RUPRECHT-KARLS –UNIVERSITÄT-HEIDELBERG

TIME SERIES RESEARCH IN PSYCHOLOGY: CONCEPTUAL AND METHODOLOGICAL ISSUES

TETIANA STADNYTSKA

Inaugural-Dissertation zur Erlangung des akademischen Grades „Doktor der Philosophie“ (Dr. Phil.) in der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Universität Heidelberg

Tag der Disputation: 14. 11. 2006

Dekan: Prof. Dr. Klaus Roth
Berater/ 1.Gutachter: Prof. Dr. Joachim Werner
2.Gutachter: Prof. Dr. Joachim Funke

DANKSAGUNG

Mein besonderer Dank gilt meinem wissenschaftlichen Betreuer, Herrn Prof. Dr. Joachim Werner, der mich für das methodische Thema begeisterte, diese Arbeit mit hilfreichen Diskussionen und konstruktiver Kritik begleitete und mir stets als interessierter Gesprächspartner zur Verfügung stand.

Für die Übernahme des zweiten Gutachtens bin ich dem Herrn Prof. Dr. Joachim Funke zu Dank verpflichtet.

Meinen Kommilitonen aus der „Zeitreihengruppe“ bin ich für die hervorragende Zusammenarbeit sehr verbunden. Besonders erwähnt seien an dieser Stelle Simone Braun und Esther Stroe-Kunold. Die zahleichen Diskussionen mit ihnen haben viel zum Fortschritt der Arbeit beigetragen.

TIME SERIES RESEARCH IN PSYCHOLOGY: CONTENTS AND METHODOLOGICAL ISSUES

The objectives of this paper are (1) demonstrate the superiority of the time series analysis over the traditional methods in dealing with dynamical phenomena; (2) discuss various possible research applications of time series procedures in psychology; and (3) solve some methodological problems occurring in applied settings. After a brief introduction into time- and frequency-domain analyses, a range of applications of time series procedures in psychology was discussed; theories and empirical studies from different fields of psychology employing time-series terminology and methods were presented. Three simulation studies designed to solve methodological problems typical for time series research in psychology, such as handling of instationary time series, identifying of appropriate dynamical models and reliable detection of long-range dependencies between successive observations in a series, represented the main field of the paper.

Keywords: time series, time-and frequency domain analyses, ARFIMA, unit root tests, automated methods for ARIMA model identification, 1/*f* noise

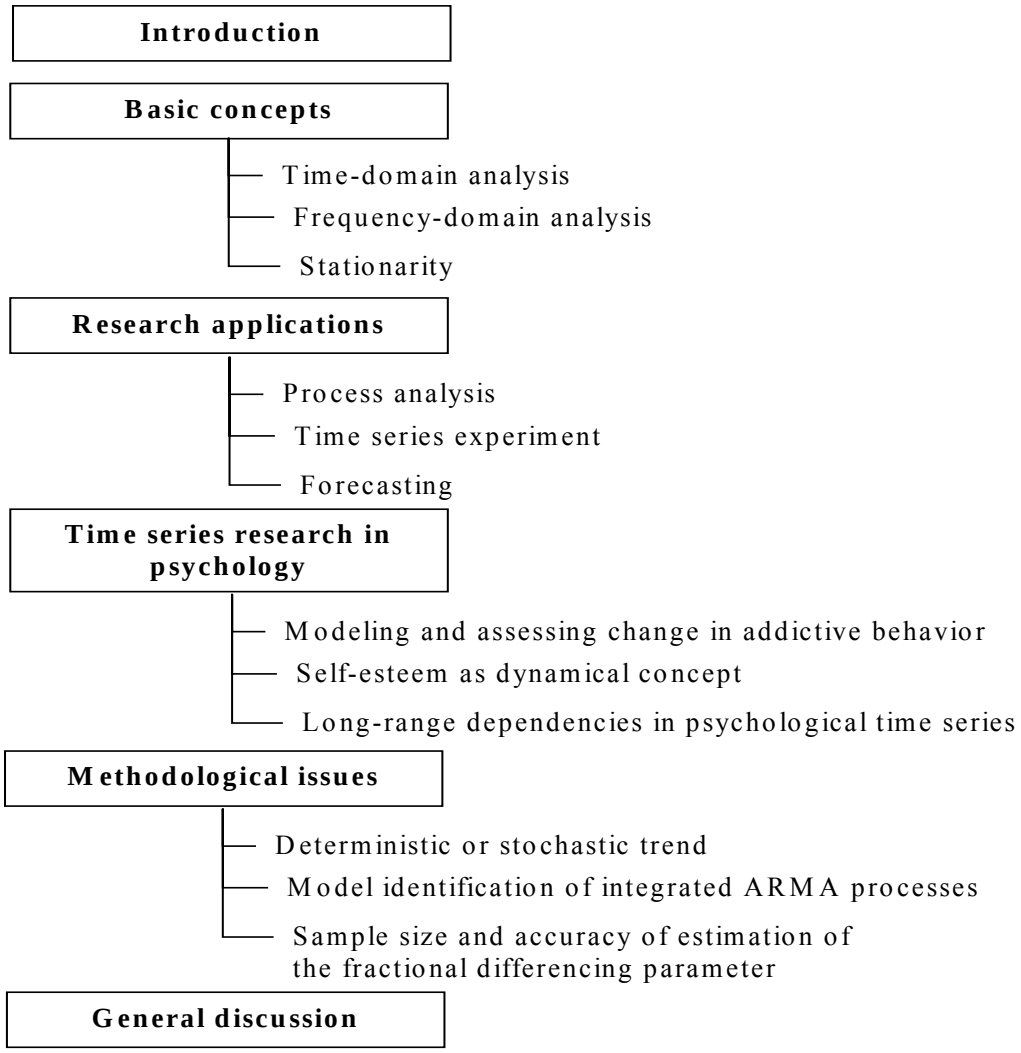
CONTENTS

1 INTRODUCTION

Time series analysis is widely used in econometrics, physic, astronomy, or seismology. To most psychologists, this methodology remains unfamiliar despite the fact that Glass, Willson, and Gottman (1975), McCleary and Hay (1980), and Gottman (1981) introduced time series procedures to social and behavioral sciences three decades ago. The standard research strategy in psychology consists in the attempt to infer general models from the average behavior of a large sample of individuals. As a result, employing classical statistics ignoring the dimension of time is characteristic of psychological research. This neglect of variation in time is rather surprising, since change, development, or growth represent typical signatures of most psychological phenomena. Traditionally, psychologists assess evolution or development through repeated measurements using mean and variance. By this procedure however, possible dependences between subsequent values remain indiscernible. Comparing means and standard deviations does not reveal the true nature of variability or change. In contrast, time series analysis is able to provide profound insight into properties of dynamical concepts. In the last few years, more and more researchers from different fields of psychology seem to recognize advantages of time series methods and increasingly apply these techniques in their empirical studies. The objectives of this thesis are to demonstrate the superiority of the time series analysis over the traditional methods in dealing with dynamical phenomena; discuss various possible research applications of time series procedures in psychology; and solve some methodological problems occurring in applied settings.

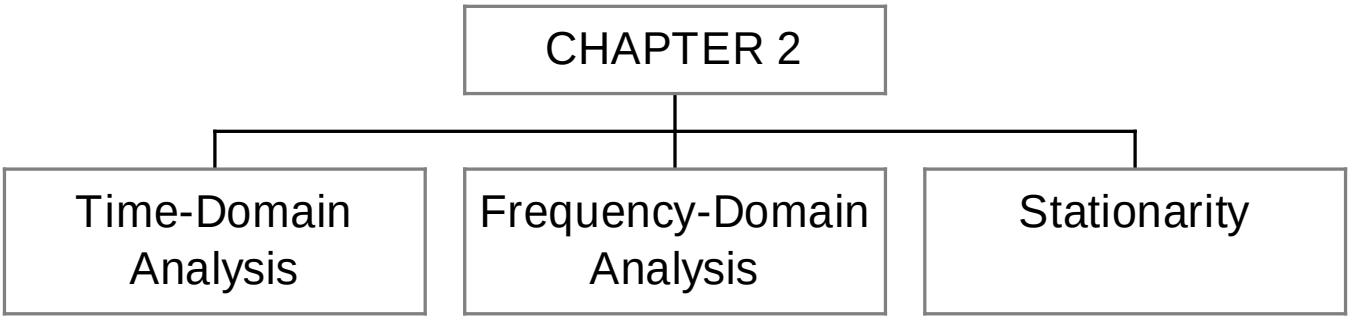
This paper is divided in six parts. Chapter 2 introduces two major approaches of the time series paradigm, time- and frequency-domains analyses, and describes their basic concepts. Chapter 3 discusses a range of applications of time series procedures in psychology,

such as process analysis, time series experiment, and forecasting. Chapter 4 focuses on theories and empirical studies from different fields of psychology employing time-series terminology and methods. Chapter 5 represents the main field of this thesis, introducing three simulation studies designed to solve methodological problems typical for time series research in psychology, such as handling of instationary time series; identifying of appropriate dynamical models; and reliable detection of long-range dependencies between successive observations in a series. General discussion with outlook and perspectives of the time series analysis in psychology completes the paper.



2 BASIC CONCEPTS

There are two major approaches in the study of time series processes, *time-domain* and *frequency-domain* analyses. Although time and frequency domains are mathematically equivalent, they examine time-series data from different perspectives and pursue different goals. In the time domain, the central concept is the memory of the series: to what extend is the present of the series predictable from its past. Memory is assessed by the so-called autocorrelation and the partial autocorrelation functions. The main goal of the frequency-domain analysis is to detect cycles in the data by means of spectral decomposition. The analysis consists of attempting to identify frequencies that explain variance in an observed time series. McCleary and Hay (1980) provide a comprehensive introduction to the time-domain analysis for social and behavioral scientists. Bloomfield (2000) gives a detailed description of the frequency-domain techniques. Warner (1998) introduces spectral analysis to the practicing researcher. For a detailed treatment and comparison of both time-domain and frequency-domain approaches, consult Gottman (1981). The objectives of this chapter are, based on the above-mentioned textbooks, to provide a brief introduction to the time- and frequency-domain analyses and to discuss the concept of *stationarity*.



2.1 Time-Domain Analysis

In the time domain, a visual plot of the data is usually the first step in the analysis of any time series. As Figure 2.1.1 illustrates, a time series is a sequence of values ordered by a time parameter (*t*). The primary goal of time series analysis is to infer from a sample of data points to the process that may have generated the sample. The terms *process* and *time ser*ies are equivalent to the concepts of *population* and *sample* in classical statistics. A process under study can consist of deterministic and stochastic components. Deterministic components are trends and deterministic cycles. A pure stochastic process is a collection of random variables ordered in time. Suppose the series in Figure 2.1.1 is a realization of a stochastic process, this implies that we observe realizations of 120 random variables ordered in time. In the majority of cases, time ordered variables can not be assumed independent, which results in the problem of correlated data. Within the scope of time series analysis, dependency is expressed by means of the autocorrelation and partial autocorrelation functions.

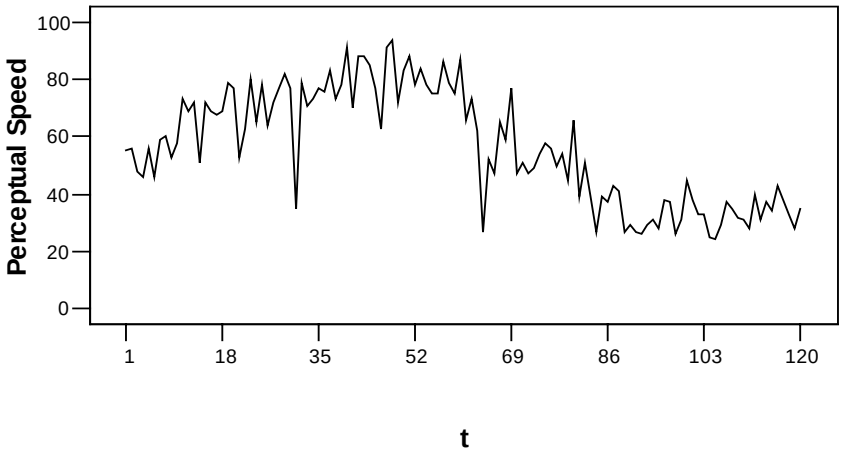


Figure 2.1.1. Perceptual speed of a schizophrenic patient for 120 successive days (Holtzman, 1963).

2.1.1 Autocorrelation and Partial Autocorrelation Functions

Kendall and Buckland (1971) define autocorrelation as correlation between members of series of observations ordered in time or in space. In the time-domain analysis, this implies correlation of a series with itself at different lags. The lag k autocorrelation is calculated as

$$r_k = \frac{\sum_{t=1}^{T-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{T}(Y_t - \overline{Y})^2}$$, where T is the length and $\overline{Y}$ is the mean of the series.

A plot of r_k against the lag length k is called the *correlogram* of the time series and gives its autocorrelation function. Since any observed series is a realization or a sample of some process, r_k is called the sample autocorrelation function. The population autocorrelation function (ACF) is defined as

ρ_k =covariance at lag k / variance

In addition to the ACF, another function, called the partial autocorrelation function (PACF), is employed to describe the memory of a series or a process. The PACF ρ_{kk} measures correlations between observations that are k time periods apart after controlling for correlations at intermediate lags. In other words, partial autocorrelation is the correlation between Y_t and Y_{t-k} after removing the effects of intermediate Y's. Analogous to the ACF, we can plot ρ_{kk} or its sample equivalent r_{kk} against k.

Within the scope of the time-domain analysis, the autocorrelation and partial autocorrelation functions are used to define various time-series models with different memory properties or dependency structures.

2.1.2 Time-Domain Models

Each stochastic time series can be described by means of three types of mathematical models: *autoregressive* (AR), *moving-average* (MA) and *integrated* (I). In an AR model, the value of the current observation depends on the values of the previous observations,

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t ,$$

where the magnitude of the dependence is quantified by ϕ, p specifies the order of the dependence und u_t is a sequence of purely independent and identically distributed random variables or innovations. A moving-average process is described by

$$Y_t = u_t - \theta_1 u_{t-1} - ... - \theta_q u_{t-q} .$$

Here performance at time t depends on a combination of the current and past error terms. A process containing both autoregressive and moving-average components is called *mixed*. An integrated process is represented by an equation

$$Y_t = Y_{t-1} + a_t ,$$

where the random part a_t can be generated by any ARMA process. The term "integrated" implies that the impact of the random component on the series does not dissipate over time. As a result, the process shows instability in level. That is why the integrated process with $a_t \sim$iid N $(0, \sigma^2)$ is also called *random walk*.

Therefore, each time series can be represented as an *Auto-Regressive Integrated Moving-Average* (ARIMA) model with three parameters p, d and q. The value of the autoregressive parameter p reflects how many preceding observations influence the current observation Y_t. The value of the moving average term q describes how many previous random shocks must be taken into account to capture the dependency present in the time series. The parameter d refers to the order of differencing that is necessary to stabilize the time series. Each ARIMA model can be defined through its theoretical ACF and PACF pattern.

Table 2.1.1 and Figure 2.1.2 show the ACF and PACF patterns of some common ARIMA processes.

Table 2.1.1. Theoretical ACF and PACF patterns.

Model	ACF	PACF
(0, 0, 0)	0	0
(p, 0, 0)	Decays exponentially or with damped sine wave or both	Significant spikes through lags p, 0 after p
(0, 0, q)	Significant spikes through lags q, 0 after q	Declines exponentially
(p, 0, q)	Declines exponentially	Declines exponentially
(0, d, 0)	Does not decay	Does not decay

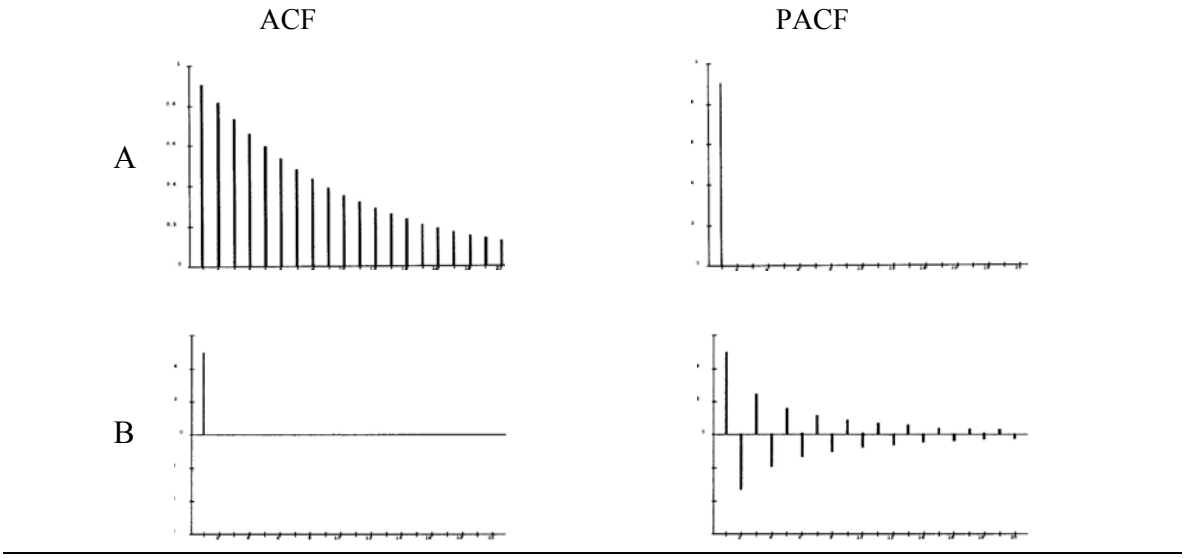


Figure 2.1.2. Theoretical ACF and PACF of selected stochastic processes: (A) AR(1) with ϕ=0.9, (B) MA(1) with θ=-0.9.

The main goal of the time-domain analysis is to provide an insight into properties of the underlying stochastic process of the series under study fitting an appropriate model to it. Once the process has been inferred, it can be used either to test some hypothesis about its generating mechanism, to forecast future values of the series, or to remove dependency from the data series so that it meets the assumptions of the general linear model for further statistical tests. There exist a number of methods for fitting suitable models to a given time series. One of the most widespread techniques is the Box-Jenkins methodology.

2.1.3 Box-Jenkins ARIMA Methodology

The ARIMA strategy proposed by Box and Jenkins (1970) is based on a three-step iterative cycle of *model identification*, *model estimation*, and *diagnostic checks* in model accuracy.

At the identification stage one chooses type and order of the model examining the behavior of the sample autocorrelation and the sample partial autocorrelation functions and comparing their shape and value with the theoretical ARIMA patterns.

At the estimation stage of the model building cycle, the parameters of the identified model are estimated. Estimates can be obtained through *conditional least squares* (CLS), *unconditional least squares* (ULS) or *full maximum likelihood* (ML) algorithms. In all three methods, the estimation procedure is based on the minimization of the difference between the observed time series and the model calculated for different values of the ARMA parameters; ϕ or θ values minimizing this difference serve as the estimators. The CLS method is based on the principle of simple least squares and called conditional because of the assumption that the initial unobserved error terms are equal to null (Box et al., 1994, p. 226). Least squares fitting turns out to be a maximum likelihood fit if distribution assumptions about the data are made. For example, if we assume normal errors in the ARMA equation. The ML algorithm maximizes the probability of the data for the given ARMA model. The maximum likelihood estimators for autoregressive and moving-average parameters are obtained by minimizing the so-called log likelihood function. In contrast to the CLS, the ULS approach includes the error term in the estimation function. That is why it is also referred to as the *exact least-squares* (ELS) method. The ULC technique is a compromise between CLS and ML methods where the unconditional sum of squares is minimized instead of the log likelihood function. Ansley and Newbold (1980, 1981) show that these three estimators are asymptotically equivalent and describe special cases, in which a particular method may be preferable. ML estimates are more expensive to compute than ULS and CLS estimates. In a recent study, Fang (2005)

compares the performance of the estimation algorithms for different models and parameter values and concludes that tests based on either ULS or ML yield more reliable inferences than CLS tests.

After the estimation stage, diagnostic checks are applied to determine whether or not the chosen model adequately represents the given set of data. The simplest check is to see if the residuals estimated from the fitted model are uncorrelated. Using the *Box-Pierce Q* or its variant the *Ljung-Box (LB) statistic*, we can test the joint hypothesis that all the autocorrelation functions up to certain lags are simultaneously equal to zero. Both Q and LB statistics are approximately chi-square distributed with *m*=lag length degrees of freedom. The LB statistics is more powerful in small-samples than the Q statistic.

Box et al. (1994), Bowerman and O'Connell (1993), Brockwell and Davis (2002) and Makridakis et al. (1998) provide a detailed treatment of the Box-Jenkins technique.

To summarize, in the time-domain analysis the primary goal is to infer from an observed time series to the process that may have generated this series. Stochastic processes can be described by means of ARIMA models. Different ARIMA models are distinguished by their memory properties or dependency structures assessed by the autocorrelation and partial autocorrelation functions. The Box-Jenkins methodology is a popular strategy for fitting ARIMA models to the data.

2.2 Frequency-Domain Analysis

The frequency-domain analysis aims to discover cyclicity in the time series data by means of spectral decomposition, in which a series is represented as a sum of independent cycles of different frequency. In time series analysis, the term *frequency* describes how rapidly things repeat themselves. Thus, there exist fast and slow frequencies. The frequency-domain analysis can be seen as a form of ANOVA where the overall variance of time series is divided into variance components due to independent cycles of different length. In frequency-domain, the variance is also called *power*. *Spectral density function* gives an amount of variance accounted for by each frequency we can measure. The theoretical spectral density function can be estimated from either the *periodogram* or the *power spectrum* depending on whether cycles composing the series are deterministic or stochastic. If the data are cyclic, there are a few so-called major frequencies that explain a great amount of the series variance (i.e., all series power is concentrated at one or some few frequencies). The major frequencies of deterministically periodic time series are fixed. Stochastically periodic series are characterized by random changes of major frequencies within certain bounds. For nonperiodic series, the variance is equally distributed across all possible frequencies. As Figure 2.2.1 shows, time series with different periodic properties cannot be distinguished visually.

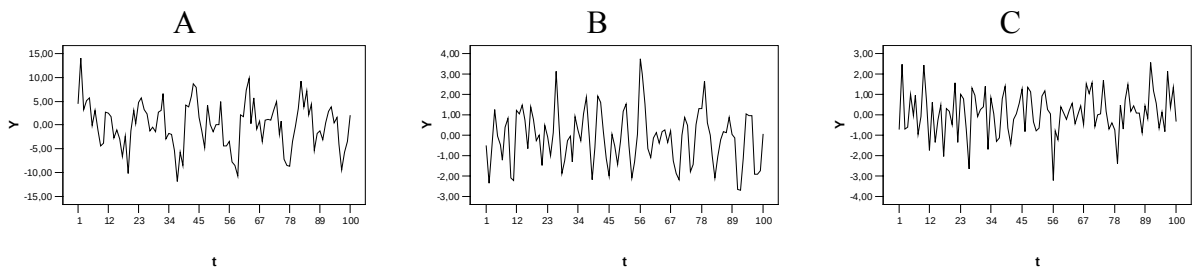


Figure 2.2.1. Time series with different periodicity: (A) deterministically periodic; (B) stochastically periodic; (C) nonperiodic.

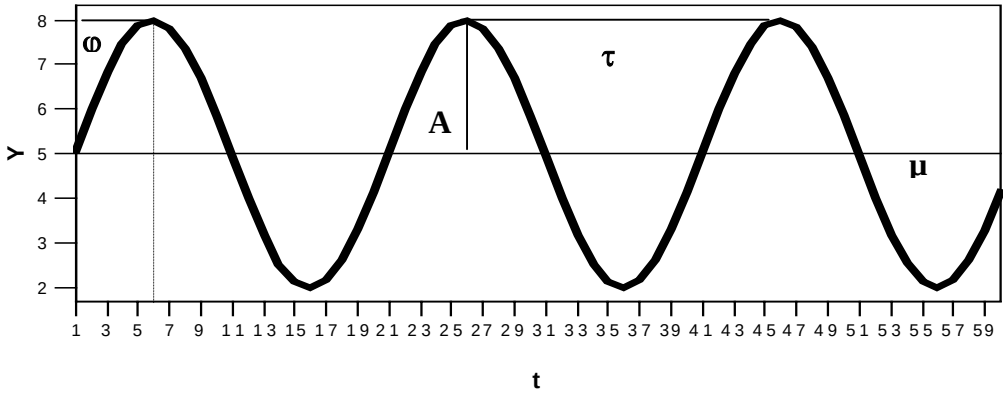


Figure 2.2.2. Graph of the function $Y_t = 5 + 3\sin(2\pi t / 20 + 6)$.

2.2.1 Modeling Repeating Phenomena

Periodic functions of the form $f(t) = f(\tau + t)$ are used to model repeating phenomena, where τ denotes the period or the length of a cycle. In the frequency-domain, the sine wave of the trigonometric sine or cosine functions serves as a model for various cycles. Different sine waves can be obtained by varying the following parameters: the mean (μ), the angular frequency (ω), the phase (φ) and the amplitude (A).

$$Y_t = \mu + A\sin(\omega t + \varphi) = \mu + A\sin(2\pi t / \tau + \varphi) = \mu + A\sin(2\pi f t + \varphi)$$

As Figure 2.2.2 shows, the period τ is the time from peak to peak. The angular frequency ω represents the number of complete cycles in 2π time units or in radians. The period and the angular frequency are linked by the equation $\omega = \frac{2\pi}{\tau}$. In time series analysis the frequency (f) is given by $f = \frac{1}{\tau} = \frac{\omega}{2\pi}$ and measures the number of cycles per unit time.

According to the *Fourier theorem* any time series with a given length can be approximated as a finite sum of sine waves of different frequencies. The series length (T) determines the number of the frequencies. For series with odd number of observations, there exist (T-1)/2 different frequencies: $f_j = \frac{j}{T}$, j = 1, 2, 3, …, $\frac{(T-1)}{2}$. These correspond to

cycles of period T, $T/2$, $T/3$,…, 2 time units and imply that the fastest frequency we can detect is $f = 1/2 = 0.5$ or $\omega = f 2\pi = 1/2 \cdot 2\pi = \pi$.

A sum of sine waves can be written in two ways[1]:

$$\sum_j A_j \sin(\omega_j t + \varphi_j) \qquad \text{or} \qquad \sum_j (a_j \cos \omega_j t + b_j \sin \omega_j t),$$

where $A_j = (a_j^2 + b_j^2)^{1/2}$ and $\sin^2 \varphi_j + \cos^2 \varphi_j = 1$. Sine and cosine functions of the same period are independent to each other. Thus, any time series[2] can be approximated as a set of orthogonal functions:

$$Y_t = \sum_j (a_j \cos \omega_j t + b_j \sin \omega_j t) + u_t, \text{ where } u_t \sim \text{iid } (0,1).$$

The unknown parameters of the series, the amplitude and the phase for each frequency w_j, can be calculated from the least-square estimates for a_j and b_j:

$$\hat{a}_j = \frac{2}{T} \sum_{t=0}^{T-1} Y_t \cos \omega_j t \qquad \hat{b}_j = \frac{2}{T} \sum_{t=0}^{T-1} Y_t \sin \omega_j t,$$

which are the covariances of the series with $cos\ \omega_j t$ and $sin\ \omega_j t$.

2.2.2 Detecting Deterministic Cycles: Periodogram

The periodogram is designed to fit a model to a time series that is the sum of deterministic waves plus noise (u_t). Figure 2.2.3 (A) illustrates the data generated by adding two sine waves with frequencies f_1=1/20=0.05 and f_2=1/10=0.1. Figure 2.2.3 (B) shows the same data where the deterministic cycles are masked by noise. The time series appears as in Figure 2.2.3 (C) if only 50 instead of 100 observations are available. The periodogram allows detecting hidden cycles.

---

[1] $A \sin(\omega + \varphi) = A(\cos \omega t \sin \varphi + \sin \omega t \cos \varphi) = a \cos \omega t + b \sin \omega t$.

[2] Note that time series with a zero or subtracted mean are used to simplify the equations.

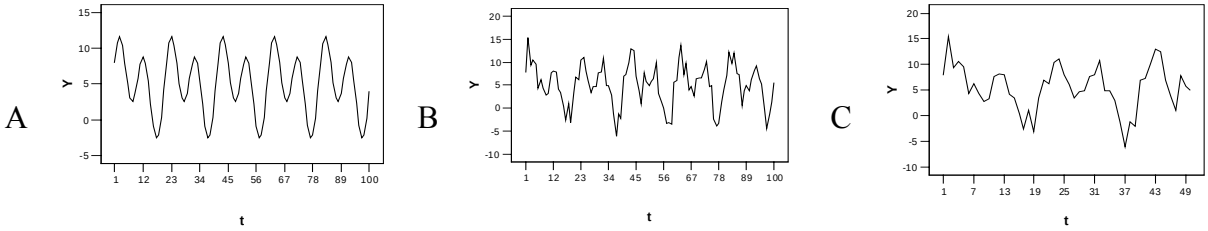


Figure 2.2.3. Example of a time series with two deterministic cycles τ_1=20 and τ_2=10 $Y_t = 5 + 3\sin(2\pi t / 20 + 6) + 5\sin(2\pi t / 10)$: (A) without noise, (B) with a noise part u_t, (C) T=50.

As Figure 2.2.4 illustrates, the periodogram gives the total sum of squares of the series distributed across either (T)/2+1 or (T-1)/2+1 different frequencies depending whether T is even or odd. The additional frequency is f=0, this is a frequency that never repeats which implies an infinite period. As noted previously, the fastest frequency we can detect is f=0.5. For these reasons, the frequencies in the periodogram always range between 0 and 0.5 (or 0 and π in radians). The total sum of squares of the series, also called the *energy,* is proportional to the number of observations ($SS = T\sigma^2$). The energy divided by T gives the *power* or the variance. For series containing deterministic cycles, the peaks of the periodogram occur at corresponding frequencies [see Figure 2.2.4 (A) and (B)]. The heights of the peaks are related to the amplitudes of the respective frequencies. As Figure 2.2.5 visualizes, the height of each peak is equal to half the square of the amplitude of each frequency (weighted with T). If the time series is a set of completely independent random numbers, also called *white noise*, then the total energy or the variance of the series is approximately equally distributed across all possible frequencies as in Figure 2.2.4 (C).

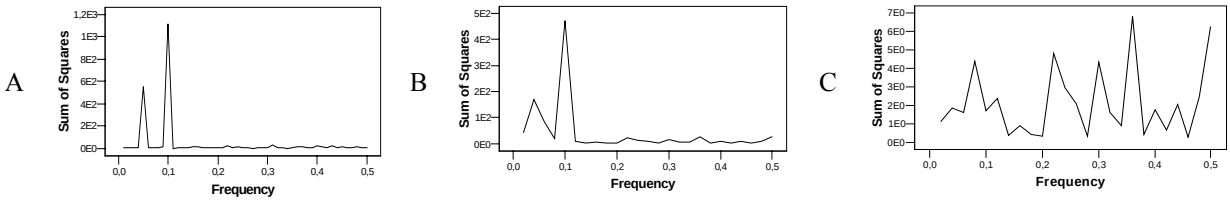
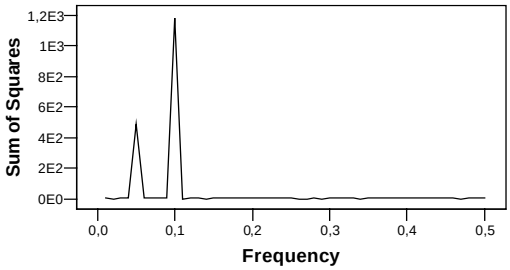


Figure 2.2.4. Periodograms: (A) T=100, $Y_t = 5 + 3\sin(2\pi t / 20 + 6) + 5\sin(2\pi t / 10) + u_t$; (B) T=50, $Y_t = 5 + 3\sin(2\pi t / 20 + 6) + 5\sin(2\pi t / 10) + u_t$; (C) u_t ~iid (0,1), T=50.

A 

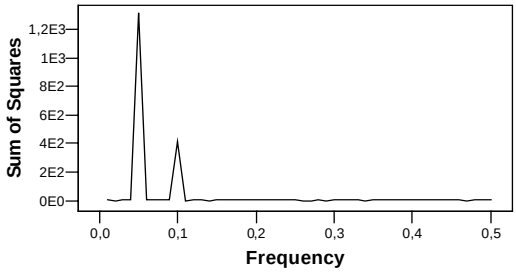
B 

Figure 2.2.5. Periodograms of series with T=100: (A) $Y_t = 3\sin(2\pi t/20) + 5\sin(2\pi t/10) + u_t$; (B) $Y_t = 5\sin(2\pi t/20) + 3\sin(2\pi t/10) + u_t$.

The sum of squares at each frequency can be obtained through

$$T\left(\frac{\hat{a}_j^2}{2} + \frac{\hat{b}_j^2}{2}\right) \text{ with } \hat{a}_j = \frac{2}{T}\sum_{t=0}^{T-1} Y_t \cos 2\pi f_j t \text{ and } \hat{b}_j = \frac{2}{T}\sum_{t=0}^{T-1} Y_t \sin 2\pi f_j t .$$

Because $\hat{a}$ and $\hat{b}$ at each frequency are independent and normal [and u_t ~iid(0,1)], their sum of squares is a chi-square with two degrees of freedom[1], under the null hypothesis a_j=b_j=0 at that frequency. Thus, the significance of periodogram values for each frequency f_j can be determined.

There exist different expressions for the periodogram[2], the most widely used are

$$I(f) = \frac{1}{4\pi}\sum_{t=0}^{T-1}(\hat{a}\cos 2\pi ft + \hat{b}\sin 2\pi ft)^2 \qquad \text{and} \qquad I(f) = \frac{1}{2\pi T}\left|\sum_{t=0}^{T-1} y_t e^{-i2\pi ft}\right|^2 .$$

The periodogram can also be derived from the autocovariance function

$$I(f) = \frac{1}{2\pi}\left(c_0 + 2\sum_{k=1}^{T-1} c_k \cos k(2\pi f)\right)$$ where c_0, …, c_{T-1} are estimates of the autocovariance.

The last equation is called the *Fourier transform* and serves as a connecting link between time- and frequency-domain representations of time series.

---

[1] *df*=2 except for the slowest and highest frequencies *f*=0 and *f*=0.5.

[2] In exponential form of the periodogram *e* is the basis for natural logarithms and $i = \sqrt{-1}$.

2.2.3 Detecting Probabilistic Cycles: Spectral Analysis

The objective of the spectral analysis is the identification of the frequencies that explain the variance in an observed time series. The spectral decomposition aims to find out how much variance is accounted for by each frequency in the series. The spectral density function gives the basic frequencies that compose the time series. The periodogram is one estimate of the spectral density where Fourier frequencies are used to approximate the real frequencies in the data. Periodogram analysis is a powerful tool for detecting a strong cyclic component. Deterministic cycles appear in the periodogram as clear peaks whose height increases with the sample size. For time series that are not strictly periodic, which implies random changes of frequencies, the periodogram analysis is associated with several problems.

