INAUGURAL - DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht - Karls - Universität
Heidelberg

vorgelegt von

M. Sc. Stefan Richter
aus Dresden / Deutschland

Tag der mündlichen Prüfung: .

Data adaptive inference for locally stationary processes

Betreuer: **Prof. Dr. Rainer Dahlhaus**
Prof. Dr. Enno Mammen

Zusammenfassung

Seit ihrer Einführung in den 1980er Jahren spielen lokal stationäre Prozesse eine wichtige Rolle in der Zeitreihenanalyse. Als Verallgemeinerung stationärer Prozesse erlauben sie es, dass Beobachtungen über die Beobachtungszeit hinweg ihre Verteilungseigenschaften ändern. In vielen lokal stationären Zeitreihenmodellen wird diese zeitliche Änderung charakterisiert durch Parameterkurven, deren Schätzung damit von zentralem Interesse ist. In dieser Arbeit entwickeln wir Methoden und zeigen theoretische Resultate für die Bandbreitenwahl bei nichtparametrischen Schätzern dieser Kurven. Wir konzentrieren uns hierbei auf lokale Maximum-Likelihood-Schätzer. Diese haben eine enge Verbindung zu Martingaldifferenzensequenzen, die in vielen Beweisen nützlich ist.

Im ersten Teil definieren wir für linear lokal stationäre Prozesse einen globalen Bandbreitenselektor, der durch die Kreuzvalidierungsmethode im nichtparametrischen Regressionsmodell motiviert ist. Wir beweisen, dass der Selektor asymptotisch optimal ist in dem Sinne, dass der Kullback-Leibler-Abstand des Modells mit diesem Selektor zum wahren Modell für lange Beobachtungszeiträume gegen den minimal möglichen Kullback-Leibler-Abstand konvergiert. In Simulationen überprüfen wir die Qualität unseres Ansatzes in der Praxis. Die Beweise basieren auf Bias-Varianz-Zerlegungen der Schätzer. Die formale Beschreibung dieser Zerlegungen ist wesentlich schwieriger im Falle nichtlinearer lokal stationärer Zeitreihenmodelle.

Im zweiten Abschnitt dieser Arbeit entwickeln wir allgemeine Approximationstechniken, um lokal stationäre Prozesse durch stationäre Prozesse anzunähern, womit solche Zerlegungen erhalten werden können. Im Zuge dessen führen wir so genannte Ableitungsprozesse ein und geben Bedingungen an, unter welchen Existenz und Eindeutigkeit gegeben sind. Ein zentrales Ergebnis ist eine Taylor-Entwicklung von lokal stationären Prozessen. Diese Resultate sind von unabhängigem Interesse für weitere Forschung in diesem Gebiet. Wir unterstreichen dies, indem wir die erhaltenen Ergebnisse nutzen, um neue Versionen einiger Standardtheoreme wie ein Gesetz der großen Zahlen und einen zentralen Grenzwertsatz für lokal stationäre Prozesse unter minimalen Momentannahmen zu beweisen.

Im letzten Teil dieser Arbeit definieren wir für eine große Klasse von lokal stationären Prozessen einen lokalen Bandbreitenselektor, der auf einem Kontrastminimierungsansatz basiert, welcher zuerst auf nichtparametrische Regressionsmodelle angewandt wurde. Wir zeigen, dass der Selektor bzgl. dem Euklidischen- und dem Kullback-Leibler-Abstand minimax-optimal bis auf einen logarithmischen Faktor ist, der typisch für lokale Modellauswahlprozeduren ist. Für die Beweise greifen wir auf die vorher entwickelten Approximationstechniken zurück. In einer Simulation untersuchen wir das Verhalten der Auswahlprozedur in verschiedenen Zeitreihenmodellen.

Die Resultate dieser Arbeit zur Bandbreitenwahl können als Verallgemeinerung der ursprünglichen Methoden im nichtparametrischen Regressionsmodell aufgefasst werden, da dieses Modell stets als Spezialfall enthalten ist. Durch die verallgemeinerte Formulierung liefert diese Arbeit daher einen Beitrag dazu, ein tieferes Verständnis dieser Methoden zu gewinnen.

Abstract

Since their introduction in the 1980s, locally stationary time series play an important role in time series analysis. As a generalization of stationary processes, they allow the observations to change their distribution properties over observation time. In many locally stationary time series models this change over time is characterized by parameter curves, whose estimation is of essential interest. In this work we develop methods and prove theoretical results for bandwidth selection for nonparametric estimators of these curves. We focus on local maximum likelihood estimators. Their strong connection to martingale difference sequences is fundamental in many of our proofs.

In the first part of this dissertation we define a global bandwidth selector for linear locally stationary processes which is motivated by the cross validation method that was first introduced in the nonparametric regression model. We prove that the selector is asymptotically optimal in the sense that the Kullback Leibler distance of the model connected with this selector to the true model converges to the minimal possible Kullback Leibler distance as the observation time increases to infinity. In simulations we analyze the quality of the method. The proofs are based on bias-variance decompositions of the estimators. The formal discussion of these decompositions gets harder in the case of nonlinear locally stationary time series models.

In the second part of this dissertation we develop general techniques to approximate locally stationary processes by stationary processes. These techniques allow us to obtain the decompositions mentioned above. We introduce so called derivative processes and give conditions under which existence and uniqueness can be guaranteed. An important result is a Taylor-like expansion of locally stationary processes. These findings are of independent interest for further research. We emphasize this point by using the approximation techniques to obtain new versions of standard theorems like a law of large numbers and a central limit theorem for locally stationary processes under minimal moment assumptions.

In the last part of this thesis we define a local bandwidth selector for a large class of locally stationary processes which is based on a contrast minimization approach which was first applied to nonparametric regression models. We show that our selector is minimax optimal up to a logarithmic factor (which is typical for local model selection procedures) with respect to the Euclidean distance and the Kullback-Leibler distance. For the proofs we use the approximation techniques which were discussed before. In a simulation we analyze the behavior of the selection routine for different time series models.

The findings of this thesis regarding bandwidth selection routines can be interpreted as a generalization of the original methods in the nonparametric regression model, because this model is included as a special case. Due to the more general formulation this thesis makes a contribution to understand these methods more deeply.

Acknowledgments

I want to thank Prof. Rainer Dahlhaus for introducing me to probability theory and statistics 7 years ago when I started my studies at the Heidelberg University. He is a great teacher and one of the main reasons why I got interested in this field. As my supervisor during the last three years, he provided me a comfortable working place with best research conditions and supported me in my progress, for instance by allowing me to go to many conferences and get in contact with other researchers in statistics. He also gave me the possibility to have a research stay at the University at Chicago and working together with Prof. Wei Biao Wu which I enjoyed very much. He proposed the topic of bandwidth selection for locally stationary processes and further gave me the freedom to pursue my own interests in the field of time series analysis.

I would also like to thank Prof. Enno Mammen for inviting me to workshops in Moscow to present my work. During my research, Prof. Jan Johannes and Claudia Strauch always had an open door for me so that I could talk with them about my research problems. For this I want to thank you both.

At Heidelberg University I got to know a lot of amazing people in our research group, I especially want to thank Alexander Kreiß, Cathrine Aeckerle, Xavier Loizeau, Dominic Edelmann, Johannes Dueck, Sebastian Schweer, Cornelia Wichelhaus, Sophon Tunyavetchakit, Nopporn Thamromgrat and Axel Bücher for the joint activities and the interesting discussions which I enjoyed, in particular at lunch time. It was always a pleasure to go to the secretariat of Marion Münster, Dagmar Neubauer, Elke Carlow, Daniela Strand and Evelyn Wilhelm who gave me the feeling that we are one big family in the statistics group.

I want to thank Johannes who shared a flat with me the last three years and created a point of rest in the sometimes stressful times. I will always accept your offer to cook another lasagna. I also enjoyed the evenings where we played games together with Alex, Steffen and Jens which were the times where I could relax most efficiently. I also want to mention Eva Dölger, Maik Rudloff, Stefan Werner and David Strauß who often invited me to activities and kept the contact to me.

Finally, a big thanks goes to my family. My parents and grandparents gave me the best possible start into life. I think they often held back their own wants, for instance having me home for a longer time, to raise me into a self-reliant man.

Funding by the *Deutsche Forschungsgemeinschaft* (German Research Foundation) within the programme "Spatio/Temporal Graphical Models and Applications in Image Analysis", grant GRK 1653, is also gratefully acknowledged.

Contents

Chapter 1

Introduction

Nonstationary processes. Stationary processes are characterized by the fact that there distribution does not change over time. They play an important role in time series analysis and lots of models and powerful methods for analyzing them were introduced during the last decades. However, more recently models have become popular which allow the observations to change their distribution properties smoothly over time. A special focus lies on so-called locally stationary processes which behave like stationary processes in small observation periods. The idea goes back to Priestley (1965) and Priestley (1981) who proposed to generalize the spectral representation of stationary processes by making it time-varying. Because of its structure, this formulation did not allow for rigorous asymptotic considerations. In two papers (cf. Dahlhaus (1996) and Dahlhaus (1997)) Dahlhaus improved the representation with an infill asymptotics scheme, meaning that the time is rescaled from $t = 1, ..., n$ to the interval $[0, 1]$ by considering the 'local time' $\frac{t}{n}$ for $t = 1, ..., n$. He obtained the spectral representation

$$X_{t,n} = \mu\Big(\frac{t}{n}\Big) + \int_{-\pi}^{\pi} A_{t,n}^{\circ}(\lambda)\ \mathrm{d}\xi(\lambda), \quad t = 1, ..., n$$

where $A_{t,n}^{\circ}(\lambda)$ is the transfer function and ξ a stochastic process on $[-\pi, \pi]$ and $\mu(\cdot)$ the mean function. Here it was assumed that $A_{t,n}^{\circ}(\lambda)$ can be approximated by some function $A(\frac{t}{n}, \lambda)$ uniformly in t, λ. In this formulation, asymptotic results were obtained by fixing a local time $u \in [0, 1]$ and considering only observations $X_{t,n}$ with $|\frac{t}{n} - u| \ll 1$. The representation of $X_{t,n}$ in the time domain is given by

$$X_{t,n} = \mu\Big(\frac{t}{n}\Big) + \sum_{k=-\infty}^{\infty} a_{t,n}(k)\varepsilon_{t-k}, \tag{1.0.1}$$

where $a_{t,n}(k)$ are deterministic sequences and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. errors. A famous example is the time-varying autoregressive moving average (tvARMA) process which is recursively defined via

$$\sum_{k=0}^{r} \alpha_k\Big(\frac{t}{n}\Big) \cdot X_{t-k,n} = \sum_{j=0}^{s} \beta_j\Big(\frac{t}{n}\Big)\varepsilon_{t-k} \tag{1.0.2}$$

with parameter curves $\alpha_k, \beta_j : [0,1] \to \mathbb{R}$. Latest publications (see Dahlhaus and Polonik (2009)) allow for even more general linear models. For a review, we refer to Dahlhaus (2011). Besides the development of a theory for linear locally stationary models, there were introduced a lot of nonlinear nonstationary processes by mimicking the infill asymptotics approach and replacing constant parameters in stationary processes by parameter curves evaluated at the rescaled time $\frac{t}{n}$. Here, we mention the tvAR process (cf. Dahlhaus and Giraitis (1998)), the tvARCH process (cf. Dahlhaus and Subba Rao (2006), Fryzclewicz, Sapatinas and Subba Rao (2008)), random coefficient models (cf. Subba Rao (2006)) or general recursively defined locally stationary processes (cf. Zhou and Wu (2009), section 4). Note that the most of processes are nonlinear and thus do not fit into the scheme (1.0.1).

