modeling psychotherapy process by time-series panel analysis (tspa). *Psychotherapy Research*, 19 (4-5), 469 - 481.
- Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. New York, NY: Springer.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass.: Addison-Wesley.
- Wagatha, M. (2005). *Kointegrationskonzepte für die Kreditrisikomodellierung* (1. Aufl.). Wiesbaden: Deutscher Universitätsverlag.
- Wagenmakers, E.-J., Farrell, S. & Ratcliff, R. (2004). Estimation and interpretation of $1/f^\alpha$ noise in human cognition. *Psychonomic Bulletin & Review*, 11, 579-615.
- Wagenmakers, E.-J., Farrell, S. & Ratcliff, R. (2005). Human cognition and a pile of sand: A discussion on serial correlations and self-organized criticality. *Journal of Experimental Psychology: General*, 134 (1), 108 - 116.
- Werner, J. (1997). *Lineare Statistik*. Weinheim: Beltz, Psychologie-Verlags-Union.
- Werner, J. (Hrsg.). (2005). *Zeitreihenanalysen mit Beispielen aus der Psychologie*. Berlin: Logos Berlin.
- Willi, J. (1982). *Couples in collusion*. New York: Jason Aronson.
- Willi, J. (1984). The concept of collusion: A combined systemic-psychodynamic approach to marital therapy. *Family Process*, 23 (2), 177-185.
- Witt, R. & Witte, A. (2000). Crime, prison, and female labor supply. *Journal of Quantitative Criminology*, 16, 69-85.
- Wold, H. O. A. (1938). *A study in the analysis of stationary time series*. Stockholm: Almqvist & Wiksell.
- Wolter, H. & Dahn, B. I. (2000). *Analysis individuell*. Berlin; Heidelberg: Springer.
- Wyatt, R. C. & Meyers, L. S. (1987). Psychometric Properties of Four 5-Point Likert Type Response Scales. *Educational and Psychological Measurement*, 47 (1), 27-35.
- Yoo, B. S. (1987). *Co-integrated time series: Structure, forecasting and testing*. Unveröffentlichte Dissertation, University of California, San Diego.
- Yule, G. (1926). Why do we sometimes get nonsense-correlation between time series? *Journal of the Royal Statistical Society*, 89, 1-64,.