The most serious failure of the periodogram is the large sampling error associated with the estimates of the sum of squares. Figure 2.2.6 shows the periodogram of white noise for different sample sizes. The theoretical spectral density of white noise is a straight line. As the figure illustrates, the variance of the sample periodogram estimates does not decrease as the number of observations increases. As a result, the peaks of the periodogram do not smooth out with increasing *T*.

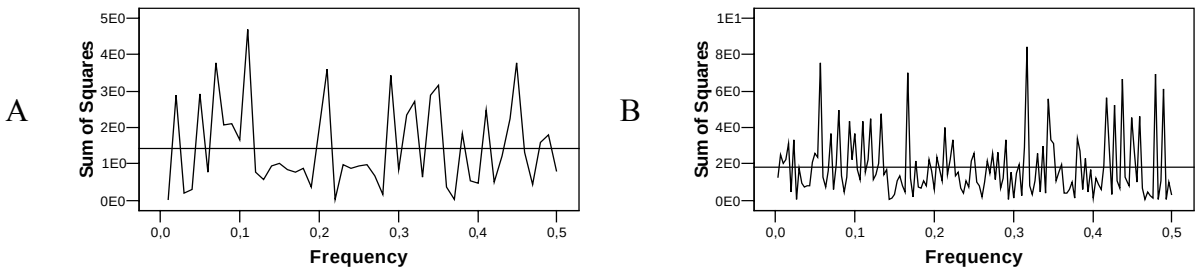
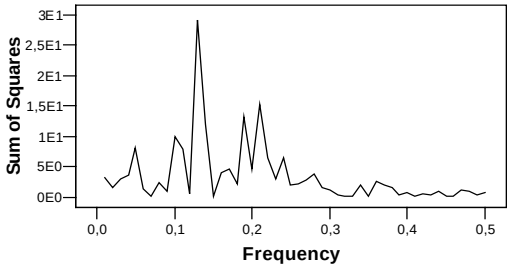


Figure 2.2.6. Spectral density function of white noise u_t ~iid(0,1) estimated from the periodogram, the flat line is the theoretical spectral density: (A) T=100; (B) T=300.

A



B

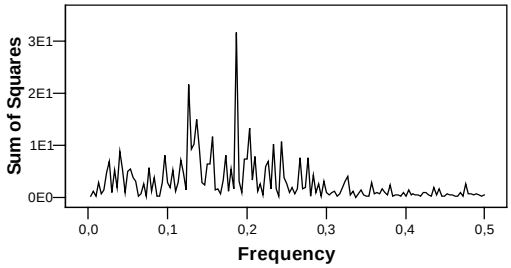


Figure 2.2.7. Periodograms of AR(2)-Process $Y_t = 0.6Y_{t-1} - 0.4Y_{t-2} + u_t$: (A) T=100; (B) T=300.

Figure 2.2.7 visualizes another problem of the periodogram analysis. Both periodograms A and B are estimates for the spectral density function of the AR(2)-process with ϕ_1=0.6 and ϕ_2=-0.4. For T=100, the periodogram result suggests one cyclic component at frequency f=0.13, which implies a cycle length of about 7.7 time units. Increasing the sample size to 300 observations gives two major frequencies 0.19 and 0.13 or the periods about 5.3 and 7.7 time units, respectively. The autocorrelation function of the series $Y_t = 0.6Y_{t-1} - 0.4Y_{t-2} + u_t$ shows that the underlying autoregressive process is nondeterministically periodic (see Figure 2.2.8). The theoretical spectral density of the AR(2) model can be computed as $p(f) = \frac{\sigma_u^2}{2\pi} \frac{1}{(1+\phi_1^2+\phi_2^2) - 2\phi_1(1-\phi_2)\cos 2\pi f - 2\phi_2 \cos 4\pi f}$, where $\frac{\sigma_u^2}{2\pi}$ is the spectral density of noise. The spectral density of an AR(2) process can have only one peak at $f = \frac{1}{2\pi}\arccos\left[\frac{|\phi_1|}{2(-\phi_2)^{1/2}}\right]$ (Gottman, 1981, p. 233). For the process with ϕ_1=0.6 and ϕ_2=-0.4, this implies the true frequency $f \approx 0.17$ or the period of about 5.9 time units. The above example is designed to illustrate two points. First, the cyclic component in the AR(2) series is not fixed which means random changes in frequencies within certain bounds. Second, the periodogram analysis breaks down when applied to time series with stochastic cycles. The power spectrum technique reduces the described problems.

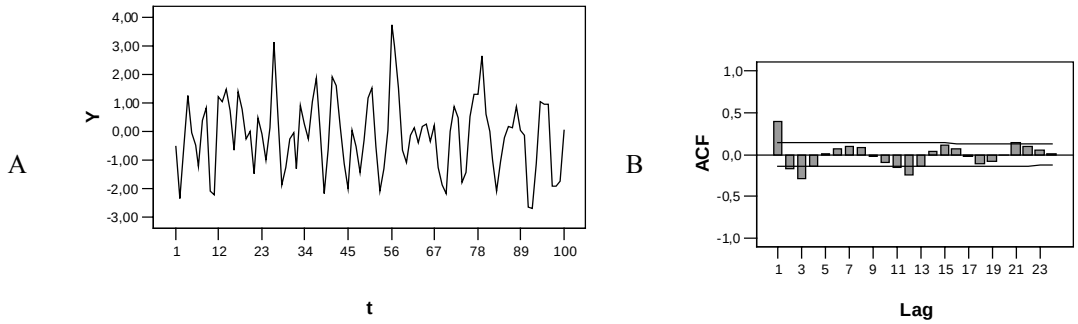


Figure 2.2.8. (A) Realization of the AR(2) process with ϕ_1=0.6 and ϕ_2=-0.4; (B) its ACF.

The *power spectrum* is employed to estimate the spectral density function of time series with stochastic cycles. The power spectrum is a smoothed version of the periodogram. In a spectrum the sum of squares of neighboring frequencies of a periodogram are averaged together to provide a more reliable estimate for this frequency band. Therefore, for nondeterministically periodic series, spectral density provides an estimate of the proportion of variance that is accounted for by a particular frequency band. There exist various smoothing functions, also called *windows,* which differ in the number of averaged frequencies and the weights used for computations. The total number of neighboring frequencies included in the weighted average is called the width of the window. The window is always symmetrical around some central frequency. That is why the width is 2*m*+1, where *m* is the number of terms on each symmetrical half of the window. The choice of *m* represents a basic dilemma in employing spectral windows. On the one hand, the sampling error decreases as *m* increases. On the other hand, using wide windows can prohibit detecting distinct cyclic components. The weights used for smoothing can be equal for all included frequencies such as in the *Daniell window*. Some procedures, as for example the *Tukey-Hamming smoothing*, give more weight to frequencies near the center of the window.

When a spectral window is used, significance tests for the spectral density estimates are performed employing "equivalent degrees of freedom" abbreviated *EDF*. In the Daniell case *EDF*=4*m*+2, this is the degrees of freedom of the sum of 2*m*+1 $\chi^2_{df=2}$ random variables.

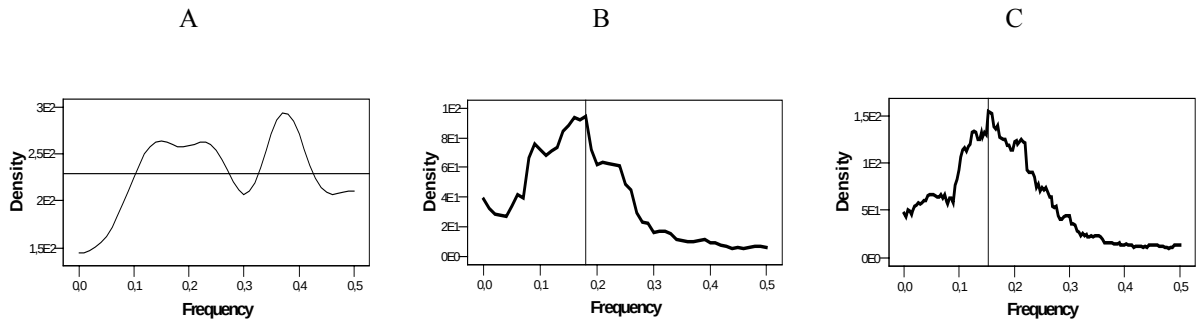


Figure 2.2.9. Spectral densities of (A) white noise series with T=100 estimated using Tukey-Hamming window with m=10, the flat line is the theoretical spectral density; (B) AR(2)-series with ϕ_1=0.6, ϕ_2=-0.4 and T=100 estimated using Daniell smoothing with m=5; (C) AR(2)-series with ϕ_1=0.6, ϕ_2=-0.4 and T=300 estimated using Daniell window with m=10.

Figure 2.2.9 shows that employing the power spectrum instead of the periodogram gives either a smoother estimate for the spectral density function of white noise or provides quite good estimates for the true stochastic frequency f=0.17 in the AR(2) series with ϕ_1=0.6 and ϕ_2=-0.4.

To summarize, the main goal of the frequency-domain analysis is the identification of major cyclic components that explain variance in an observed time series. The spectral density function gives these basic frequencies. For series containing deterministic cycles, the periodogram is an appropriate estimate of the spectral density where either (T)/2+1 or (T-1)/2+1 Fourier frequencies are used to approximate the real frequencies in the data. If time series are nondeterministically periodic, the power spectrum, which is a smoothed version of the periodogram, provides a more reliable estimate of the spectral density function.

2.3 Stationarity

The concept of stationarity is central for time-series modeling. Stationarity means stability of a process or a series. Assumption of stationarity is essential for inference and forecasting, because the absence of stability prohibits any reasonable prediction. Thus, stationarity conditions require the studied process to be stable over time in some statistical sense. One distinguishes between *strictly* and *weakly* stationary assumptions. A process is said to be strictly stationary if all its moments are invariant over time. Stationarity is called weak or second-order if just the first moments of the probability distribution of a process under study, mean and variance, are constant over time. Most nonstationary series in psychology have a time-varying mean or a time-varying variance or both (see Figure 2.3.1).

For further statistical analyses, nonstationary time series have to be transformed to make them stationary. (Another possibility for larger samples is to examine separate pieces of the studied series, each of which is stationary). The transformation method depends on the cause of nonstationarity. Special procedures called *unit root tests* were developed to test stationarity conditions (see Chapter 5.1 for details). Before formal testing, it is always advisable to plot the time series under study against time. As Figure 2.3.1 shows, such a plot can reveal the nature of the observed series. Instability in level can be also detected by means of the correlogram or the power spectrum.

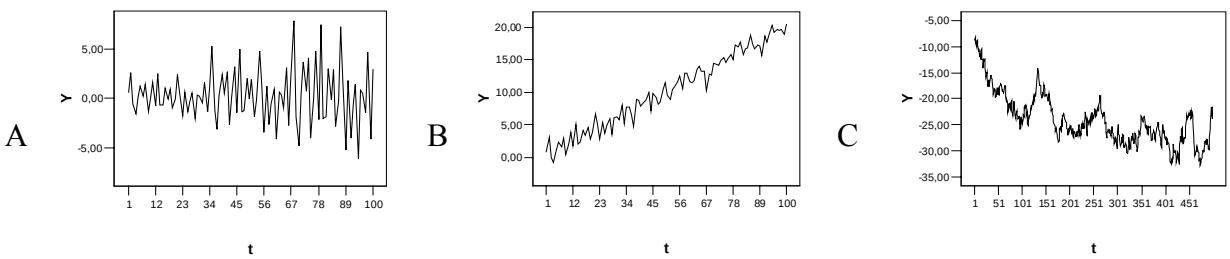


Figure 2.3.1. Nonstationary processes: (A) time-varying variance; (B) time-varying mean; (C) time-varying mean and variance.

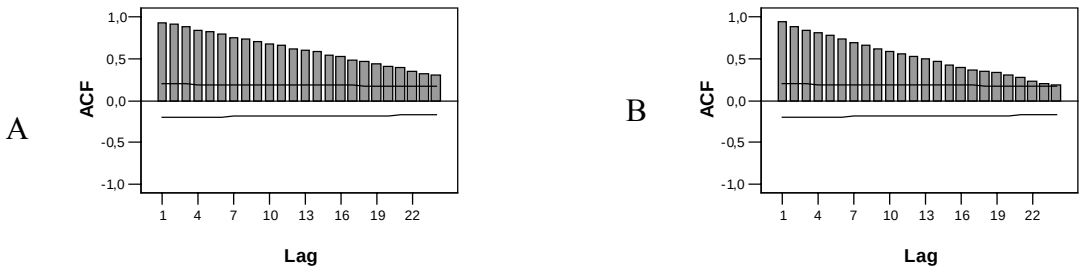


Figure 2.3.2. Autocorrelation functions of nonstationary time series: (A) time series with a linear trend $Y_t = 0.2t + u_t$, T=100; (B) integrated time series $Y_t = Y_{t-1} + u_t$, T=100.

As Figure 2.3.2 illustrates, the correlogram of nonstationary series is characterized by large or significant autocorrelation coefficients up to a lag of about one-quarter the length of the time series. Figure 2.3.3 shows that the power of series with a changing mean is concentrated at zero frequency implying an infinite period, which makes sense for series with a trend component. Further properties of instationary series are discussed in Chapter 3.

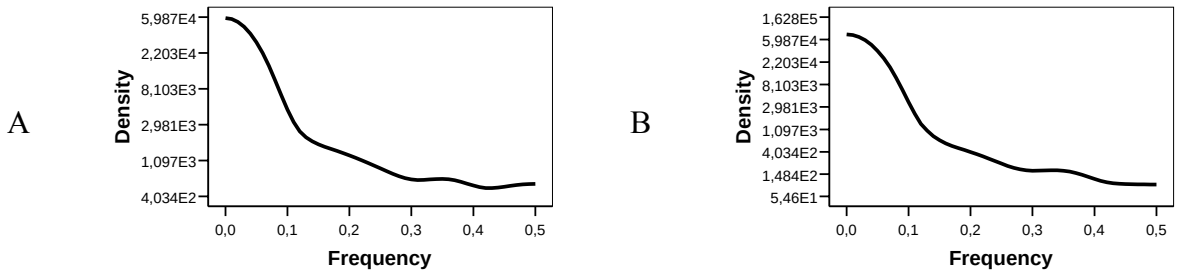
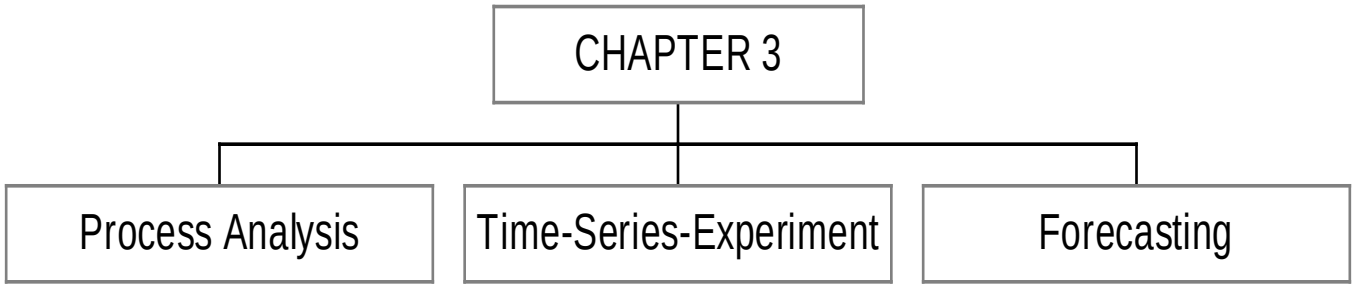


Figure 2.3.3. Spectral density of nonstationary time series obtained using Tukey-Hamming window with m=10: (A) time series with a linear trend $Y_t = 0.2t + u_t$, T=100; (B) integrated time series $Y_t = Y_{t-1} + u_t$, T=100.

3 RESEARCH APPLICATIONS

This chapter discusses a range of applications of time series analysis in psychology or other behavioral sciences. For longitudinal designs generating repeated observations on a single unit, time series analysis is the appropriate statistical methodology allowing either to understand the process under investigation or to measure different forms of intervention effects occurring in experimental designs. In the first place, time series analysis provides an insight into properties of the underlying stochastic process of the variable under study. Once the process has been inferred, it can be used either to test some hypothesis about its generating mechanism or to forecast future values of the series. Thus, there are three major research applications of time series analysis: (1) inference about the data generating process, (2) time-series-experiments, and (3) forecasting. Velicer and Fava (2003) discuss time series analysis as it is commonly employed in psychological research. Glass et al. (1975), Gottman (1981) and McCleary and Hay (1980) provide a comprehensive introduction into time-series experiments for social scientists. Bowerman and O'Connell (1993) and Makridakis et al. (1998) describe various forecasting techniques for the practicing researcher. The above-mentioned books build the basis for this chapter.



3.1 Process Analysis

Time series analysis always aims at a basic understanding of the process under investigation. Various processes can be described and categorized using the following dimensions: stability, memory, and dependency structure.

3.1.1 Stability

As noted earlier, within the scope of time series analysis stable processes are called stationary. Unstable or instationary processes are characterized by changes in mean or variance or covariance structure with historical time. The nature of instability can be *deterministic* or *stochastic*.

Time series with a deterministic change in level are said to have a *trend*. This usually implies a linear trend, but it could be quadratic or a polynomial of higher order. Trends can be detected and analyzed employing regression techniques where time serves as the explanatory variable. Figure 3.1.1 shows realizations of three processes with the same linear trend component but different disturbance terms a_t and their estimated regression lines. The estimated equations prove that the ordinary least square estimation method ensures good estimates of both the intercept and the slope for series with either independent or autocorrelated error terms[1].

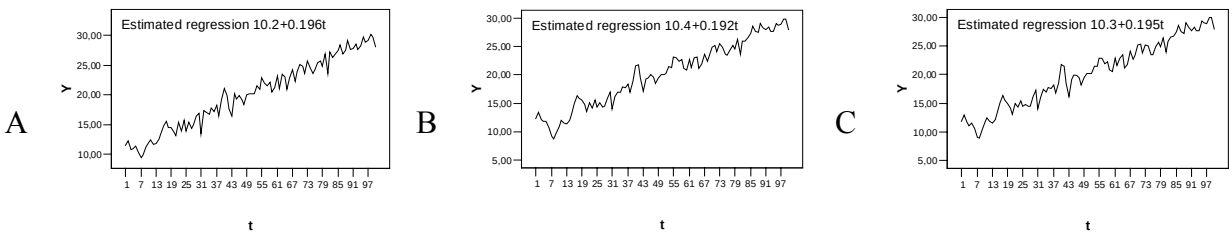


Figure 3.1.1. Time series with a linear trend $Y_t = 10 + 0.2t + a_t$ with T=100: (A) a_t ~iid(0,1); (B) $a_t = 0.5a_{t-1} + u_t$; (C) $a_t = u_t + 0.5u_{t-1}$ with u_t ~iid(0,1).

[1] Note that significance tests require uncorrelated error terms.

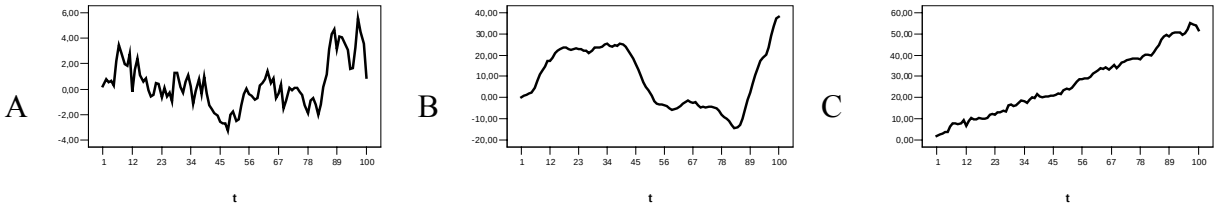


Figure 3.1.2. Integrated time series with *T*=100: (A) ARIMA (0, 1, 0); (B) ARIMA (0, 2, 0); (C) ARIMA (0, 1, 0) with the drift parameter α=0.5.

Integrated processes exhibit stochastic or random changes in level [see Figure 3.1.2 (A) and (B)]. The ARIMA parameter *d* denotes the order of integration. Most psychological time series are integrated of order 1. That is, they generally become stationary or stable after taking their first differences. As noted previously, the random walk series $Y_t = Y_{t-1} + u_t$ is an example of an integrated series. After differencing $\Delta Y = Y_t - Y_{t-1} = u_t$ it becomes white noise or a stationary series. Integrated processes can also exhibit a positive or a negative stochastic trend. In the time series literature, a stochastic trend is called *drift*. The name drift comes from the fact that if we modify the random walk equations as follows:

$$Y_t = \alpha + Y_{t-1} + u_t, \qquad \Delta Y = Y_t - Y_{t-1} = \alpha + u_t,$$

it shows that Y_t drifts upward or downward, depending on the parameter α being positive or negative [see Figure 3.1.2. (C)].

In the case of the deterministic trend, the deviations from the trend line are purely random and dissipate quickly; they do not contribute to the long run development of the series. In the case of the stochastic trend, on the other hand, the random component affects the long-run course of the series. In practice, we usually deal with relatively short realizations of processes, therefore it is difficult to distinguish visually between deterministic and stochastic trends, or even between stationary and instationary time series. The so-called unit root tests are designed to answer the question whether a process under study is stationary or not and to reveal the nature of instability for nonstationary series (see Chapter 5.1 for details).

3.1.2 Memory

In time-series analysis, memory means predictability from the past of a series to its current values. Memory properties of each process can be described employing the following dimensions: persistence and strength. Persistence describes whether a series has a long or a short memory. Strength quantifies the magnitude of the dependence and reflects which proportion of the previous component still affects the current observation.

In the time-domain, memory characteristics of a process are inferred from the autocorrelation function. Within the scope of ARIMA terminology, the parameter *d, p* and *q* reflect whether the process has a long or a short memory. The autoregressive and moving average prediction weights ϕ and θ quantify the memory strength[1]. In the case of the ARIMA (0, 0, 0) or white noise, we deal with a process without memory. This implies that knowledge of the value of a white noise series at any point in time does not improve prediction. Integrated processes (*p*, 1, *q*) can be represented as the sum of random terms $Y_t = \sum a_t$. As a result, we observe a strong persistence of random errors in the autocorrelation function, which means that the impact of a particular random term does not dissipate with the time. Thus, the ARIMA (*p*, 1, *q*) processes have an infinite memory. ARIMA (*p*, 0, *q*) models with small *p* or *q* possess a short memory and are predictable only from their immediate past. Their autocorrelations decay quickly as the number of intervening observations increases. If *p* or *q* are large, we deal with persistent autocorrelations implying a long memory of the data generating process or, in other words, statistical dependence between observations separated by a large number of time units. Long-range dependencies can be parsimoniously captured through the differencing parameter *d*, if we allow it to take any real value between 0 and 1. Time series with continuous

[1] In some cases, however, the value of ϕ determines the memory persistence. For instance, AR(1) model with ϕ=.99 can generate autocorrelations over hundreds of trials. (Recall that ARIMA (0, 1, 0) can be represented as a AR(1) model with ϕ=1).

d values are called fractionally integrated or ARFIMA. Processes with the finite long memory have *d* between 0 and 0.5 (see Chapter 5.3 for further details).

In the frequency domain, the power spectrum plotted on a log-log scale provides an insight into the memory properties of a series under study. For a purely random process, the power spectrum is an approximately straight line with a slope of zero. In random walk, also called *brown noise* because of its relation to Brownian motion, low frequency components predominate (see Chapter 2.3). Therefore its power falls off rapidly with increasing frequency. As a result, the log-log power spectra of the random walk series are *linear* and have negative slopes. Theoretically, random walks follow a power spectrum function $1/f^2$, which implies a straight line with a slope of -2. Generally, differencing increases the power spectrum slope by 2, and integrating decreases the slope by 2. Differencing of a random walk gives white noise. Thus, a power spectrum of white noise is $1/f^0$. Time series with a long memory are called *pink, flicker or burst noise*. Their log-log power spectrum follows a straight line with slope –1 implying a power spectrum function of $1/f$. Denoting the power spectrum function $1/f^a$, where *a* is called the *power exponent*, we obtain for processes without memory a=0, for process with an infinite memory a=2. In the case of a long memory, *a* can vary from 0.5 to 1.5. For short-memory processes, the log-log power spectrum in *not* a straight line because the linear relation between power and frequency breaks down at the low frequencies where random variation appears. As a result, a flat plateau (the zero slope of white noise) dominates low frequencies in spectral plots. Figure 3.1.3 shows theoretical power spectra for processes with different memory properties.

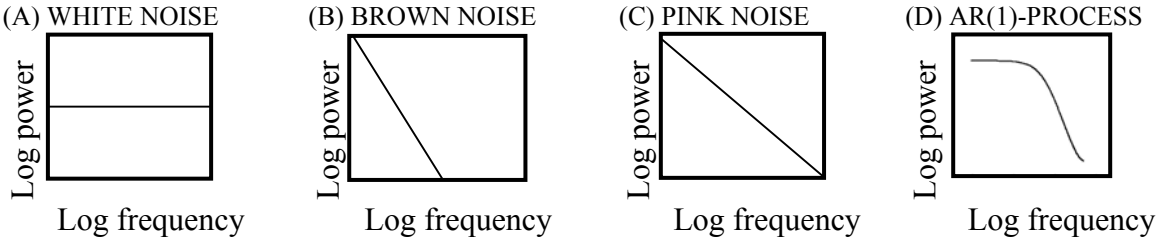


Figure 3.1.3. Theoretical power spectra: (A) $1/f^0$; (B) $1/f^2$; (C) $1/f$; (D) AR(1) with ϕ=0.7.

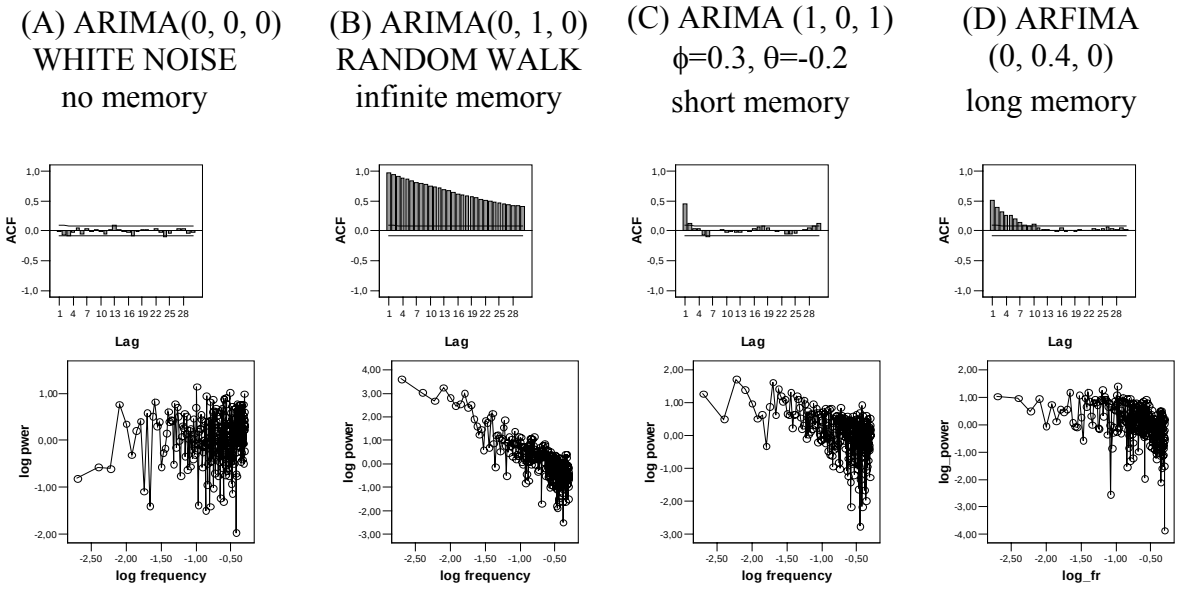


Figure 3.1.4. ACF and log-log power spectra (logarithms are base 10) for series with *T*=500.

For empirical time series, the regression slope of a linear function on a log-log plot can serve as an estimate of the power exponent. Beran (1994), Gilden (2001), Handel and Chung (1993), Kasdin (1995), and Pilgram and Kaplan (1998) provide a detailed description of the technique. Figure 3.1.4 depicts autocorrelation functions and log-log power spectra of empirical time series with different memory properties. Wagenmakers et al. (2004; 2005) and Thornton and Gilden (2005) point out that the log-log power spectrum of an ARMA(1, 1) series can resemble the spectrum of $1/f$ noise (compare Figure 3.1.4 C and D). This implies that, in frequency domain, short memory time series may mimic the statistical properties of the long memory process. To solve this problem, several procedures have been proposed for rigorous distinguishing of series with different memory properties. Wagenmakers et al. (2004) present a method in which the ARMA model, representing short-range processes, are competitively tested against the ARFIMA model, representing long-memory processes. Specifically, Wagenmakers et al. suggest determining the maximum likelihood of a time series under the ARMA and ARFIMA models, and then selecting the appropriate representation using Akaike's information criterion (AIC). Thornton and Gilden (2005) propose a spectral classifier procedure, in which the likelihood of a time series is estimated by

comparing its power spectrum with spectra of the competing memory models. Farrell et al. (2006; in press) conduct simulation experiments comparing the *spectral classifier method* of Thornton and Gilden with the *ARFIMA approach* of Wagenmakers et al.. Both procedures prove to be equally effective in discriminating between long- and short memory series. Despite the vast similarity of the approaches, Farrell et al. advocate the ARFIMA method on the following reasons. First, the ARFIMA modeling is commonly available in statistical packages such as Ox, R or S-Plus, whereas the spectral classifier is not freely available. Furthermore, the ARFIMA procedures are easily extended to different sample sizes and higher order models, whereas the spectral classifier requires generation of a new library for each new model or time series length. Finally, the theoretical properties of the ARFIMA models are well known, whereas those of the spectral classifiers have yet to be explored.

3.1.3 Dependency Structure

Time-series analysis distinguishes two types of processes with different dependency structures: *autoregressive* and *moving-average*. The MA(q) process is called q-dependent, because the dependency only lasts for q successive time units, and is zero thereafter. This is different from the autoregressive process, where the dependency, expressed as autocorrelations on different lags, decreases exponentially (see Figure 3.1.5). Autoregressive models are characteristic of systems containing internal temporal regularity, whereas moving average models are typical for unstable systems depending on external and occasional events.

Autoregressive and moving-average models are dual. This implies that most stationary time series can be approximated as either a MA or an AR model. The conditions needed to guarantee that we can transform from an autoregressive to a moving-average representation of a series are called *stationarity conditions*, and from MA to AR representations are called *invertibility conditions*. The AR(1) model can be written as an MA(∞) model under

stationarity conditions ($|\phi| < 1$). The MA(1) is the AR of infinite order under the invertibility condition that $|\theta| < 1$. Therefore, the same dependency structure can be approximated using different models. In practice, the parsimonious representation is preferred.

In time-series analysis, one distinguishes *positive* and *negative* dependencies. If autocorrelations are positive the ACF shows exponential decay as in Figure 3.1.5 (A). For a negative dependency, the autocorrelation function still decays, but it oscillates around zero [see Figure 3.1.5 (B)]. Time series with positive autocorrelation move upward or downward over extended time periods as in Figure 3.1.6 (A). Negative autocorrelations cause a constant up-and down movement such as in Figure 3.1.6 (B). Bürgy and Werner (2005) and Velicer and Fava (2003) discuss psychological models for positive and negative dependency structures (see also Chapter 4.1 for details).

Autoregressive and moving-average processes with positive and negative weights can also be distinguished employing the power spectrum function. The moving-average process is a linear filter of white noise (u_t). The spectral density of the MA(1) process is thus derived from the spectral density function of white noise ($\frac{\sigma_u^2}{2\pi}$):$1 + \theta^2 + 2\theta\cos 2\pi f)\frac{\sigma_u^2}{2\pi}$. As Figure 3.1.5 (C) and (D) shows, the negatively weighted moving average positively weights lower frequencies, the negatively weighted moving-average does exactly the opposite[1]. The spectral density function of an AR(p) process can be derived from the moving-average spectrum. For an AR(1)-process, the equation $\frac{\sigma^2}{2\pi}\frac{1}{1+\phi^2-2\phi\cos\lambda}$ with $\lambda = 2\pi f$ gives the theoretical power spectrum function. As Figure 3.1.5 (A) and (B) illustrates, the shape of this function is similar to the spectrum of the MA(1) process but much steeper. For further details, consult Gottman (1981), pp. 228-235.

[1] Note that here moving-average equation is written as $Y_t = u_t - \theta_1 u_{t-1} - \ldots - \theta_q u_{t-q}$.

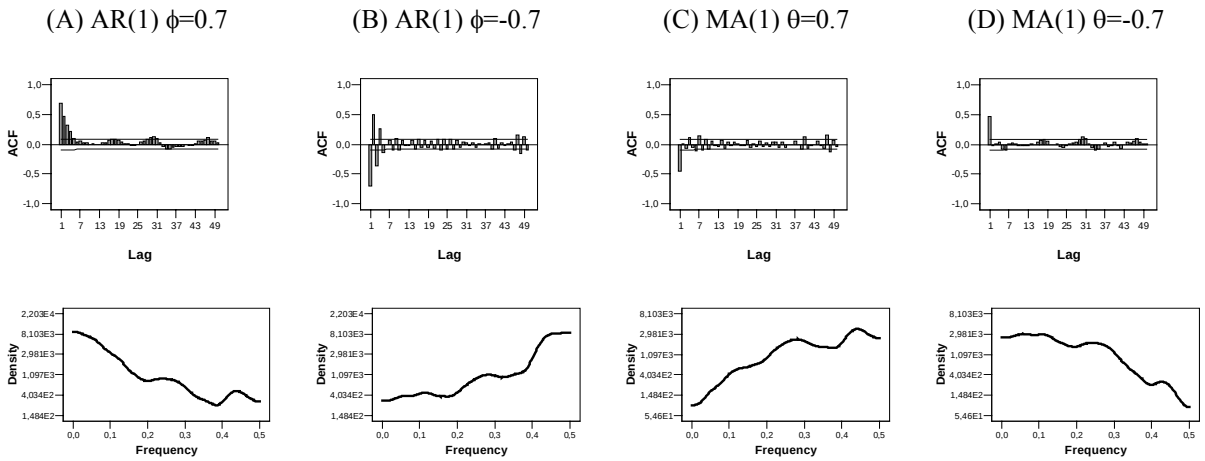


Figure 3.1.5. ACF and power spectra (Tukey-Hamming window with *m*=20) for series with *T*=500.