More recently, Zhou and Wu (2009) and Karmakar and Wu (2016) among others proposed a representation of locally stationary processes by Bernoulli shifts, namely $X_{t,n} = J_{t,n}(\varepsilon_t, \varepsilon_{t-1}, ...)$ with measurable functions $J_{t,n}$, by generalizing a similar approach for stationary processes introduced in Wu (2005). This approach covers both linear and recursively defined processes. More abstract formulations based on approximation properties were given by numerous authors, for instance Vogt (2012).

As mentioned above there is a large class of locally stationary processes whose evolution over time is mainly described by parameter curves, for instance α_k, β_j in (1.0.2). A central objective in inference of such processes is estimation of these curves. Besides parametric approaches (cf. Dahlhaus (1997)), a large literature for nonparametric estimation via quasi Maximum Likelihood methods is available in special cases like the tvAR process (cf. Dahlhaus and Giraitis (1998)), tvARCH process (cf. Dahlhaus and Subba Rao (2006)) or linear processes in general (cf. Dahlhaus and Polonik (2009)).

Bandwidth selection in nonparametric estimation. As can be seen in these publications, there is a strong connection to nonparametric estimation of i.i.d. regression models (see the monograph Tsybakov (2009) for an introduction) which have the form

$$X_{t,n} = \mu(\frac{t}{n}) + \varepsilon_t, \quad t = 1, ..., n. \tag{1.0.3}$$

In (1.0.3), a standard approach to estimate μ from the observations $X_{t,n}$ is the so-called Nadaraya-Watson estimator

$$\hat{\mu}_b(u) := \frac{\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big) X_{t,n}}{\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)}, \quad u \in [0,1], \tag{1.0.4}$$

where $K : \mathbb{R} \to \mathbb{R}$ is a probability density (the so-called kernel function) and $b = b_n$ is the bandwidth which may depend on the number of observations n. Since $\hat{\mu}_b(u)$ can be obtained through a quasi Maximum Likelihood approach by minimizing $\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)\big(X_{t,n} - \mu\big)^2$ in μ, this estimator can be seen as a special case of quasi Maximum Likelihood estimators in (1.0.1) and it seems worthwhile to transfer asymptotic results to the more general case.

The main issue which is connected to the form of the estimator (1.0.4) is to choose the right window size b. If the true function $\mu(\cdot)$ is twice continuously differentiable and K is symmetric around 0, it is possible to calculate the mean squared error (MSE) and obtain a so called bias-variance decomposition

$$\mathbb{E}|\hat{\mu}_b(u) - \mu(u)|^2 = \mathrm{Var}(\hat{\mu}_b(u)) + |\mathbb{E}\mu_b(u) - \mu(u)|^2 \approx \frac{V_0}{nb} + B_0\frac{b^4}{4} \tag{1.0.5}$$

with constants V_0, B_0. Minimization of this term in b leads to the (MSE optimal) bandwidth choice $b_0 = (\frac{V_0}{B_0})^{1/5}n^{-1/5}$. Although b_0 is a good starting point for further investigation, it is not useful in practice since B_0, V_0 heavily depend on the unknown function $\mu(\cdot)$. During the last decades, many efforts have been made to find natural and 'good' estimators of the MSE-optimal bandwidth b_0. Popular methods are selectors based on Cross validation (cf. Rice (1984)), plugin approaches (originally from Woodroofe (1970)) and, more recently, contrast minimization approaches (cf. Lepski, Mammen and Spokoiny (1997), Lepski and Spokoiny (1997) and, more general, in Goldenshluger and Lepski (2011)) among others. All three methods mentioned are very general and therefore have applications in many other fields of statistics.

Combining the two topics. In several special cases of locally stationary processes (cf. Dahlhaus and Giraitis (1998) for tvAR, Dahlhaus and Polonik (2009) for tvARMA and Dahlhaus and Subba Rao (2006), Fryzclewicz, Sapatinas and Subba Rao (2008) for tvARCH processes), asymptotic properties and similar bias-variance decompositions as in (1.0.5) have been obtained for quasi Maximum Likelihood estimators of the corresponding parameter curves. Especially for nonlinear processes the analysis of these estimators is much harder than in i.i.d. regression (1.0.3) since an explicit representation of the estimator may not be available. Theoretical properties and results regarding practical behavior of bandwidth selection for locally stationary processes however are still unavailable unless in very special cases (cf. Arkoun (2010)).

In this thesis, we start at this point. Our goal is to shed light on the theoretical behavior of bandwidth selectors for large classes of locally stationary processes. We will focus on selectors based on cross validation and contrast minimization approaches. As a byproduct, we will obtain consistency results with rates for quasi Maximum likelihood estimators in these models. Since we allow the unknown parameter curves to map into a d-dimensional parameter space with $d \geq 1$, a natural question is how to measure distances between two elements of this space. Here, we will use the Euclidean norm as a standard measure in $\mathbb{R}^d$ as well as a weighted Euclidean norm which we will show to be interpretable as the Kullback-Leibler divergence between two time series models.

1.1 Outline and Contribution

Outline. Let us briefly sketch the outline of the remainder of this thesis and adduce the main contributions. In Chapter 2 we focus on global bandwidth selection for quasi Maximum Likelihood estimation in linear locally stationary time series models. We adopt a leave-one-out cross validation method from Rice (1984) for the i.i.d. regression model (1.0.3). In our method, the interpretation of the term which is omitted in the leave-one-out estimator will change: We do not omit the t-th observation but the t-th projection error which may be generated by all past observations before time t.

We prove that the obtained bandwidth selector is asymptotically optimal in the sense that the Kullback Leibler divergence of the model connected with this selector to the true model converges to its minimal possible value as the observation time increases to infinity. We use simulations to analyse the behavior of the method in practice for different time series models.

In Chapter 3 we develop general techniques to approximate recursively defined locally stationary processes by stationary processes. We introduce so called derivative processes and give conditions under which existence and uniqueness can be guaranteed. An important result is a Taylor-like expansion of locally stationary processes. The approximation techniques are then used to obtain new versions of standard theorems like a law of large numbers and a central limit theorem for locally stationary processes under minimal moment assumptions. Finally we apply the results to a obtain consistency and asymptotic normality results for Maximum Likelihood estimators.

In Chapter 4 we define a local bandwidth selector for a large class of locally stationary processes which is based on the contrast minimization approach from Goldenshluger and Lepski (2011). We show that this selector is minimax optimal up to a logarithmic factor (which is typical for local model selection procedures) with respect to the Euclidean distance and the Kullback-Leibler distance. For the proofs we use the approximation techniques from Chapter 3. We apply the method to various time series models. Finally, Chapter 5 summarizes the work and gives an outlook into possible future work.

Contributions. The main contributions are:

- Definition of a global bandwidth selector via cross validation for linear locally stationary processes and proof of its asymptotic optimality, see Chapter 2
- Definition of a local bandwidth selector via contrast minimization for a large class of locally stationary processes and proof of its minimax optimality, see Chapter 4
- Creating the set up for a general approximation and maximum likelihood theory for recursively defined locally stationary processes, see Chapter 3

The findings of this thesis regarding bandwidth selection routines can be interpreted as a generalization of the original methods in the nonparametric regression model, because this model is included as a special case. Due to the more general formulation this thesis makes a contribution to understand these methods more deeply.

1.2 Notation and Preliminaries

Here we introduce some basic notation that will be used throughout this thesis.

With $|x|$ we denote the absolute value of real vectors $x \in \mathbb{R}^d$, applied componentwise. For real numbers $q > 0$ we define $|x|_q := \left(\sum_{i=1}^d |x_i|^q\right)^{1/q}$ to be the ℓ^q-norm. Especially $|x|_\infty := \max_{i=1,\dots,d} |x_i|$ denotes the maximum norm. For another vector $y \in \mathbb{R}^d$, we use $\langle x, y\rangle := \sum_{i=1}^d x_i y_i$ to denote the standard scalar product in $\mathbb{R}^d$. We also apply $|\cdot|_q$ to matrices $x \in \mathbb{R}^{d\times d}$ which then means that the matrix is vectorized before $|x|_q$ is evaluated. For instance, $|x|_2 = (\sum_{i,j=1}^{p} |x_{ij}|^2)^{1/2}$ is the Frobenius norm.

For some fixed vector of nonnegative values w, we define the weighted ℓ^q-norm $|x|_{w,q} := \big(\sum_{i=1}^{d} w_i|x_i|^q\big)^{1/q}$.

We use $\lambda_{min}(x)$, $\lambda_{max}(x)$, $|x|_{spec}$ and $\mathrm{tr}(x)$ to denote the minimal / maximal eigenvalue, spectral norm and the trace of a matrix $x \in \mathbb{R}^{d\times d}$, respectively. We write $x \succeq y$ or $x \succ y$ for matrices $x, y \in \mathbb{R}^{d\times d}$ if $x - y$ is positive semidefinite or positive definite, respectively. We use the prime symbol x' to denote the transpose of matrices.

As long as $q \geq 1$, $|\cdot|_{w,q}$ is a norm. For real numbers a, b we use the notation $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$ to denote their maximum and minimum.

For real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$, $(x_1, ..., x_d) \mapsto f(x_1, ..., x_d)$ we use $\partial_i^k f$ or $\partial_{x_i}^k f$ to denote the k-th derivative with respect to the i-th component.

Bias-variance decompositions in nonparametric statistics as well as properties of the corresponding bandwidth selectors are usually stated under the assumption that the true curve is in some function class. In this thesis we will formulate results with the class $\Sigma(\beta, L)$ of Hoelder continuous functions with exponent β,

$$\begin{aligned}\Sigma(\beta, L) := \{g : [0,1] \to \mathbb{R} \quad | \quad & g \text{ is } l_\beta - \text{times differentiable and} \\ & \forall x, y \in T : |g^{(l_\beta)}(x) - g^{(l_\beta)}(y)| \leq L|x-y|^{\beta - l_\beta}\},\end{aligned}$$

where $l_\beta := \max\{k \in \mathbb{N}_0 : k < \beta\}$.

We will sometimes use Landau's notation $a_n = O(b_n)$ and $a_n = o(b_n)$ to determine how real sequences $(a_n), (b_n)$ asymptotically behave with respect to each other. The definitions are as follows:

$$\begin{aligned} a_n = o(b_n) \quad &\Leftrightarrow \quad \lim_{n\to\infty} \left|\frac{a_n}{b_n}\right| = 0, \\ a_n = O(b_n) \quad &\Leftrightarrow \quad \limsup_{n\to\infty} \left|\frac{a_n}{b_n}\right| < \infty. \end{aligned}$$

We say that an event holds a.s. (almost surely), if it is true with probability 1.