To summarize, processes are usually described employing three dimensions: stability, memory, and dependency. Stable stochastic processes are called stationary. One distinguishes between stochastic and deterministic causes of instability or instationarity. Memory properties of time series can be captured using either the autocorrelation function of the log-log power spectrum. Autoregressive models are characteristic of systems containing internal temporal regularity, whereas moving average models are typical for unstable systems depending on external and occasional events. Both AR(1) and MA(∞) models are appropriate approximations of high dependency stationary processes. Positive and negative dependencies have different impact on time series development.

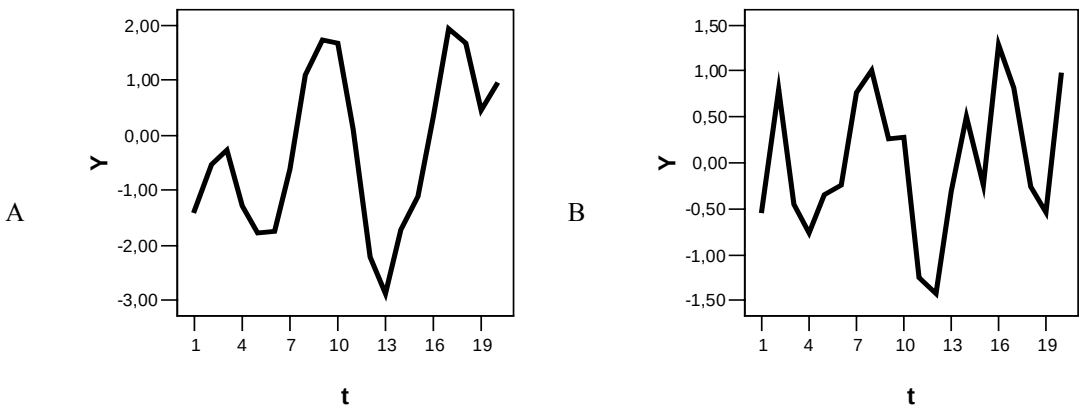


Figure 3.1.6. AR(1) series with *T*=20: (A) ϕ=0.7; (B) ϕ=-0.7.

3.2 Time-Series Experiment

One of the enduring tasks for social and behavioral scientists is to measure the effects of some planned or not planned intervention occurring in experiments or quasi-experiments. Many intervention effects can be detected only in longitudinal designs where a dependent variable is studied over time. According to Glass et al. (1975), the most important advantage of the time-series experiment over other designs is the possibility to account for *different* intervention effects. Figure 3.2.1 illustrates that the time-series design actually "offers a unique perspective on the evaluation of intervention (or "treatment") effects" (Glass et al., 1975, p. 4).

In longitudinal designs, we deal with repeated measurements on a single research unit producing autocorrelated data. Crosbie (1993) compares different possibilities to measure intervention effects in longitudinal designs: visual inference, ANOVA, Box-Jenkins intervention analysis and Interrupted Time-Series Experiment (ITSE).

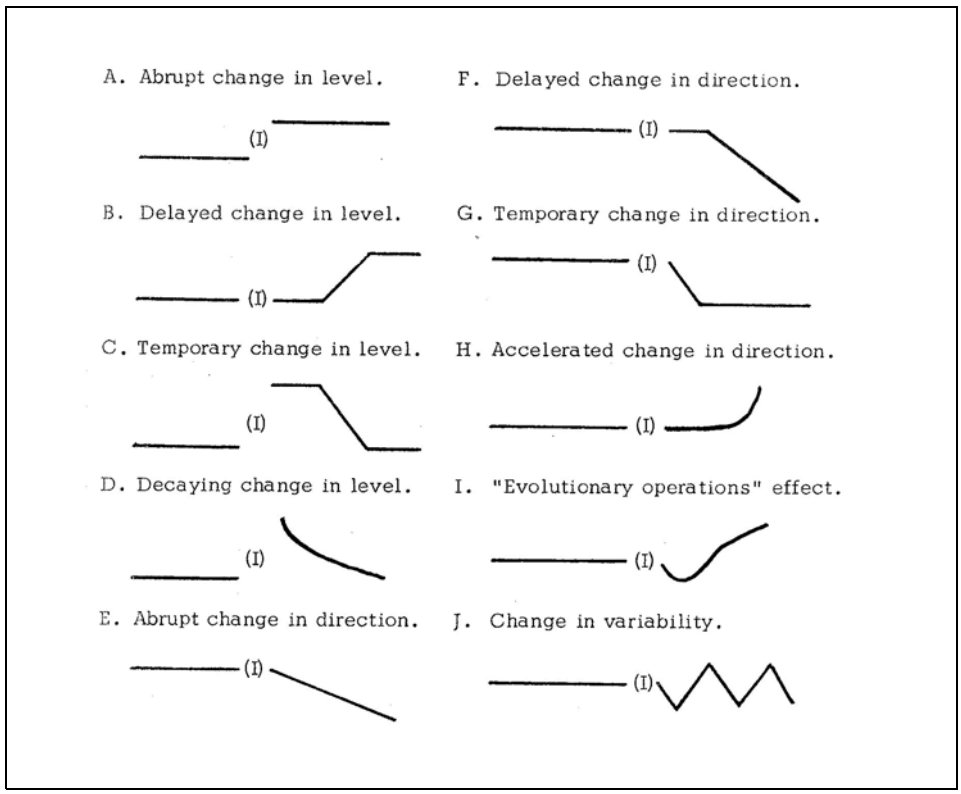


Figure 3.2.1. Different forms of intervention effects (Glass et al., 1975, p. 44).

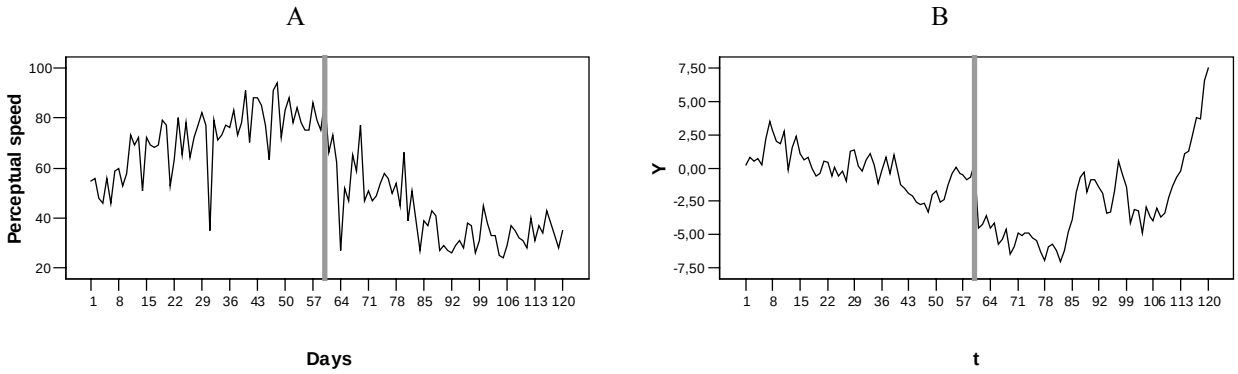


Figure 3.2.2. Time series with abrupt change in level at *t*=60: (A) ARIMA(0, 1, 1) with estimated intervention effect –21.83; (B) simulated ARIMA(0, 1, 0) with true intervention effect –5.

Some researchers (e.g., Kazdin, 1982) propose visual inference as a conservative test for assessing change in longitudinal designs. They argue that strong, robust and reliable intervention effects ought to be seen with the "naked eye". Figure 3.2.2 (A) shows that even a very strong effect is not always visually obvious. Furthermore, it is also important to be able to detect small but significant effects. Gottman (1981) points out that in social sciences we always deal with settings producing a lot of noise in the data. Thus, one of the most frequent tasks for researcher is "to see a change in the "signal" over and above the noise present in the data" (p. 58). Figure 3.2.2 (B) is designed to illustrate that it is impossible to detect the stable downward shift in level of –5 from *t*=60 present in the random walk series visually. Moreover, several studies have shown that visual inference is unreliable and cannot control Type I error (Glass et al., 1975; Jones et al., 1977; Matyas & Greenwood, 1990; Ottenbacher, 1986).

Huitema (1985) argues for the use of traditional *t* tests and ANOVA models when testing for the presence of intervention effects within typical behavioural data, which are low autocorrelated. Sharpley and Alavosius (1988) show, however, that Huitema's suggestion is based on two *incorrect* premises: (a) low levels of autocorrelation can be neglected; (b) time-series procedures are inapplicable with short data series. Glass et al. (1975), Gottman (1981) and Sharpley and Alavosius (1988) demonstrate that the use of traditional ANOVA is

unjustified when there is *any* autocorrelation at all in the data, because even very low levels of dependency can lead to unacceptable distortion of *F* values from traditional procedures. As a result, ANOVA cannot control Type I error and is, consequently, invalid with autocorrelated data.

The Box-Jenkins intervention analysis is the most common methodology to measure intervention effects in time-series experiments. The procedure consists of several consecutive steps. First, the ARIMA model for the data is identified, and *p*, *d* and *q* parameters are estimated (see Chapter 2.1.3). Accurate model identification is necessary to determine the specific transformation matrix to be used to remove dependency from the series so that it meets the assumptions of the general linear model. Then, the GLM *t* tests are employed for uncorrelated residuals to determine whether the post-intervention scores differ significantly in slope and level from the pre-intervention scores. The main difficulty with the Box-Jenkins approach is that the accurate ARIMA model identification requires times series with at least 50 observations both before and after intervention, which can be prohibitive in applied settings.

Some researchers point out that accurate model identification is not always necessary for time-series experiments (Algina & Swaminathan, 1977, 1979; Crosbie, 1993; Harrop & Velicer, 1985; Simonton, 1977; Velicer & McDonald, 1984). In experimental settings, the ARIMA modeling is used to remove dependency from the data. It has been shown that an autoregressive model with one to five parameters can capture successfully all autocorrelations in most empirical series. Therefore, the model identification step can be successfully abandoned in interrupted time series analysis. As a result, different alternatives to the Box-Jenkins procedure, which can be used with short series of 10 to 20 observations per phase, have been proposed. The following techniques prove to be simple and reliable for assessing change with short autocorrelated series: the ITSE of Gottman (1981), the ITSACORR of

Crosbie (1993), and the General Transformation Approach of Velicer and McDonald (1984, 1991).

To summarize, accurate measurement of intervention effects in time-series experiments implies control of autocorrelation in the data. Interrupted time-series experiments do not require very long series. Larger samples of 50 to 100 observations are necessary for accurate model identification and forecasting. When testing for intervention effects, even very short time-series with 20 to 40 data points can be successfully used.

3.3 Forecasting

Forecasting is one of the most important parts of econometric analysis. For economists, prediction of variables such as stock prices, unemployment rates or inflation is an every day task. For psychologists, however, forecasting seems to be less in demand compared to both research applications described earlier. In some research fields, for example, in clinical psychology, prediction of those phenomena as headache pain for migraine patients, depressive or psychotic episodes for patients with psychic disorders can represent an important issue.

Gujarati (2003) distinguishes five forecasting approaches based on time series data: exponential smoothing, single-equation regression methods, simultaneous-equation regression models, ARIMA approach, and vector autoregression. The Box-Jenkins ARIMA methodology is the most popular technique, which has been increasingly used among psychologists. Unlike regression models, in which k regressors explain a dependent variable Y_t, the Box-Jenkins models explain Y_t by past values of Y itself and stochastic error terms. In other words, the future values of the series are predicted from its past values.

The Box-Jenkins forecasting strategy just extends the three-stage model identification cycle, as described in Chapter 2.1.3, to the fourth step used for forecasting. As usual, the procedure begins by examining the series for stationarity. Instationary series are transformed

to stabilize them. The ACF and PACF of the stationary series are employed for identification of a tentative ARMA(p, q) model. Parameters of the tentative model are then estimated. As the next step, the residuals from the tentative model are examined to find out if they are white noise. If the residuals are independent, the tentative model is accepted as an appropriate approximation to the underlying stochastic process. For autocorrelated residuals, the model identification and estimation steps are started again. Therefore, the Box-Jenkins method is called iterative. The model finally selected is used for forecasting. Figure 3.3.1 summarizes the Box-Jenkins iterative process.

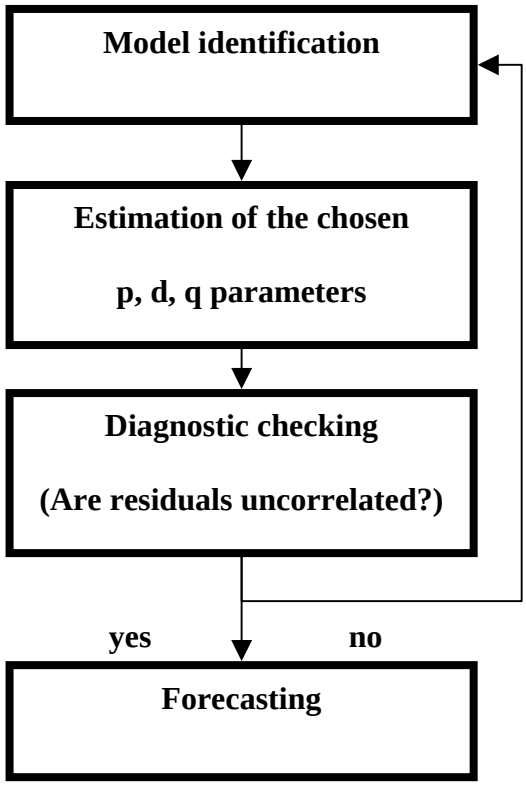
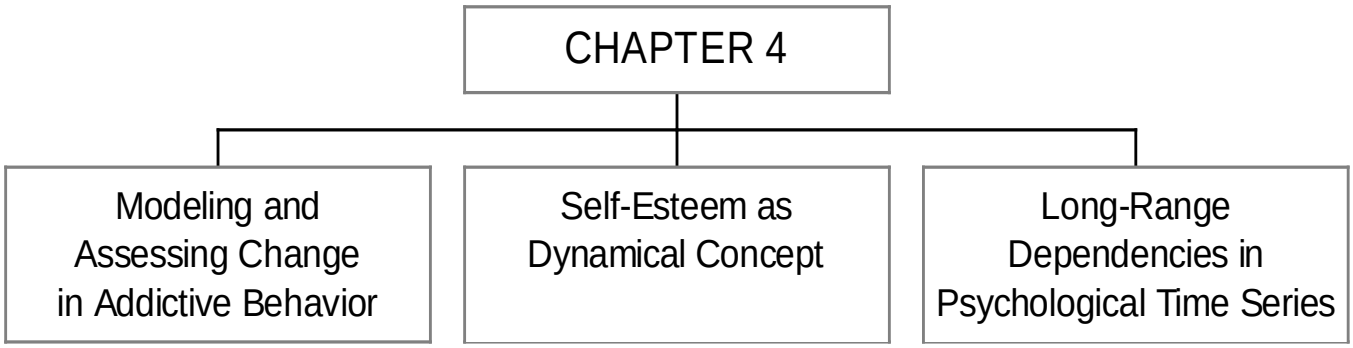


Figure 3. 3. 1. The Box-Jenkins forecasting procedure.

4 TIME-SERIES RESEARCH IN PSYCHOLOGY

The standard research strategy of most psychologists is primarily nomothetic and in the majority of cases consists in the attempt to infer general models from the average behavior of a large sample of individuals. This procedure has been recently challenged in different fields of psychology. An increasing number of cognitive, social and clinical psychologists employ an alternative experimental strategy obtaining their knowledge from time-series data of some few subjects. Emphasizing common dynamics in human behavior, time-series analysis opens a new perspective for psychological research, where the understanding of development and change of psychological processes are in the focus of attention. Delignières et al. (2004) point out that, in this *dynamical* approach, the quality of individual time-series data is more important then the number of individuals in a particular sample. In various research cases, some few time-series can provide a deeper insight into phenomena under study than averages from a large amount of independent observations. The aim of this chapter is to introduce theories and empirical studies from different fields of psychology employing time-series terminology and methods. The chosen theoretical models and studies are designed to illustrate the advantages of the time-series approach for psychological research. First examples show how to model and to assess change in addictive behavior using longitudinal data. Next, dynamical view on self-esteem is discussed. Finally, studies from the 1/*f* noise paradigm challenging traditional models in cognitive and social psychology are presented.



4.1 Modeling and Assessing Change in Addictive Behavior

In health and clinical psychology, the efficacy of a treatment is often measured employing longitudinal data (Crosbie, 1993; Velicer & Colby, 19997; Velicer & Fava, 2003). The following examples aim to show that, in these research fields, time-series analysis is not limited to the evaluation of intervention effects but also enables testing psychological theories explaining development and maintenance of problematic behaviors.

4.1.1 Testing Theories Explaining Smoking Habits

An empirical study of Velicer, Redding, Richmond, Greeley and Swift (1992) provides an excellent example how time-series methods can be used in theory testing. The study is designed to determine which of three popular tobacco-consumption models best represent most smokers. The alternative theoretical models used in the study are the *fixed effect model*, the *nicotine regulation model*, and the *multiple regulation model* (Leventhal & Cleary, 1980).

The *fixed effect model* assumes that smoking maintenance is primarily due to positive effects of nicotine. According to this model, nicotine simulates specific reward inducing centers of the nervous system. The inhalation of nicotine causes either autonomic arousal or a feeling of mental alertness and relaxation or both. This implies that occasional above-averaged nicotine consumption at time t must cause an increasing smoking at time t+1. In terms of time-series analysis, the described dependency structure means positive autocorrelation in the first lag or AR(1) model with a positive ϕ value.

The *nicotine regulation model* assumes the optimal personal nicotine level. This stable set point is under biological control and determines tobacco consumption of an individual. According to this model, all variations in smoking are random and due to the environment. This means that there is no dependency between subsequent smoking occasions and smokers

differ only in their set points. In terms of time-series analysis, no dependency implies a white noise model.

The main assumption of the *multiple regulation model* is that smoking behavior reflects the attempts of the smoker to regulate emotional states. The smoker is balancing between positive and negative (craving) reactions caused by nicotine. According to this model, an occasional increase in smoking rate at time *t* implies a subsequent decrease at time *t*+1, and vice versa. For time-series data, this means a negative autocorrelation at the first lag (between observations at *t* and *t*+1) and alternating positive and negative autocorrelations at subsequent lags.

Velicer et al. (1992) studied smoking habits of 4 male and 6 female subjects. The data were collected by means of self-monitoring twice daily for 2 months. Number of cigarettes served as the main dependent measure. Employing the traditional ARIMA methodology as well as different automated procedures for model selection, a first-order autoregressive model with a moderate to high degree of negative dependence (–0.3 to –0.8) was identified in seven cases. Three individuals demonstrated either a zero or low positive dependence. The results indicate that smoking behavior of the majority of the participants is consistent with the multiple regulation model.

Rosel and Elósegui (1994) conducted a similar study examining the daily cigarette-consumption of 9 male and 20 female smokers over a 12-week period. 75% of the series were identified as the first order autoregressive models. This is an indication of internal regularity in smoking, which is expected for subjects who have smoked for a long period of time. Moving average models were not found in this sample. MA patterns indicate the presence of external influence on the subject behavior and are typical for occasional smokers. The behavior of 13 subjects contained a seasonal cycle of 7 days. The majority of the time series (21) exhibited positive autocorrelation supporting the fixed effect model. In the data of 7

participants, no significant autocorrelation was detected which implies a nicotine regulation model. In contrast to Velicer et al. (1992), only one subject fit the multiple regulation model.

According to Velicer and Fava (2003), the main reason for the discrepancy in the results of the two studies is the different time intervals used (twice daily or daily): a negative autocorrelation at the first lag turns out to be a positive autocorrelation at the second lag, since $r_1^2 = r_2$ [1]. Thus, r_2 values of Velicer et al. (1992) and r_1 coefficients of Rosel and Elósegui (1994) reflect the same time period; contrasting these statistics would provide a more direct comparison of the studies. Subsequent analysis of the r_1 and r_2 values revealed that the two are both positive and of comparable magnitude. Therefore, for the *daily* cigarette-consumption both studies indicate the fixed effect model. Velicer and Fava (2003) point out that these results highlight an important methodological issue of the choice of the "correct" time interval between observations. The conclusion about the appropriate nicotine regulation model turned out to be affected by the frequency of measurement. Therefore, it is critical to pay attention to the time interval when interpreting time series studies.

4.1.2 Assessing Change in Addictive Behavior

Time is an essential part of theories analyzing dynamics of human behavior. Thus, in the *Transtheoretical Model* (Velicer, DiClemente, Rossi & Prochaska, 1990) explaining change in health behavior, the time component represents the crucial factor. The model postulates four different stages of change for different theoretical constructs labeled the processes of change. Both behavioral dependent measures, such as abstinence, and hypothetical constructs, such as self-efficacy or decisional balance, are proposed to describe the process of change in addictive behavior. *Precontemplation*, *Contemplation*, *Action,* and *Maintenance* constitute the

[1] The autocorrelation function of AR(1) process is given by $\rho_j = \phi^j$ with j = lag, which implies ρ_1=0.5 and ρ_2=0.5^2 for ϕ=0.5 (Hamilton, 1994, pp. 53-56); see also Chapter 2.1.

stages of change. The model postulates different levels of dependent measurements for each stage of change. In addition to the process of change, the model assumes the existence of external and internal independent variables specific to the problem area. According to the model, any intervention or changes in the natural environment can serve as external variables. Personality characteristics, cognitive abilities, or available recourses are possible internal factors.

Evaluation of the model using time-series methods confirmed clear stage differences for various processes of change (DiClemente et al., 1991; Fava et al., 1994). Furthermore, time-series analysis revealed the dynamics of different processes of change across the stages of change. For example, Prochaska et al. (1991) showed that cognitive processes predominate in the Contemplation Stage and behavioral processes in the Action and Maintenance Stages. The greatest empirical support for the model comes from the area of smoking cessation. Prochaska et al. (1994) demonstrated, however, that the same pattern of change across the stages is typical for a broad range of problem behaviors including weight control, sun exposure, or HIV risk reduction.

Velicer, Rossi, Prochaska and DiClemente (1996) extended the Transtheoretical Model to the *Three Construct Model* using Positive Evaluation Strength, Negative Evaluation Strength and Habit Strength as dependent measurements. The first two constructs are designed to measure the cognitive aspects and the last one the behavioral aspects of phenomena under study. The hypothesized pattern of change for each construct is described across four stages of change. The model postulates two thresholds representing the ability of the environment to modify the processes of change: Asthenic Threshold and Abstention Threshold. The following example illustrates the hypothesized pattern of change for the behavioral aspects of smoking. The Habit Strength in smoking is modeled as an autoregressive process of order one. The autoregressive pattern reflects internal regularity in

addictive behavior of long-term smokers. Two thresholds symbolize the influence of the environment on smoking. If the level of the series is above the Asthenic Threshold, the person will smoke regardless of the environment and make efforts to control it. Buying an adequate number of cigarettes and avoiding no smoking situation are possibilities for controlling the environment. If the level of the series is below the Abstention Threshold, the person will avoid smoking regardless of the environment. People between the two thresholds will be influenced by their personal and physical environments. The model relates the level of the time series to the stage of change. In the Precontemplation stage, it is located above the Asthenic Threshold. In the Contemplation stage (thinking about quitting smoking), the level is between the two thresholds. According to the model, people in the Action stage remain at about the same level as people in the Contemplation stage even though they are not smoking. The authors point out that the cessation of smoking is often the result of environmental control, rather than a lack of desire. People in the Maintenance stage have a reduced Habit Strength. They are at level lower then the Abstention Threshold, which means that they generally are not at risk. Figure 4.1.1 summarizes the model assumptions for the construct Habit Strength.

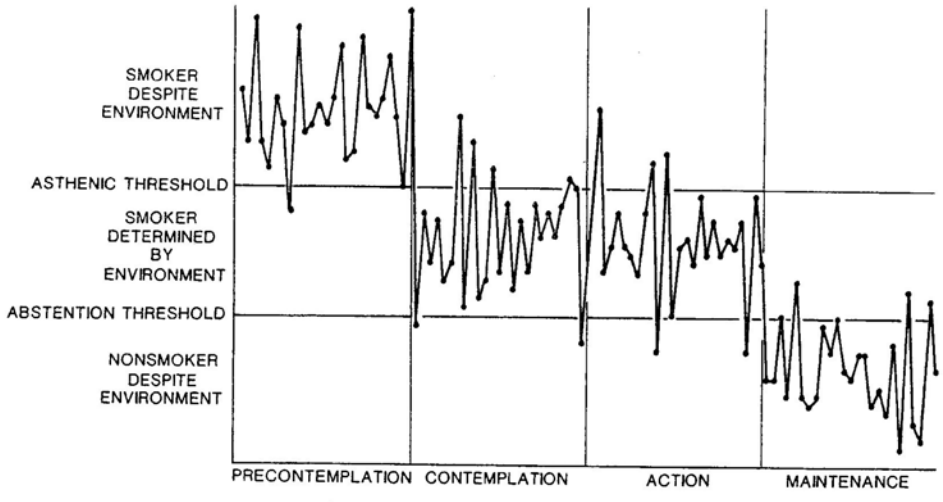


Figure 4.1.1. Model for Habit Strength (Velicer et al., 1996, p. 563).

Velicer, Rossi, Prochaska and DiClemente (1996) demonstrated that empirical profiles of various addictive behaviors are in accordance with those of their Three Construct Model. Furthermore, time series analysis for different dependent measures including number of cigarettes, craving, or carbon monoxide proved an ARIMA (1, 0, 0) as the underlying model for processes of change. This autoregressive pattern proves internal regularity in addictive behavior.

Once again, understanding development and change in behavior or cognition is impossible without regard to the time component. Time-series methods allow modeling different dependency structures for constructs under study and comparison of theoretical and empirical change profiles. Furthermore, hypothesized stage differences in theoretical models can be tested by means of the time-series experiment.

4.2 Self-Esteem as Dynamical Concept

From the classical viewpoint, *self-esteem* (i.e., a favorable global evaluation of oneself) is a stable personality trait not greatly affected by daily events (Mischel, 1969). Empirical research, comparing repeatedly measured means of individuals, supports a relative stability of self-esteem *level* for various psychological phenomena, including anger or hostility proneness (see Greenier, Kernis, & Waschull, 1995, for review).

It has been widely recognized, however, that some specific life events such as professional success or failure can cause meaningful short-term instabilities in self-esteem, and that such aspects of self-esteem as self-worth incline especially strong to fluctuations (Rosenberg, 1986). Kernis (1993) accentuates self-esteem lability and argued that people differ not only in level but also in *stability* of their self-esteem. Nezlek and Plesko (2001) showed daily fluctuations in self-esteem. Greenier et al. (1999) demonstrated that negative and positive events have a greater impact on the self-feelings of individuals with unstable as

opposed to stable self-esteem. In these studies, standard deviation is used for assessing the stability of self-esteem.

The so-called *hierarchical models* of self-concept combine the ideas from theories focused on trait stability with concepts of cross-situational variability of self-esteem (Brown, 1998; Fox, 1997; Marsh & Craven, 1997). According to these models, self-esteem constitutes the apex of the hierarchical system. Diverse domains of competence such as social, physical or familial with their more specific subdomains build the subordinate levels of the hierarchy (see, for example, Figure 4.2.1). The hierarchical concepts imply either linear dependencies or nonlinear influences between levels. Fox (1997) assumes higher stability for global dimension of the hierarchy and stronger variability for the more specific subdomains. Marsh and Yeung (1998) and Amorose (2001) demonstrated, however, that global self-concept measures were less stable than more specific scales.

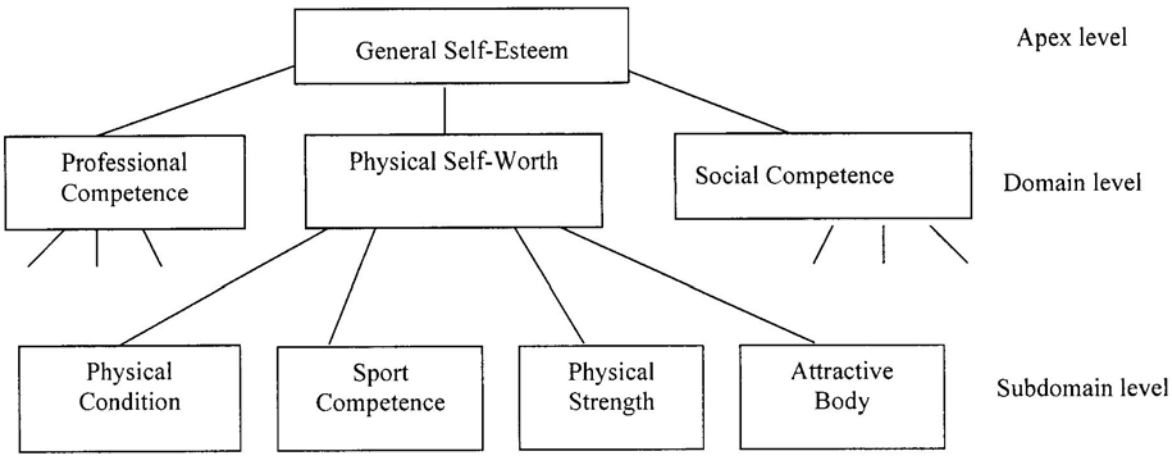


Figure 4.2.1. Hierarchical model of self-esteem (Ninot et al., 2001, p. 206).

New conceptions of the self define self-esteem as a transitory state, depending simultaneously on previous states and current disturbances (Marks-Tarlow, 1999, 2002; Nowak et al, 2000; Vallacher et. al., 2002). From this *dynamical* viewpoint, understanding of self-esteem is impossible without the analysis of its evolution (Fortes et al., 2005). In other words, the mechanism underlying the development of self-esteem over time are more important than it's level or the magnitude of variability. Considering self-esteem as a dynamical construct requires a special methodological approach, based on time-series analysis. For this purpose, Ninot et al. (2001) developed a new assessment tool, allowing continuous monitoring of self-esteem and physical self. The *Physical-Self Inventory* is composed on six single items, measuring global self-esteem, physical self-worth and its four subdomains: physical condition, sport competence, physical strength and attractive body. Therefore, the inventory assesses all three levels of the hierarchical model of self-esteem (compare Figure 4.2.1). Daily or twice daily completion of this instrument over a long period allows the collection of time series.

In the recent study, Fortes, Delignières and Ninot (2004) analyzed psychological processes that underline the dynamics of global self-esteem and physical self over time. For 228 consecutive days, seven participants completed twice daily the Physical-Self Inventory. Each of six individual series was then modeled by means of ARIMA procedures. The analysis revealed large inter-individual differences in the global shape of time series. Global self-esteem development of two subjects looked rather stationary. Time series of the majority of participants were characterized by marked instability in level or variance. Figure 4.2.2 illustrates the seven individual time series collected for global self-esteem.

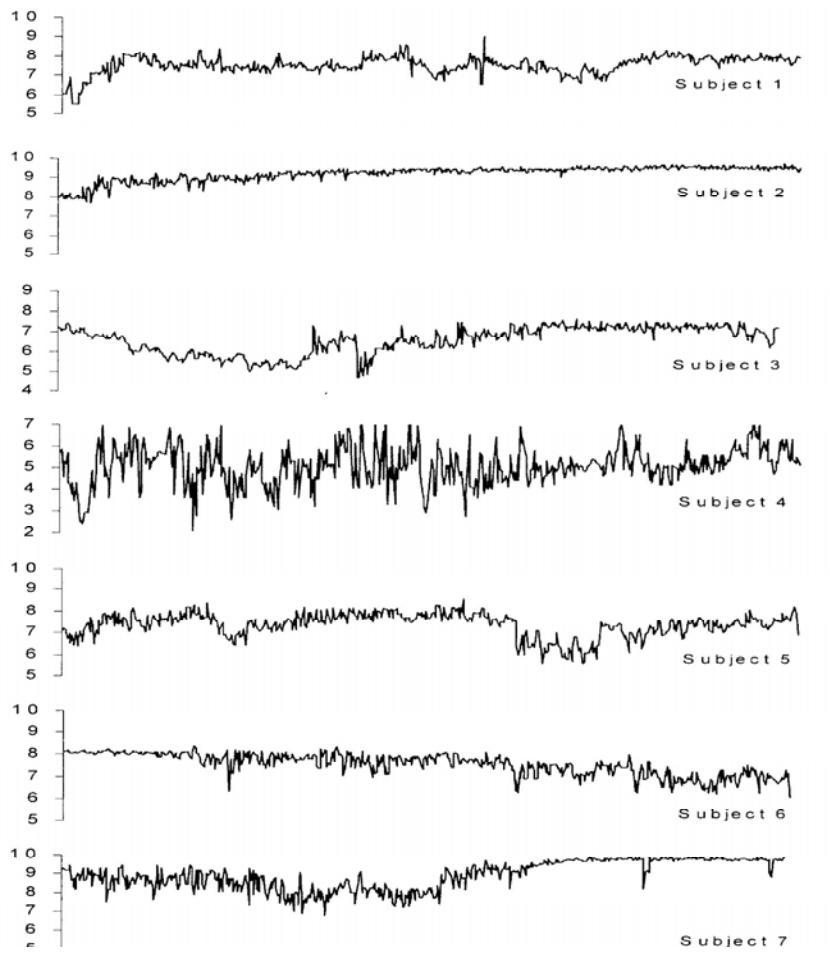


Figure 4.2.2. Time series for global self-esteem (Fortes et al., 2004, p. 741).

The analysis of the ACF of the series showed the presence of significant positive autocorrelations for more than 10 lags. Fortes et al. (2004) point out that this result invalidates the use of means and standard deviation in dynamical self-esteem research because the use of these statistics supposes uncorrelated data. Most of the series (83%) were identified as the ARIMA(0, 1, 1) with a positive moving-average term ranging from 0.44 to 0.88. In seven from 42 cases (7 subjects x 6 items), (0, 1, 2) or (0, 1, 3) models were fitted. The identified ARIMA structures were quite similar among the six dimensions of each subject.

In short, the main result of the study is that a differenced first-order moving-average model represents the best fit in most cases. According to Spray and Newell (1986), this model is typical for time series that exhibit noisy fluctuations around a slowly varying mean. The time series from different levels of self-esteem (global, domain and subdomain) were not stationary implying the instability of the studied dimensions. Fortes et al. (2004) suggest that the observed dynamics of self-esteem reflect a combination of two opposite processes: *preservation*, which tends to restore the previous value after a disturbance, and *adaptation*, which tends to inflect the series in the direction of the perturbation. As stated previously, autoregression is characteristic for internal regularity, whereas moving-average pattern reflects external influences. In combination with $\phi=1$, the positive moving average coefficient θ determines the balance between the preservation and adaptation processes. Based on the results of their study, Fortes et al. argue that the dynamics of self-esteem processes can be understood as organized around a locally stable reference value, which evolves progressively under the influence of life events.