1.2.1 The functional dependence measure

For some real-valued random variable Z we define $\|Z\|_q := (\mathbb{E}|Z|^q)^{1/q}$. Let L^q denote the space of real-valued random variables Z with $\|Z\|_q < \infty$.

During the last decades, several measures of dependence for stochastic processes have been invented, for instance mixing properties or joint cumulants. In Chapters 3 and 4 we will make use of a new approach, the (uniform) functional dependence measure which was introduced in Wu (2005) and Liu, Xiao and Wu (2013).

For a sequence of independent and identically distributed (i.i.d.) random variables ε_t, $t \in \mathbb{Z}$ we define the shift process $\mathcal{F}_t := (\varepsilon_t, \varepsilon_{t-1}, ...)$. For $t \geq 0$, let $\mathcal{F}_t^{*(t-k)} := (\varepsilon_t, ..., \varepsilon_{t-k+1}, \varepsilon_{t-k}^*, \varepsilon_{t-k-1}, \varepsilon_{t-k-2}, ...)$, where ε_{t-k}^* is a random variable which has the same distribution as ε_1 and is independent of all ε_t, $t \in \mathbb{Z}$. For a process $Y_t = H_t(\mathcal{F}_t) \in$

L^q with deterministic $H_t : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ we define $Y_t^{*(t-k)} := H_t(\mathcal{F}_t^{*(t-k)})$ and the uniform functional dependence measure

$$\delta_q^Y(k) := \sup_{t \in \mathbb{Z}} \|Y_t - Y_t^{*(t-k)}\|_q, \tag{1.2.1}$$

as well as $\Delta_{m,q}^Y := \sum_{k=m}^{\infty} \delta_q^Y(k)$. If Y_t is stationary, (1.2.1) reduces to the functional dependence measure

$$\delta_q^Y(k) = \|Y_k - Y_k^{*0}\|_q.$$

Furthermore let us define the projection operator $P_j \cdot := \mathbb{E}[\cdot|\mathcal{F}_j] - \mathbb{E}[\cdot|\mathcal{F}_{j-1}]$. It can be shown (cf. Wu (2005), Theorem 1(i) and (ii)) that for $q \geq 1$ it holds that

$$\|P_{t-k} Y_t\|_q \leq \delta_q^Y(k).$$

Chapter 2

Global bandwidth selection with cross validation

In this chapter we discuss adaptive estimation of a multidimensional parameter curve $\theta_0 : [0,1] \to \Theta \subset \mathbb{R}^d$ with cross validation in locally stationary processes. The technical core of the chapter are several results for quadratic statistics needed in this context, meaning that we also restrict ourselves to a quasi-Gaussian likelihood and to linear processes. An advantage of the linear model is that we can formulate all smoothness assumptions with respect to the (time-varying) spectral density of the process which therefore can be easily verified for some standard processes like the tvARMA process.

In Section 2.1 we introduce the locally stationary time series model and formalize the partition of these processes into parametric stationary processes and (unknown) parameter curves. We propose estimators for these curves and define the cross validation procedure. Finally we introduce integrated / averaged squared error type distance measures which are connected to the Kullback-Leibler divergence and will be used to state our results.

In Section 2.2 we state the main result of this chapter, which is the asymptotic optimality of the cross validation procedure with respect to the distance measures defined before. We give an overview of the proof, which is similar to the methods used in Härdle and Marron (1985) and Härdle, Hall and Marron (1988). The result is stated under weak assumptions on the unknown parameter curves, that are Hoelder continuity and bounded variation in each component.

In Section 2.3 we analyze the performance of the method in the case of tvARMA processes in simulations. Some concluding remarks are drawn in Section 2.4. Some lemmas and most of the proofs are deferred without further reference to Section 2.5.

2.1 Introduction

2.1.1 The Model

We start with the definition of the linear locally stationary time series model. Recall that $\Sigma(\beta, L)$ is the class of Hoelder continuous functions with exponent β.

Assumption 2.1.1 (Locally stationary time series model)**.** *Suppose that the observations $X_{t,n}$, $t = 1, ..., n$ have a moving average representation*

$$X_{t,n} = \sum_{j=0}^{\infty} a_{t,n}(j)\varepsilon_{t-j}, \tag{2.1.1}$$

where $a_{t,n}(j)$ are deterministic coefficients and $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}\varepsilon_t^2 = 1$ and existing moments of all orders. We set $\kappa_4 := cum_4(\varepsilon_t)$.

Furthermore, we assume that

$$\sup_{t,n} |a_{t,n}(j)| \leq \frac{C}{\chi(j)} \tag{2.1.2}$$

with

$$\chi(j) := \mathbb{1}_{\{|j|\leq 1\}} + |j| \log^{1+\kappa} |j| \cdot \mathbb{1}_{\{|j|>1\}}$$

for some $\kappa > 0$, and that there exist functions $a(\cdot, j) : [0, 1] \to \mathbb{R}$ with

$$\sup_{j\geq 0} \sum_{t=1}^{n} \left| a_{t,n}(j) - a\Big(\frac{t}{n}, j\Big) \right| \leq C. \tag{2.1.3}$$

We assume that the time dependence of $a(\cdot, j) : [0, 1] \to \mathbb{R}$ is solely via a finite dimensional parameter curve $\theta_0(\cdot)$ whose components are of bounded variation and lie in $\Sigma(\beta, L)$ for some $L, \beta > 0$, i.e. $a(\cdot, j)$ is of the form $a(\cdot, j) = a_{\theta_0(\cdot)}(j)$ with some functions $a_{.}(j)$.

It is well known that the function

$$f_{\theta_0(u)}(\lambda) = \frac{1}{2\pi} |A_{\theta_0(u)}(\lambda)|^2$$

with

$$A_{\theta_0(u)}(\lambda) := \sum_{j=-\infty}^{\infty} a_{\theta_0(u)}(j) \exp(-i\lambda j)$$

then is the time varying spectral density of the process.

Examples/Remark:

(i) As Dahlhaus and Polonik (2009) point out, the complicated construction with different coefficients $a_{t,n}(j)$ and $a(t/n, j)$ is necessary to include important examples such as tvAR - processes. The assumption is fulfilled by tvARMA(p, q) processes (cf. Dahlhaus and Polonik (2009), Proposition 2.4), i.e. by the process

$$X_{t,n} + \sum_{j=1}^{p} \alpha_j\big(\frac{t}{n}\big) X_{t-j,n} = \sum_{k=0}^{q} \beta_k\big(\frac{t}{n}\big) \sigma\big(\frac{t-k}{n}\big)\varepsilon_{t-k}$$

where $\theta(\cdot) = \big(\alpha_1(\cdot), \ldots, \alpha_p(\cdot), \beta_1(\cdot), \ldots, \beta_q(\cdot), \sigma^2(\cdot)\big)'$ consists of the coefficient functions.

(ii) Other important examples are e.g. models with shape- and transition curves (cf. Dahlhaus and Polonik (2009), Proposition 2.4), a simple model being

$$X_{t,n} - 2\, r \cos\big(\phi(\frac{t}{n})\big) X_{t-1,n} + r^2 X_{t-2,n} = \sigma(\frac{t}{n})\, \varepsilon_t \tag{2.1.4}$$

with $\theta(\cdot) = \big(\phi(\cdot), \sigma(\cdot)\big)'$ and $r \in (0,1)$ which models a time varying frequency- and a time varying amplitude-behavior of oscillations.

(iii) We conjecture that the assumption on the existence of all moments of ε_t can be dropped - but the calculations would be very tedious without much additional insight.

2.1.2 The quasi maximum likelihood estimator

As an estimator of $\theta_0(\cdot)$ we consider local conditional Gaussian likelihood estimators weighted by kernels, that is

$$\hat{\theta}_b(u) := \mathrm{argmin}_{\theta \in \Theta}\, L_{n,b}(u, \theta), \tag{2.1.5}$$

where

$$L_{n,b}(u, \theta) := \frac{1}{n} \sum_{t=1}^{n} K_b\Big(\frac{t}{n} - u\Big) \ell_{t,n}(\theta) \tag{2.1.6}$$

and

$$\ell_{t,n}(\theta) := -\log p_\theta(X_{t,n} | X_{t-1,n}, ..., X_{1,n}, X_{0,n} = 0, X_{-1,n} = 0, ...) \tag{2.1.7}$$

is the infinite past likelihood with constant parameter $\theta \in \Theta$ (localized in $L_{n,b}(u, \theta)$ by the kernel K). $K : \mathbb{R} \to \mathbb{R}$ fulfills $\int K = 1$, and $b \in (0, \infty)$ is the bandwidth. We use the common abbreviation $K_b(x) := \frac{1}{b} K(\frac{x}{b})$.

Remark:
(i) For example for tvAR(p) processes one usually replaces $L_{n,b}(u, \theta)$ by $\frac{1}{n} \sum_{t=p+1}^{n} K_b\big(\frac{t}{n} - u\big) \ell_{t,n}(\theta)$. The results of this chapter also hold with this likelihood.

(ii) Using instead the finite past conditional likelihood $\log p_\theta(X_{t,n} | X_{t-1,n}, ..., X_{1,n})$ corresponds to the exact likelihood which usually is much more difficult to calculate and more difficult to investigate theoretically.

It is possible to derive an explicit form of $\ell_{t,n}(\theta)$:

Proposition 2.1.2. *Suppose that Assumption 2.1.1 holds, and $|A_\theta(\lambda)| \geq \delta_A > 0$ uniformly in $\theta \in \Theta$, $\lambda \in [-\pi, \pi]$ for some $\delta_A > 0$. Define the Fourier coefficients*

$$\gamma_\theta(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} A_\theta(-\lambda)^{-1} e^{i\lambda k}\, d\lambda,$$

and

$$d_{t,n}(\theta) := \sum_{k=0}^{t-1} \gamma_\theta(k) X_{t-k,n}. \tag{2.1.8}$$

Then it holds that

$$\ell_{t,n}(\theta) = -\frac{1}{2}\log\left(\frac{\gamma_\theta(0)^2}{2\pi}\right) + \frac{1}{2}\Big[d_{t,n}(\theta)\Big]^2. \tag{2.1.9}$$

2.1.3 Distance measures

In the following, let ∇ denote the derivative with respect to $\theta \in \Theta$. As global distance measures we use the averaged and the integrated squared error (ASE/ISE) weighted by the local Fisher information

$$I(\theta) := \frac{1}{4\pi}\int_{-\pi}^{\pi} (\nabla \log f_\theta(\lambda))(\nabla \log f_\theta(\lambda))' \, \mathrm{d}\lambda. \tag{2.1.10}$$

(cf. Dahlhaus (1996), Theorem 3.6). In addition the weight function $w(\cdot) = w_n(\cdot) := \mathbb{1}_{[b/2,1-b/2]}(\cdot)$ is needed to exclude boundary effects. Since the proof is the same for other weights $w(\cdot)$ we allow in Assumption 2.2.1 for more general weights.