The main conclusion from the presented study is that analysis of dynamical concepts such as self-esteems requires the adequate methodological approaches. Traditionally, psychologists assess evolution or development through repeated measurements. By this procedure, however, possible dependences between subsequent values remain indiscernible. According to Slifkin and Newell (1998), classical statistics such as mean or standard deviation are not able to reflect the true nature of variability or change in living systems. Revealing the autocorrelation structure of dynamical concepts, the time-series methods allow to conceive the development of various psychological phenomena.

4.3 Long-Range Dependencies in Psychological Time Series

In traditional psychological research, autocorrelations in the data are often ignored. This reflects the implicit assumption that most psychological phenomena can be viewed as randomly distributed in time around a more or less stable mean. Researchers who paid attention to serial correlations in human performance considered them to be small and transient (Laming, 1968; Huitema, 1985). Recent empirical studies from cognitive, social, and biological psychology are questioning this popular view. Employing time-series methods such as the log-log power spectrum or the ARFIMA modeling (see Chapter 3.2.1), persistent long-range dependencies have been detected in various psychological time series. These very slowly decaying autocorrelations are known as 1/*f noise*. The objectives of this chapter are: to review empirical studies on 1/*f* noise and to discuss theoretical models accounting for long-range dependencies in psychological time series.

4.3.1 Review of Empirical Findings

Time series exhibiting persistent autocorrelations have been observed in physics, biology, hydrology, economics, sociology, and other disciplines. A summary of interdisciplinary 1/*f* noise literature can be found under: *http://www.nslij-genetics.org/wli/1fnoise/*. The bibliography contains more than 800 publications demonstrating 1/*f* noise in a wide range of natural phenomena such as heart beat rhythmus, brain activity, human coordination, music, or speech. The primary objective of the bibliography is, therefore, to show the ubiquity of 1/*f* noise.

In psychology, correlated noise was initially detected in *controlled cognitive performances* (Gilden, Thornton & Mallon, 1995). Using spectral analysis techniques, Gilden and his colleagues demonstrated in experiments including mental rotation, lexical decision, shape and color discrimination, or visual search that persistent autocorrelations account for

even more variability in the data than most standard manipulations in cognitive psychology (Gilden, 1997, 2001; Gilden & Wilson, 1995; Gilden et al., 1995). Serial correlations observed in these studies were not only persistent but also relatively large in absolute magnitude. Wagenmakers et al. (2004) confirmed these findings employing in their analysis the ARFIMA methodology.

Van Orden et al. (2003), Wagenmakers et al. (2004), and Ward and Richard (2001) found long-range dependencies in *automatic cognitive performances* such as word naming or simple reaction times. The intensity of the dependence was here lower than in tasks requiring cognitive control. This implies that the persistence and magnitude of serial dependence is not based purely on temporal contiguity. According to Wagenmakers et al. (2005), empirical support for the existence of persistent autocorrelations in automatic tasks is not so strong as for controlled cognitive performance. Therefore, competitive assumption of short-range dependencies cannot be definitely excluded from consideration.

Chen et al. (1997, 2001), Delignières et al. (2004) and Ding et al. (2002) observed persistently correlated noise in human *rhythmic activities* such as tapping or other tasks requiring coordination or synchronization of motor and cognitive activities. The evidence for the long-range pattern was rather strong with estimated power exponents[1] ranging from 0.49 to 0.87 (Ding et al., 2002) or even from 0.61 to 1.68 (Delignières et al., 2004). Moreover, the log-log power spectra of most series were fitted with quite perfect straight lines reflecting pink noise in the data. These results are inconsistent with the assumption of a simple autoregressive error correction model for synchronization tasks proposed by Pressing and Jolley-Rogers (1997). Specifically, the model postulates that linear autoregressive error correction processes of order 1 predominate in interactive human performance requiring synchronization. In other words, the noise component of rhythmical performance arises

[1] Recall that, for a long memory, the power exponent a in $1/f^{a}$ can vary from 0.5 to 1.5 (see also Chapter 3.1.2).

primarily from an AR(1) process, a second order process may be used if task demands of coordination and speed require greater accuracy. In synchronized tapping experiments, Pressing and Jolley-Rogers (1997) observed exponential spectral patterns typical for short-range dependencies. According to Pressing and Jolley-Rogers, the interplay of cognitive and motor components in tapping produces these spectral curves. The motor component concentrates its power at higher frequencies, whereas the cognitive component is dominant for lower frequencies. Recall that concentration of the power at zero frequency implies instationarity. Thus, for AR(1) series, more power at low frequencies implies ϕ coefficients near 1. Predominance of high frequencies is typical for AR(1) models with negative ϕ values (see Chapter 3.1.3). Therefore, the findings concerning dependency structure of rhythmic activities are contradictory: Chen et al. (1997, 2001), Delignières et al. (2004) and Ding et al. (2002) observed a strong evidence for $1/f$ noise, whereas Pressing and Jolley-Rogers (1997) found only short-range dependencies in their data. Gilden (2001) points out that the pink noise spectrum ($1/f$) can result from a combination of white noise ($1/f^{0}$) and random walk ($1/f^{2}$) spectra. It is possible that cognitive processes in tapping are instationary due to fluctuations in attention or fatigue, and simple motor responses such as key pressing introduce white noise into the data. The interplay of these factors can simulate the $1/f$ like behavior. Therefore, further research is required to answer the question whether time series from rhythmic activities are the real pink noise or if we deal here with $1/f$ type effects due to fluctuations in speed of performance or attention. New methodologies such as the spectral classifier procedure of Thornton and Gilden (2005) and the ARFIMA approach of Wagenmakers et al. (2004) can help to solve this problem.

Aks and Sprott (2003) detected $1/f$ noise in *visual perception*. The timing of perceptual reversals of Necker cubes served as an independent variable. While viewing the Necker cube, subjects pressed a key each time they perceived a change in the cube's orientation.

Experiments were performed under three sets of binocular disparity conditions. Viewing duration was extended, moderate, or brief (60, 30, or 15 minutes). Spectral analysis of 40 series detected 1/*f* noise in 80% of the cases. Regression slopes in log-log plots varied from –0.6 to –0.9. More disparity and reduced viewing time produced steeper slopes in the spectra (larger power exponents). According to Aks and Sprott (2003), the last finding reveals the stabilizing function of binocular disparity in perception. Disparity may either filter out extraneous information or signal the system to rely more on previous percepts. More studies and methodologies employing rigorous testing of alternative hypotheses to 1/*f* noise are necessary to verify this assumption.

Delignières, Fortes and Ninot (2004) reported long-range dependencies in time series of *self-esteem and physical self*. Twice a day for 512 consecutive days, four adults completed Physical-Self Inventory (Ninot et al., 2001). Persistent autocorrelations were detected employing different methods. For example, spectral analysis revealed for each series a straight line in the double logarithmic plot of power against frequency. No traces of flattening of the plot in the low-frequency region, as expected for short memory processes, were found. The power exponent values appeared close to 1, suggesting that the series behaved like 1/*f* noise.

Fractal analysis, accounting for the degree of self-similarity in time series, represents another methodology used in the study. This approach differentiates two types of series with long-range dependencies: persistent and anti-persistent. The series is said to be persistent if an increasing trend in the past is likely to be followed by an increasing trend in the future. In anti-persistent series, an increasing trend in the past is followed by a decreasing trend. Mathematically, processes with different autocorrelation properties can be characterized using the following scaling law: $<\Delta Y> \propto \Delta t^H$, implying that the expected increment $<\Delta Y>$ is a power

function of the time interval (Δt) over which this increment is observed[1]. *H* is called the *scaling, fractal or Hurst exponent* and can be any real number in the range $0<H<1$. (Mandelbrot & van Ness, 1968; Torre, Delignières, & Lemoine, in press). The aim of fractal analysis is to check whether this scaling law holds for experimental series. A theoretical scaling exponent of random walk is 0.5; this value constitutes the frontier between anti-persistent ($H<0.5$) and persistent ($H>0.5$) series. For $1/f$ noise, fractal exponents lower than 0.5 are expected. The estimates of *H* obtained by Delignières et al. were located in a quite narrow range, between 0.18 and 0.40, suggesting an anti-persistent long-range correlation process. Delignières et al. compared their findings with the results obtained by Fortes, Delignières and Ninot (2004) for shorter time series. (Recall that observations of this study were collected twice daily for 228 days). They found that the moving average coefficients of shorter series are negatively related to *H* estimates. According to Delignières et al., this suggests that $1/f$ noise and the moving average model possess similar properties, characterized by a subtle balance between the preservation of a reference value and an adaptation to events. This balance is not simply achieved over the short term, as implied by the ARIMA models, but occurs at multiple time scales, in a self-similar way. Delignières et al. formulated an interesting hypothesis concerning the relationship between moving average coefficients and the scaling exponent: low moving average coefficients are related to weakly anti-correlated series close to the random walk pattern with *H* values about 0.5, and higher coefficients correspond to series closer to $1/f$ noise with lower *H* exponents. Empirical evidence for the postulated relationship between moving average values and fractal exponents could be of great practical importance in applied settings, since, in contrast to spectral methods or the ARFIMA methodology, ARIMA procedure can work with relatively short times series.

---

[1] Recall that successive increments of random walk $\Delta Y = Y_t - Y_{t-1} = u_t$ are uncorrelated.

Long-range dependencies are also found in human gait (Hausdorff et al., 1997; 1999); force production tasks (Pressing, 1999); brain activity (Linkenkaer-Hansen, 2002); heart rate fluctuations or other biological phenomena (Hausdorff & Peng, 1996). Table 4.3.1 presents an overview of the studies.

Table 4.3.1. Overview of empirical studies reporting 1/*f* noise.

Controlled Cognitive Performance	Gilden, Thornton & Mallon (1995) Gilden & Wilson (1995) Gilden (1997, 2001) Wagenmakers et al. (2004)
Automatic Cognitive Performances	Van Orden et al. (2003) Wagenmakers et al. (2004) Ward & Richard (2001)
Rhythmic Activities	Chen et al. (1997, 2001) Delignières et al. (2004) Ding et al. (2002)
Visual Perception	Aks & Sprott (2003)
Self-Esteem and Physical Self	Delignières, Fortes & Ninot (2004)
Human Gait	Hausdorf, Zemany, Peng, & Goldberger (1999)
Force Production Tasks	Pressing (1999)
Brain Activity	Linkenkaer-Hansen (2002)
Heart Rate Fluctuations	Hausdorff & Peng, 1996

4.3.2 Explanations for Long-Range Dependencies

Numerous explanations for the observed long-range dependencies in psychological time series have been proposed. Roughly, two different approaches can be distinguished. One perspective is characterized by the idea that 1/*f* noise patterns can arise from *aggregation of more simple models*. Another perspective is embedded in the framework of the *nonlinear dynamical system theory*.

As describes previously, a *combination of white noise and random walk spectra* can imitate the 1/*f* noise behavior. Thus, long-range dependencies in psychological time series can be caused by instationarity in time series due to fluctuations in attention or shifts in the strategy. Busey and Townsend (2001), Gilden (2001), and Wagenmakers et al. (2004) use this explanation.

Chen et al. (2001), Ding et al. (2002), and Wagenmakers et al. (2005) argue that the mathematical model of the long memory process proposed by Granger (1980) for economic measures can also work for psychological data. Granger hypothesized that pink noise can arise via *aggregation of multiple component processes* that separately generate transient correlations. In other words, a simple summation of independent AR(1) processes results in persistent serial correlations of 1/*f* type. This model can be applied to human cognition if we assume that the observed series is an aggregation of the behavior of many independent groups of neurons, each with their own different autoregressive parameter.

A similar idea for biological systems comes from Hausdorff and Peng (1996). Their *multiscaled randomness model* assumes that 1/*f* noise pattern can result from summation of short-range processes with different characteristic time scales. Hausdorff and Peng argued that in many biological series overall behavior is influenced by systems operating on widely different time scales. For example, heart rate fluctuations are short-term regulated via the

autonomic nervous system; long-term influences come from circadian rhythm via hormonal systems.

The assumption that the presence of *multiple time scales* can yield 1/*f* noise effects is also shared by Pressing (1999). According to Pressing, multiple time scales explain long-range correlations in human cognitive control and attentional fluctuations. 1/*f* spectra are typical for systems that feature multiple discrete time scales, as in relaxation, processing, or production. Simulation experiments evaluating this model demonstrated that summation of random processes with different time scales can yield 1/*f* spectra with exponents ranging from 0.5 to 1.5; three or even two series are sufficient for this effect. To achieve a 1/*f* like spectrum, slow process must have greater amplitude (weight) than faster process (Pressing, 1999, p. 6).

Another perspective associates 1/*f* noise with *deterministic nonlinearity* operating in intermittency or chaotic regimes (Schuster, 2005). The most popular model from this approach is the *Self-Organized Criticality* (SOC), introduced by Bak, Tang and Wiesenfeld (1987). Using a pile of sand as a metaphor, Bak et al. explain the large-scale dynamics of various phenomena. In the same way as grains of sand added slowly and randomly to a sandpile cause avalanches, barely detectable movements of the earth's crust can cause the devastating earthquakes, or small random changes in stock prices can lead to financial crashes (Bak, 1996; Jensen, 1998). A system is said to be self-organized when its structure emerges without explicit influences from outside the system. To state it differently, non-randomness in the system emerges from random initial condition and random input. The term *self-organized* implies this internal dynamic. The term *criticality* is analogous to the critical point of equilibrium systems. Thus, SOC characterizes systems that naturally evolve to a critical state in which a minor event starts a chain reaction that can affect any number of elements in the system. Such systems are said to have a critical point as an attractor. Long-range autocorrelations (pink noise, fractal dynamics) constitute the typical signature of complex

systems in a critical self-organized state. The important feature of 1/*f* noise is its *scale-invariance* or *self-similarity* across different levels of the system structure. In other words, self-similar processes look roughly the same when viewed at different levels of magnification. From this viewpoint, the rare events such as financial crashes or devastating earthquakes are the consequence of the same mechanism, causing small changes to stock prices every minute or harmless movements of the earth's crust every day.

From the nonlinear perspective, the 1/*f* behavior is conceived as the typical intrinsic dynamics of complex systems acting at the edge of chaos, a transitional state between unpredictability and predictable order. Thus, as an intermediate between white noise and brown noise, 1/*f* noise is an indication of both stability and adaptability of a system. The intrinsic stability of 1/*f* noise is primarily due to the relative independence of the underlying processes acting at different time scales. Hence, 1/*f* noise behavior is more adaptive to endogenous and exogenous perturbations and can be viewed as the typical signature of young, healthy, and adaptive systems (West & Shlesinger, 1990). Marks-Tarlow (1999) argued that psychological health resides at the edge of chaos. In this state systems possess enough stability to maintain consistency but sufficient randomness to ensure adaptability and creativity. The behavior of disabled systems is, to the contrary, either more unpredictable or rigid. Marks-Tarlow expects, for instance, the white noise pattern for behavior of hysterical patients and brown noise for people with obsessive-compulsive disorder. There is empirical evidence providing an indirect support for this assumption. Hausdorf et al. (1997) observed 1/*f* noise in the gait of healthy adults. For elderly people or patients with Huntington's disorder, behavior close to white noise was typical. Ninot, Delignières and Varray (2003) showed that the variability of self-esteem time series was more random in patients suffering from chronic pulmonary disease than in healthy subjects (see also Delignières et al., 2004). Gottschalk, Bauer and Whybrow (1995) demonstrated that mood variation of patients with

bipolar disorder was more organized and characterized by a loss of complexity compared to the mood pattern of healthy participants.

To put it briefly, from the nonlinear perspective self-organized criticality is assumed to be a typical state of dynamical complex systems acting at the edge of chaos. Scale-invariant 1/*f* behavior representing a balance between stability and adaptability is characteristic of such systems.

Van Orden, Holden and Turvey (2003; 2005) and Van Orden, Moreno and Holden (2003) suggest SOC as an alternative to the current paradigms in cognitive psychology. They argue that pink noise observed in psychological time series suggests processes of mind and body that change each other's dynamics. According to Van Orden and his colleagues, self-organized criticality supplies a very plausible metaphor for self-control. "Near critical points, interaction-dominant dynamics coordinate activity across the multiple time scales of embodied fluctuations. Context sensitivity near critical points situates behavior within the flow of circumstances. An actor situated in this sense reflects previous and oncoming circumstances directly as purposive behavior" (Van Orden, Holden & Turvey, 2003, p.347). Moreover, self-organized criticality points out the internal dynamics of human behavior and characterizes "human beings as intentional beings".

Aks and Sprott (2003) share the ideas of SOC and explain the observed 1/*f* noise in visual tasks as an expression of complexity and adaptability of the human perceptual system. They argue, for example, that depth information guides the perceptual system into a self-organized state to assist us in resolving ambiguous information.

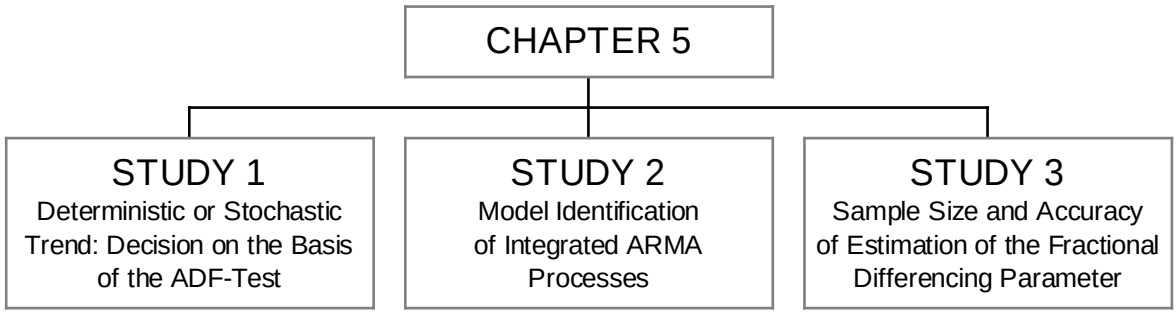
According to Delignières et al. (2004), the evidence of 1/*f* noise in self-esteem supports dynamical conceptions of the self proposed by Nowak et al. (2000) or Vallacher et al. (2002). These models consider self-esteem as a self-organized dynamical system. Delignières et al. (2004) point out the discovered self-similarity in behavior of different levels

of self-esteem and hypothesize that this fractal pattern reflects the intrinsic dynamics of global self-esteem for healthy adults.

In sum, long-range dependencies or 1/*f* noise have been observed in time-series representing cognitive, motor, perceptual, and biological processes or self-esteem development. Some researchers explain this phenomenon *linear* as a result of aggregation of different simpler models. Another perspective conceives 1/*f* noise *nonlinear* as a signature of complex dynamic self-organized systems. Argumentation and exchange of views between advocates of these two paradigms can be found in articles of Wagenmakers et al. (2005) and Van Orden et al. (2005).

5 METHODOLOGICAL ISSUES

Analysis of empirical research employing time-series techniques revealed several methodological issues remaining to be clarified. The main goal of this thesis is to treat some of them. This chapter introduces three simulation studies dealing with the following topics: transformation of instationary time series; model estimation; and sample size requirements for reliable detection of long-range dependencies or 1/*f* noise. The objectives of the first study are to develop testing strategies allowing to distinguish between different causes of instationarity and to evaluate these strategies by means of Monte Carlo experiments. The second study compares the performance of ESACF and SCAN, automated methods for ARMA model identification commonly available in current versions of SAS for Windows, as identification tools for various integrated processes. The last study examines sample size requirements for the accurate estimation of the long-memory parameter *d*, and documents the quality of the conditional sum of squares estimates for time series of different length in various (0, *d*, 0) and (1, *d*, 1) models.



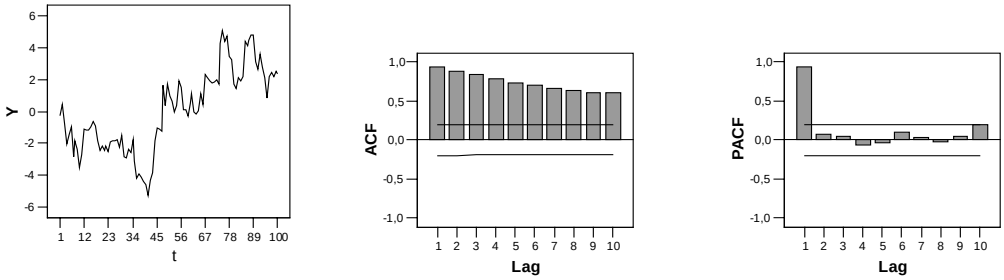
5.1 Study 1: Deterministic or Stochastic Trend: Decision on the Basis of the Augmented Dickey-Fuller Test

5.1.1 Introduction

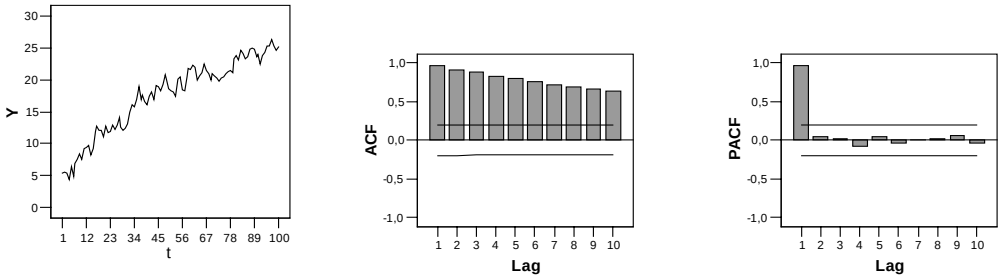
In social and behavioral sciences, the goal of time series analysis is usually to measure the effects of an intervention, as in an interrupted time series experiment, to forecast future values of the series under consideration, or to determine the nature of the process that describes an observed behavior. In the first two cases, *stationarity* of the series under study is required. A process is said to be stationary if all its moments are constant over time. Most nonstationary series in psychology have a time-varying mean or a time-varying variance or both. Nonstationary time series have to be transformed to make them stationary. The proper transformation method depends on the cause of nonstationarity. The consequences of a false treatment can be rather serious. Unfortunately, the last issue is not emphasized in the time series textbooks used among psychologists (Glass et al., 1975; Gottman, 1981; McCleary & Hay, 1980, and Warner, 1998). Some descriptions even suggest that two popular methods for stabilizing nonstationary series, *differencing* and *ordinary least squares regression*, are interchangeable, and that the choice of the transformation method is simply a matter of researcher's preferences (see, for example, Warner, 1998, p. 39). One of the objectives of this chapter is to demonstrate the importance of the proper stationarity transformation for empirical time-series research.

Figure 5.1.1 shows three common nonstationary processes and their autocorrelation and partial autocorrelation functions (ACF and PACF, respectively).

A $Y_t = Y_{t-1} + u_t \quad u_t \sim \text{IID}(0, \sigma^2)$



B $Y_t = 0.2 + Y_{t-1} + u_t \quad u_t \sim \text{IID}(0, \sigma^2)$



C $Y_t = 0.1t + 0.2Y_{t-1} + a_t \qquad a_t = 0.5a_{t-1} + u_t \qquad u_t \sim \text{IID}(0, \sigma^2)$

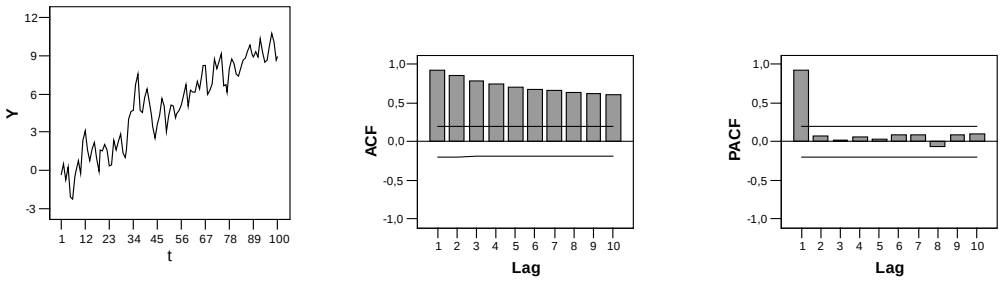


Figure 5. 1. 1. Nonstationary processes and their autocorrelation and partial autocorrelation functions: (A) pure random walk, (B) random walk with drift, (C) deterministic time trend.

As described previously, the process $Y_t = Y_{t-1} + u_t$ with $u_t \sim$ IIDN (0, σ^2) is called *a pure random walk*. The mean of this process is equal to its initial value but its variance increases indefinitely over time. A pure random walk process can also be represented as the sum of random shocks $Y_t = \sum u_t$. As a result, the impact of a particular shock does not dissipate, and the random walk remembers the shock forever. That is why a random walk is said to have an infinite memory. If a constant term is present in the equation $Y_t = \alpha + Y_{t-1} + u_t$, Y_t is called *random walk with drift*, where α is known as the drift parameter. Depending on α being negative or positive, Y_t exhibits a negative or positive *stochastic* trend. For a random walk with drift, the mean as well as the variance increase over

time. Random walk processes are nonstationary, but their first differences $\Delta Y_t = Y_t - Y_{t-1}$ are stationary. Hence, both types of random walks are called *difference stationary* (DS) processes. Random walk models are also known in the time series literature as *unit root processes*. A situation of nonstationarity is called the unit root problem, if in the first order autoregressive model $Y_t = \rho Y_{t-1} + u_t$ ρ is 1. The name unit root is due to the fact that ρ=1. (The autoregressive model can also be written as $(1-L)Y_t = u_t$. The term unit root refers to the root of the polynomial in the lag operator[1].) Random walk is a specific case of a more general class of stochastic models known as *integrated processes*. An integrated process of first order is represented by an equation $Y_t = \alpha + Y_{t-1} + a_t$, where any stationary ARMA (*p*, *q*) process can generate the random part a_t. Therefore, random walk processes are integrated of order 1, denoted as *I(1)* or (0, 1, 0) in terms of Box-Jenkins ARIMA methodology. In general, if a time series has to be differenced *d* times to make it stationary, that series is called integrated of order *d*. In the time series literature the terms nonstationarity, random walk and unit root are often treated as synonymous.

The development of the third process in Figure 5.1.1, represented by the equation $Y_t = 0.1t + 0.2Y_{t-1} + a_t$ with $a_t = 0.5a_{t-1} + u_t$, is determined by a *deterministic* time trend. This process has a constant variance and a changing mean. In a more simple case of the deterministic trend without a stationary AR(1) component $Y_t = \beta_1 + \beta_2 t + u_t$, the mean is $\beta_1 + \beta_2 t$. If we subtract this mean from Y_t, the resulting series will be stationary (this procedure is known as polynomial detrending). Hence, processes of this type are called *trend stationary* (TS). In contrast to integrated series, processes with a deterministic trend do not exhibit an infinite memory. The deviations from the trend line do not contribute to the long-

[1] Lag operator *L*: $LY_t = Y_{t-1}$, $L^2Y_t = Y_{t-2}$ and so on. If (1-*L*)=0, we obtain, *L*=1, hence the name unit root

run development of the series. In the case of a stochastic trend, however, the random component a_t affects the long-run course of the series.

Therefore, the proper transformation method of nonstationary data crucially depends on the *data generating process* (DGP). If an empirical time series is a realization of the random walk process, the solution here is to take the first differences of the time series. If a series is stationary around the trend line, the correct way to transform such a time series is to regress it on time. The residuals from this regression will then be stationary. As Figure 5.1.1 shows, realizations of DS and TS processes can appear very similar. It is especially difficult to decide whether the trend in a time series is stochastic or deterministic (cases B and C in Figure 5.1.1). As a result inappropriate transformations are not unusual in practice. Different studies have shown that consequences of mis-specifying trend characteristics of the data can be rather serious.

Chan, Hayya and Ord (1977) analyzed the effects of a wrong transformation on the autocorrelation and the power spectral density functions. The authors showed that the ACF of residuals from linear regression of a random walk series on time are not stationary and tend to exhibit cycles of increasing length and amplitude around a fitted trend line as sample size gets larger. That is why erroneous detrending of DS series is also called *underdifferencing*. The residuals of inappropriate differenced TS series follow a noninvertible moving-average process. This is known as *overdifferencing.* There has been some debate in the literature arguing that overdifferencing is a less serious error than underdifferencing (see Maddala & Kim, 1998, for an overview). Nelson and Kang (1981, 1984) detected artificial periodicity in inappropriately detrended time series and presented several ways in which investigators could obtain misleading results using underdifferenced series. Diebold and Senhadji (1996) showed that applying difference stationary and trend stationary models to the same time series could result in very different predictions. Schenk-Hoppè (2001) and Psaradakis and Sola (2003) demonstrated dramatic consequences of inappropriate detrending within the scope of business

cycle research. In a recent paper, Dagum and Giannerini (2006) investigated the impact of erroneous transformations on tests detecting non-linearity and a unit root. The authors concluded that either a wrong differencing or a wrong detrending lead to biased results. Ashley and Verbrugge (2004, 2006) studied the effects of false transformations in the context of ordinary parameter inference in simple linear models. Underdifferencing yielded seriously over-sized tests. Overdifferencing, in its turn, produced very distorted estimated impulse response functions. Distortions became increasingly severe as sample size gets larger.

The following example demonstrates the impact of inappropriate detrending on the outcome of an interrupted time series experiment. Consider a time series with a stochastic trend as presented in Figure 5.1.1 (B). Suppose that at t=50 an intervention takes place causing an abrupt change in level (+3 units) of the series. As Figure 5.1.2 illustrates, the intervention effect cannot be detected by means of visual inference because the original series drifts slightly downward just at t=50. Table 5. 1. 1 contains the results of the statistical analysis obtained applying the methodology proposed by Glass, Willson and Gottman (1975). In the case of differencing, the ARIMA (0, 1, 0) model accounts for the observed dependency in the series. In the case of detrending, the dependency is captured through the following model: $y = 6.97 + 0.24t + a_t$, $a_t = 0.8a_{t-1} + u_t$, $u_t \sim$ iid N(0, 1). As Table 5.1.1 shows, applying the appropriate transformation to the time series allows detecting the actual change in level (t=2.28, df=97, p<.05). After the erroneous detrending, however, the change in level is heavily underestimated 1.24 (t=1.58, df=97, n.s.)

Table 5.1.1. Example of an interrupted time series analysis.

Method	Estimated Change in Level	t	df
Differencing	2.49	2.28*	97
Detrending	1.24	1.58	97

* p<.05

A Process Development with Intervention at t=50, Intervention effect +3

B Process Development without Intervention

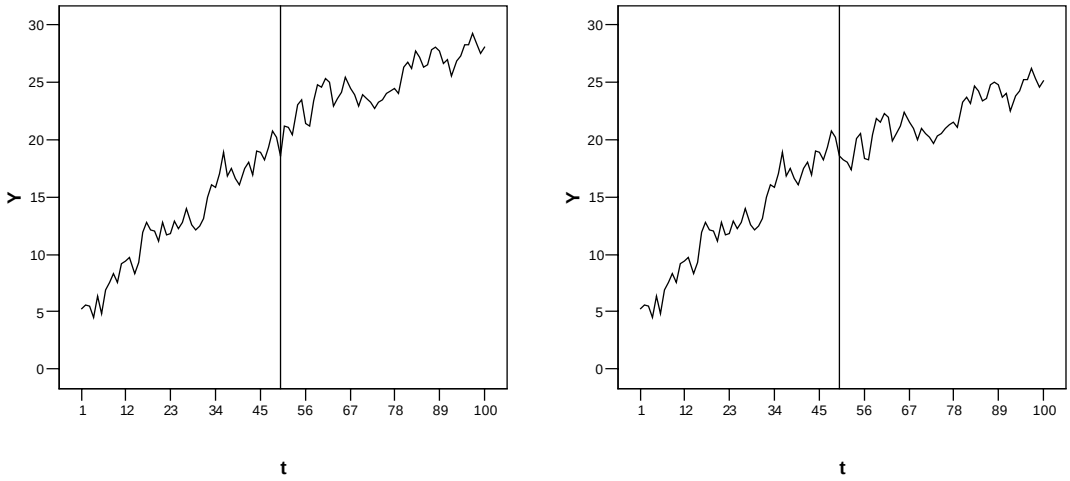


Figure 5.1.2. DS Process $Y_t = 0.2 + Y_{t-1} + u_t$ with and without intervention at t=50.

To summarize, memory characteristics and long-range development of a time series crucially depend on whether its trend component is deterministic or stochastic. A deterministic trend is completely predictable and not variable, a stochastic trend is not predictable. Series with a stochastic trend have to be differenced to make them stationary. For series with a deterministic trend, polynomial detrending is a correct transformation to achieve stationarity. Inappropriate transformation is consequential for subsequent analysis and should be avoided. It is impossible to distinguish between stochastic and deterministic alternatives visually or analyzing ACF and PACF of the series under consideration. The objectives of this study are to develop unit root testing strategies allowing to determine whether an observed time series is TS or DS and to evaluate them by means of Monte Carlo experiments.

5.1.2 Unit Root Testing

Testing for stationarity with *unit root tests* is common practice in econometrics. In the popular time-series textbooks designed for social and behavioral scientists, this topic is not treated. Therefore, the basic concepts of the unit root approach are presented at the beginning (see Gujarati, 2003, for an introduction; see Maddala & Kim, 1998, for an overview; consult Hamilton, 1994, at the advanced level).

There exist numerous unit root tests, one of the most popular among them is the *Augmented Dickey-Fuller* (ADF) test. In the simplest case of an uncorrelated error term, the test begins by estimating the equation $Y_t = \rho Y_{t-1} + u_t$. For theoretical reasons, this equation is written in the form $\Delta Y_t = (\rho - 1)Y_{t-1} + u_t = \delta Y_{t-1} + u_t$ with $\Delta Y_t = Y_t - Y_{t-1}$. The null hypothesis is $\delta=0$, that means $\rho=1$ or there is a unit root, implying the time series under consideration is nonstationary. Dickey and Fuller (1981) have shown that under the null hypothesis the estimated *t* value of δ follows the τ statistics and computed the critical τ values on the basis of Monte Carlo simulations. The authors also introduced a competing *F* test with the usual *F* statistic but using special critical values. Elder and Kennedy (2001a) showed in their recent paper, however, that, in testing for stationarity, the *t* test is preferable.

The actual procedure of implementing the ADF test involves several decisions. To allow for various possibilities of nonstationarity, the ADF test is estimated in three different forms:

(1) Y_t is a random walk $\Delta Y_t = \delta Y_{t-1} + u_t$

(2) Y_t is a random walk with drift $\Delta Y_t = \alpha + \delta Y_{t-1} + u_t$

(3) Y_t is a random walk with drift around a deterministic trend $\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + u_t$.