More precisely we set (with $|x|_A^2 := \langle x, Ax\rangle$ for $x \in \mathbb{R}^d$ and a $d \times d$ -matrix A)

$$d_A(\hat\theta_b, \theta_0) := \frac{1}{n}\sum_{t=1}^{n} \left|\hat\theta_b\Big(\frac{t}{n}\Big) - \theta_0\Big(\frac{t}{n}\Big)\right|^2_{I(\theta_0(t/n))} w\left(\frac{t}{n}\right) \tag{2.1.11}$$

and

$$d_I(\hat\theta_b, \theta_0) := \int_0^1 \left|\hat\theta_b(u) - \theta_0(u)\right|^2_{I(\theta_0(u))} w(u) \, \mathrm{d}u. \tag{2.1.12}$$

It can be shown for $w \equiv 1$ that $2d_A$ and $2d_I$ are an approximation of the Kullback-Leibler divergence between models with parameter curves $\hat\theta_b(\cdot)$ and $\theta_0(\cdot)$.

In Theorem 2.2.4 we will prove that $d_A(\hat\theta_b, \theta_0)$ can be approximated uniformly in b by a deterministic distance measure $d_M^{**}(b)$, which has a unique minimizer $b_0 = b_{0,n} \sim n^{-1/5}$. b_0 can be seen as the (deterministic) optimal bandwidth.

2.1.4 The cross validation approach

We now choose the bandwidth b by a generalized cross validation method. We define a 'quasi-leave-one-out' local likelihood

$$L_{n,b,-s}(u,\theta) := \frac{1}{n}\sum_{t=1, t\neq s}^{n} K_b\Big(\frac{t}{n} - u\Big)\ell_{t,n}(\theta) \tag{2.1.13}$$

and a 'quasi-leave-one-out' estimator of θ_0 by

$$\hat\theta_{b,-s}(u) := \mathrm{argmin}_{\theta\in\Theta}\, L_{n,b,-s}(u,\theta). \tag{2.1.14}$$

Here, 'leave-one-out' does not mean that we ignore the s-th observation of the process $(X_{t,n})_{t=1,\dots,n}$, but that we ignore the term which is contributed by the likelihood $\ell_{t,n}$ at time step s. Because of that, we refer to the estimator as a quasi-leave-one-out method.

We then choose $\hat{b}$ via minimizing the cross validation functional

$$CV(b) := \frac{1}{n}\sum_{s=1}^{n} \ell_{s,n}\left(\hat{\theta}_{b,-s}\left(\frac{s}{n}\right)\right) w\left(\frac{s}{n}\right). \tag{2.1.15}$$

It is important to note that such a minimizer $\hat{b}$ of $CV(b)$ does not need to exist, because continuity of $CV(b)$ can not be shown. When b varies it is possible that the location of the minimum of $L_{n,b,-s}(u,\theta)$ changes and therefore $\hat{\theta}_{b,-s}(u)$ makes a jump. Thus we choose $\hat{b}$ such that

$$CV(\hat{b}) - \inf_{b\in B_n} CV(b) \leq \frac{1}{n}, \tag{2.1.16}$$

where B_n is a suitable subinterval of $(0,1)$ which covers all relevant values of b.

2.2 Main results

In this section we present our main results concerning the bandwidth $\hat{b}$ chosen by cross validation. We prove in Theorem 2.2.3 that $\hat{b}$ is asymptotically optimal with respect to d_A, i.e.

$$\lim_{n\to\infty} \frac{d_A(\hat{\theta}_{\hat{b}}, \theta_0)}{\inf_{b\in B_n} d_A(\hat{\theta}_b, \theta_0)} = 1 \quad a.s.,$$

and in Theorem 2.2.5 that $\hat{b}$ is consistent in the sense that $\hat{b}/b_0 \to 1$ a.s., where b_0 is the deterministic optimal bandwidth defined in (2.2.7). In Assumption 2.2.1 we summarize the smoothness conditions on the model class and in Assumption 2.2.2 the conditions on the estimation procedure.

Assumption 2.2.1. *Suppose that*

(i) $\theta \in \Theta$ *is identifiable from* A_θ *(i.e.,* $A_\theta(\lambda) = A_{\theta'}(\lambda)$ *for all* λ *implies* $\theta = \theta'$*) and* $\theta_0(u)$ *lies in the interior of the compact parameter space* $\Theta \subset \mathbb{R}^d$ *for all* $u \in [0,1]$.

(ii) There exists some $\delta_A > 0$ *such that uniformly in* $\theta \in \Theta, \lambda \in [-\pi,\pi]$, $|A_\theta(\lambda)| \geq \delta_A$. $A_\theta(\lambda)$ *is* $\max\{4, l_\beta+1\}$*-times continuously differentiable in* $\theta \in \Theta$*. The derivatives fulfill* $\nabla^k A_\theta(\cdot) \in \Sigma(\beta_A, L_A)$ *uniformly in* $\theta \in \Theta$ *for some* $L_A > 0, \beta_A > 1$.

(iii) The minimal eigenvalue of

$$I(\theta) := \frac{1}{4\pi}\int_{-\pi}^{\pi} (\nabla \log f_\theta(\lambda))(\nabla \log f_\theta(\lambda))' \, d\lambda$$

is bounded away from 0 uniformly in $\theta \in \Theta$.

Assumption 2.2.2. *Suppose that:*

(i) The weight function $\tilde{w} : [0,1] \to \mathbb{R}_{\geq 0}$ *is bounded, has bounded variation and compact support* $\subset [0,1]$ *with nonempty interior. Set* $w(\cdot) := \tilde{w}(\cdot)\mathbb{1}_{[\frac{b}{2},1-\frac{b}{2}]}(\cdot)$.

(ii) For $n \in \mathbb{N}$ *let* $B_n = [\underline{b}, \bar{b}]$, *where* $\underline{b} \geq c_0 n^{\delta-1}$ *and* $\bar{b} \leq c_1 n^{-\delta}$ *for some constants* $c_0, c_1, \delta > 0$.

(iii) $K : \mathbb{R} \to \mathbb{R}$ *fulfills* $\int K(x)\ dx = 1$ *and is Lipschitz continuous with compact support* $[-\frac{1}{2}, \frac{1}{2}]$. *Furthermore,* K *is of order* l_β, *i.e.* $\int x^k K(x)\ dx = 0$ *for* $k = 1, ..., l_\beta$.

We now show that the cross validation bandwidth $\hat{b}$ is asymptotically optimal.

Theorem 2.2.3 (Asymptotic optimality of cross validation)**.** *Under assumptions 2.1.1, 2.2.1 and 2.2.2 the bandwidth* $\hat{h}$ *chosen by cross validation is asymptotically optimal in the sense that*

$$\lim_{n\to\infty} \frac{d(\hat{\theta}_{\hat{b}}, \theta_0)}{\inf_{b\in B_n} d(\hat{\theta}_b, \theta_0)} = 1,$$

where d *is* d_A *or* d_I.

Under stronger smoothness assumptions on $\theta_0(\cdot)$ we will prove (in Theorem 2.2.5 below) that $\hat{b}$ is asymptotically equivalent to the asymptotically optimal bandwidth b_0 (ao-bandwidth for short) which we now define. We know from standard asymptotics that (cf. the proof of Corollary 2.2.7)

$$\hat{\theta}_b(u) - \theta_0(u) \approx -\nabla_\theta^2 L_{n,b}(u, \bar{\theta}(u))^{-1}\, \nabla_\theta L_{n,b}(u, \theta_0(u)) \approx -I(\theta_0(u))^{-1}\, \nabla_\theta L_{n,b}(u, \theta_0(u))$$

which motivates the following approximations to $d_A(\hat{\theta}_b, \theta_0)$ and $d_I(\hat{\theta}_b, \theta_0)$:

$$d_A^*(\hat{\theta}_b, \theta_0) \;:=\; \frac{1}{n}\sum_{t=1}^{n} \Big|\nabla_\theta L_{n,b}\Big(\frac{t}{n}, \theta_0\Big(\frac{t}{n}\Big)\Big)\Big|^2_{I(\theta_0(\frac{t}{n}))^{-1}} w\Big(\frac{t}{n}\Big), \tag{2.2.1}$$

$$d_I^*(\hat{\theta}_b, \theta_0) \;:=\; \int_0^1 \big|\nabla_\theta L_{n,b}(u, \theta_0(u))\big|^2_{I(\theta_0(u))^{-1}} w(u)\ \mathrm{d}u. \tag{2.2.2}$$

As a deterministic approximation of the above distances, we set

$$d_M^*(\hat{\theta}_b, \theta_0) := \mathbb{E}[d_I^*(\hat{\theta}_b, \theta_0)].$$

If θ_0 is twice continuously differentiable, Proposition 2.5.5 implies the usual bias-variance decomposition for d_M^*:

$$d_M^*(\hat{\theta}_b, \theta_0) = \frac{V_0}{nb} + \frac{b^4}{4} B_0 + o((nb)^{-1}) + o(b^4), \tag{2.2.3}$$

uniformly in $b \in B_n$, where (d is the dimension of the parameter space, $\Theta \subset \mathbb{R}^d$)

$$V_0 := \mu_K \int_0^1 \left[d + \kappa_4 \left| \frac{1}{4\pi} \int_{-\pi}^{\pi} \nabla_\theta \log f_{\theta_0(u)}(\lambda) \, \mathrm{d}\lambda \right|^2_{I(\theta_0(u))^{-1}} \right] w(u) \, \mathrm{d}u > 0, \quad (2.2.4)$$

$$B_0 := d_K^2 \int_0^1 \left| \frac{1}{4\pi} \int_{-\pi}^{\pi} \partial_u^2 f(u,\lambda) \cdot \nabla_\theta (f_{\theta_0(u)}(\lambda)^{-1}) \, \mathrm{d}\lambda \right|^2_{I(\theta_0(u))^{-1}} w(u) \, \mathrm{d}u \geq 0, \quad (2.2.5)$$

where $\mu_K := \int K(x)^2 \, \mathrm{d}x$ and $d_K := \int x^2 K(x) \, \mathrm{d}x$, leading to the definition of the asymptotically optimal bandwidth in the following theorem.

Theorem 2.2.4 (Approximation of distance measures)**.** *Let the assumptions of Theorem 2.2.3 hold. Assume that $\theta_0(\cdot)$ is twice continuously differentiable, i.e. $\beta \geq 2$, and define*

$$d_M^{**}(b) := \frac{V_0}{nb} + \frac{b^4}{4} B_0 \quad (2.2.6)$$

If the bias B_0 is not degenerated, i.e. $B_0 > 0$, then it holds

$$\sup_{b \in B_n} \left| \frac{d(\hat{\theta}_b, \theta_0) - d_M^{**}(b)}{d_M^{**}(b)} \right| \to 0 \qquad a.s.$$

where d is d_A or d_I.

Theorem 2.2.5 (consistency of the cross validation bandwidth)**.** *Let the assumptions of Theorem 2.2.4 hold. Then the bandwidth $\hat{b}$ chosen by cross validation fulfills*

$$\frac{\hat{b}}{b_0} \to 1 \qquad a.s.$$

where

$$b_0 = \left(\frac{V_0}{B_0} \right)^{1/5} n^{-1/5}. \quad (2.2.7)$$

is the unique minimizer of $d_M^{**}(b)$.

Proofs.