In each case, the null hypothesis is δ=0. The alternative hypothesis is that δ is less than zero, that is the time series is stationary. If the null hypothesis is rejected, it means in the first case that Y_t is a stationary series with zero mean. In the second case, Y_t is stationary with nonzero mean, and in the third case, Y_t is stationary around a deterministic trend. The critical τ values are *different* for each of the three preceding specifications of the ADF test. If the error term in the model is autocorrelated, the ADF test is conducted by "augmenting" the preceding three equations by adding the lagged values of the dependent variable ΔY_t. Therefore, the third model from the ADF family is written as

$$\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta Y_{t-i} + a_t \ .$$

The number of lagged differenced terms (p) is determined empirically using information criteria such as the AIC and the BIC, the idea being to include enough terms so that the error part a_t is serially uncorrelated. In general, a small p is adequate for autoregressive errors or ARMA processes with small moving-average components. For error terms with moving-average coefficients near ±1, a large p is necessary (see Ng & Perron, 1995, 2001; Lopez, 1997; Stock, 1994, for further details).

As the majority of the unit root procedures, the ADF test suffers from size distortion through misspecification of H_0 or an inappropriate selection of lagged terms (the true significance level exceeds the usual nominal level of α error such as .05 or .01). Another drawback is the low power in smaller samples and in cases of ρ near 1 (the probability of the erroneous maintenance of H_0 or the β error is very large).

5.1.3 Deterministic or Stochastic Trend

This study focuses on a special issue of *growing* time series. As noted previously, such growth could occur due to a deterministic or a stochastic trend. In the former case, there is no unit root, meaning the series under consideration is *I(0)*. In the latter case, we have a unit root with drift or, in other words, the *I(1)* process. The simultaneous existence of a unit root *and* a deterministic trend is thought to be unrealistic (see Elder & Kennedy, 2001b; Holden & Perman, 1994; Perron, 1988, for explanations). Therefore, the first hypothesis from the ADF family (a pure random walk) can be ruled out because it implies that the DGP is not growing. Hence, we have to decide between two competing hypotheses: (1) there is a unit root with drift $\delta=0$ ($\rho=1$), $\alpha\neq0$, and (2) there is a time trend without unit root $\delta<0$ ($\rho<1$), $\beta\neq0$.

An informal testing strategy used among practitioners is to reject a unit root when at least one of the ADF tests rejects. In other words, the null hypothesis of the unit root is dismissed if a test with an intercept does not reject the null, and one including an additional time trend term does (Kim et al., 2004). Perron (1988) introduced a sequential testing strategy and recommended to start from the most general trend specification and to test down to more restricted specifications. Ayat and Burridge (2000) also advocate a sequential testing starting with the highest trend degree maintained. Elder and Kennedy (2001b) propose to begin with the hypothesis of a unit root with drift $\delta=0$ ($\rho=1$), $\alpha\neq0$. If this null is not rejected, there is a stochastic trend. In case of rejection, the series under study is probably stationary around a deterministic time trend because two other possibilities ($\delta\neq0$, $\beta=0$ or $\delta=0$, $\beta\neq0$) are unreasonable and could be ruled out. There exist reasonable doubts, however, that the last strategy works. West (1987) demonstrated that if a linear trend term is incorrectly avoided from the estimating equation, the rejection probability of H_0 converges to 0. In other words, one almost never rejects a unit root hypothesis. It is noteworthy that the inclusion of a

redundant time trend parameter leads to a considerable loss in power of the test if the series under consideration is stationary about an intercept alone (Dickey, 1984). For purposes of the present study, however, this case is irrelevant because the H_1 implies that the DGP is not growing.

In general, there are two competing testing strategies. Both procedures require testing of hypotheses (2) and (3) from the ADF family. The *first strategy* suggests starting with the most specific H_0 of unit root with drift around a deterministic trend (3) and then continuing testing with a more general H_0 of a unit root with drift component only (2). If this strategy works, the ADF test for simulated TS series will reject the third null hypothesis because the series are stationary around a deterministic trend. In other words, a small Type II error probability β or a high power $(1-\beta)$ of the ADF test is expected. Testing of the second H_0 serves as a check. No power at all is expected in this case because the ADF test is known to treat the growth in the series as a signature of the random walk (West, 1987). For simulated random walks with drift, the correct test decision is to maintain the both null hypotheses because the series contain a unit root. Thus, the frequency of the false rejection of the null hypotheses (2) and (3) (i.e., test size) is expected to correspond with the nominal Type I error level. In sum, for TS series different test powers and for DS series similar test sizes in two subtests are expected.

The *second strategy* proposes to begin with the hypothesis of a unit root with drift (2). For TS series, the rejection and, for DS series, the maintenance of the null hypothesis represents the correct test decisions. Therefore, the percentage of significant ADF tests (rejection of H_0) not greater than the nominal significance level (e.g., 5%) is expected for simulated DS series. For series with a linear trend, this percentage reflects the power of the test and should be considerably higher. Within the scope of this strategy, testing of the third null hypothesis serves as a check. For the simulated DS series, the percentage of H_0 rejections

is expected to remain on the same level because the series contain a unit root. For the simulated TS series, a large amount of significant ADF tests is predicted in both trials. In sum, similar test sizes for DS series and comparable powers for TS series are expected in two subtests. Table 5.1.2 presents an overview of the testing strategies.

The following Monte Carlo experiments are designed to evaluate the proposed testing strategies.

Table 5.1.2. Choosing one of two competing models for growing time series: (1) a unit root with drift $\delta=0$ ($\rho=1$), $\alpha\neq0$ and (2) a time trend without unit root $\delta<0$ ($\rho<1$), $\beta\neq0$.

STRATEGY I	**STRATEGY II**
Start with H_0 (3): $\delta=0$ ($\rho=1$), $\alpha\neq0$, $\beta\neq0$ $\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta Y_{t-i} + a_t$ Continue with H_0 (2): $\delta=0$ ($\rho=1$), $\alpha\neq0$ $\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta Y_{t-i} + a_t$	Start with H_0 (2): $\delta=0$ ($\rho=1$), $\alpha\neq0$ $\Delta Y_t = \alpha + \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta Y_{t-i} + a_t$ Continue with H_0 (3): $\delta=0$ ($\rho=1$), $\alpha\neq0$, $\beta\neq0$ $\Delta Y_t = \alpha + \beta t + \delta Y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta Y_{t-i} + a_t$
EXPECTED TEST PERFORMANCE FOR TS SERIES (no unit root): $Y_t = \beta t + \rho Y_{t-1} + a_t$ [$\delta<0$ ($\rho<1$), $\beta\neq0$]	
Rejection of H_0 (3) ➔ high power Maintenance of H_0 (2) ➔ low power Distinct change in power	Rejection of H_0 (2) Rejection of H_0 (3) No considerable change in power
EXPECTED TEST PERFORMANCE FOR DS SERIES (unit root): $Y_t = \alpha + \rho Y_{t-1} + a_t$ [$\delta=0$ ($\rho=1$), $\alpha\neq0$]	
Maintenance of H_0 (3) Maintenance of H_0 (2) Test size close to the nominal significance level	Maintenance of H_0 (2) Maintenance of H_0 (3) Test size close to the nominal significance level

5.1.4 Method

The null hypotheses of a unit root with drift ($\delta=0$, $\alpha\neq0$) and of the unit root with drift around a stochastic trend ($\delta=0$, $\alpha\neq0$, $\beta\neq0$) from the family of ADF tests are examined on 6 DS and 12 TS models. The results are computed for the first 6 lags to demonstrate the influence of the lag length selection on the performance of the ADF test. Series with a stochastic trend are generated applying the following DGP: $Y_t = \alpha + Y_{t-1} + a_t$ with two different values of the drift parameter $\alpha=0.2$ and $\alpha=0.5$. The DGP of series with a deterministic trend is $Y_t = 0.1t + \rho Y_{t-1} + a_t$ with ρ equal to 0, 0.2, 0.5 and 0.8 to account for different possible degrees of serial dependency. To limit the scope of the study, only three types of the error term a_t [white noise or ARMA(0, 0); a first order autoregressive model with $\phi=0.5$ or ARMA(1, 0), and a first order moving-average model with $\theta=-0.5$ or ARMA(0, 1)] are used in DS and TS cases. Such simple error structures are most common in practice. Each simulated time series consists of 100 observations and is replicated 1000 times. Within the scope of the ARIMA methodology, samples of 100 observations represent an optimum length. For effects of sample size on the performance of the ADF test, consult DeJong et al. (1992). The percentage of significant decisions of the ADF test serves as a dependent measurement.

All computations and the generation of independent N(0, 1) innovations u_t are performed with the SAS System for Windows Version 9.1. For the ADF test, the following statement of PROC ARIMA is used: “identify var=*name* stationarity= (adf=6)”. The 5 percent critical τ values for the null hypotheses $\delta=0$, $\alpha\neq0$ and $\delta=0$, $\alpha\neq0$, $\beta\neq0$ are –2.89 and –3.45, respectively.

5.1.5 Results

Results obtained for models with a *stochastic* trend are considered first. In this case, the DGP contains a unit root. Thus, a correct test decision is to maintain the null hypotheses in the both subtests. Table 5.1.3 presents the rejection percentages of the ADF test at a nominal significance level of 5%. Although the inclusion of a redundant time trend parameter (β) in the estimating equation enhances the number of false test decisions, for all simulated models, however, the frequency of H_0 rejections or the test size is close to the nominal level of significance, provided a correct lag length is selected. Recall that for models containing error terms with a moving-average part employing larger lags is appropriate. This explains the values greater than 5% for lags 1 to 5 in the moving-average case (0, 1). The obtained results are in accordance with the both testing strategies since they expect for DS series the adherence to the nominal level of significance and no noticeable discrepancies in the sizes of the two subtests.

Table 5.1.4 contains the percentage of significant ADF tests for models with a *deterministic* trend. In this case, the DGP does not contain a unit root, therefore the rejection of H_0 represents a correct test decision. It can be seen from the left sections of Table 5.1.4 that omitting a time trend term from the estimating regression leads to the lack of any test power. In other words, there is no possibility to reject the null hypothesis of a unit root. Including a deterministic parameter, on the other hand, ensures quite good results. This enormous discrepancy in the power of the two subtests serves as a distinct evidence in favor of the first strategy. In addition to this, the right section of Table 5.1.4 shows that the quality of test decisions is better for low than for high ρ values. As expected, the worst results are obtained for ρ near 1. Recall that $\rho=1$ implies a unit root. For all TS models, the power of the ADF test decreases with the number of lagged terms used.

Table 5.1.3. Percentage of significant decisions of the ADF-Test at the nominal 5%-level of significance for difference stationary series with DGP: $Y_t = \alpha + Y_{t-1} + a_t$; T=100; 1000 replications.

	H_0: δ=0 (ρ=1) α≠0 unit root with drift						H_0: δ=0 (ρ=1) α≠0 β≠0 unit root with drift around a deterministic trend					
a_t Lag:	1	2	3	4	5	6	1	2	3	4	5	6
						α=0.2						
(0, 0)	2.6	2.4	2.0	2.3	3.0	2.7	4.5	5.7	5.0	4.6	4.5	5.3
(1, 0)	4.2	4.0	4.2	4.7	5.1	4.7	5.5	5.6	5.5	4.5	5.1	5.5
(0, 1)	0.2	0.8	0.7	0.6	0.8	0.9	29.0	12.0	7.8	6.2	6.1	4.7
						α=0.5						
(0, 0)	0.6	0.2	0.4	0.6	0.8	0.8	4.5	5.7	5.0	4.6	4.5	5.3
(1, 0)	1.9	1.6	1.7	2.1	2.5	2.3	5.5	5.6	5.5	4.5	5.1	5.5
(0, 1)	0	0.4	0.5	0.3	0.5	0.2	29.0	12.0	7.8	6.2	6.1	4.7

Table 5.1.4. Percentage of significant decisions of the ADF-Test at the nominal 5%-level of significance for trend stationary series with DGP: $Y_t = 0.1t + \rho Y_{t-1} + a_t$; T=100; 1000 replications.

	H_0: δ=0 (ρ=1) α≠0						H_0: δ=0 (ρ=1) α≠0 β≠0					
a_t Lag:	1	2	3	4	5	6	1	2	3	4	5	6
						ρ=0.0						
(0, 0)	0	0	0	0	0	0	100	100	99.1	96.4	86.4	69.0
(1, 0)	0	0	0	0	0	0	99.9	97.2	89.3	75.7	62.6	47.3
(0, 1)	0	0	0	0	0	0	99.9	99.0	98.0	89.9	79.1	62.7
						ρ=0.2						
(0, 0)	0	0	0	0	0	0	100	100	98.4	92.1	80.4	63.7
(1, 0)	0	0	0	0	0	0	99.5	94.5	84.6	69.6	57.4	43.4
(0, 1)	0.2	0	0	0	0	0	99.9	98.2	96.6	84.4	72.8	56.9
						ρ=0.5						
(0, 0)	0	0	0	0	0	0	99.9	97.2	89.3	75.7	62.6	47.3
(1, 0)	1.9	0.2	0.3	0.1	0	0	92.0	78.9	68.2	52.1	43.6	32.6
(0, 1)	2.8	0	0	0	0	0	99.7	87.1	86.4	65.9	56.8	44.8
						ρ=0.8						
(0, 0)	0	0	0	0	0	0	56.8	41.8	35.6	29.8	24.0	18.0
(1, 0)	11.5	7.9	6.0	4.9	4.5	3.5	38.6	30.7	27.2	21.4	18.6	15.7
(0, 1)	11.9	2.2	2.8	1.9	2.7	1.3	76.2	28.0	39.5	24.8	25.8	20.9

5.1.6 Conclusions

In this study the ADF test has been applied to determine whether a trend component in a time series is deterministic or stochastic. Testing for a unit root always implies a testing strategy and not a mere calculation of a single test statistic. The first step in the strategy is to plot data against time and to rule out unreasonable hypotheses on the basis of theoretical considerations (see Elder & Kennedy, 2001b). In the case of growing time series, there exist two plausible competing hypotheses: a unit root with drift and a time trend without a unit root. According to the simulation results, the recommended algorithms is to begin testing with the third most general hypothesis from the ADF family (a unit root with drift and a time trend) and then to continue with the more restricted case of a unit root with drift. The rejection of the null hypothesis in the first test can be treated as a strong evidence for a deterministic trend. If the null is not rejected in both tests, the growth in the observed series is probably due to a stochastic trend. Furthermore, the reported findings confirm that the choice of lag length represents an important issue in the unit root testing.

The presented analysis is limited to rather simple models. Using more complex error structures or data generating processes with nonlinear or broken trends could lead to divergent results (see Kim et al., 2004).

5.2 Study 2: Model Identification of Integrated ARMA Processes

5.2.1 Introduction

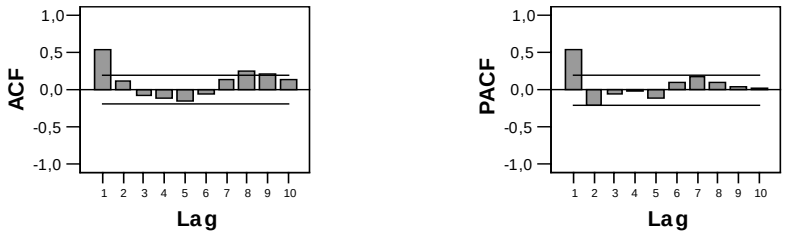
There exist a number of methods for fitting suitable models to a given time series. One of the most widespread techniques is the Box-Jenkins methodology. As described previously (Chapter 2.1.3), this strategy is based on a three-step iterative cycle of model identification, model estimation, and diagnostic checks in model accuracy. Recall that at the identification stage one chooses type and order of the model examining behavior of the sample autocorrelation and the sample partial autocorrelation functions (ACF and PACF). Within the scope of the Box-Jenkins methodology, this step is the most important and problematic issue. Model identification represents the primary goal of the analysis, if the eventual aim of researchers is to construct a model from the empirical data that has similar properties to those of the underlying stochastic process. If the goal of research is to determine the efficacy of a specific intervention, as in interrupted time series analyses, model identification is necessary to remove dependency from the data series so that it meets assumptions of the general linear model.

Table 5.2.1. Theoretical ACF and PACF patterns

Model	ACF	PACF
(0, 0, 0)	0	0
(p, 0, 0)	Decays slowly	0 after p
(0, 0, q)	0 after q	Decays slowly
(p, 0, q)	Decays slowly	Decays slowly
(0, d, 0)	Does not decay	Does not decay

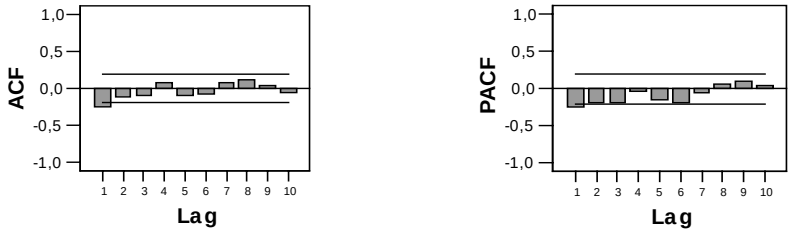
A

(1, 0, 0) ϕ=0.5



B

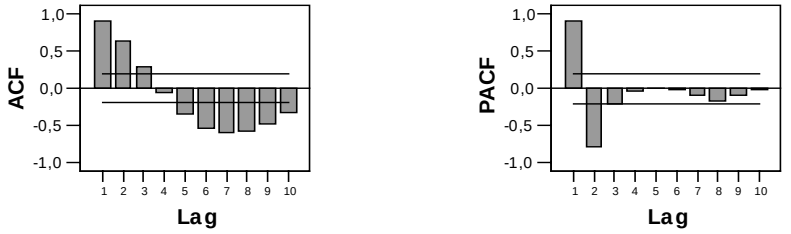
(0, 0, 1) θ=0.5



C

(2, 0, 0)

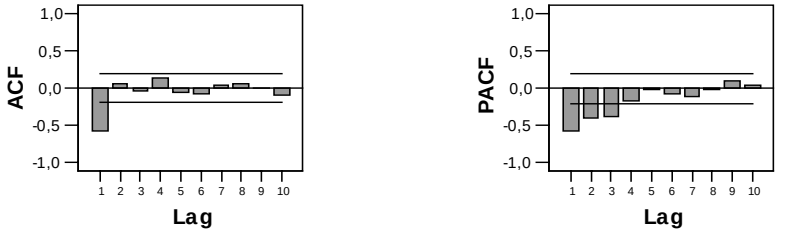
ϕ_1=1.8 ϕ_2=-0.97



D

(0, 0, 2)

θ_1=1.8 θ_2=-0.97



E

(1, 0, 1)

ϕ=0.6 θ=-0.4

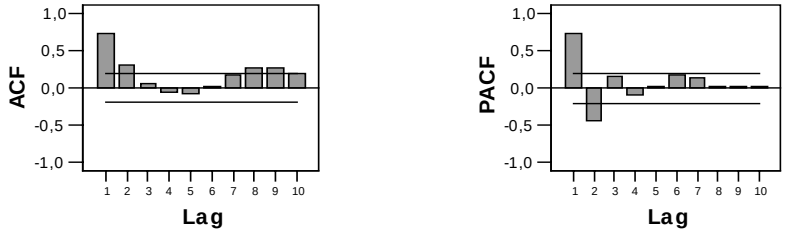


Figure 5.2.1. ACF and PACF of simulated models with T=100.

Table 5.2.1 describes theoretical ACF and PACF patterns of some common ARMA processes. Figure 5.2.1 illustrates that *estimates* of the ACF and PACF, received from finite samples, can be rather ambiguous, even under ideal conditions. Furthermore, the quality of empirical ACF und PACF crucially depends on the number of observations in a series and is sensitive to outliers or other departures from ideal assumptions. As can be seen from Figure 5.2.1(E), the Box-Jenkins method is not very useful for identifying mixed ARMA models if p and q are unequal 0. The reason for the difficulty is that the ACF and PACF of mixed models tail off to infinity rather than cut off at a particular lag. Therefore, model identification using the Box-Jenkins approach is a complicated and problematic task requiring many data points and a great deal of expertise from a researcher.

Velicer and Harrop (1983) evaluated the performance of the Box-Jenkins model identification technique employing 12 extensively trained subjects and found a disappointingly low overall accuracy rate of 28%. The factors affecting the quality of identification were the length of a time series (the increase from 40 to 100 observations improved the percentage of correct identification from 20 to 36), the degree of dependency (higher dependency was favorable) and the complexity of an examined model (accuracy was better for simple models). Identifying *integrated* models turned out to be the most complicated issue. Extensively trained judges were only able to correctly identify ARIMA (0, 1, 0) and (0, 1, 1) models 4% and 13% of the time, respectively.

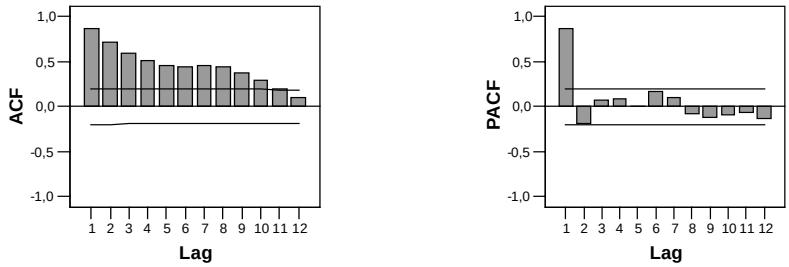
As a response to the described drawbacks of the Box-Jenkins technique, alternative procedures for removal of dependency from the data series have been proposed (Algina & Swaminathan, 1977, 1979; Simonton, 1977; Velicer & McDonald, 1984). It has been shown that the model identification step can be successfully abandoned in interrupted time series analysis (see Chapter 3.2). In various research cases, however, model identification represents the primary goal of the analysis and cannot be avoided (see studies of Fortes et al., 2004;

Rosel & Elósegui, 1994, and Velicer et al., 1992, from Chapter 4). Recall that autoregressive models are characteristic of systems containing internal temporal regularity, whereas moving average models are typical for unstable systems depending on external and occasional events. In a study analyzing travel behavior of different population groups, Fraschini and Axausten (2001) found out that the AR models predominated in the age classes between 35 and 65 years old, reflecting a more regular behavior probably caused by a fixed employment and a more settled lifestyle. On the other hand, the MA patterns were mainly identified for younger participants, indicating the presence of external influences on the subject behavior.

Integrated series constitute a widespread phenomenon among behavioral or psychological series. Glass et al. (1975) reported that out of 95 time series taken from a wide range of application in the social sciences 44 were nonstationary in level. Recall that the ARIMA (0, 1, 1) is characteristic of time series exhibiting noisy fluctuations around a slowly varying mean (Spray & Newell, 1986). In combination with a unit root, the positive moving average coefficient θ seems to determine the balance between the preservation and adaptation processes (Fortes et al., 2004). Integrated models are typical for processes with an infinite memory.

For integrated processes, the identifying procedure consists of two stages. The first step is to decide whether differencing is necessary or not. As Figure 5.2.2 illustrates, it can be a rather distressing task since a unit root series (ϕ=1) is scarcely distinguishable from a stationary autoregressive series with ϕ=0.9. Depending on the outcome of the first stage, the next step is to infer the ARMA model by inspecting ACF and PACF of either the original or the differenced series.

A Autoregressive Process



B Random Walk Process

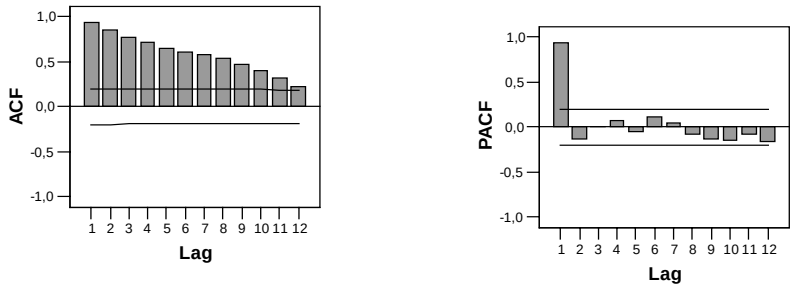


Figure 5.2.2. ACF and PACF of stationary and integrated processes: (A) Autoregressive Process with ϕ=0.9, (B) Random Walk Process.

Numerous *automated* alternatives to the Box-Jenkins approach have been developed during the last three decades in order to make the model identification process more accurate and less subjective. Choi (1992) published a survey cataloguing many of the procedures proposed in the literature. Some of the described techniques, such as *Extended Sample Autocorrelation Function* (ESACF) and *Smallest Canonical Correlation Method* (SCAN) proposed by Tsay and Tiao (1984; 1985), can be applied for identifying both stationary and nonstationary models. This means that using ESACF and SCAN eliminates the need to determine the order of differencing necessary to produce stationarity in ARIMA modeling. Besides, SCAN and ESACF are commonly available in current versions of SAS for Windows.

Table 5.2.2.Theoretical ESACF table for ARIMA (1, 1, 1)[1].

Autoregressive Order (p+d)	Moving-Average Order (q)				
	0	1	2	3	4
0	X	X	X	X	X
1	X	X	X	X	X
2	X	**0**	0	0	0
3	X	X	0	0	0
4	X	X	X	0	0

X = significant value, 0 = insignificant value.

The idea behind the standard ESACF approach is to identify the orders of an ARMA process employing iterated least squares estimates of the autoregressive parameters. First, estimates for the autoregressive components are obtained and removed from the data. Then the order of the remaining moving average component is determined from the transformed series. Since the order of the autoregressive component is never known in advance, an array of autocorrelations from series for which AR (*m*) components have been removed must be calculated. These autocorrelations are termed extended sample autocorrelations (ESAC). The order of the time series is tentatively determined from the shape of the zero and nonzero elements in an overall ESAC array. The vertex of the triangle of zero values identifies the order of the process. Table 5.2.2 depicts the theoretical pattern associated with an ARIMA (1, 1, 1) series. If an empirical ESACF table contains more than one triangular region in which all elements are insignificant (zero values), SAS gives out several recommendations ordered by the number of insignificant terms contained in the triangle. The first

[1] The same pattern applies to ARIMA (2, 0, 1), since both ARIMA (1, 1, 1) and ARIMA (2, 0, 1) represent the same ARMA [*P*=*p*+*d*; *q*]: [2, 1] model (see p. 78 for further explanations).

recommendation is a model with a maximal triangular pattern. For a detailed description of the ESACF method, see Choi (1992), Tsay and Tiao (1984, 1990), and SAS (1999).

The SCAN method of Tsay and Tiao (1985) employs canonical correlation for ARMA model identification. The procedure consists in analyzing eigenvalues of the correlation matrix of the ARMA process. The first step is to compute the (m+1) x (m+1) matrix containing covariances and variances of the vectors $\mathbf{y}_{m,t}$ and $\mathbf{y}_{m,t-j-1}$, where t ranges from j+m+2 to T. $m = p_{\min}, \ldots, p_{\max}$ is the autoregressive test order, $j = q_{\min}, \ldots, q_{\max}$ is the moving-average test order. $\mathbf{y}_{m,t}=(\tilde{Y}_t, \tilde{Y}_{t-1},..., \tilde{Y}_{t-m})'$, where $\tilde{Y}_t$ is a mean corrected series $\tilde{Y}_t = Y_t - \mu$ with $1 \leq t \leq T$. The *smallest* eigenvalue of this matrix serves as the squared canonical correlation estimate for model (m, j). Finding a rectangular pattern in which the canonical correlation estimates are insignificant for all specified test orders ($m \geq p+d$, $j \geq q$) then identifies the ARMA model. Table 5.2.3 shows the theoretical pattern associated with an ARIMA (1, 1, 1) process. If there is more than one zero rectangular in an empirical SCAN table, parsimony and the number of insignificant items in the rectangular pattern determine the model order. For more details about the SCAN method, consult Box et al. (1994), Choi (1992), Tsay and Tiao (1985), Werner (2005) and SAS (1999).

Table 5. 2. 3. Theoretical SCAN table for ARIMA (1, 1, 1).

Autoregressive Order (p+d)	Moving Average Order (q)				
	0	1	2	3	4
0	X	X	X	X	X
1	X	X	X	X	X
2	X	**0**	0	0	0
3	X	0	0	0	0
4	X	0	0	0	0

X = significant value, 0 = insignificant value.

For integrated series, both ESACF and SCAN can be applied to the original nondifferenced data. As a consequence, an ARIMA (p, d, q) process is identified as an ARMA [P, q][1] model with $P=p+d$. This implies that different ARIMA processes such as (1, 1, 1) and (2, 0, 1) represent the same ARMA [2, 1] model in terms of ESACF and SCAN. To decide whether a series is stationary or not, Tsay and Tiao (1984; 1985) suggest examining the iterated AR estimates for given specified values of p and q. Suppose [2, 0] model is identified, then one can obtain the estimates of the autoregressive parameters. For example, the parameters ϕ_1=1.81 and ϕ_2=-0.82 strongly indicate the presence of a unit root in the series or the (1, 1, 0) model (Tsay & Tiao, 1984, p. 88). Wei (1990, 1994) argues, however, that identifying nonstationarity through this approach is generally difficult and recommends to apply ESACF and SCAN to a properly transformed stationary series.

In addition to ESACF and SCAN, another method called MINIC is available in SAS for automated ARMA model selection. MINIC follows an algorithm suggested by Hannan and Rissanen (1982). It combines the regression technique and the penalty functions AIC and BIC for modeling *stationary* and *invertible* ARMA (p, q) processes. This implies that the integrated series must be differenced before applying MINIC to it. The MINIC procedure consists of three steps. The first step is to fit a high order AR model to the observations. The choice of the autoregressive parameter is determined by the order that minimizes the AIC. The second step is to apply the OLS method to the series and the estimated innovations of the fitted AR model. As a result, m autoregressive and j moving average OLS estimates are obtained ($m = p_{min}, \ldots, p_{max}$ is the autoregressive test order, $j = q_{min}, \ldots, q_{max}$ is the moving-average test order). As the last step, the BIC is computed for each of m x j ARMA models. A model with the smallest BIC is used as a MINIC recommendation. For a detailed treatment of

[1] In the following square brackets are used for [P, q] models where $P=p+d$.

the Hannan and Rissanen's method, consult Choi (1992), Hannan and Rissanen (1982), and SAS (1999).

Dickey (2004) compared the performance of ESACF, SCAN, and MINIC on 600 stationary ARIMA (1, 0, 1) series each of length T=500. SCAN showed the best results with 461 correctly identified series. ESACF did slightly worse (441 correct identifications). Both SCAN and ESACF were superior to MINIC (252 correct classifications). The methods almost never *under*estimated p or q when T=500. The same experiment with series of length 50 resulted in 203, 65, and 53 correct identification for SCAN, ESACF, and MINIC, respectively. Besides, the simulation study showed that the complexity of the model affected the performance of the identification methods. For 600 replicates of the ARMA model with ϕ_4=0.5 and θ_1=-0.3 using T=50, the correct model was rarely chosen by any technique. The number of correct selections was 10 for MINIC, 52 for ESACF, and no correct choices for SCAN.

Koreisha and Yoshimoto (1991) conducted Monte Carlo experiments with three identification procedures including ESACF, the Corner Method, and the Autoregressive Order Determination Criterion (ARCRI), an approach similar to MINIC. Only stationary models were considered. The ARCRI outperformed the other two methods regardless of sample size and model structure. The ESACF method performed poorly and showed the tendency to overestimate the order of the process. The results were better for MA than for AR models. ESACF's power in selecting the order of mixed processes was low, the percentage of correct identifications was between 21 and 62. The increase in number of observations did not improve the performance of the ESACF approach. In many of the cases the percentage of correctly identified structures was even higher for smaller sample sizes.

Stadnytska, Braun and Werner (2006) compared the performance of SCAN, ESACF, and MINIC for stationary models with different parameterizations, degrees of dependency,

and sample sizes. For autoregressive models, MINIC achieved the best results. SCAN was superior to the other two procedures for mixed models. For moving-average processes, ESACF obtained the most correct selections. For all three methods, model identification was less accurate for low dependency than for medium or high dependency processes. The effect of sample size was more pronounced for MINIC than for SCAN and ESACF. MINIC and SCAN had difficulty in identifying moving-average models. ESACF demonstrated low power in autoregressive cases. SCAN and ESACF showed the tendency to select higher order mixed structures. For autoregressive processes of second order, both SCAN and ESACF performed better in narrowband than in broadband cases. All three methods were superior to subjective judgments as reported by Velicer and Harrop (1983). As Table 5.2.4 shows, the superiority of automated methods was especially pronounced for autoregressive models of second order.

To my knowledge, there are no studies evaluating the performance of SCAN, ESACF and MINIC for integrated models.

Table 5.2.4. Percentage of correct model identifications.

Type of Model	T	Subjective Judgments	Automated Methods*		
			MINIC	SCAN	ESACF
	40	19	47	64	54
(1, 0, 0)	100	46	79	72	50
	mean	32	63	68	52
	40	46	25	46	61
(0, 0, 1)	100	67	51	63	78
	mean	56	38	54	69
	40	0	50	55	62
(2, 0, 0)	100	4	82	69	54
	mean	2	66	62	58
		30	***56***	***61***	***60***

* Only similar models and parameterizations are considered.

To summarize, the model identification is one of the most challenging issues in the ARIMA modeling, especially if the analyzed series is integrated. Two automated procedures, SCAN and ESACF, can be used for identifying both stationary and nonstationary models. The aims of this study are: (a) to evaluate SCAN and ESACF as model identification tools for integrated processes, (b) to compare the performance of SCAN and ESACF with the results of MINIC for differenced (i.e. stationary) series, (c) to examine the influence of sample size, complexity of a model and degree of dependency in a time series on the efficiency of these methods.

5.2.2 Method

The performance of the identification procedures is studied on six types of nonstationary ARMA models: first and second order autoregressive, first and second order moving-average, mixed processes and random walk. Higher order series are not considered in the Monte Carlo experiments because they have been rarely found in the social and behavioral sciences (Glass et al., 1975; Marsh & Shibano, 1984; Rankin & Marsh, 1985; Revenstorf et al., 1980). In the first order models, ϕ or θ values of ±0.2, ±0.5 and ±0.8 are used to represent different possible degrees of dependency. The parameter values of the second order models are the ones used in the evaluation study of Koreisha and Yoshimoto (1991). The parameters ϕ and θ of the simulated mixed models have opposite signs, so that the corresponding autoregressive and moving-average terms in the equation

$$a_t = \phi_1 a_{t-1} + u_t - \theta_1 u_{t-1}$$

do not approach cancellation. For example, at ϕ=0.8, θ=0.8 cancellation would take place and the apparent ARMA (1, 1) process would be in fact ARMA (0, 0). In cases of near-cancellation, the process may be well approximated by a parsimonious equivalent mathematical representation leading to difficulty in evaluating the model identification

methods. Such cases are ruled out by our design. The length of the simulated time series is varied between 50 and 200. The traditional Box-Jenkins approach recommends as a guideline a minimum sample size of 50 observations (Box & Pierce, 1970; Glass et al., 1975; Granger & Newbold, 1986; Ljung & Box, 1978; McCleary & Hay, 1980; Velicer & Fava, 2003). Within the scope of the ARIMA methodology samples of 100 or 200 observations represent an optimum length. Each simulated model is replicated 1000 times.