Here we present the structure of the proofs of Theorems 2.2.3, 2.2.4 and 2.2.5. The technical details including the proofs of the lemmata are postponed to the appendix. From now on, we assume that Assumptions 2.1.1, 2.2.1 and 2.2.2 hold. All convergences stated here are with respect to $n \to \infty$. The following Lemma shows that the approximated distances d_I^*, d_A^* are close to d_M^*.

Lemma 2.2.6. *We have almost surely*

$$\sup_{b \in B_n} \left| \frac{d_I^*(\hat{\theta}_b, \theta_0) - d_M^*(\hat{\theta}_b, \theta_0)}{d_M^*(\hat{\theta}_b, \theta_0)} \right| \to 0, \qquad \sup_{b \in B_n} \left| \frac{d_A^*(\hat{\theta}_b, \theta_0) - d_M^*(\hat{\theta}_b, \theta_0)}{d_M^*(\hat{\theta}_b, \theta_0)} \right| \to 0.$$

As a consequence of Lemma 2.2.6 also the distances d_I, d_A are close to d_M^*:

Corollary 2.2.7. *We have almost surely*

$$\sup_{b\in B_n}\left|\frac{d_I(\hat\theta_b,\theta_0)-d_M^*(\hat\theta_b,\theta_0)}{d_M^*(\hat\theta_b,\theta_0)}\right|\to 0, \qquad \sup_{b\in B_n}\left|\frac{d_A(\hat\theta_b,\theta_0)-d_M^*(\hat\theta_b,\theta_0)}{d_M^*(\hat\theta_b,\theta_0)}\right|\to 0.$$

To get a connection between the distance measure d_M^* and the cross validation functional $CV(b)$, we define

$$\overline{d}_A(\hat\theta_b,\theta_0) := \frac{1}{n}\sum_{t=1}^{n}\left|\hat\theta_{b,-t}\left(\frac{t}{n}\right)-\theta_0\left(\frac{t}{n}\right)\right|^2_{I(\theta_0(t/n))}.$$

The next two lemmata show that $\overline{d}_A$ is close both to d_M^* and $CV(b)$:

Lemma 2.2.8. *We have almost surely*

$$\sup_{b\in B_n}\left|\frac{\overline{d}_A(\hat\theta_b,\theta_0)-d_M^*(\hat\theta_b,\theta_0)}{d_M^*(\hat\theta_b,\theta_0)}\right|\to 0.$$

Lemma 2.2.9. *We have almost surely*

$$\sup_{b\in B_n}\left|\frac{CV(b)-\frac{1}{n}\sum_{t=1}^{n}\ell_{t,n}\left(\theta_0\left(\frac{t}{n}\right)\right)w(t/n)-\overline{d}_A(\hat\theta_b,\theta_0)}{d_M^*(\hat\theta_b,\theta_0)}\right|\to 0. \tag{2.2.8}$$

With the help of these results, we can now prove Theorems 2.2.3, 2.2.4, 2.2.5:

Proof of Theorem 2.2.3. An immediate consequence of Lemma 2.2.9 is (use $\frac{x_1+x_2}{y_1+y_2}\le\frac{x_1}{y_1}+\frac{x_2}{y_2}$ for positive numbers $x_1,x_2,y_1,y_2>0$)

$$\sup_{b,b'\in B_n}\left|\frac{\overline{d}_A(\hat\theta_b,\theta_0)-\overline{d}_A(\hat\theta_{b'},\theta_0)-(CV(b)-CV(b'))}{d_M^*(\hat\theta_b,\theta_0)+d_M^*(\hat\theta_{b'},\theta_0)}\right|\to 0 \quad a.s.$$

almost surely. Now, using Corollary 2.2.7 and Lemma 2.2.8 it is easy to see that

$$\sup_{b,b'\in B_n}\left|\frac{d_A(\hat\theta_b,\theta_0)-d_A(\hat\theta_{b'},\theta_0)-(CV(b)-CV(b'))}{d_A(\hat\theta_b,\theta_0)+d_A(\hat\theta_b,\theta_0)}\right|\to 0 \quad a.s.$$

Choosing $b=\hat b$ and b' such that

$$d_A(\hat\theta_{b'},\theta_0)-\inf_{b\in B_n}d_A(\hat\theta_b,\theta_0)\le\frac{1}{n}$$

yields

$$\begin{aligned}0 \quad\leftarrow\quad & \frac{d_A(\hat\theta_{\hat b},\theta_0)-d_A(\hat\theta_{b'},\theta_0)-(CV(\hat b)-CV(b'))}{d_A(\hat\theta_{\hat b},\theta_0)+d_A(\hat\theta_{b'},\theta_0)}\\ \ge\quad & \frac{d_A(\hat\theta_{\hat b},\theta_0)-\inf_{b\in B_n}d_A(\hat\theta_b,\theta_0)-(\inf_{b\in B_n}CV(b)-CV(b'))}{d_A(\hat\theta_{\hat b},\theta_0)+\inf_{b\in B_n}d_A(\hat\theta_b,\theta_0)+\frac{1}{n}}+\frac{\frac{2}{n}}{d_A(\hat\theta_{\hat b},\theta_0)+d_A(\hat\theta_{b'},\theta_0)}\end{aligned}$$

almost surely. Because of Corollary 2.2.7 and (2.2.3) we have $\sup_{b\in B_n} \frac{\frac{1}{n}}{d_A(\hat{\theta}_b,\theta_0)} \to 0$ a.s. Thus,

$$\frac{d_A(\hat{\theta}_{\hat{b}},\theta_0) - \inf_{b\in B_n} d_A(\hat{\theta}_b,\theta_0)}{d_A(\hat{\theta}_{\hat{b}},\theta_0) + \inf_{b\in B_n} d_A(\hat{\theta}_b,\theta_0)} \to 0 \quad a.s.,$$

from which

$$\frac{d_A(\hat{\theta}_{\hat{b}},\theta_0)}{\inf_{b\in B_n} d_A(\hat{\theta}_b,\theta_0)} \to 1 \quad a.s.$$

follows. The same can be done for d_I. □

Proof of Theorem 2.2.4. Because of $B_0 > 0$ and (2.2.3), we have

$$\sup_{b\in B_n} \left| \frac{d_M^*(\hat{\theta}_b,\theta_0) - d_M^{**}(b)}{d_M^{**}(b)} \right| \to 0 \quad a.s. \tag{2.2.9}$$

Application of Corollary 2.2.7 finishes the proof. □

Proof of Theorem 2.2.5. As in the proof of Theorem 2.2.4, we show (2.2.9). This result in combination with Lemma 2.2.8 and Lemma 2.2.9 gives almost surely

$$\sup_{b\in B_n} \left| \frac{CV(b) - \frac{1}{n}\sum_{t=1}^{n} \ell_{t,n}(\theta_0(t/n))w(t/n) - d_M^{**}(b)}{d_M^{**}(b)} \right| \to 0.$$

Using the same methods as in the proof of Theorem 2.2.3, we have almost surely

$$\frac{d_M^{**}(\hat{b})}{d_M^{**}(b_0)} = \frac{d_M^{**}(\hat{b})}{\inf_{b\in B_n} d_M^{**}(b)} \to 1$$

The structure of $d_M^{**}(b)$ implies $\hat{b}/b_0 \to 1$ a.s. □

2.3 Simulations and Examples

As mentioned below, our results hold for tvARMA-processes and the time varying frequency model defined in (2.1.4). For these models it is also straightforward to check Assumption 2.2.1 (for more details see Dahlhaus and Polonik (2009), Proposition 2.4).

For our simulations we use the following models with $\varepsilon_t \sim N(0,1)$:

$$X_{t,n} = \theta_0\Big(\frac{t}{n}\Big)X_{t-1,n} + \varepsilon_t, \qquad \theta_0(u) = 0.9\,\sin(2\pi u) \tag{2.3.1}$$

$$X_{t,n} = \varepsilon_t - \theta_0\Big(\frac{t}{n}\Big)\varepsilon_{t-1}, \qquad \theta_0(u) = 0.3 + 0.4\,\sin(2\pi u) \tag{2.3.2}$$

and

$$X_{t,n} = 2\,r\cos\Big(\phi\Big(\frac{t}{n}\Big)\Big)X_{t-1,n} - r^2 X_{t-2,n} + \sigma\Big(\frac{t}{n}\Big)\varepsilon_t, \qquad \theta_0(u) = (\phi(u),\sigma(u))', \tag{2.3.3}$$

with $\phi(u) = \frac{\pi}{2} + \frac{\pi}{4}\sin(2\pi u)$, $\sigma(u) = 1.0 + 0.7\sin(2\pi u)$ and $r = 0.9$.

We do not want to go into details on the specific forms on the estimators. We just mention that for (2.3.1) (and for tvAR(p)-models $X_{t,n} = \theta_1(t/n)X_{t-1,n} + ... + \theta_p(t/n)X_{t-p,n} + \sigma(t/n)\varepsilon_t$ in general) the estimator is a Yule-Walker type estimator of the form $\hat{\theta}_b(u) = -\hat{\Gamma}_b(u)^{-1}\hat{\gamma}_b(u)$ and $\hat{\sigma}_b(u) = \frac{1}{n}\sum_{t=p+1}^{n} K_b\left(\frac{t}{n} - u\right) \cdot (X_{t,n} - \hat{\theta}_{b,1}(u)X_{t-1,n} - ... - \hat{\theta}_{b,p}(u)X_{t-p,n})^2$ with covariances

$$\hat{\Gamma}_b(u) := \frac{1}{n}\sum_{t=p+1}^{n} K_b\left(\frac{t}{n} - u\right) \cdot Y_{t-1,n}Y_{t-1,n}', \qquad \hat{\gamma}_b(u) := \frac{1}{n}\sum_{t=p+1}^{n} K_b\left(\frac{t}{n} - u\right) \cdot X_{t,n}Y_{t-1,n}$$

where $Y_{t-1,n} := (X_{t-1,n}, ..., X_{t-p,n})'$. For (2.3.2) we have $A_\theta(\lambda) = 1 - \theta e^{i\lambda}$ leading to $\gamma_\theta(k) = 2\pi\theta^k \cdot \mathbb{1}_{\{k\geq 0\}}$ and therefore to

$$\ell_{t,n}(\theta) = \text{const} + \frac{1}{2}\left(\sum_{k=0}^{t-1} \theta^k X_{t-k,n}\right)^2$$

which we have to minimize numerically. For the model (2.3.3) we obtain

$$\hat{\phi}_b(u) = \cos^{-1}\left(\frac{\hat{\gamma}_{b,1}(u) - r^2\hat{\Gamma}_{b,12}(u)}{2r\hat{\Gamma}_{b,11}(u)}\right)$$

and $\hat{\sigma}_b(u)^2 = \frac{1}{n}\sum_{t=p+1}^{n} K_b\left(\frac{t}{n} - u\right) \cdot (X_{t,n} - 2r\cos(\hat{\phi}_h(u))X_{t-1,n} - r^2X_{t-2,n})^2$.
We performed a Monte Carlo study by generating in each case $N = 1000$ realizations of time series with length $n = 500$.