The identification procedure is applied to both nonstationary and differenced series. As a quality criterion, the percentage of correct model selections is computed. In addition, a detailed analysis of incorrect model identifications is conducted. For SCAN and ESACF, only first recommendations in SAS output table are considered.

All computations and the generation of independent N(0, 1) innovations u_t are performed with SAS System for Windows Version 9.1. In the IDENTIFY statement of PROC ARIMA default dimensions of SAS are used.

5.2.3 Results

Tables 5.2.5 and 5.2.6 display the percentage of correct model selections of SCAN and ESACF for (1, 1, 0) and (0, 1, 1) processes. In the first place, two observations can be made. One is that the performance of both methods is less accurate for low dependency than for medium or high dependency models. Second, the accuracy of SCAN and ESAC does not appear to be strongly dependent on sample size. For positive and negative parameterizations, similar results are obtained.

As Table 5.2.6 shows, the performance of ESACF is better in moving-average than in autoregressive cases. There is no manifest distinction between integrated and differenced models. For SCAN, however, the interaction of the independent variables is apparent (see Table 5.2.5). For high-valued parameterizations, the percentage of correct selections is about

70% for both differenced and integrated models independent of the ARMA structure. For low and medium degrees of dependency, differencing is favorable in autoregressive cases. In moving-average cases, on the contrary, better results are achieved for integrated than for differenced models.

Table 5.2.5. Accuracy (percentage correct) of SCAN, based on 1000 replications.

Type of Model	T	ϕ,θ:	Differenced Models -0.2	-0.5	-0.8	Integrated Models -0.2	-0.5	-0.8
	50		24.4	70.7	73.9	4.0	18.9	67.1
(1, 1, 0)	100		42.8	73.8	72.5	5.3	40.9	73.8
	200		62.1	74.0	71.9	9.9	66.0	72.1
	50		8.8	31.1	63.1	29.9	71.1	74.0
(0, 1, 1)	100		7.2	50.2	73.3	42.9	75.9	75.8
	200		10.7	67.6	76.0	62.7	77.8	79.1

Table 5.2.6. Accuracy (percentage correct) of ESACF, based on 1000 replications.

Type of Model	T	ϕ,θ:	Differenced Models -0.2	-0.5	-0.8	Integrated Models -0.2	-0.5	-0.8
	50		10.6	42.9	65.3	8.6	33.6	56.4
(1, 1, 0)	100		7.7	35.2	57.7	5.9	30.3	51.9
	200		7.5	29.8	52.2	8.4	28.6	50.7
	50		10.8	51.7	78.2	17.5	54.4	78.8
(0, 1, 1)	100		29.5	71.7	78.4	31.3	71.7	75.6
	200		53.8	77.9	78.0	52.8	75.6	74.8

Table 5.2.7 illustrates the accuracy of MINIC for differenced models of first order. The percentage of correct selections is distinctly higher for structures with medium and large parameter values. Furthermore, the performance of MINIC is better in autoregressive than in moving-average cases. For all models and parameterizations, the increase in number of observations has a much more pronounced effect for MINIC than for SCAN or ESACF.

Table 5.2.8 compares the accuracy of the identification methods for different first-order series with T=100. For autoregressive series, better results are achieved for differenced than for integrated models: MINIC and SCAN outperformed ESACF (34%) with more than 60 % of correct selections (average value for low, medium and high degrees of dependency). In the moving-average case, the best accuracy is obtained by ESACF, the results are similar for integrated and differenced models.

Table 5.2.7. Accuracy (percentage correct) of MINIC for differenced models, based on 1000 replications.

Type of Model	T	Degree of Dependency		
		Low (-0.2)	Medium (-0.5)	High (-0.8)
(1, 1, 0)	50	18.2	52.9	66.1
	100	31.0	74.7	83.4
	200	46.1	87.2	90.9
(0, 1, 1)	50	9.1	31.4	33.9
	100	17.3	49.5	53.0
	200	32.7	66.3	64.2

Table 5. 2. 8. Percentage of correct model identifications for first-order models with T=100, based on 1000 replications.

Type of Model	Dependency	Differenced Models			Integrated Models	
		MINIC	SCAN	ESACF	SCAN	ESACF
(1, 1, 0)	low	31.0	42.8	7.7	5.3	5.9
	medium	74.7	73.8	35.2	40.9	30.3
	high	83.4	72.5	57.7	73.8	51.9
(0, 1, 1)	low	17.3	7.2	29.5	42.9	31.5
	medium	49.6	50.2	71.7	75.9	71.7
	high	53.0	73.3	78.4	75.8	75.6

It can be seen from the upper section of Table 5.2.9 that SCAN and ESACF are relatively accurate in identifying the random walk process. On the average, the percentage of correct selections is about 70%. As Table 5.2.9 shows, there is no manifest distinction between integrated and differenced models for the (2, 1, 0) processes. It is noteworthy that the performance of MINIC for differenced (2, 1, 0) models visibly improves with the increase in sample size. For *T*=200, MINIC is distinctly superior to both SCAN and ESACF with more than 90% of correct selections. The accuracy of MINIC and SCAN in the (0, 1, 2) cases is disappointing regardless of sample size and parameterization (maximum 17% and 31.7 % of correct selections for MINIC and SCAN, respectively). The performance of ESACF for the (0, 1, 2) processes is better for differenced than for integrated models and appears to be dependent on sample size. In the (1, 1, 1) case the results are similar for both parameterizations. SCAN and ESACF perform slightly better for differenced than for integrated series. For mixed models, SCAN yields the best accuracy.

Figure 5.2.3 visualizes the performance of SCAN and ESACF in differenced and integrated cases. On the average, similar results are obtained for integrated and differenced series.

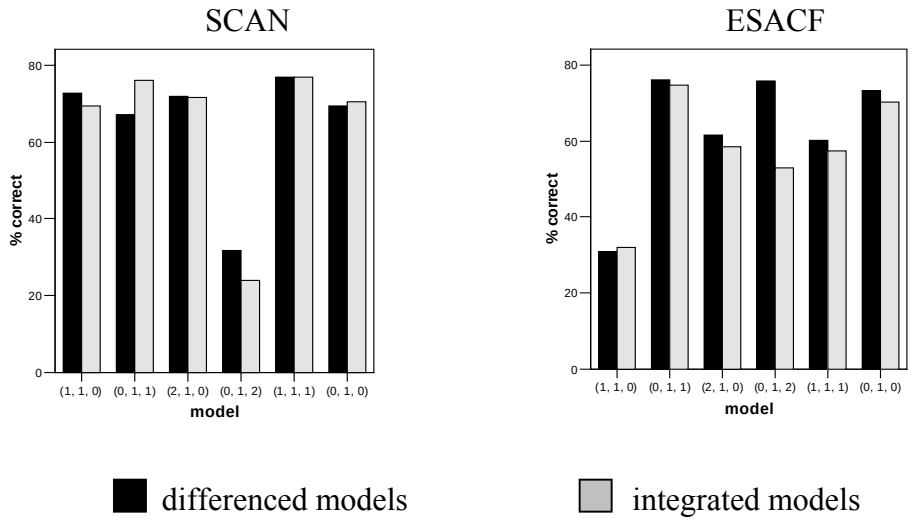


Figure 5.2.3. Comparison results for differenced and integrated models, T=200.

Table 5.2.9. Comparison results (percentage correct) for random walk, second-order, and mixed models, based on 1000 replications.

		Differenced Models			Integrated Models	
Type of Model	T	MINIC	SCAN	ESACF	SCAN	ESACF
(0, 1, 0)	50	67.4	67.0	80.0	67.8	70.8
	100	84.2	67.6	74.6	68.3	67.5
	200	92.1	69.4	73.3	70.6	70.3
(2, 1, 0) ϕ_1=1.8 ϕ_1=-0.9	50	65.7	73.9	71.6	68.6	67.7
	100	81.3	73.3	65.9	73.9	63.7
	200	91.5	71.8	61.7	71.7	58.5
(2, 1, 0) ϕ_1=1.42 ϕ_2=-0.73	50	61.7	66.8	57.8	24.6	52.7
	100	84.3	75.0	48.0	60.7	45.3
	200	91.3	71.5	44.5	71.8	40.9
(0, 1, 2) θ_1=1.8 θ_2=-0.9	50	2.0	7.6	13.2	2.0	0.8
	100	2.2	10.3	47.4	7.0	6.5
	200	2.1	11.1	69.1	15.0	31.3
(0, 1, 2) θ_1=1.42 θ_2=-0.73	50	5.0	10.3	15.7	5.7	5.0
	100	8.7	19.3	60.3	13.4	27.7
	200	17.0	31.7	75.7	24.0	52.8
(1, 1, 1) ϕ=0.8 θ=-0.7	50	22.0	70.9	60.7	69.9	51.5
	100	47.6	79.0	66.2	75.0	60.0
	200	63.8	77.0	60.1	76.9	57.4
(1, 1, 1) ϕ=-0.8 θ=0.7	50	25.7	77.4	68.5	20.5	36.8
	100	48.0	80.2	66.9	67.9	58.3
	200	64.2	79.0	60.6	79.6	58.0

Figures 5.2.4 to 5.2.5 present the frequency distribution of identified models for integrated processes. In addition to the correct identifications, the figures provide a detailed analysis of incorrect selections. As can be seen from the left section of Figure 5.2.4, the most common error of both SCAN and ESACF for the ARIMA(1, 1, 0) with low-valued parameterizations is the underestimation of $P=p+d$. Recall that in terms of SCAN and ESACF the [2, 0] model represents the correct identification for the (1, 1, 0) process. In the (1, 1, 0) case with $\phi=\pm0.2$, about 70% of the simulated series are identified as either [1, 0] or [1, 1] models. For $\phi=\pm0.5$, the [1, 1] model serves as the most frequent incorrect selection. In the (1, 1, 0) case with $\phi=\pm0.8$, both SCAN and ESACF show the tendency to overparametrization with [2, 1] and [3, 1] models as the most frequent incorrect selections. For the (0, 1, 1) process with $\theta=\pm0.2$, both methods select the [1, 0] structure as the most common identification alternative to the true model. Note that in the moving-average case, this structure does not underestimate the order of P, since the correct selection is [1, 1]. For medium and high degrees of dependency, both SCAN and ESAC show the tendency to overestimate the order of P with [P=2, q] structures as the most frequent incorrect selections.

It can be seen from the bottom of Figure 5.2.5 that the performance of SCAN and ESACF for the random walk process appears to be very similar. The most incorrect selections for both methods are [1, 1] or [2, 1] structures. As Figure 5.2.5 shows, SCAN outperformed ESACF in the (1, 1, 1) case. For SCAN, the [2, 2] model represents the most frequent incorrect selection. Depending on sample size, ESACF selected [1, 1], [2, 2] or [3, 2] structures as the most common identification alternatives to the true model. For the second order models, SCAN performs better in the autoregressive than in the moving-average case. For the (2, 1, 0) model, increasing the number of observation markedly improves the accuracy of SCAN. In smaller samples, the most widespread incorrect selections are either [1, 2] or [2, 2] structures. For T=200, the overestimation of P represents the most common error. In the

moving-average case (0, 1, 2), the accuracy of SCAN appears to be extremely low. The majority of series from samples of 50 and 100 observations is incorrectly identified as either [0, 1] or [1, 1] structures. About 30% of series with T=200 are classified as [2, 1] models. Increasing time series length has a positive impact on the performance of ESACF in the (0, 1, 2) case. Identification of the correct structure improves dramatically from 5.0% to 52.8%. The most frequently selected incorrect model is [0, 1]. For the (2, 1, 0) process, however, the effect of sample size is much less pronounced: the increase in the number of observations reduces the percentage of correct identifications from 57.8% to 44.5%. Depending on sample size ESACF selected [2, 1], [3, 1] and [3, 2] or [4, 1] structures as the most common identification alternatives to the true model.

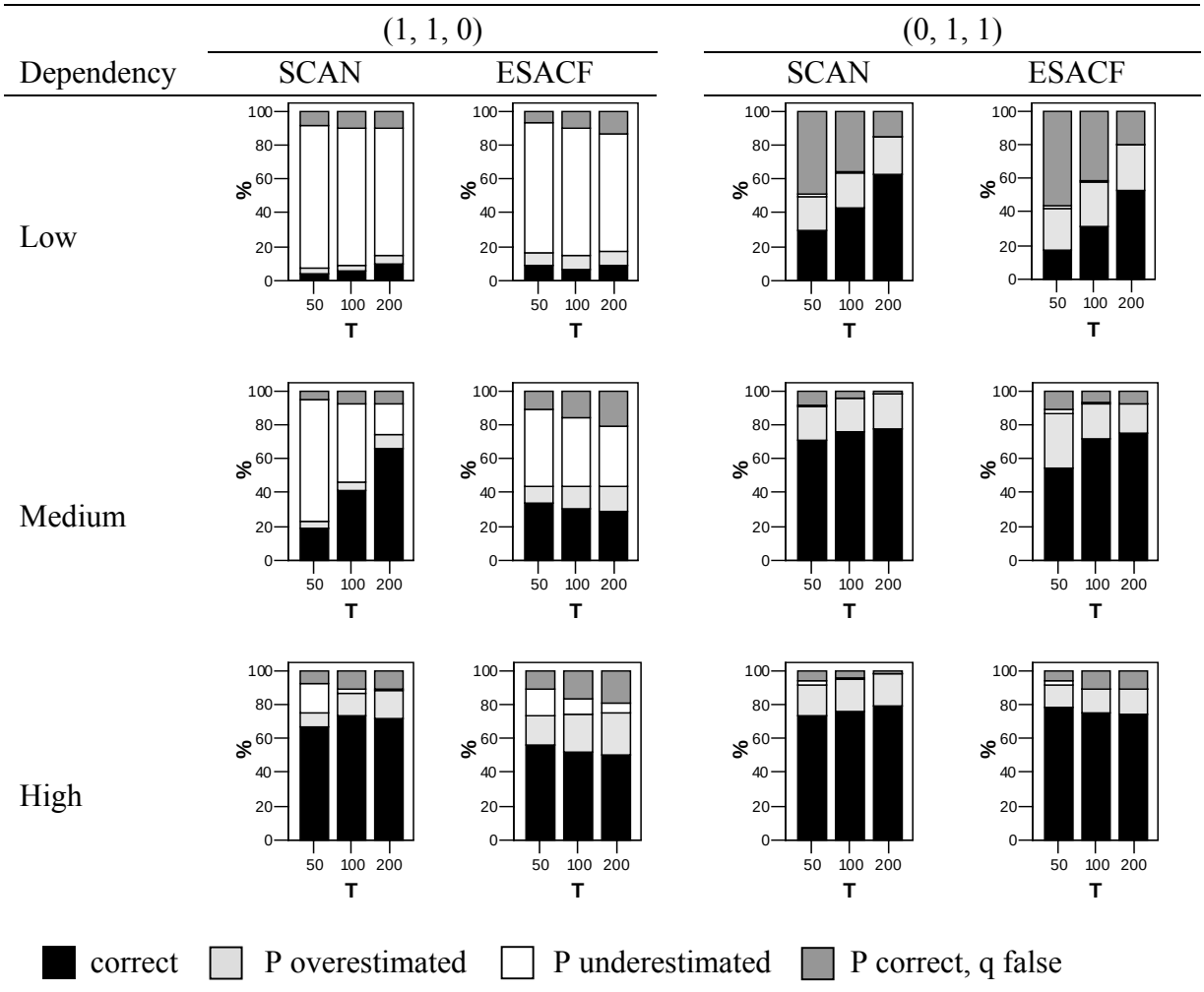


Figure 5.2.4. Results for integrated models of first-order, based on 1000 replications.

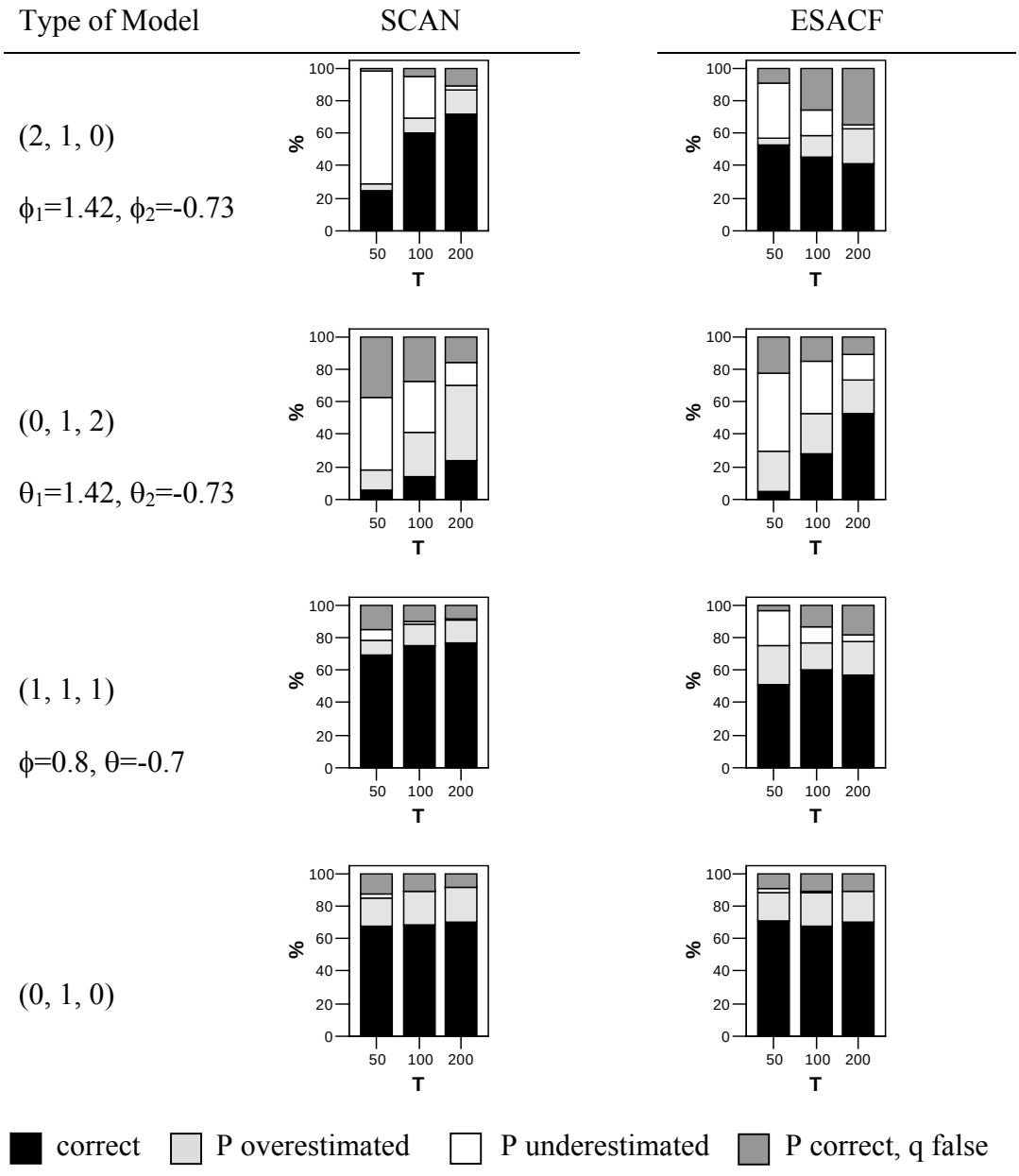


Figure 5.2.5. Results for various integrated models, based on 1000 replications.

Figures 5.2.6 to 5.2.9 present the frequency distribution of identified models for *differenced* series. As Figure 5.2.6 illustrates, all three methods have difficulty in identifying autoregressive structures with low-valued parameterizations. The procedures select either white noise or a MA(1) process as identification alternatives to the simulated structure. Note that white noise and a moving-average process of first order are parsimonious nearly equivalent mathematical representations of an AR(1) model with ϕ values near zero. Increasing the sample size visibly refines the accuracy of MINIC and SCAN. The

performance of ESACF, however, remains disappointing. It is noteworthy that the number of observations exercises an influence on the type of incorrect selections of ESACF. For instance, in smaller samples (T=50) 60% of series are identified as white noise and only 14% as MA(1) models. For T=200, the MA (1) model represents the most widespread incorrect selection (55%) followed by white noise (14%). For autoregressive processes with large parameter values, MINIC and SCAN outperform ESACF. As compared to MINIC, SCAN and ESACF select a distinctly larger percentage of higher order mixed structures.

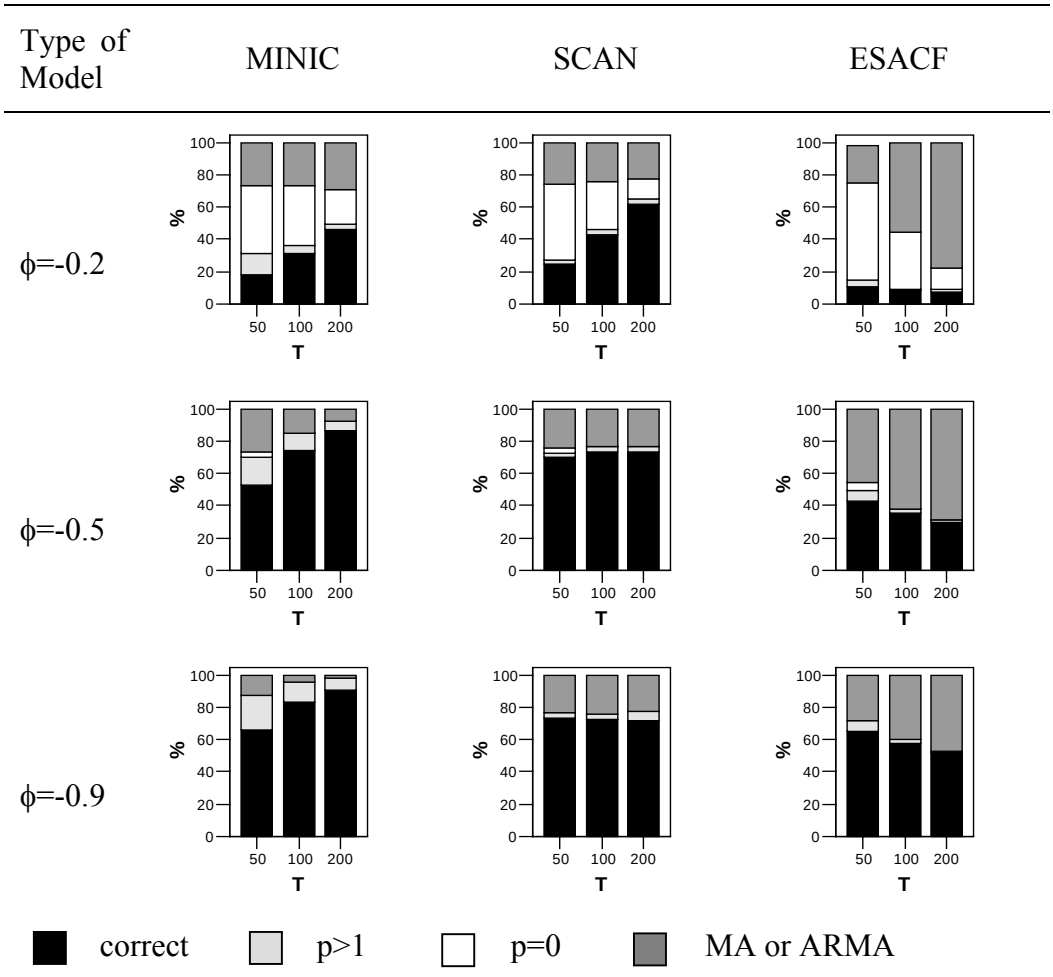


Figure 5. 2. 6. Results for differenced (1, 1, 0) models, based on 1000 replications.

As Figure 5.2.7 shows, in the moving-average case ESACF outperforms the two other procedures for differenced series. For θ=-0.2, MINIC and SCAN select in the majority of the cases either white noise or an AR(1) process to describe the behavior of the data. For processes with medium and large parameter values, all approaches identify pure autoregressive structures of higher order as the most common alternatives to the (0, 0, 1) model. For larger samples, SCAN and ESACF show the tendency to select mixed structures.

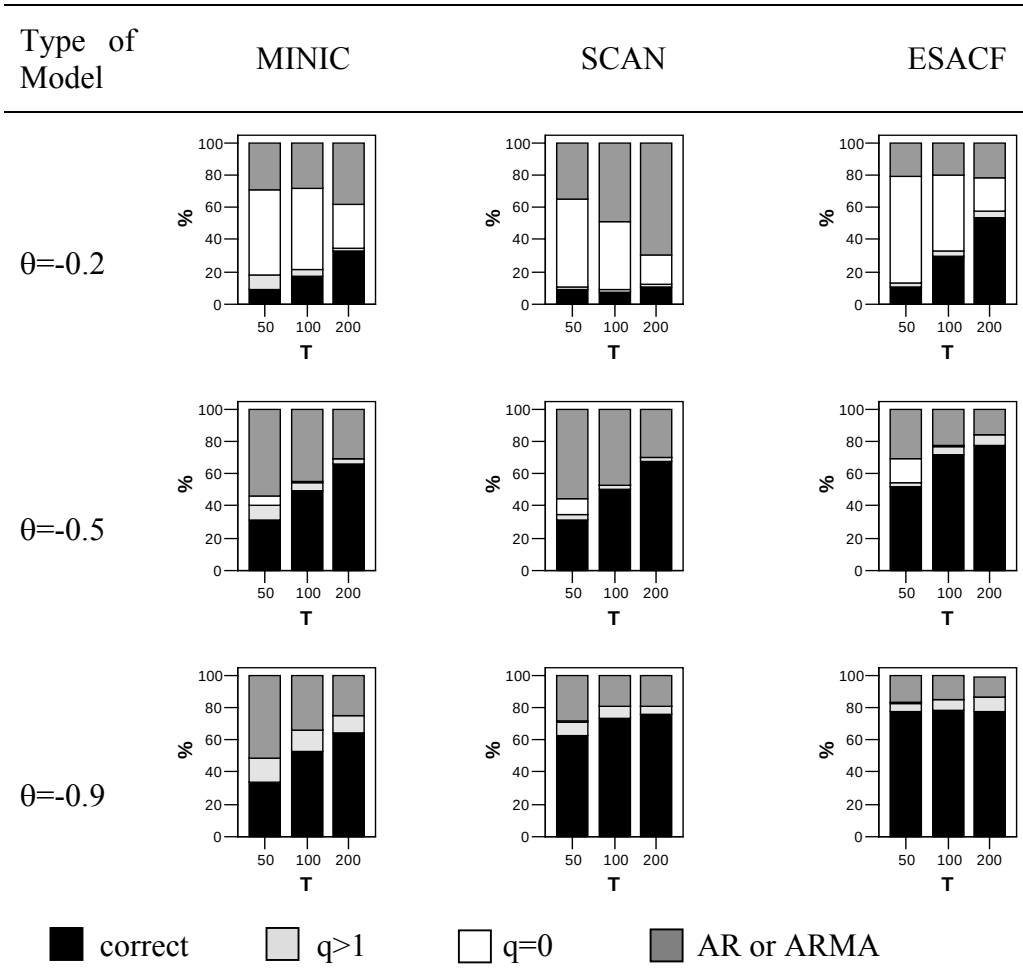


Figure 5. 2. 7. Results for differenced (0, 1, 1) models, based on 1000 replications.

Figure 5.2.8 presents results for differenced models of second order. All procedures perform better in autoregressive than in moving-average cases. For the AR(2) process, MINIC is superior to both SCAN and ESACF. The quality of model identifications of the MINIC approach becomes more exact with increasing sample size, the number of incorrect selections practically vanished. The accuracy of SCAN and ESACF does not appear to be dependent on sample size. Both methods demonstrate the tendency to overparametrization. In the MA(2) case, ESACF outperforms two other procedures. The MA(1) model represents the most frequent incorrect selection of ESACF. In smaller samples SCAN identifies the majority of (0, 0, 2) series as (0, 0, 1) models. The most widespread incorrect selection of SCAN for T=200 is the ARIMA(1, 0, 1). MINIC tends to select autoregressive structures of higher order in place of the MA(2) model.

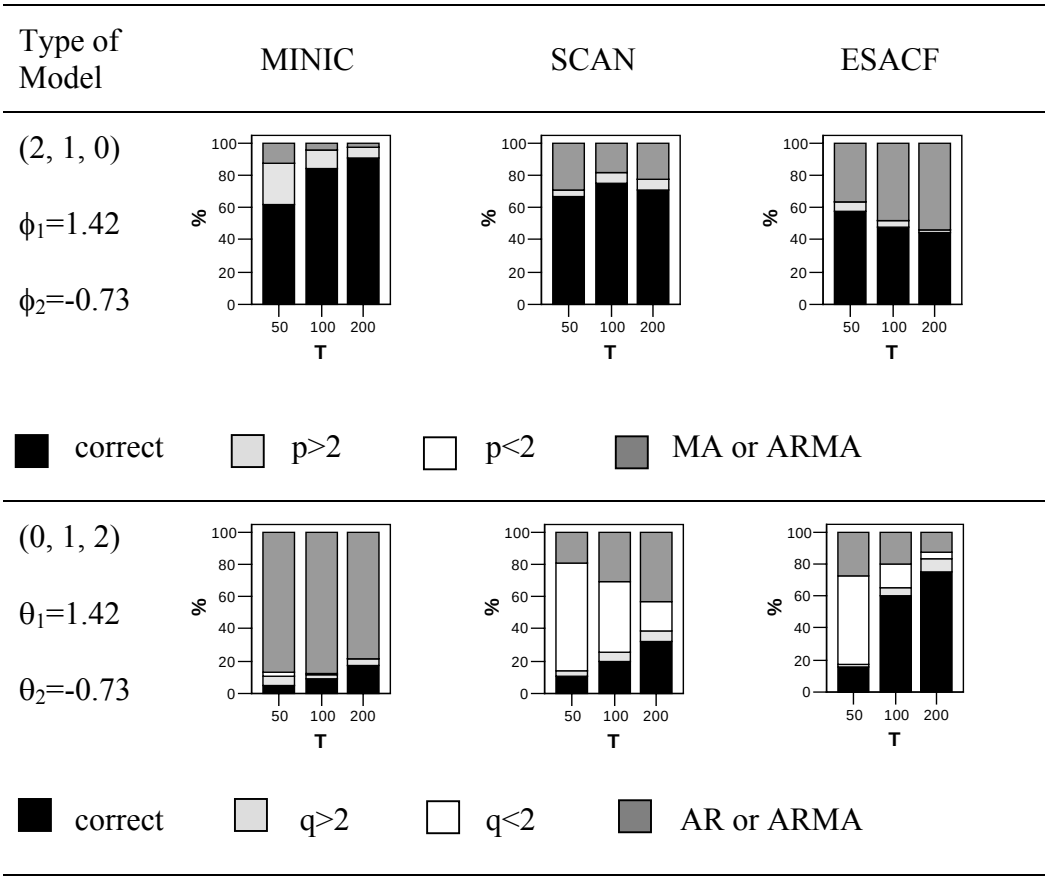


Figure 5.2.8. Results for differenced models of second-order, based on 1000 replications.

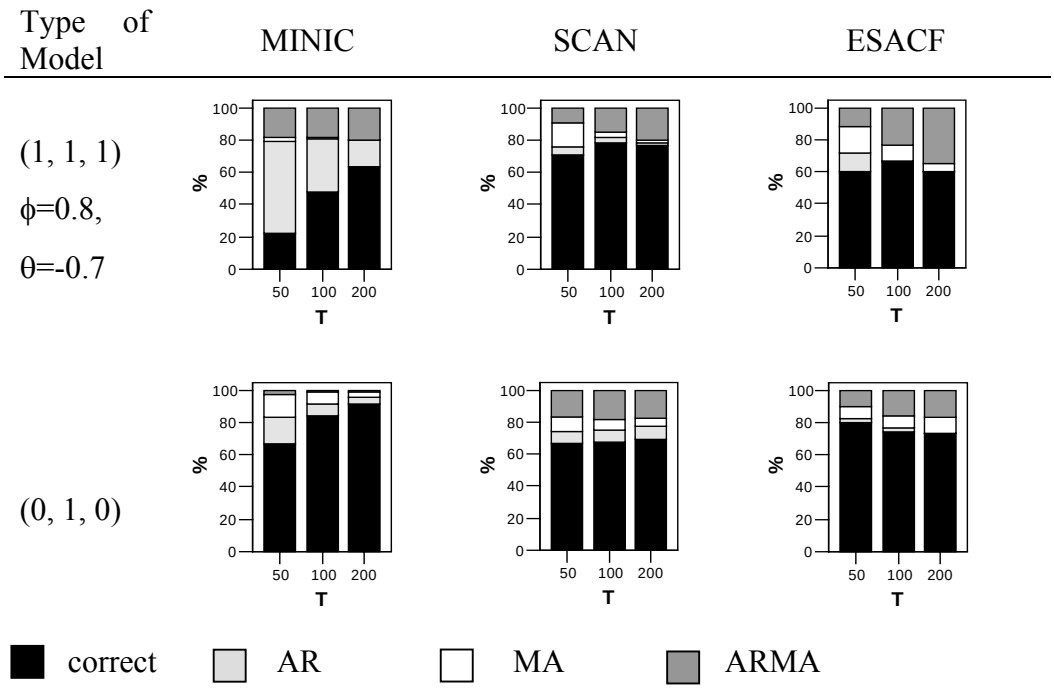


Figure 5.2 9. Results for differenced (1, 1, 1) and (0, 1, 0) models, based on 1000 replications.

As Figure 5.2.9 illustrates, SCAN is superior to both MINIC and ESACF in the ARMA (1, 1) case. Pure autoregressive models of higher order represent the most frequent incorrect selections of MINIC. SCAN and ESACF tend to select higher order mixed structures. For the white noise process, MINIC outperforms two other procedures. In samples of 200 observations the MINIC approach correctly identifies the (0, 0, 0) process above 90% of the time.

5.2.4 Conclusions

The performance of SCAN and ESACF as identification tools for various integrated processes has been empirically evaluated by means of Monte Carlo simulations. On the average, both procedures either correctly identify the simulated integrated structures or select parsimonious nearly equivalent mathematical representations in about 70% of the trials conducted. SCAN performs better than ESACF in pure autoregressive cases. ESACF is superior for MA(2) models. The results of MINIC, SCAN and ESACF for differenced models are consistent with

the findings for stationary processes reported in the literature. For all three procedures, model identification is less accurate for low dependency than for medium or high dependency models. MINIC and SCAN have difficulty in identifying moving-average models. ESACF demonstrates low power in autoregressive cases. SCAN and ESACF show the tendency to select higher order mixed structures. For some structures and parameterizations, this tendency is especially conspicuous in larger samples. For SCAN and ESACF, no manifest distinction between integrated and differenced models has been observed.

The effect of sample size is more pronounced for MINIC than for SCAN and ESACF. The increase in the number of observations refines the performance of MINIC. For SCAN and ESACF, however, in some of the cases studied (e.g. (1, 1, 0) and (2, 1, 0)) the increase in sample size reduces the percentage of correctly identified structures. The decrease is stronger for ESACF than for SCAN. To investigate this effect, additional tests were performed on some of the simulated structures using samples of 500 and even 10000 observations. The performance of SCAN remained stable, neither substantial worsening nor improvement was observed. For ESACF, the stabilization was not achieved until T=1000. The reduction in performance from T=200 to T=1000 was in the most cases not large, but nevertheless quite noticeable, depending on models and parameterizations. Koreisha & Yoshimoto (1991) reported a similar tendency of ESACF for stationary time series. For some simulated structures, even rather strong effects emerged. In the ARIMA (1, 0, 1) case, for instance, the decrease in the percentage of correct identifications changed from 58% when T=50 to 13% when T=1000 (Koreisha & Yoshimoto, 1991, p. 51).

If we estimate the value of some parameter, consistent statistics tend to be closer to the population value as the sample size increases. Consistency is an important and desirable property of an estimator. It is noteworthy that model identification methods cannot be unequivocally characterized in terms of consistency. The point is that the same dependency

structure can be approximated using different models. Note that notwithstanding the fact that the percentage of correct selections of SCAN and ESACF for ARIMA (1, 1, 0) with $\phi=\pm 0.2$ is extremely low, the performance of the automated procedures can be classified as good, since the most alternatives to the correct model selection [2, 0] are the following two parsimonious nearly equivalent mathematical representations: [1, 0] or [1, 1]. In other words, the decrease in the percentage of correctly identified structures with the increase in sample size is for the model identification less problematic than for the parameter estimation. An efficient model selection procedure, however, should improve its performance with an increasing number of observations. Therefore, the failure of SCAN and ESACF to refine their identification with increasing T can be viewed as a drawback of these methods. As noted previously, SCAN and ESACF exhibit the tendency to select higher order mixed models in larger samples. This effect is probably due to the fact that the ESACF or SCAN tables only provide information on the *maximum* values of p and q (Tsay & Tiao, 1984, p. 95). The identified overparametrized structures are not necessarily incorrect, since they may possess characteristics similar to those of the original formulation. However, including irrelevant parameters reduces estimation efficiency and raises the probability of selecting senseless models.

Despite of the described shortcomings of SCAN and ESACF, all evaluated methods are superior to subjective judgments as reported in the literature. For instance, in the study of Velicer and Harrop (1983) extensively trained judges were only able to correctly identify (0, 1, 0) and (0, 1, 1) models 4% and 13% of the time, respectively.

The reported findings could help to choose an appropriate identification procedure, especially if some knowledge about properties of the stochastic process under study is available. In economics and engineering sciences, for example, mixed models predominate (see Granger & Newbold, 1986, for explanations). Physiological processes such as heart rate or brain activity are often autoregressive. In the social and behavioral sciences, the most

widespread processes are autoregressive and moving average of first or second order (Glass et al., 1975; Marsh & Shibano, 1984; Revenstorf et al., 1980). For a researcher using SCAN and ESACF, the choice of an ARIMA structure hidden behind the identified [$P=p+d$, q] model remains the most difficult task. The decision whether differencing is necessary or not, however, appears to be an equally complicated issue, especially for series with roots close to the unit circle.

5.3 Study 3: Sample Size and Accuracy of Estimation of the Fractional Differencing Parameter[1]

5.3.1 Introduction

As described in Chapter 4.3.1, long-range dependencies have been observed in various psychological time-series representing cognitive, motor, perceptual and biological processes or self-esteem development. Persistent autocorrelations imply a *long memory* of the data generating process or, in other words, statistical dependence between observations separated by a large number of time units. In contrast, if a time series has a *short memory* and is predictable only from its immediate past, autocorrelations decay quickly as the number of intervening observations increases. Recall that a sequence of time-ordered uncorrelated random variables u_t, sometimes termed random shocks or innovations, is called white noise or the *process without memory*. Its counterpart, the *process with an infinite memory*, is known as random walk. A random walk process can be represented as the sum of random shocks $Y_t = \sum u_t$. As a result, the impact of a particular shock does not dissipate, and random walk remembers the shock forever. Figure 5.3.1 shows time series with different memory characteristics and their autocorrelation functions.

There exist a number of methods for modeling and estimating the memory property of an observed time series (see Chapter 3.1.2). As described in Chapter 4.3.1, the long-range dependence in psychological time series was initially detected using the spectral method (Gilden et al., 1995). It has been recently argued, however, that this method does not enable precise discrimination between the long- and short-memory processes and is inadequate for time series exhibiting both long- and short-term dependence (see Farrell et al., 2005;

[1] The results of this study are published in: Stadnytska, T., & Werner, J. (2006). Sample size and accuracy of estimation of the fractional differencing parameter. *Methodology*, *4*, 135-141.

Wagenmakers et al., 2004, for further details). In contrast to the spectral analysis, *Autoregressive Fractionally Integrated Moving Average* (ARFIMA) methodology suggested by Granger and Joyeux (1980) and Hosking (1981) allows rigorous discrimination and simultaneous estimation of short- and long-memory components of a time series by means of just three parameters *p*, *d* and *q*. ARFIMA (*p, d, q*) modeling is a straightforward extension of the Box-Jenkins ARIMA method.

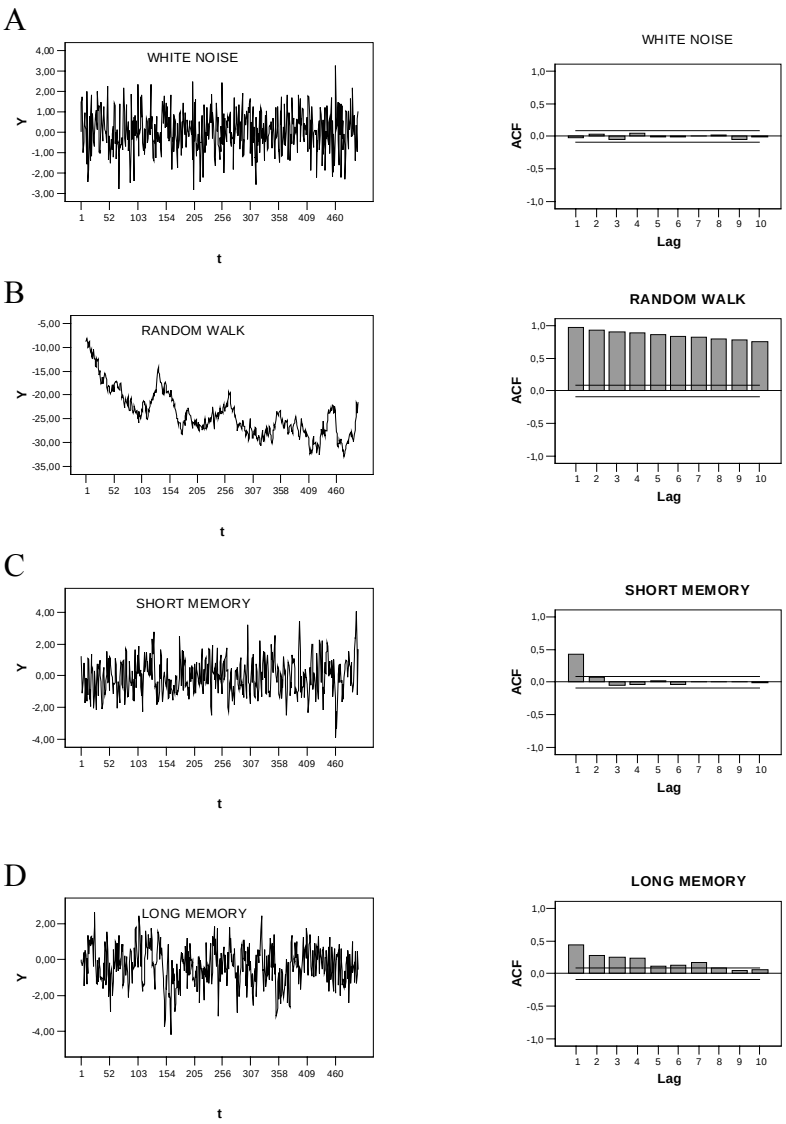


Figure 5.3.1. Processes with different memory properties and their autocorrelation functions (ACF): (A) White Noise, (B) Random Walk, (C) Short-Memory ARMA (1, 1) Process with ϕ=0.4 and θ=-0.1, (D) Long-Memory ARFIMA (0, d, 0) Process with d=0.3.

Within the scope of the ARFIMA methodology the short-memory components of a time series can be modeled through the autoregressive (AR) parameters *p* and the moving-average (MA) parameters *q*. Long-range dependencies can be captured through the fractional differencing parameter *d*, which can take any real value between 0 and 1. Within the Box-Jenkins ARIMA framework, *d* is an integer and refers to the order of differencing that is necessary to stabilize a time series. The ARFIMA modeling extends the traditional Box-Jenkins approach by allowing the differencing parameter *d* to take on *continuous* values. This enables ARFIMA models to give parsimonious descriptions of any long-range dependencies in time series. The finite long memory as expected in psychological time series can be modeled with $0 < d < 0.5$. For $0.5 \leq d \leq 1$, the process is nonstationary with noninvertible ARMA representations. For a detailed treatment of the ARFIMA model, see Baillie (1996), Brockwell and Davis (1991), Granger and Joyeux (1980) and Hosking (1981).

Different estimation procedures have been suggested for ARFIMA models. Some of them, such as the exact maximum likelihood (EML) method proposed by Sowell (1992a) or the conditional sum of squares (CSS) approach introduced by Chung (1996), allow the joint estimation of the short-memory ARMA and long-memory *d* parameters, and solve a potential finite-sample problem of the biased overestimation of *d* in time series which contain both long-range and short-range components (see Sowell, 1992b, for details).

To summarize, the memory property of an ARFIMA process depends crucially on the value of the fractional differencing parameter *d.* The process displays short memory for $d=0$. For $0<d<0.5$ the process is said to exhibit long memory. In light of the recent interest in exploring long-term dependencies in psychological time series and considering the restricted range of possible values of *d*, accurate estimation of the fractional differencing parameter *d* is of great practical relevance. The aim of this study is to examine sample size requirements for

the precise estimation of *d*. The quality of the long memory estimates as a function of time series length is investigated by means of Monte Carlo experiments.

5.3.2 Method

First, samples from a selection of ARFIMA (0, *d*, 0) and (1, *d*, 1) processes are generated, and then the quality of estimates of the long memory parameter *d* obtained with different sample sizes (*T*) is examined. The number of observations in simulated time series is varied between 100 and 2500. In both models four different values of *d* are used: *d*=0.1 for "weak long memory", *d*=0.2 for "moderate long memory", *d*=0.3 for "strong long memory" and *d*=0.4 for "very strong long memory". In the multiple-parameter case, the same value of the MA parameter θ=0.3 is combined with four values of the AR parameter ϕ=0.2, ϕ=0.4, ϕ=0.6 and ϕ=0.8. Note that the short-memory parameters ϕ and θ have the same sign in the equation

$$Y_t = \phi_1 Y_{t-1} + u_t + \theta_1 u_{t-1}.$$

Recall that cancellation would take place and the apparent ARFIMA (1, *d*, 1) process would be in fact ARFIMA (0, *d*, 0) for ϕ and θ with opposite signs. In cases of near-cancellation, the process may be well approximated by models with substantially different parameter values leading to difficulty in evaluating estimates.

For the estimation method, the conditional sum of squares procedure as proposed by Chung (1996) is employed. In one case the CSS results are compared with estimates obtained from the exact maximum likelihood approach as suggested by Sowell (1992a). The computational cost of the repeated inversion of *T* x *T* covariance matrices associated with EML becomes prohibitive for more comparisons (see Baillie, 1996; Sowell, 1992a, for further details). Each simulated model is replicated 1000 times. In the pure fractionally integrated case, estimates obtained from two different fitting models (0, *d*, 0) and (1, *d*, 1) are compared. To the series generated by the (1, *d*, 1) process, an ARFIMA (1, *d*, 1) model is fitted. The

mean (*M*), the *standard error* (*SE*) and the *mean square error* (*MSE*) computed from 1000 estimates serve as quality criteria. Additionally the percentage of estimates, which deviate more then +/-0.1 from the corresponding true parameter, is calculated.

All computations and the generation of independent N (0,1) innovations u_t are performed with IML subroutines FARMASIM and FARMAFIT from SAS for Windows Version 9.1.

5.3.3 Results

The pure fractionally integrated case is considered first. Tables 5.3.1 to 5.3.4 contain results for the ARFIMA (0, *d*, 0) processes. In the first place two observations can be made. One is that the results do not differ greatly between models with varying values of the long-memory parameter, the estimates are only slightly worse at d=0.4. Second, the quality of the estimates obtained from the same sample sizes depends on the fitted model.

It can be seen from the right sections of the Tables 5.3.1 to 5.3.4 that fitting the (0, *d*, 0) model to the pure fractionally integrated process already leads to excellent results in moderately long time series. For example, a sample size of 300 observations provides standard errors not greater than 0.048, *MSE* values about 0.002 and less than 5% of estimates that deviate more than +/-0.1 from the true values of *d*. Note that standard error values less than 0.051 confine the width of the 95% confidence interval to 0.2. If the long memory is "weak" or "moderate", the extension of time series length to 400 leads to the *SE*=0.039 and *MSE*=0.0016, which limits the width of the 99% confidence interval to 0.2 and allows only about 1% of estimates differ more than +/-0.1 from the true parameter value. In the cases with "strong" and "very strong" long memory 500 observations are needed for the same estimation accuracy.

Table 5.3.1. Results for model (0, d, 0) with d=0.1, based on 1000 replications.

	Fitted Model (1, *d*, 1)				Fitted Model (0, *d*, 0)			
T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$
100	.073	.360	.1303	60.9	.090	.083	.0070	22.9
200	.046	.298	.0914	47.3	.093	.056	.0032	8.2
300	.085	.189	.0359	41.0	.094	.046	.0021	3.1
400	.087	.149	.0225	34.0	.098	.039	.0016	1.2
500	.094	.116	.0136	25.6	.097	.036	.0013	1.1
600	.080	.139	.0197	25.0	.097	.032	.0010	0.4
700	.083	.102	.0107	25.9	.102	.028	.0008	0
800	.090	.096	.0093	21.8	.099	.029	.0008	0
900	.088	.079	.0063	16.4	.099	.027	.0007	0
1000	.090	.077	.0060	16.2	.100	.025	.0006	0

Table 5.3.2. Results for model (0, d, 0) with d=0.2, based on 1000 replications.

	Fitted Model (1, *d*, 1)				Fitted Model (0, *d*, 0)			
T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$
100	.180	.350	.1227	60.6	.194	.083	.0069	22.9
200	.148	.301	.0931	48.4	.195	.056	.0031	7.8
300	.189	.186	.0348	42.4	.196	.046	.0021	3.2
400	.195	.143	.0205	34.9	.200	.039	.0016	1.2
500	.200	.116	.0135	24.2	.198	.036	.0013	1.1
600	.193	.124	.0155	26.7	.198	.032	.0010	0.4
700	.191	.099	.0099	25.2	.203	.028	.0008	0
800	.195	.087	.0076	17.7	.200	.029	.0008	0
900	.197	.076	.0060	17.4	.200	.027	.0007	0
1000	.199	.072	.0055	16.6	.201	.025	.0006	0

The comparison of the CSS and the EML estimators presented in Table 5. 3. 3 shows that EML estimates are slightly better at d=0.3. However, d is constrained by the EML optimization algorithm to be not greater than 0.5, whereas CSS is not restricted in this way. EML estimation requires that the process be stationary, or transformed to stationarity. EML therefore benefits in models with "strong long memory" from being constrained to lie in a region around the true value, which is fairly tightly bounded on one side. In general, both techniques provide quite similar results.

Although the congruence of the fitted model with the data generating process ensures the best results, in practice the true model is unknown and must be identified from an observed time series. Determination of the correct model order to use is one of the most important and difficult practical considerations in application of ARFIMA models. An ARFIMA(1, d, 1) is the more plausible fitting model for the rigorous discrimination of short- and long-memory components, since it accounts for any possible short-range dependencies in a series and isolates them from the estimate of d. Fitting (1, d, 1) instead of (0, d, 0) model to the data implies, however, two additional parameter to estimate. This requires more observations for the same precision of parameter estimation. Clearly, the greater the number of parameters we have to estimate, the larger sample size we need for an adequate representation of the process generating a time series. As Tables 5.3.1 to 5.3.4 show, in the (1, d, 1) case an acceptable precision of the estimates ($SE \leq .077$, $MSE \leq .006$) is not attained for times series shorter than 900-1000 observations. $MSE \leq .006$ provides at least 80% of estimates within the interval [d-0.1, d+0.1] and $SE \leq .077$ confines the width of the 95% confidence interval to 0.3. Considering a restricted range for the fractional differencing parameter d, a confidence interval larger than [$\hat{d}$ - 0.15, $\hat{d}$ + 0.15] implies an unacceptable amount of uncertainty about the true parameter value.

Table 5.3.3. Results for model (0, d, 0) with d=0.3, based on 1000 replications.

	Fitted Model (1, *d*, 1)			Fitted Model (0, *d*, 0)					
				CSS estimation			EML estimation		
T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$
100	.302	.353	.1248	.302	.083	.0069	.286	.073	.0055
200	.273	.289	.0844	.300	.056	.0032	.291	.052	.0028
300	.303	.178	.0316	.300	.046	.0021	.293	.043	.0019
400	.300	.142	.0201	.303	.040	.0016	.297	.038	.0014
500	.300	.120	.0145	.300	.036	.0013	.296	.034	.0012
600	.303	.114	.0129	.300	.032	.0010	.296	.030	.0009
700	.298	.100	.0099	.305	.029	.0008	.298	.030	.0009
800	.302	.085	.0073	.303	.029	.0008	.297	.027	.0007
900	.301	.077	.0060	.301	.027	.0007	.298	.025	.0006
1000	.300	.070	.0054	.302	.025	.0006	.298	.024	.0006

Table 5. 3. 4. Results for model (0, d, 0) with d=0.4

	Fitted Model (1, *d*, 1)				Fitted Model (0, *d*, 0)			
T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$
100	.430	.341	.1170	64.0	.421	.086	.0078	26.6
200	.397	.283	.0801	52.4	.412	.059	.0036	9.7
300	.427	.165	.0279	39.2	.410	.048	.0024	4.9
400	.422	.147	.0222	39.2	.411	.041	.0018	1.5
500	.422	.113	.0132	29.1	.407	.037	.0014	1.2
600	.415	.111	.0126	27.1	.407	.032	.0011	0
700	.419	.098	.0101	28.9	.410	.030	.0010	0
800	.412	.109	.0121	21.7	.408	.029	.0009	0
900	.421	.085	.0076	20.1	.406	.028	.0008	0
1000	.405	.076	.0062	17.2	.407	.026	.0007	0

Tables 5.3.5 to 5.3.8 present results for the simulated ARFIMA (1, *d*, 1) models. The results reveal that a type of short-memory component influences the estimation of the long-memory parameter. In particular, the combination of low values of *d* with low autoregressive parameters provides much faster accurate estimation than the cases with high *d* and ϕ values. The pattern of *SE*'s and *MSE*'s for (0.2, 0.1, 0.3) and (0.4, 0.2, 0.3) models is broadly similar to that observed in the (0, *d*, 0) cases. $SE \leq .077$ and $MSE \leq .006$ require at least 800-1100 observations. In the (0.6, 0.3, 0.3) case, however, the same estimation accuracy is not reached until the time series length of 1500-1700 observations, and in the (0.8, 0.4, 0.3) model even not before 2400.

Table 5.3.5. Results for model (1, d, 1) with d=0.1 ϕ=0.2 θ=0.3, based on 1000 replications.

T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{\phi}}$	$MSE_{\hat{\phi}}$	$M_{\hat{\theta}}$	$MSE_{\hat{\theta}}$
100	.015	.338	.1215	63.1	.251	.1638	.319	.1196
200	.030	.201	.0453	45.6	.233	.0857	.331	.0296
300	.048	.168	.0309	36.7	.234	.0587	.316	.0175
400	.071	.122	.0157	25.1	.213	.0368	.313	.0126
500	.074	.111	.0129	22.4	.224	.0318	.303	.0106
600	.080	.085	.0076	15.6	.212	.0205	.305	.0078
700	.084	.079	.0064	14.9	.209	.0192	.307	.0077
800	.085	.065	.0045	12.0	.208	.0159	.305	.0062
900	.094	.059	.0035	9.3	.206	.0138	.301	.0060
1000	.091	.052	.0030	6.3	.206	.0106	.302	.0046
1100	.090	.052	.0028	5.6	.206	.0100	.303	.0044
1200	.091	.049	.0025	5.1	.206	.0097	.302	.0040
1300	.095	.049	.0024	4.4	.206	.0093	.300	.0040
1400	.095	.046	.0021	3.5	.202	.0076	.304	.0034
1500	.093	.044	.0020	2.6	.206	.0079	.301	.0036

Table 5.3.6. Results for model (1, d, 1) with d=0.2 ϕ=0.4 θ=0.3, based on 1000 replications.

T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{\phi}}$	$MSE_{\hat{\phi}}$	$M_{\hat{\theta}}$	$MSE_{\hat{\theta}}$
100	.107	.299	.0981	68.8	.418	.1209	.356	.0544
200	.132	.211	.0489	52.5	.419	.0626	.341	.0160
300	.137	.190	.0400	46.6	.435	.0447	.323	.0095
400	.163	.150	.0239	33.0	.415	.0299	.318	.0065
500	.168	.134	.0190	30.2	.421	.0248	.311	.0052
600	.176	.112	.0131	23.6	.411	.0173	.310	.0041
700	.178	.108	.0121	22.9	.411	.0164	.310	.0040
800	.183	.088	.0080	18.7	.406	.0127	.308	.0030
900	.191	.084	.0071	14.9	.407	.0115	.303	.0029
1000	.188	.078	.0061	12.2	.406	.0097	.305	.0025
1100	.189	.069	.0049	10.5	.405	.0082	.305	.0023
1200	.190	.066	.0045	9.4	.404	.0078	.304	.0019
1300	.194	.064	.0041	9.7	.404	.0074	.302	.0020
1400	.195	.057	.0033	6.4	.401	.0059	.305	.0018
1500	.193	.055	.0030	6.3	.403	.0059	.303	.0018
1600	.193	.052	.0028	6.3	.403	.0053	.303	.0016
1800	.197	.049	.0024	4.0	.400	.0046	.302	.0015
2000	.196	.045	.0021	3.0	.402	.0041	.301	.0012

Table 5.3.7. Results for model (1, d, 1) with d=0.3 ϕ=0.6 θ=0.3, based on 1000 replications.

T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{\phi}}$	$MSE_{\hat{\phi}}$	$M_{\hat{\theta}}$	$MSE_{\hat{\theta}}$
100	.294	.281	.0790	70.7	.542	.0909	.345	.0453
200	.278	.205	.0424	63.3	.573	.0468	.340	.0119
300	.260	.185	.0359	55.1	.607	.0294	.328	.0073
400	.281	.156	.0247	45.9	.594	.0221	.320	.0051
500	.282	.140	.0198	43.1	.603	.0173	.315	.0039
600	.286	.127	.0163	36.6	.596	.0140	.313	.0031
700	.279	.128	.0167	34.5	.605	.0135	.314	.0032
800	.287	.114	.0132	32.0	.598	.0114	.312	.0023
900	.290	.111	.0124	30.2	.603	.0106	.307	.0021
1000	.289	.103	.0107	25.7	.600	.0097	.309	.0019
1100	.290	.096	.0092	22.3	.600	.0081	.308	.0017
1200	.291	.093	.0088	21.8	.600	.0078	.307	.0014
1300	.294	.090	.0082	20.3	.601	.0073	.305	.0014
1400	.297	.080	.0063	16.4	.597	.0059	.306	.0013
1500	.296	.078	.0060	15.4	.599	.0057	.305	.0012
1700	.299	.072	.0053	14.5	.597	.0052	.304	.0010
1900	.297	.068	.0046	13.1	.598	.0044	.305	.0008
2000	.296	.066	.0044	11.2	.600	.0042	.302	.0008
2400	.298	.058	.0034	8.4	.598	.0034	.303	.0007
2500	.301	.056	.0031	7.2	.597	.0032	.302	.0007

Table 5.3.8. Results for model (1, d, 1) with d=0.4 ϕ=0.8 θ=0.3, based on 1000 replications.

T	$M_{\hat{d}}$	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$\%\left\|d-\hat{d}\right\|>0.1$	$M_{\hat{\phi}}$	$MSE_{\hat{\phi}}$	$M_{\hat{\theta}}$	$MSE_{\hat{\theta}}$
100	.609	.301	.1340	75.8	.583	.1578	.296	.0508
200	.546	.234	.0758	70.7	.644	.0919	.306	.0141
300	.493	.189	.0442	62.2	.710	.0389	.295	.0074
400	.479	.171	.0357	58.6	.726	.0281	.293	.0054
500	.459	.155	.0274	54.1	.750	.0178	.293	.0045
600	.463	.146	.0253	53.0	.744	.0166	.291	.0038
700	.446	.132	.0196	50.0	.760	.0117	.295	.0035
800	.446	.128	.0184	44.4	.759	.0116	.294	.0028
900	.440	.123	.0168	43.7	.769	.0094	.292	.0027
1000	.441	.112	.0143	39.6	.766	.0085	.293	.0023
1100	.438	.110	.0136	39.5	.768	.0075	.294	.0022
1200	.436	.108	.0130	38.2	.770	.0071	.293	.0020
1300	.431	.103	.0116	33.9	.776	.0061	.293	.0020
1400	.432	.097	.0104	32.4	.775	.0054	.295	.0018
1600	.424	.090	.0086	28.9	.780	.0043	.296	.0016
1800	.429	.085	.0080	25.7	.778	.0042	.294	.0014
2000	.424	.081	.0071	23.4	.782	.0034	.293	.0013
2400	.422	.076	.0062	20.1	.783	.0031	.295	.0011
2500	.422	.072	.0057	18.6	.784	.0028	.295	.0010

To explore this issue further, time series with different combinations of *d* and ARMA values are simulated. For all models, a sample size of 1000 observations is used. The results presented in Table 5.3.9 reveal that higher autoregressive coefficients complicate the estimation of the fractionally differencing parameter. Regardless of the values of *d*, models with ϕ=0.2 or ϕ=0.4 provide much better estimates of *d* as time series with ϕ=0.6 or ϕ=0.8. It is noteworthy that the opposite is true for the quality of the estimates of the short-memory

parameters. Models with ϕ=0.6 or ϕ=0.8 provide better estimates of autoregressive and moving-average parameters than models with ϕ=0.2 or ϕ=0.4. In general, the higher the value of ϕ, the worse is the quality of the estimates of *d*. The observed phenomenon is not surprising in finite samples, since ϕ=1 implies *d*=1 or an integrated process with an infinite memory. Therefore, the long-memory parameter and the short-memory autoregressive terms near 1 can both contribute to similar patterns of autocorrelation. This explains the difficulty in discriminating these two components at smaller samples.

Table 5.3.9. SE's and MSE's of CSS estimates of the ARFIMA parameters from series of length 1000, received from 500 replications.

Simulated Model (d, ϕ, θ)	$SE_{\hat{d}}$	$MSE_{\hat{d}}$	$SE_{\hat{\phi}}$	$MSE_{\hat{\phi}}$	$SE_{\hat{\theta}}$	$MSE_{\hat{\theta}}$
0.1, 0.2, 0.1	.058	.0035	.164	.0268	.128	.0167
0.1, 0.4, 0.1	.071	.0052	.174	.0404	.103	.0108
0.1, 0.6, 0.1	.108	.0121	.097	.0093	.053	.0029
0.1, 0.8, 0.1	.117	.0138	.074	.0055	.061	.0039
0.2, 0.2, 0.1	.058	.0034	.163	.0268	.129	.0168
0.2, 0.4, 0.1	.080	.0065	.110	.0121	.070	.0050
0.2, 0.6, 0.1	.106	.0117	.096	.0093	.053	.0029
0.2, 0.8, 0.1	.118	.0139	.075	.0056	.061	.0039
0.3, 0.2, 0.1	.058	.0034	.168	.0285	.132	.0178
0.3, 0.4, 0.1	.080	.0064	.110	.0121	.070	.0051
0.3, 0.6, 0.1	.104	.0108	.095	.0091	.052	.0028
0.3, 0.8, 0.1	.121	.0146	.076	.0060	.063	.0040
0.4, 0.2, 0.1	.061	.0038	.171	.0307	.134	.0188
0.4, 0.4, 0.1	.082	.0068	.112	.0131	.072	.0053
0.4, 0.6, 0.1	.102	.0106	.095	.0094	.053	.0029
0.4, 0.8, 0.1	.127	.0168	.081	.0072	.065	.0043

5.3.4 Conclusions

The accuracy of estimates of the fractional differencing parameter *d* has been empirically investigated as a function of sample length. The conditional sum of squares methodology proposed by Chung (1996) allowing the joint estimation of the short- and long-memory parameters has been employed as estimation method. This method provided consistent und sufficient estimates for *d* and showed good finite-sample performance, comparable to that of EML. CSS estimation is computationally much easier and more convenient in choosing starting values compared to techniques such as ML. Moreover, the CSS method does not require knowledge of an appropriate transformation to apply and is available in the current versions of the SAS software for Windows.

Although fitting the (0, *d*, 0) model to the pure fractionally integrated process already provided very accurate estimates in moderate sample sizes of 400-500 observations, the (1, *d*, 1) case appears to be the more plausible fitting model for any observed times series, since it accounts for possible short-range dependencies in a series and isolates them from the estimate of *d.* The reported simulation results suggest that applying ARFIMA (1, *d*, 1) as a fitting model to the data requires at least 1000-1600 observations for acceptable estimation accuracy.

6 GENERAL DISCUSSION

The emphasis of this paper is certainly on methodological issues. The primary research tasks of the thesis were: (1) to develop and evaluate testing strategies revealing the nature of growing time series, (2) to analyze the efficiency of automated procedures for ARMA model identification commonly available in current versions of SAS for Windows, and (3) to examine sample size requirements for the accurate estimation of the long-memory parameter *d* in ARFIMA models. Besides, the initial chapters have aimed to demonstrate broad possibilities of time series procedures to deal with *dynamical* psychological phenomena.

Time series approach allows ambitious handling of those indispensable features of dynamical concepts such as memory, stability or dependency structure. Systems containing internal temporal regularity can be distinguished from unstable systems depending on external and occasional events. Complex behavior such as balancing between preservation and adaptation or nonlinear dynamic of hierarchical structures can be represented by means of rather simple time series models. Processes with different memory properties become distinguishable. These examples make clear that in most research cases time series methods cannot be viewed as substitutes or alternatives to the traditional statistical procedures. Time series analysis rather opens a new perspective for psychological research, where the understanding of development and change of psychological processes are in the focus of attention.

Within the scope of experimental settings, time series techniques represent the proper analysis tools for longitudinal designs producing dependent measurements. In this case, time series analysis is an alternative to the traditional methods such as ANOVA. It is important to understand that the classical statistical procedures assume independent and uncorrelated data. That is why they are no longer relevant for autocorrelated data. The widespread opinion that

time series analysis requires a lot of points is not valid for time series experiments. By the way, the sample size problem is not unique for time series analysis. For classical methods such as analysis of variance, larger samples imply more power to detect differences between means. Nobody hopes seriously to reveal existing distinctions between treatments using only, say, five subjects in each cell. For time series data, more points permit identification of more sophisticated models and allow detecting smaller departures from the process after intervention (see also Gottman, 1981, pp. 57-59).

This paper has focused on analysis of a single time series. Handling of interrelations of multiple processes represents another interesting research field for psychologists. *Transfer functions* enable to reveal the relationship pattern of two or more interacting subjects. For example, we can study whether one of two playing children is dominant or whether a mother tries to adjust her behavior to the infant's behavior in social interactions. *Cross-spectral methods* allow to reveal cyclicity in social interplay and to assess the mother's sensitivity to her child's rhythms of attention and inattention. *Cointegration analysis* enables researcher to detect a meaningful relation between two or more integrated series such as income and consumption in economics or mood patterns of a married couple in psychology. Distinguishing between cointegration and *spurious regression* (i.e., meaningless relation between integrated series) represents a challenging methodological issue.

Solving methodological problems occurring in applied settings was the main objective of this thesis. One of the most difficult and important tasks in psychological time series research is to determine the cause of instationarity in growing time series. Testing strategy developed and evaluated in this paper is one possibility to distinguish between trend and difference stationary processes. The Augmented Dickey-Fuller test was employed in the study because it is commonly available in current versions of SAS for Windows. Developing and incorporating in software packages new testing procedures with more power, appropriate for

smaller samples and using the null hypothesis of no unit root remain to be solved. Furthermore, the presented strategy was designed for rather simple models characteristic of the majority of time series in psychology. Procedures for more specific cases such as models with complex error structures or nonlinear and broken trends are still necessary.

Results obtained for SCAN and ESACF as identification tools for integrated processes demonstrated the pronounced superiority of the automated procedures over subjective judgments. The reported findings could help to choose an appropriate identification procedure in cases where some knowledge about properties of the stochastic process under study is available. It is noteworthy that, for some models and parameterizations, the performance of SCAN and ESACF was disappointing. This strongly supports the recommendation of Box, Jenkins and Reinsel (1994) to use automated methods as supplementary guidelines in the model selection process and not as a substitute for critical examination of the ACF, PACF, and the model residuals. In other word, model identification implies a strategy consisting of different consecutive steps; employing automated procedures is one of these steps.

Despite of the contributions of Thornton and Gilden (2005) and Wagenmakers et al. (2004; 2005), rigorous discrimination of processes with different memory properties remains one of the most urgent methodological issues. The findings reported in this paper revealed sample size requirements on the accurate estimation on the long memory parameter d within the scope of the time domain analysis. It makes sense to assume that, depending on the error structure of the studied processes, up to 1600 observations are necessary in the frequency domain as well for reliable distinguishing between long-range and short-range dependency processes. For parsimonious planning of psychological experiments, it is important to clarify how the spectral classifier method of Thornton and Gilden (2005) and the ARFIMA procedure of Wagenmakers et al. (2004; 2005) respond to sample size variations.

Handling of missing values and performance of tests or automated procedures for model identification in the presence of outliers or with non-metric data represent further methodological topics relevant for applied time series research.

REFERENCES

Aks, D. J., & Sprott, J. C. (2003). The role of depth and 1/f dynamics in perceiving reversible figures. *Nonlinear Dynamics, Psychology, and Life Sciences, 7*, 161-180.

Algina, J., & Swaminathan, H. A. (1977). A procedure for the analysis of time series designs. *Journal of Experimental Education, 45*, 56-60.

Algina, J., & Swaminathan, H. A. (1979). Alternatives to Simonton's analysis of interrupted and multiple-group time series designs. *Psychological Bulletin, 86*, 919-926.