We chose $B_n = [0.01, 1]$ and calculated the cross-validation bandwidth $\hat{b}$, the ao-bandwidth b_0 ('plugin bandwidth') from Theorem 2.2.5 and the optimal bandwidth

$$b^* = \operatorname{argmin}_{b\in B_n} d_A(\hat{\theta}_b, \theta_0),$$

Note that $\hat{b}, b^*$ depend on the current realization while b_0 is deterministic and fixed (and remember that b^* and b_0 depend on the unknown true curve $\theta_0(\cdot)$ and are not available in practice).

Figure 2.1 shows on the right side histograms of the chosen bandwidths $\hat{b}$ ('Cross-validation'), b^* ('Optimal') for the three models (2.3.1), (2.3.2) and (2.3.3) respectively. We also marked the bandwidth b_0 via a grey vertical dashed line. The variability of the optimal bandwidth b^* reflects nicely the dependence on the specific data-set. $\hat{b}$ has a bigger variance than b^* which is not unexpected since $\hat{b}$ has to compensate the fact that it does not use the unknown parameter curve θ_0. We find it however remarkable that $\hat{b}$ is quite close to b^*.

In the plot on the left hand side of each figure we have visualized the values of $d_A(\hat{\theta}_b, \theta_0)$ for $b \in \{\hat{b}, b_0, b^*\}$ ('Crossvalidation','Plugin','Optimal'). This is perhaps the more important plot since it shows how close the fitted model is to the true one. It can be seen that the estimator based on h_0 behaves nearly optimal. The distances produced by the estimator based on the cross validation procedure are of course greater in average, but they still look quite satisfying in our opinion. Even the models are not directly comparable, we can see that in the case of the tvMA(1) process, the variance of the bandwidth selector is much higher than in the case of the tvAR processes. We conjecture that the main reason for this is the higher variance of the maximum likelihood estimator in tvMA processes, which then leads to higher variances in bandwidth selection.

2.4 Concluding remarks

In this chapter we have introduced a cross validation procedure for linear locally stationary processes which is applicable under weak conditions on the underlying process. The idea of leave-one-out estimators based on omitting the t-th projection error is a general concept which we believe is also applicable in generalizations of the model we discussed. We conjecture that similar results can be shown for multivariate time series as well as nonlinear locally stationary processes. The method works well in simulations, but its quality is connected to the quality of the corresponding maximum likelihood estimator.

An alternative would be a plugin estimator where the bandwidth is (iteratively) estimated by the formula (2.2.7) based on estimates of B_0 and V_0 in (2.2.5) and (2.2.4) respectively. Such estimators are generally regarded as less stable. Furthermore it is much more difficult to estimate these terms in the present situation since the occurring terms are difficult to calculate explicitly.

Based on the simulation results we conjecture that $n^{-1/10} \cdot (\hat{b} - b_0)$ is asymptotically normal if θ_0 is twice continuously differentiable, like Härdle, Hall and Marron (1988) showed in the i.i.d. regression case. This raises the question if there are improved cross validation methods like Hall, Marron and Park (1992) or Chiu (1991) presented in the i.i.d. regression or the kernel regression case that attain better rates if further smoothness assumptions on θ_0 are supposed. However, most of these methods are not applicable in our situation because the unknown parameter curve θ_0 in our model is strongly connected with the stochastic part of the observations $X_{t,n}$, which is not the case in i.i.d. nonparametric regression.

2.5 Lemmas and Proofs

Recall that $|x|_A^2 = x'Ax$ for $A \in \mathbb{R}^{d\times d}, x \in \mathbb{R}^d$ and $|x|_A = \sum_{i,j,k=1}^d A_{ijk}x_i x_j x_k$ for $A \in \mathbb{R}^{d\times d\times d}, x \in \mathbb{R}^d$. For $\theta \in \Theta$, we define a stationary approximation of $X_{t,n}$ by

$$\tilde{X}_t(\theta) := \sum_{k=0}^{\infty} a_\theta(k)\varepsilon_{t-k}. \tag{2.5.1}$$

Furthermore, for a function $g : [0,1] \to \mathbb{R}$ we define the variation of g by

$$V(g) := \sup\Big\{ \sum_{k=1}^{m} |g(x_k) - g(x_{k-1})| : 0 \le x_0 < ... < x_m \le 1, m \in \mathbb{N}\Big\}.$$

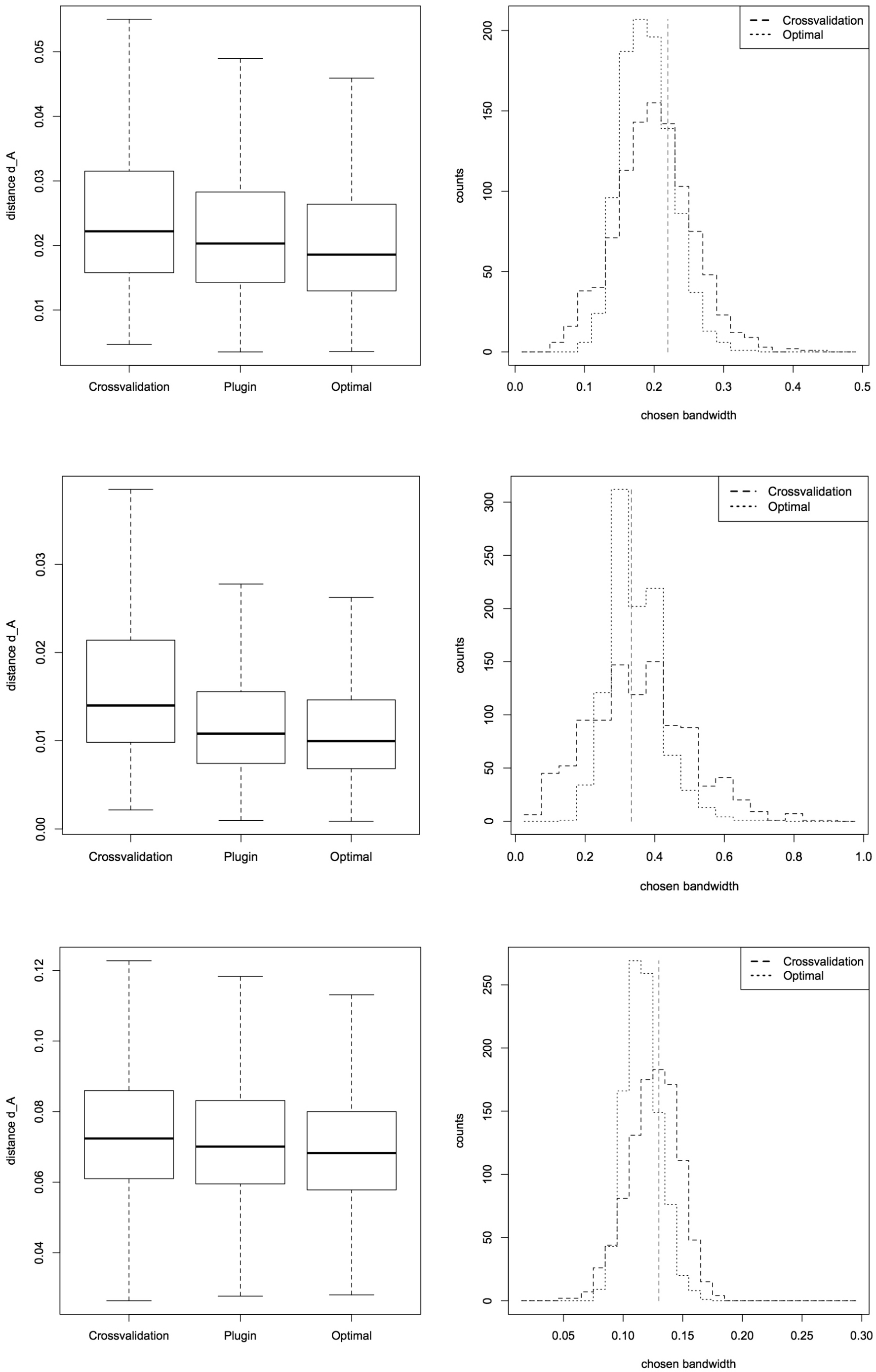

Figure 2.1: Simulation results for the tvAR(1) (top row), the tvMA(1) (second row) and the tvAR(2) (third row) models. Left: Boxplots of the distances $d_A(\hat{\theta}_b, \theta_0)$ obtained with the different procedures. Right: Histogram of bandwidths obtained with the cross validation selector $\hat{b}$ and the (unknown) optimal selector b^*, respectively. The vertical dashed line is the ao-bandwidth b_0.

If $g : [0,1] \to \mathbb{R}^d$ is multivariate, we denote by $V(g)$ the vector $(V(g_i))_{i=1,\ldots,d}$ of variations, applied component-wise.

2.5.1 Coefficient bounds

Here we prove that under Assumption 2.2.1, the coefficients $a(\cdot,k), \gamma_\theta(k), a_\theta(k)$ (recall Assumption 2.1.1 and Proposition 2.1.2 for their definition) are uniformly bounded in θ by absolutely summable sequences and fulfill some further smoothness assumptions. We will use these results in the following Lemmata without further reference.

Lemma 2.5.1. *Let Assumption 2.2.1 hold. For $q = 0,1,2,3,4$ we have with a constant $C > 0$ independent of k,*

$$\sup_{\theta\in\Theta} |\nabla^q \gamma_\theta(k)|_\infty \le \frac{C}{\chi(k)}. \tag{2.5.2}$$

Moreover, we have

$$V(a(\cdot,k)) \le \frac{C}{\chi(k)}, \qquad \sup_u |a(u,k)| \le \frac{C}{\chi(k)}. \tag{2.5.3}$$

Proof. (2.5.2) and a similar result

$$\sup_{\theta\in\Theta} |\nabla^q a_\theta(k)|_\infty \le \frac{C}{\chi(k)} \tag{2.5.4}$$

with some constant $C > 0$ independent of k are consequences of Assumption 2.2.1(ii) (see Katznelson (2004), chapter I, section 4). Because of $a(u,k) = a_{\theta_0(u)}(k)$, the second assertion in (2.5.3) follows from (2.5.4). By uniform continuity of θ_0, it is easily seen that $\theta_0(u_1), \theta_0(u_2)$ lie in some open convex ball included in Θ if $|u_1 - u_2| \le \delta_1$, where $\delta_1 > 0$. Use the mean value theorem to write

$$A_{\theta_0(u_1)}(\lambda) - A_{\theta_0(u_2)}(\lambda) = \langle \nabla A_{\bar\theta(u_1,u_2)}(\lambda), \theta_0(u_1) - \theta_0(u_2)\rangle.$$

with some $\theta(u_1,u_2) \in \Theta$. Note that the variation increases the finer the partition is. We consider only partitions $0 = t_0 < \ldots < t_n = 1$ with $\max_{i=1,\ldots,n} |t_i - t_{i-1}| \le \delta_1$. Then we have

$$\begin{aligned}
\sum_{i=1}^n |a(t_i,k) - a(t_{i-1},k)| &= \frac{1}{2\pi}\sum_{i=1}^n \left| \int \Big(A_{\theta_0(t_i)}(\lambda) - A_{\theta_0(t_{i-1})}(\lambda)\Big) e^{i\lambda k}\, \mathrm{d}\lambda \right| \\
&\le \frac{1}{2\pi}\sum_{i=1}^n |\theta_0(t_i) - \theta_0(t_{i-1})|_1 \cdot \left| \int_{-\pi}^{\pi} \nabla A_{\bar\theta(t_{i-1},t_i)}(\lambda) e^{i\lambda k}\, \mathrm{d}\lambda \right|_\infty \\
&\le \frac{C\cdot d}{2\pi\chi(k)} \cdot |V(\theta_0(\cdot))|_\infty.
\end{aligned}$$

This shows the first assertion in (2.5.3). □

2.5.2 Prediction of linear stationary processes

In the next proposition, we derive the best predictor of $\tilde{X}_t(\theta)$ given the past as well as the corresponding prediction error. The results are then used to prove Proposition 2.1.2.