Amorose, A. J. (2001). Intraindividual variability of self-evaluations in the physical domain: Prevalence, consequences, and antecedents. *Journal of Sport and Exercise Psychology, 23*, 222-244.

Ansley, C. F, & Newbold, P. (1980). Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics, 13,* 159-183.

Ansley, C. F, & Newbold, P. (1981). On the bias in estimates of forecast mean squared error. *Journal of the American Statistical Association, 76*, 569-578.

Ashley, R., & Verbrugge, R. J. (2004). To difference or not to difference: A Monte Carlo investigation of inference in vector autoregressive models. VPI Economic Department Working Paper. *http://ashleymac.econ.vt.edu/working_papers/9915.pdf*

Ashley, R., & Verbrugge, R. J. (2006). Comments on „A critical investigation on detrending procedures for non-linear processes". *Journal of Macroeconomics, 28*, 192-194.

Ayat, L., & Burridge, P. (2000). Unit root tests in the presence of uncertainty about the non-stochastic trend. *Journal of Econometrics, 95*, 71-96.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: an explanation of 1/f noise. *Physical Review Letters, 59*, 381-384.

Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York: Copernicus-Springer.

Baillie, R. T. (1996). Long memory process and fractional integration in econometrics. *Journal of Econometrics*, 73, 5-59.

Barge-Schaapveld, D. Q., Nicolson, N.A., Berkhof, J., & DeVries, M. W. (1999). Quality of life in depression: Daily life determinates and variability. *Psychiatry Research, 88*, 173-189.

Beran, J. (1994). *Statistics for long-memory processes.* New York: Chapman & Hall.

Bloomfield, P. (2000). *Fourier analysis of time series*. New York: Willey.

Bowerman, B.L. & O'Connell, R.T. (1993). *Forecasting and time series: An applied approach.* Belmont: Duxbury Press.

Box, G.E.P., & Jenkins, G.M. (1970). *Time series analysis, forecasting and control*. San Francisco, CA: Holden Day.

Box, G.E.P., Jenkins, G.M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Box, G. E. P., & Pierce, W. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of American Statistical Association, 65*, 1509-1526.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods (2nd Ed.).* New York: Springer.

Brown, J. D. (1998). *The self*. Boston, MA: McGraw Hill.

Busey, T. A. & Townsend, J. T. (2001). Independent sampling vs. inter-items dependencies in whole report processing: Contributions of processing architecture and variable attention. *Journal of Mathematical Psychology, 45*, 283-323.

Bürgy, R., & Werner, J. (2005). Modellinterpretation von ARIMA-Modellen. In J. Werner (Ed.), *Zeitreihenanalysen mit Beispielen aus der Psychologie* (pp. 185-222). Berlin: Logos.

Chan, K.H., Hayya, J.C., & Ord, J.K. (1977). A note on trend removal methods: The case of polynomial regression versus variate differencing. *Econometrica, 45*, 737-744.

Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long memory processes ($1/f^{\alpha}$ type) in human coordination. *Physical Review Letters, 79,* 4501-4504.

Chen, Y., Ding, M., & Kelso, J. A. S. (2001). Origins of time errors in human sensorimotor coordination. *Journal of Motor Behavior, 33*, 3-8.

Choi, B. S. (1992). *ARMA model identification*. New York: Springer.

Chung, C. F. (1996). A generalized fractionally integrated ARMA process. *Journal of Time Series Analysis, 2,* 111-140.

Crosbie, J. (1993). Interrupted time-series analysis with brief single-subject data. *Journal of Consulting and Clinical Psychology*, *61*, 966-974.

Dagum, E. B., & Giannerini, S. (2006). A critical investigation on detrending procedures for non-linear processes. *Journal of Macroeconomics, 28*, 175-191.

DeJong, D. N., Nankervis, J. C, Savin, N. E., & Whiteman, H. (1992). The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics, 53,* 323-343.

Delcor, L., Cadopi, M., Delignières, D., & Mesure, S. (2002). Dynamics of the memorization of a morphokinetic movement sequence. *Neuroscience Letters, 336*, 25-28.

Delignières, D., Fortes, N., & Ninot, G. (2004). The fractal dynamics of self-esteem and physical self. *Nonlinear Dynamics in Psychology and Life Sciences, 8*, 479-510.

Delignières, D., Lemoine, L., &Torre, K. (2004). Time intervals production in tapping oscillatory motion. *Human Movement Science, 23*, 87-103.

Dickey, D. A. (1984). Power of unit root tests. *Proceedings of Business and Economic Statistics Section of ASA*, 489-493.

Dickey, D. A. (2004). Applied time series notes. *http://www.stat.ncsu.edu/people/dickey/st730/Notes/pg_38_55.pdf.*

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica, 49*, 1057-72.

DiClemente, C. C., Prochaska, J. O., Fairhurst, S. K., Velicer, W. F., Velasquez, M. M., & Rossi, J. S. (1991). The process of smoking cessation: An analysis of precontemplation, contemplation, and preparation stages of change. *Journal of Consulting and Clinical Psychology, 59*, 295-304.

Diebold, F. X., & Senhadji, A. S. (1996). The uncertain unit root in real GNP: Comment. *American Economic Review, 86*, 1291-1298.

Ding, M., Chen, Y., & Kelso, J. A. S. (2002). Statistical analysis of timing errors. *Brain and Cognition, 48*, 98-106.

Elder, J., & Kennedy, P. E. (2001a). F versus t test for unit roots. *Economic Bulletin, 3*, 1-6.

Elder, J., & Kennedy, P. E. (2001b). Testing for unit roots: What should students be taught? *Journal of Economic Education*, *32*, 137-146.

Fang, Y. (2005). The effect of the estimation on Goodness-of-Fit tests in time series models. *Journal of Time Series Analysis, 26,* 527-541.

Farrell, S., Wagenmakers, E-J., & Ratcliff, R. (2005). ARFIMA time series modeling of serial correlations in human performance. *Manuscript submitted for publication*.

Farrell, S., Wagenmakers, E-J., & Ratcliff, R. (2006). Methods for detecting 1/*f* noise. *Unpublished manuscript*.

Farrell, S., Wagenmakers, E-J., & Ratcliff, R. (in press). 1/f noise in human cognition: Is it ubiquitous, and what dose it mean? *Psychonomic Bulletin & Review*.

Fava, J. L., Velicer, W. F., & Prochaska, J. O. (1994). Applying the transtheoretical model to a large representative smoking sample. *Addictive Behaviors, 19*, 189-203.

Fortes, M., Ninot, G., & Delignières, D. (2004). The dynamics of self-esteem and physical self: Between preservation and adaptation. *Quality and Quantity, 38*, 735-751.

Fortes, M., Ninot, G., & Delignières, D. (2005). The auto-regressive integrated moving average procedures implications for adapted physical activity research. *Adapted Physical Activity Quarterly, 22*, 221-236.

Fox, K. R. (1997). *The physical self.* Champaign, IL: Human Kinetics.

Fraschini E. M., & Axausten, K. W. (2001). Day on day dependencies in travel: First results using ARIMA modeling, Working paper. *http://e-collection.ethbib.ch.*

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, 10*8, 33-56.

Gilden, D. L., & Wilson, S. G. (1995). On the nature of streaks in signal detection. *Cognitive Psychology, 28*, 17-64.

Gilden, D. L., Thornton, T., & Mallon, M.W. (1995). 1/*f* noise in human cognition. *Science, 267*, 1837-1839.

Gilden, D. L., & Wilson, S. G. (1995). Streaks in skilled performance. *Psychonomic Bulletin & Review, 2,* 260-265.

Glass, G. V., Willson, V. L., & Gottman, J. M. (1975). *Design and analysis of time-series experiments*. Boulder, CO: Colorado Associated University Press.

Gottman, J. M. (1981). *Time-Series Analysis*. New York: 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. J (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics, 14*, 227-238.

Granger, C. W. J., & Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis, 1*, 15-30.

Granger, C.W.J., & Newbold, P. (1986). *Forecasting economic time series*. San Diego, CA: Academic Press.

Greenier, K. D., Kernis, M. H., & Waschull, S. B. (1995). Not all height (or low) self-esteem people are the same: Theory and research on stability of self-esteem. In M. H. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 51-71). New York: Plenum Press.

Greenier, K. D., Kernis, M. H., McNamara, C. W., Waschull, S. B., Berrry, A. J., Herlocker, C. E., & Abend, T. A. (1999). Individual differences in reactivity to daily events: Examining the roles of stability of self-esteem. *Journal of Personality, 61*, 187-208.

Gujarati, D. (2003). *Basic econometrics*. (4th ed.). New York: McGraw-Hill.

Handel, P. H., & Chung, A. L. (Eds.) (1993). *Noise in physical systems and 1/f fluctuations*. New York: American Institute of Physics.

Hannan, E. J., & Rissanen, J. (1982). Recursive estimation of mixed autoregressive moving average order. *Biometrika, 69*, 81-94.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, CA: Princeton University Press.

Harrop, J. W., & Velicer, W. F. (1985). A comparison of three alternative methods of time series model identification. *Multivariate Behavioral Research, 20*, 27-44.

Hausdorf, J. M., & Peng, C. K. (1996). Multiscaled randomness: A possible source of 1/f noise in biology. *Physical Review E, 54*, 2154-2157.

Hausdorf, J. M., Mitchell, S.L., Firtion, R., Peng, C. K., Cudkowicz, M. E., Wei, J. Y., & Goldberger, A. L. (1997). Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. *Journal of Applied Physiology, 82*, 262-269.

Hausdorf, J. M., Zemany, L., Peng, C. K., & Goldberger, A. L. (1999). Maturation of gait dynamics: Stride-to-stride variability and its temporal organization in children. *Journal of applied Physiology, 86*, 1040-1047.

Holden, D., & Perman, R. (1994). Unit roots and cointegration for the economist. In B. B. Rao (Ed.) *Cointegration for the applied economist* (pp. 47-112). New York: St. Martin's.

Holtzman, W. (1963). Statistical models for the study of change in the single case. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 199-211). Madison, WI: University of Wisconsin Press.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika, 68*, 165-176.

Huitema, B. E. (1985). Autocorrelation in applied behavioral analysis: A myth. *Behavioral Assessment, 7*, 107-118.

Jensen, H. J. (1998). *Self-organized criticality*. Cambridge, MA: University Press.

Jones, R. R., Vaught, R. S., & Weinrott, M. (1977). Time-series analysis in operant research. *Journal of Applied Behavior Analysis, 10,* 151-166.

Kasdin, N. J, (1995). Discrete simulation of colored noise and stochastic processes and $1/f^{a}$ power low noise generation. *Proceedings of the IEEE, 83*, 802-827.

Kazdin, A. E. (1982). *Single-case research design: Methods for clinical and research settings*. New York: Oxford University Press.

Kendall, M. G., & Buckland, W. R. (1971). *A dictionary of statistic terms*. New York: Hafner Publishing Company.

Kim, T., Leybourne, S. J., & Newbold, P. (2004). Behaviour of Dickey-Fuller unit root tests under trend misspecification. *Journal of Time Series Analysis, 25*, 755-764.

Koreisha, S. G., & Yoshimoto, G. (1991). A comparison among identification procedures for autoregressive moving average models. *International Statistical Review, 59*, 37-57.

Laming, D. R., J. (1968). *Information theory of choice-reaction times.* London: Academic Press

Leventhal, H., & Cleary, P. D. (1980). The smoking problem: A review of the research and theory in behavioural risk modification. *Psychological Bulletin, 88*, 370-405.

Linkenkaer-Hansen, K. (2002). *Self-organized criticality and stochastic resonance in the human brain.* Dissertation. Helsinki University of Technology Laboratory of Biomedical Engineering.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika, 65*, 297-303.

Lopez, J. H. (1997). The power of the ADF test. *Economics Letters, 57*, 5-10.

Maddala, G. S., & Kim, I. (1998). *Unit roots, cointegration and structural change*. Cambridge: University Press

Makridakis, S., Wheelwright, S.C., & Hyndman, R.J. (1998). *Forecasting: Methods and applications*. New York: Wiley.

Mandelbrot, B. B., & van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review, 10*, 422-437.

Marks-Tarlow, T. (1999). The self as a dynamical system. *Nonlinear Dynamics, Psychology and Life Sciences, 3*, 311-345.

Marks-Tarlow, T. (2002). Fractal dynamics of the psyche. *Dynamical Psychology: An International, Interdisciplinary Journal of Complex Mental Processes*, http://goertzel.org/dynacon.html

Marsh, J. C., & Shibano, M. (1984). Issues in the statistical analysis of clinical time-series data. *Social Work Research and Abstracts, 20*, 7-12.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: learning, achievement and adjustment* (pp. 131-198). Thousand Oaks, CA: Sage

Marsh, H. W., & Yeung, A. S. (1998). Top-down, bottom-up, and horizontal models: The direction of causality in multidimensional, hierarchical self-concept models. *Journal of Personality and Social Psychology, 75*, 509-527.

Matyas, T. A., & Greenwood, K. M. (1990). Problems in the estimation of autocorrelations in brief time series and some implications for behavioral data. *Behavioral Assessment, 13*, 137-157.

McCleary, R., & Hay, R.A. Jr. (1980). *Applied time series analysis for the social sciences.* Beverly Hills, CA: Sage Publications.

Mischel, W. (1969). Continuity and change in personality. *American Psychologist, 24*, 1012-1018.

Nelson, C.R., & Kang, H. (1981). Spurious periodicity in inappropriately detrended time teries. *Econometrica, 49*, 741-751.

Nelson, C.R., & Kang, H. (1984). Pitfalls in the use of time as an explanatory variable in regression. *Journal of Business and Economic Statistics, 2*, 73-82.

Nezlek, J. B., & Plesko, R. M. (2001). Day-to day relations smong self-concept clarity, self-esteem, daily events, and moods. *Personality and Social Psychology Bulletin, 27*, 201-211.

Ng, S., & Perron, P. (1995). Unit root tests in ARMA models with data dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association, 90*, 268-281.

Ng, S., & Perron, P. (2001). Lag length selection and the construction of the unit root test with good size and power. *Econometrica, 69*, 1519-1554.

Ninot, G., Fortes, M., & Delignières, D. (2001). A psychometric tool of the assessment of the dynamics of the physical self. *European Journal of Applied Psychology, 51*, 205-216.

Ninot, G., Delignières, D., & Varray, A. (2003). *Individual differences in stability of physical self: Examining the role of chronic obstructive pulmonary disease*. Unpublished manuscript.

Nowak, A., Vallacher, R. R., Tesser, A., & Borkowski, W. (2000). Society of self: The emergence of collective properties in self-structure. *Psychological Review, 107*, 39-61.

Ottenbacher, K. J. (1986). Reliability and accuracy of visually analyzing graphed data from single-subject designs. *American Journal of Occupational Therapy, 40*, 464-469.

Perron, P. (1988). Trends and random walks in macroeconomic time series. *Journal of Economic Dynamics and Control, 12*, 297-332.

Pilgram, B., & Kaplan, D. T. (1998). A comparison of estimator for 1/f noise. *Physica D, 114,* 108-122.

Pressing, J. (1999). S*ources for 1/f noise effects in human cognition and performance.* Proceedings of the 4th Conference of the Australasian Cognitive Science Society, Newcastle.

Pressing, J., & Jolley-Rogers, G. (1997). Spectral properties of human cognition and skill. *Biological Cybernetics, 76*, 339-347.

Prochaska, J. O., Velicer, W. F., DiClemente, C. C., Guadagnoli, E., & Rossi, J. S. (1991). Patterns of change: Dynamic topology applied to smoking cessation. *Multivariate Behavioral Research, 26*, 83-107.

Prochaska, J. O., Velicer, W. F., Rossi, J. S., Goldstein, M. G., Marcus, B.H., Rakowski, W., Fjore, C., Harlow, L. L., Redding, C. A., Rosenbloom, D., & Rossi, S. R. (1994). Stages of change and decisional balance for twelve problem behaviors. *Health Psychology, 13*, 39-46.

Psaridakis, Z., & Sola, M. (2003). On detrending and cyclical asymmetry. *Journal of Applied Econometrics, 18,* 271-289.

Rankin, E. D., & Marsh, J. C. (1985). Effects of missing data on the statistical analysis of clinical time series. *Social Work Research and Abstracts, 21*, 13-16.

Revenstorf, D., Kessler,A. Schindler, L., Hahlweg, K., & Bluemner, E. (1980). Time series analysis: Clinical applications evaluating intervention effects in diaries. In O. D. Anderson (Ed.), *Analyzing time series: Proceedings of the international conference held on Guernsey, Channel Islands, October 1979* (pp. 291-312). Amsterdam: North-Holland.

Rosel, J., & Elósegui, E. (1994). Daily and weekly smoking habits: A Box Jenkins analysis. *Psychological Reports, 75*, 1639-1648.

Rosenberg, M. (1986). Self-concept from middle childhood through adolescence. In J. Sules & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (vol. 3, pp. 107-136). Hillsdale, NJ: Erlbaum.

SAS Institute Inc. (1999). *SAS/ETS User's Guide*, Version 8. Cary, NC.

Schenk-Hoppè, K. (2001). Economic growth and business cycles: A critical comment on detrending time series. *Studies in Nonlinear Dynamics and Econometrics*, 5.

Schuster, H. G. (2005). *Deterministic chaos: An introduction.* Cambridge: Weinheim.

Sharpley, C. F., & Alavosius, M. P. (1988). Autocorrelation in behavioral data: An alternative perspective. *Behavioral Assessment, 10*, 243-251.

Simonton, D. K. (1977). Cross-sectional time-series experiments. Some suggested statistical analysis. *Psychological Bulletin, 84*, 489-502.

Slifkin, A. B., & Newell, K. M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science, 7*, 170-177.

Sowell, F. (1992a). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics, 53*, 165-188.

Sowell, F. (1992b). Modeling long-run behavior with the fractional ARFIMA model. *Journal of Monetary Economics, 29*, 277-302.

Spray, J. A., & Newell, K. M. (1986). Time series analysis of motor learning: KP versus no-KP. *Human Movement Science, 5*, 59-74.

Stadnytska, T., Braun, S., & Werner, J. (2006). Comparison of Automated Procedures for ARMA Model Identification. *Manuscript submitted for publication*..

Stock, J. H. (1994). Unit roots, structural breaks, and trends. In R. Engle, & D. McFadden (Eds.), *Handbook of Econometrics*, volume IV (pp. 2740-2843). Amsterdam: Elsener.

Thornton, T. L., & Gilden, D. L. (2005). Provenance of correlation in psychological data. *Psychonomic Bulletin & Review, 12,* 409-441.

Torre, K., Delignières, D., & Lemoine (in press). Detection of long-range dependence and estimation of fractal exponents through ARFIMA modeling. *British Journal of Mathematical and Statistical Psychology*.

Tsay, R. S., & Tiao, G. C. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of American Statistical Association, 79*, 84-96.

Tsay, R. S., & Tiao, G. C. (1985). Use of canonical analysis in time series model identification. *Biometrika, 72*, 299-315.

Tsay, R. S., & Tiao, G. C. (1990). Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Annals of Statistics, 18*, 220-250.

Vallacher, R. R., Nowak, A., Froehlich, M., & Rockloff, M. (2002). The dynamics of self-evaluation. *Personality and Social Psychology Review, 6*, 370-379.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology, 3*, 331-350.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2005). Human cognition and 1/f scaling. *Journal of Experimental Psychology: General, 134*, 117-123.

Van Orden, G. C., Moreno, M. A., & Holden, J. G. (2003). A proper metaphysics for cognitive performance. N*onlinear Dynamics, Psychology, and Life Sciences, 7*, 49-60.

Velicer, W. F., & Colby, S. M. (1997). Time series analysis for prevention and treatment research. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 211-249). Washington, DC: American Psychological Association.

Velicer, W. F., DiClemente, C. C., Rossi, J. S, & Prochaska, J. O. (1990). Relapse situations and self-efficacy: An integrative model. *Addictive Behaviors, 15*, 271-283.

Velicer, W. F., & Fava, J. l. (2003). Time series analysis. In J. Schinka & W. F. Velicer (Eds.), *Research methods in psychology* (pp. 581-606). New York: John Wiley.

Velicer, W. F. & Harrop, J. W. (1983). The reliability and accuracy of time series model identification. *Evaluation Review,* 7, 551-560.

Velicer, W. F., & McDonald, R. P. (1984). Time series analysis without model identification. *Multivariate Behavioral Research, 19*, 33-47.

Velicer, W. F., & McDonald, R. P. (1991). Cross-sectional time series design: A General Transformation Approach. *Multivariate Behavioral Research, 26*, 247-254.

Velicer, W. F., Redding, C. A., Richmond, R., Greeley, J., & Swift, W. (1992). A time series investigation of tree nicotine regulation models. *Addictive Behaviors, 17*, 325-345.

Velicer, W. C., Rossi, J. S., Prochaska, J. O., & DiClemente, C.C. (1996). A comparison measurement model for health behavior change. *Addictive Behaviors, 21*, 555-584.

Wagenmakers, E.-J., Farrell, S., & Ratcliff, R. (2004). Estimation an 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,* 108-116.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.

Ward, L. M., & Richard, C. M. (2001). *$1/f^{\alpha}$ noise and decision complexity*. Unpublished manuscript, University of British Columbia, Vancouver, British Columbia, Canada.

Warner, R. M. (1998). *Spectral analysis of time- series data*. New York: Guilford.

Wei, W. W. S. (1990, 1994). T*ime series analysis, univariate and multivariate methods.* Redwood city, CA: Addison-Wesley Publishing Company. Reprinted with corrections, February 1994.

Werner, J. (Ed.) (2005). *Zeitreihenanalysen mit Beispielen aus der Psychologie.* Berlin: Logos.

West, K. D. (1987). A note of the power of least squares tests for a unit root. *Economics Letters, 24*, 249-252.

West, B. J., & Shlesinger, M.F. (1990). The noise in natural phenomena. *American Scientist, 78*, 40-45.

APPENDIX

Study 1: Deterministic or Stochastic Trend: Decision on the Basis of the Augmented Dickey-Fuller Test

Simulation of 1000 series with T=100 where DGP is $Y_t = 0.2 + Y_{t-1} + e_t$ *with* $e_t \sim IIDN\ (0, \sigma^2)$

```
*************************************************************************
data probe;
keep a3;
do i=1 to 1000;
      a1=ranuni (54893);
      a2=a1*71474839;
      a3=round(a2);
      output;
end;
run;
proc transpose data=probe out=seeds;
run;

data dat.dswn02;
set seeds;
keep t y1-y1000;
array cols (1000) col1-col1000;
array y (1000) y1-y1000;
array a (1000) a1-a1000;
array e (1000) e1-e1000;

do i=1 to 1000;
      a(i)=0; y(i)=0; e(i)=0;
end;
do t=-5 to 100;
      do i=1 to 1000;
            call rannor (cols(i), e(i));

            y(i)= 0.2+a(i) +e(i);
            a(i)=y(i);
      end;
if t gt 0 then output;
end;
run;
*************************************************************************
```

Simulation of 1000 series with T=100 where DGP is $Y_t = 0.2 + Y_{t-1} + e_t$ *with* $e_t=0.5e_{t-1}+u_t$ *and* $u_t \sim IIDN\ (0, \sigma^2)$

```
**************************************************************************

data probe;
keep a3;
do i=1 to 1000;
      a1=ranuni (54893);
      a2=a1*71474839;
      a3=round(a2);
      output;
end;
run;
proc transpose data=probe out=seeds;
run;

data dat.ds02ar;
set seeds;
keep t y1-y1000;
phi=0.5;
array cols (1000) col1-col1000;
array y (1000) y1-y1000;
array a (1000) a1-a1000;
array e (1000) e1-e1000;
array b (1000) b1-b1000;
array c (1000) c1-c1000;
do i=1 to 1000;
      a(i)=0; y(i)=0; e(i)=0; b(i)=0;c(i)=0;
end;
do t=-5 to 100;
      do i=1 to 1000;
            call rannor (cols(i), e(i));
            b(i)=phi*c(i)+e(i);
            c(i)=b(i);
            y(i)= 0.2+a(i) +b(i);
            a(i)=y(i);
      end;
if t gt 0 then output;
end;
run;
**************************************************************************
```

Simulation of 1000 series with T=100 where DGP is $y_t = 0.2y_{t-1} + 0.1t + e_t$ *with* $e_t=0.5e_{t-1}+u_t$ *and* $u_t \sim IIDN(0, \sigma^2)$

```
**************************************************************************
data probe;
keep a3;
do i=1 to 1000;
      a1=ranuni (54893);
      a2=a1*71474839;
      a3=round(a2);
      output;
end;
run;
proc transpose data=probe out=seeds;
run;


data ar05;
set seeds;
keep t x1-x1000;
phi=0.5;
array cols (1000) col1-col1000;
array x (1000) x1-x1000;
array a (1000) a1-a1000;
array e (1000) e1-e1000;
array b (1000) b1-b1000;
array c (1000) c1-c1000;
do i=1 to 1000;
      a(i)=0; x(i)=0; e(i)=0; b(i)=0;c(i)=0;
end;
do t=-5 to 100;
      do i=1 to 1000;
            call rannor (cols(i), e(i));
            b(i)=phi*c(i)+e(i);
            c(i)=b(i);
            x(i)= phi*a(i) +b(i);
            a(i)=x(i);
      end;
if t gt 0 then output;
end;
run;

data dat2.tsar08;
set ar08;
keep t y1-y1000;
array x (1000) x1-x1000;
array y (1000) y1-y1000;

do i=1 to 1000;
y(i)=0.1*t+x(i);
end;
run;

**************************************************************************
```

Simulation of 1000 series with T=100 where DGP is $y_t = 0.2y_{t-1} + 0.1t + e_t$ *with* $e_t=u_t - 0.5u_{t-1}$ *and* $u_t \sim IIDN(0, \sigma^2)$

```
*************************************************************************
data probe;
keep a3;
do i=1 to 1000;
      a1=ranuni (54893);
      a2=a1*71474839;
      a3=round(a2);
      output;
end;
run;
proc transpose data=probe out=seeds;
run;

data dat.dsma02;
set seeds;
keep t y1-y1000;
teta=0.5;
array cols (1000) col1-col1000;
array y (1000) y1-y1000;
array a (1000) a1-a1000;
array e (1000) e1-e1000;
array b (1000) b1-b1000;

do i=1 to 1000;
      a(i)=0; y(i)=0; e(i)=0; b(i)=0;
end;
do t=-100 to 100;
      do i=1 to 1000;
            call rannor (cols(i), e(i));
            y(i)= 0.2+a(i) +e(i)-teta*b(i);
            b(i)=e(i);
            a(i)=y(i);

      end;
if t gt 0 then output;
end;
run;

*************************************************************************
```

Performing ADF test

```
*************************************************************************
options nonumber center nodate label pagesize=100 ls=64;
filename routed 'C:\...\dswn.dat';

%macro adf;
%do ii=1 %to 1000;
      proc printto print=routed;
      run;
      ods select StationarityTests;
      ods noproctitle;

      proc arima data=dat.dswn;
            identify var=y&ii stationarity=(adf=6);
      run;
      proc printto print=print;   run;
%end;
%mend adf;
%adf;

data dswn;
infile routed;
input word1 $ 1-11 @;
if word1='Trend' then  do;
input
      lag0 rho0 prob0 tau0 probt0
      lag1 rho1 prob1 tau1 probt1
      lag2 rho2 prob2 tau2 probt2
      lag3 rho3 prob3 tau3 probt3
      lag4 rho4 prob4 tau4 probt4
      lag5 rho5 prob5 tau5 probt5
      lag6 rho6 prob6 tau6 probt6
      ;
keep tau1 tau2 tau3 tau4 tau5 tau6;
output;
end;
run;

data adf.dswn;
set dswn;
      if tau0< -3.45 then sig0='sig';
      else sig0='n.s';
      if tau1< -3.45 then sig1='sig';
      else sig1='n.s';
      if tau2< -3.45 then sig2='sig';
      else sig2='n.s';
      if tau3< -3.45 then sig3='sig';
      else sig3='n.s';
      if tau4< -3.45 then sig4='sig';
      else sig4='n.s';
      if tau5< -3.45 then sig5='sig';
      else sig5='n.s';
      if tau6< -3.45 then sig6='sig';
      else sig6='n.s';
keep sig0 sig1 sig2 sig3 sig4 sig5 sig6;
run;
*************************************************************************
```

Study 2: Model Identification of Integrated ARMA Processes

Simulated series with T=200

****

Random Walk ARIMA (0,1,0)

```
data bib.rw;
keep y e;
Y=0; e=0;u=0; a=0;g=0;
do t=-50 to 200;
e=rannor(59837);
y=a+e;
a=y;
if t gt 0 then output;
end;
run;
```

ARIMA (1,1,1)

```
data bib.arima;
keep y e a;
Y=0; e=0; teta=0.7; phi=-0.8;u=0; a=0;g=0;b=0;
do t=-50 to 200;
e=rannor(59837);
y=b+a;
a=phi*u+e-teta*g;
g=e;
u=a;
b=y;
if t gt 0 then output;
end;
run;
```

ARIMA (1,1,0)

```
data bib.ari;
keep y e a;
Y=0; e=0;phi=0.5;u=0; a=0;g=0;b=0;
do t=-50 to 200;
e=rannor(59837);
y=b+a;
a=phi*u+e;
u=a;
b=y;
if t gt 0 then output;
end;
run;
```

****

****

ARIMA (0,1,1)

```
data bib.ima;
keep y e a;
Y=0; e=0; teta=0.5;u=0; a=0;g=0;b=0;
do t=-50 to 200;
e=rannor(59837);
y=b+a;
a=e-teta*g;
g=e;
b=y;
if t gt 0 then output;
end;
run;
```

ARIMA (2,1,0)

```
data bib.ari_2;
keep y e a;
Y=0; e=0; phi1=1.8; phi2=-0.9;u=0; a=0;g=0;b=0;
do t=-50 to 200;
e=rannor(59837);
y=b+a;
a=phi1*u+phi2*g+e;
g=u;
u=a;
b=y;
if t gt 0 then output;
end;
run;
```

ARIMA (0,1,2)

```
data bib.ima_2;
keep y e a;
Y=0; e=0; teta1=1.8; teta2=-0.9;u=0; a=0;g=0;b=0;
do t=-50 to 200;
e=rannor(59837);
y=b+a;
a=e-teta1*g-teta2*u;
u=g;
g=e;
b=y;
if t gt 0 then output;
end;
run;
```

****

Analysis steps

```
**************************************************************************
```

1. Step

```
proc arima data=bib.ari_2;
identify var=y(1) minic esacf scan;
run;
```

2. Step

```
proc arima data=bib.ari_2;
identify var=y minic esacf scan;
run;
**************************************************************************
```

Macros

```
**************************************************************************
%macro scan_eig(datei);
options nonumber center nodate label pagesize=15 ;
%do ii=1 %to &wdh;
      proc arima data=&datei;
      identify var= col&ii(1) scan p=(0:5) q=(0:5);
      ods select TentativeOrders;
      ods output TentativeOrders = rscan&ii;
      data rscan&ii;
      set rscan&ii;
      t+1;
      run;
      proc append base =ar02.rsca&tt  data=rscan&ii ;
      run;
      quit;
%end;
%mend scan_eig;

%macro esac_eig(datei);
options nonumber center nodate label pagesize=15 ;
%do ii=1 %to &wdh;
      proc arima data=&datei;
      identify var= col&ii(1) esacf p=(0:5) q=(0:5);
      ods select TentativeOrders;
      ods output TentativeOrders = resac&ii;
      data resac&ii;
      set resac&ii;
      t+1;
      run;
      proc append base =ar02.resa&tt  data=resac&ii ;
      run;
      quit;
%end;
%mend esac_eig;

**************************************************************************
```

Study 3: Sample Size and Accuracy of Estimation of the Fractional Differencing Parameter

Simulation of 1000 (0, 0.2, 0) series with T ranging from 600 to 1200 and estimating d

```
*******************************************************************************
proc iml;
%macro wn;
%mend wn;
%macro ar;
p=1;
%mend ar;
%macro ma;
q=1;
%mend ma;
%macro arma;
p=1 q=1;
%mend arma;
      %let modell = %wn;
      %let d = (.2);
      %let phi = 0.0;
      %let theta = 0.0;
      %let seed = 1235;
      %let wdh = 1000;
      %let tmin = 600;
      %let tmax = 1200;
      %let tstep = 100;
      %let yt = yt;

%macro fit;
 seed = -&seed;;
 %do i = 1 %to &wdh;
      n=&ii;
      ods output D_AR_MA_S_N =est&i&ii;
      print 'D-Wert vor call farmasim', &d;
      call farmasim (yt, &d, &phi, &theta) n=n seed = seed;
      call farmafit (d, ar, ma, s, yt) &modell opt=0;
      print d  ar ma s  N;
%end;
%mend fit;

%macro wdh;
      %do ii = &tmin %to &tmax %by &tstep;
            %fit;
      %end;
%mend wdh;
%wdh;
```

```
%macro data;
%do ii = &tmin %to &tmax %by &tstep;
      %do i=1 %to &wdh;
            data est&i&ii;
            set est&i&ii;
            Dspez=&d; ARspez= &phi; MAspez= &theta;
            run;
            proc append base=dat.par&ii data=est&i&ii force;
      %end;
%end;
%do ii = &tmin %to &tmax %by &tstep;
    proc means data =dat.par&ii maxdec=4;
      var d ar ma;
      ods output summary=means&ii;
      data means&ii;
      set means&ii;
      rename D_N = wdh; label D_N=wdh;
      label D_mean=D_mean; label D_StdDev = D_StdDev;
      label D_min=D_min; label D_max= D_max;
      MSE_D=(&d-D_mean)**2 + D_StdDev**2;
      label AR_mean=AR_mean; label AR_StdDev = AR_StdDev;
      label AR_min=AR_min; label AR_max= AR_max;
      MSE_MA=(&theta-MA_mean)**2 + MA_StdDev**2;
      label MA_mean=MA_mean; label MA_StdDev = MA_StdDev;
      label MA_min=MA_min; label MA_max= MA_max;
      drop AR_N MA_N  VName_D VName_AR VName_MA;
      MSE_AR=(&phi-AR_mean)**2 + AR_StdDev**2;
      Dspez=&d; ARspez= &phi; MAspez= &theta; t=&ii;
      drop AR_N MA_N  VName_D VName_AR VName_MA;
      proc append base=gesamt data=means&ii force;
%end;
      data gesamt;
      retain t wdh Dspez ARspez MAspez
      D_mean D_StdDev MSE_D D_min D_max
      AR_mean AR_StdDev MSE_AR AR_min AR_max
      MA_mean MA_StdDev MSE_MA MA_min MA_max;
      set gesamt;
      run;
      data dat.d02_2;
      set gesamt;
      keep t wdh Dspez
      D_mean D_StdDev MSE_D D_min D_max;
      run;

%mend data;
%data;


**************************************************************************
```