Proposition 2.5.2 (Prediction of $\tilde{X}_t(\theta)$ given the past). *Assume that*

$$\sup_{\theta\in\Theta}|a_\theta(k)|, \quad \sup_{\theta\in\Theta}|\gamma_\theta(k)| \leq \frac{C}{\chi(k)}$$

for some constant C independent of k. Moreover assume that $A_\theta(\lambda) \geq \delta_A > 0$ uniformly in $\theta \in \Theta, \lambda \in [-\pi,\pi]$. Then the stationary approximation $\tilde{X}_t(\theta)$ of $X_{t,n}$ fulfills for all $\theta \in \Theta$:

$$\sum_{k=0}^{\infty} \gamma_\theta(k)\tilde{X}_{t-k}(\theta) = \varepsilon_t.$$

Moreover, the linear prediction of $\tilde{X}_t(\theta)$ given the past is

$$\mathbb{E}[\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\}] = -\frac{1}{\gamma_\theta(0)}\sum_{k=1}^{\infty} \gamma_\theta(k)\tilde{X}_{t-k}(\theta),$$

with prediction error

$$Var(\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\}) = \frac{1}{\gamma_\theta(0)^2}.$$

The following formula holds:

$$-\log\left(\frac{\gamma_\theta(0)^2}{2\pi}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_\theta(\lambda)\ d\lambda.$$

Proof of Proposition 2.5.2: We have

$$\sum_{k=0}^{\infty} \gamma_\theta(k)\tilde{X}_{t-k}(\theta) = \sum_{d=0}^{\infty}\left(\sum_{k=0}^{d}\gamma_\theta(k)a_\theta(d-k)\right)\varepsilon_{t-d} = \varepsilon_t, \tag{2.5.5}$$

because for each $d \geq 0$ it holds (using Parseval's equality)

$$\sum_{k=0}^{d}\gamma_\theta(k)a_\theta(d-k) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{A_\theta(-\lambda)} \cdot A_\theta(-\lambda)e^{i\lambda d}\,\mathrm{d}\lambda = \begin{cases}1, & d = 0\\ 0, & \text{else}\end{cases}.$$

(2.5.5) together with (2.5.1) implies that $\{\tilde{X}_s(\theta) : s < t\}$ and $\{\varepsilon_s : s < t\}$ generate the same linear closed subspaces of the space of square-integrable random variables L^2. In the case of Gaussian ε_t, the linear prediction of $\tilde{X}_t(\theta)$ given $\{\tilde{X}_s(\theta) : s < t\}$ is

$$\mathbb{E}[\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\}] = -\frac{1}{\gamma_\theta(0)}\sum_{k=1}^{\infty} \gamma_\theta(k)\tilde{X}_{t-k}(\theta),$$

the linear prediction error of $\tilde{X}_t(\theta)$ given $\{\tilde{X}_s(\theta) : s < t\}$ is

$$\sigma_f^2 := \mathrm{Var}(\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\}) = \mathrm{Var}\left(\frac{\varepsilon_t}{\gamma_\theta(0)}\right) = \frac{1}{\gamma_\theta(0)^2}.$$

By Kolmogorov's formula (see Brockwell and Davis (1987), chapter 5.8 therein), we have

$$-\log\left(2\pi\gamma_\theta(0)^2\right) = \log\left(\frac{\sigma_f^2}{2\pi}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_\theta(\lambda)\ \mathrm{d}\lambda.$$

□

Proof of Proposition 2.1.2: Assume $\theta_0(\cdot) = \theta$ with fixed $\theta \in \Theta$. If it holds that $a_{t,n}(k) = a_\theta(k)$, we have $X_{t,n} = \tilde{X}_t(\theta)$ in this case. The negative log Gaussian conditional likelihood of $\tilde{X}_t(\theta)$ given the past has the form

$$\ell_{t,n}(\theta) = \frac{1}{2}\frac{\left(\tilde{X}_t(\theta) - \mathbb{E}[\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\}]\right)^2}{\mathrm{Var}(\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\})} + \frac{1}{2}\log 2\pi \mathrm{Var}(\tilde{X}_t(\theta)|\{\tilde{X}_s(\theta) : s < t\}).$$

Plugging in the results from Proposition 2.5.2, we obtain the claimed result. □

2.5.3 The bias-variance decomposition of $\nabla L_{n,b}(u, \theta_0(u))$

In this chapter, we prove a bias-variance decomposition for $\nabla L_{n,b}(u, \theta_0(u))$.

For a function $\phi(u, \lambda) : [0, 1] \times [-\pi, \pi] \to \mathbb{C}$, let $\hat{\phi}(u, k) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(u, k)e^{i\lambda k}\ \mathrm{d}\lambda$ denote its Fourier coefficients with respect to λ. Define

$$\tilde{\phi}(k) := \max\{\sup_{u\in[0,1]} |\hat{\phi}(u, k)|, \sup_{u\in[0,1]} |\hat{\phi}(u, -k)|\}.$$

Put $r(u) = \mathbb{1}_{(0,1]}(u)$. We will use this function as a data taper, and $X_{t,n}^{(r)} := X_{t,n} \cdot r(\frac{t}{n})$ will denote the tapered version of $X_{t,n}$. Let $c(u, k)$ denote the fourier coefficients of the time-varying spectral density $f(u, \lambda)$, i.e. $c(u, k) = \int_{-\pi}^{\pi} f(u, \lambda)e^{i\lambda k}\ \mathrm{d}\lambda$.

Lemma 2.5.3. *Let Assumption 2.1.1 hold. Let $\phi_1, \phi_2 : [0, 1]\times[-\pi, \pi] \to \mathbb{C}$ be functions satisfying the following conditions:*

$$\sum_{k\in\mathbb{Z}} V(\hat{\phi}_i(\cdot, k)) \leq C_1, \quad \sum_{k\in\mathbb{Z}} \tilde{\phi}_i(k) \leq C_2, \quad |k| \cdot \tilde{\phi}_i(k) \leq C_3, \quad |\phi_i(u, \lambda)|_\infty \leq C_4.$$

Here, the C_i $(i = 1, 2, 3, 4)$ are constants not depending on k. Define

$$L_n(\phi_1, \phi_2) = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{k_1=0}^{t-1} \hat{\phi}_1(t/n, k_1)X_{t-k_1,n}\right)\left(\sum_{k_2=0}^{t-1} \hat{\phi}_2(t/n, k_2)X_{t-k_2,n}\right).$$

Then, we have

$$\mathbb{E}L_n(\phi_1, \phi_2) = \int_0^1\int_{-\pi}^{\pi} f(v, \lambda)\phi_1(v, \lambda)\phi_2(v, -\lambda)\ d\lambda\ dv + R_n^{(1)}.$$

with $|R_n^{(1)}| \leq \frac{C}{n}$, where C does depend only on C_i $(i = 1, 2, 3, 4)$, not on the specific values of the functions ϕ_1, ϕ_2.

Proof of Lemma 2.5.3: Throughout the proof, we use a generic constant C which does only depend C_1, C_2, C_3, C_4 and not on the specific values of ϕ_i. Our proof uses similar techniques as the proofs in the appendix of Dahlhaus and Polonik (2009). We can write

$$L_n(\phi_1, \phi_2) = \frac{1}{n} \sum_{t=1}^{n} \sum_{k_1, k_2 \in \mathbb{Z}} \hat{\phi}_1(t/n, k_1) \hat{\phi}_2(t/n, k_2) X_{t-k_1,n}^{(r)} X_{t-k_2,n}^{(r)}.$$

Thus,

$$\mathbb{E} L_n(\phi_1, \phi_2) = \frac{1}{n} \sum_{t=1}^{n} \sum_{k_1, k_2 \in \mathbb{Z}} \hat{\phi}_1(t/n, k_1) \hat{\phi}_2(t/n, k_2) \mathrm{Cov}(X_{t-k_1,n}^{(r)} X_{t-k_2,n}^{(r)}). \tag{2.5.6}$$

Under Assumption 2.1.1, the following inequalities were shown in Dahlhaus and Polonik (2009) (Proposition 5.4):

$$\begin{aligned} \sum_{t=1}^{n} \left| \mathrm{Cov}(X_{t+k_1,n}^{(r)}, X_{t-k_2,n}^{(r)}) - r\left(\frac{t}{n}\right)^2 \cdot c\left(\frac{t}{n}, k_1 + k_2\right) \right| &\leq C\left(1 + \frac{|k_1|}{\chi(k_1 + k_2)}\right) & (2.5.7) \\ \sum_{j \in \mathbb{Z}} \frac{1}{\chi(j+k)\chi(j)} &\leq \frac{C}{\chi(k)}. & (2.5.8) \\ V(c(\cdot, k)) &\leq \frac{C}{\chi(k)}. & (2.5.9) \end{aligned}$$

where the constants C do not depend on n, k, k_1, k_2. We replace the covariances in (2.5.6) by $r(t/n)^2 c(t/n, k_1 - k_2)$ to get

$$\frac{1}{n} \sum_{t=1}^{n} r(t/n)^2 \sum_{k_1, k_2 \in \mathbb{Z}} \hat{\phi}_1(t/n, k_1) \hat{\phi}_2(t/n, k_2) c(t/n, k_1 - k_2). \tag{2.5.10}$$

with replacement error

$$\begin{aligned} & \frac{C}{n} \sum_{k_1, k_2 \in \mathbb{Z}} \tilde{\phi}_1(k_1) \tilde{\phi}_2(k_2) \left(1 + \frac{|k_1|}{\chi(k_1 + k_2)}\right) \\ \leq \; & \frac{C}{n} \sum_{k_1 \in \mathbb{Z}} \tilde{\phi}_1(k_1) \sum_{k_2 \in \mathbb{Z}} \tilde{\phi}_2(k_2) + \frac{CC_3}{n} \sum_{k_2 \in \mathbb{Z}} \tilde{\phi}_2(k_2) \sum_{k_1 \in \mathbb{Z}} \frac{1}{\chi(k_1 + k_2)} \leq \frac{C}{n}. \end{aligned}$$

Now we replace the sum over t in (2.5.10) by an integral. Because all terms dependent on t/n have uniformly bounded variation, the replacement error is $\leq C/n$ again. In total, we have shown that

$$\begin{aligned} \mathbb{E} L_n(\phi_1, \phi_2) &= \int_0^1 \sum_{k_1, k_2 \in \mathbb{Z}} \hat{\phi}_1(v, k_1) \hat{\phi}_2(v, k_2) c(v, k_1 - k_2) \, \mathrm{d}v + O(n^{-1}) \\ &= \int_0^1 \int_{-\pi}^{\pi} \phi_1(v, \lambda) \phi_2(v, -\lambda) f(v, \lambda) \, \mathrm{d}\lambda \, \mathrm{d}v + O(n^{-1}). \end{aligned}$$

□

Lemma 2.5.4. *Let Assumption 2.1.1 hold. Let $\psi_i : [-\pi, \pi] \to \mathbb{C}$ $(i = 1, 2, 3, 4)$ be functions with*

$$|k| \cdot |\hat{\psi}_i(k)| \leq C_1, \qquad \sum_{k \in \mathbb{Z}} |\hat{\psi}_i(k)| \leq C_2$$

and $\phi_1, \phi_2 : [0,1] \to \mathbb{R}$ functions with $V(\phi_i) \leq C_3$, $|\phi_i|_\infty \leq C_4$ for some constants $C_1, C_2, C_3, C_4 > 0$. Define

$$L_n(\phi_1, \psi_1, \psi_2) := \frac{1}{n} \sum_{t=1}^{n} \phi_1\left(\frac{t}{n}\right) \cdot \left(\sum_{k=0}^{t-1} \hat{\psi}_1(k) X_{t-k,n}\right) \cdot \left(\sum_{l=0}^{t-1} \hat{\psi}_2(l) X_{t-l,n}\right), \tag{2.5.11}$$

then we have

$$\mathbb{E} L_n(\phi_1, \psi_1, \psi_2) = \int_0^1 \phi_1(v) \int_{-\pi}^{\pi} f(v, \lambda) \cdot \psi_1(\lambda) \psi_2(-\lambda) \; d\lambda \; dv + R_n^{(1)} \tag{2.5.12}$$

and

$$\begin{aligned}
& Cov(\sqrt{n} L_n(\phi_1, \psi_1, \psi_2), \sqrt{n} L_n(\phi_2, \psi_3, \psi_4)) \\
= \;& 2\pi \int_0^1 \phi_1(v) \phi_2(v) \cdot \int_{-\pi}^{\pi} f(v, \lambda)^2 \cdot \psi_1(\lambda) \psi_2(-\lambda) \Big[\psi_3(\lambda) \psi_4(-\lambda) + \psi_3(-\lambda) \psi_4(\lambda)\Big] \; d\lambda \; dv \\
& + \kappa_4 \int_0^1 \phi_1(v) \phi_2(v) \left(\int_{-\pi}^{\pi} f(v, \lambda) \psi_1(\lambda) \psi_2(-\lambda) \; d\lambda\right) \cdot \left(\int_{-\pi}^{\pi} f(v, \lambda) \psi_3(\lambda) \psi_4(-\lambda) \; d\lambda\right) \; dv \\
& + R_n^{(2)},
\end{aligned} \tag{2.5.13}$$

where $|R_n^{(1)}|, |R_n^{(2)}| \leq \frac{C}{n}$ and the constant C depends only on C_1, C_2, C_3, C_4 and not on the specific values of the functions ϕ_i, f_i.

Proof of Lemma 2.5.4: Again, we use similar proof techniques as in Dahlhaus and Polonik (2009), Lemma 5.6. Throughout the proof, we use a generic constant C which does only depend C_1, C_2, C_3, C_4 and not on the specific values of f_i, ϕ_i. Write

$$L_n(\phi_1, \psi_1, \psi_2) = \frac{1}{n} \sum_{t=1}^{n} \phi_1\left(\frac{t}{n}\right) \cdot \sum_{k,l \in \mathbb{Z}} \hat{\psi}_1(k) \hat{\psi}_2(l) X_{t-k,n}^{(r)} X_{t-l,n}^{(r)}$$

Discussion of the expectation: We have

$$\mathbb{E} L_n(\phi_1, \psi_1, \psi_2) = \frac{1}{n} \sum_{t,k,l} \phi_1\left(\frac{t}{n}\right) \hat{\psi}_1(k) \hat{\psi}_2(l) \mathrm{Cov}(X_{t-k,n}^{(r)}, X_{t-l,n}^{(r)}). \tag{2.5.14}$$

In (2.5.14) we replace $\mathrm{Cov}(X_{t-k,n}^{(r)}, X_{t-l,n}^{(r)})$ by $r(\frac{t}{n})^2 c(\frac{t}{n}, k-l)$ to obtain

$$\frac{1}{n} \sum_{t,k,l} \phi_1\left(\frac{t}{n}\right) r\left(\frac{t}{n}\right)^2 \hat{\psi}_1(k) \hat{\psi}_2(l) c\left(\frac{t}{n}, k-l\right) \tag{2.5.15}$$

with replacement error (see 2.5.7)

$$\begin{aligned}&\frac{C_4}{n}\sum_{k,l}|\hat{\psi}_1(k)|\cdot|\hat{\psi}_2(l)|\left(1+\frac{|k|}{\chi(k-l)}\right)\\ \leq\quad&\frac{C_4C_1}{n}\sum_l|\hat{\psi}_2(l)|\cdot\sum_k\frac{1}{\chi(k-l)}+\frac{C_4}{n}\sum_k|\hat{\psi}_1(k)|\sum_l|\hat{\psi}_2(l)|\leq\frac{C}{n}.\end{aligned}$$

All terms depending on $\frac{t}{n}$ in (2.5.15) have bounded variation, therefore we can replace (2.5.15) by an integral with replacement error

$$\frac{C}{n}\Big(\|\phi_1\|_\infty+V(\phi_1)\Big)\sum_{k,l}|\hat{\psi}_1(k)||\hat{\psi}_2(l)|\leq\frac{C}{n}$$

and obtain

$$\int_0^1\phi_1(v)\sum_{k,l}\hat{\psi}_1(k)\hat{\psi}_2(l)c(v,k-l)\ \mathrm{d}v=\int_0^1\phi_1(v)\int_{-\pi}^{\pi}f(v,\lambda)\cdot\psi_1(\lambda)\psi_2(-\lambda)\ \mathrm{d}\lambda\ \mathrm{d}v.$$

Discussion of the variance: Recall $\tilde{\phi}(k)=\max\{|\hat{\phi}(k)|,|\hat{\phi}(-k)|\}$ and define $\tilde{c}(k):=\sup_{u\in[0,1]}|c(u,k)|$. Put

$$\begin{aligned}\Psi(s,t,k,l)\ &:=\ \Psi(s,t,k_1,k_2,l_1,l_2):=\phi_1\left(\frac{s}{n}\right)\phi_2\left(\frac{t}{n}\right)\cdot\hat{\psi}_1(k_1)\hat{\psi}_2(k_2)\hat{\psi}_3(l_1)\hat{\psi}_4(l_2),\\ \tilde{\Psi}(k,l)\ &:=\ C_4^2\cdot\tilde{\psi}_1(k_1)\tilde{\psi}_2(k_2)\tilde{\psi}_3(l_1)\tilde{\psi}_4(l_2),\end{aligned}$$

then we have:

$$\begin{aligned}&\mathrm{Cov}(\sqrt{n}\nabla_iL_n(\phi_1,\psi_1,\psi_2),\sqrt{n}L_n(\phi_2,\psi_3,\psi_4))\\ =\quad&\frac{1}{n}\sum_{s,t,k_1,k_2,l_1,l_2}\Psi(s,t,k,l)\cdot\mathrm{Cov}(X_{t-k_1,n}^{(r)}X_{t-k_2,n}^{(r)},X_{s-l_1,n}^{(r)}X_{s-l_2,n}^{(r)}).\end{aligned}\tag{2.5.16}$$

Now we use the formula for the fourth-order cumulant $\kappa(X,Y,Z,W)=\mathbb{E}[XYZW]-\mathbb{E}[XY]\mathbb{E}[ZW]-E[XZ]E[YW]-\mathbb{E}[XW]\mathbb{E}[YZ]$ to write (2.5.16) as

$$\begin{aligned}\frac{1}{n}\sum_{s,t,k_1,k_2,l_1,l_2}\Psi(s,t,k,l)\cdot\Big[\quad&\mathrm{Cov}(X_{t-k_1,n}^{(r)},X_{s-l_1,n}^{(r)})\cdot\mathrm{Cov}(X_{t-k_2}^{(r)},X_{s-l_2,n}^{(r)})\\ &+\mathrm{Cov}(X_{t-k_1,n}^{(r)},X_{s-l_2,n}^{(r)})\cdot\mathrm{Cov}(X_{t-k_2,n}^{(r)},X_{s-l_1,n}^{(r)})\\ &+\kappa(X_{t-k_1,n}^{(r)},X_{t-k_2,n}^{(r)},X_{s-l_1,n}^{(r)},X_{s-l_2,n}^{(r)})\Big]\end{aligned}\tag{2.5.17}$$

We look at the first summand. Define $k_3:=(t-s)-k_1+l_1=(t-k_1)-(s-l_1)$, $k_4:=(t-s)-k_2+l_2=(t-k_2)-(s-l_2)=k_3+k_1-l_1-k_2+l_2$ and $s=t-k_1+l_1-k_3$, then we can replace the first summand by

$$\frac{1}{n}\sum_{k_1,k_2,l_1,l_2,k_3,t=1}r\left(\frac{t}{n}\right)^4\cdot\Psi(t-(k_1-l_1+k_3),t,k,l)\cdot c\left(\frac{t}{n},k_3\right)c\left(\frac{t}{n},k_3+k_1-l_1-k_2+l_2\right)\tag{2.5.18}$$

with replacement error (using (2.5.7) and (2.5.8)), part one:

$$\frac{C}{n}\sum_{k_1,k_2,l_1,l_2,k_3}\tilde{\Psi}(k,l)\cdot\frac{1}{\chi(k_3)}\left(1+\frac{|k_2|}{\chi(k_3+k_1+l_1+k_2+l_2)}\right)$$
$$\leq \frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\cdot\left(1+\frac{|k_2|}{\chi(k_1+l_1+k_2+l_2)}\right)\leq\frac{C}{n}.$$

and part two (replacing the second covariance) having the same form. Now we replace $\Psi(t-(k_1-l_1+k_3),t,k,l)$ by $\Psi(t,t,k,l)$ in (2.5.18) with replacement error

$$\frac{C}{n}\sum_{k_1,k_2,l_1,l_2,k_3}\tilde{\Psi}(k,l)\tilde{c}(k_3)\tilde{c}(k_3+k_1-l_1-k_2+l_2)$$
$$\times\sum_{t=1}^{n}\left|\phi_1\left(\frac{(t-(k_1+l_1-k_3))}{n}\right)-\phi_1\left(\frac{t}{n}\right)\right|$$
$$\leq \frac{C}{n}\sum_{k_1,k_2,l_1,l_2,k_3}\tilde{\Psi}(k,l)\frac{|k_1|+|l_1|+|k_3|}{\chi(k_3)\chi(k_3+k_1-l_1-k_2+l_2)}\leq\frac{C}{n}.$$

Now change the sum over t in (2.5.18) to an integral. All terms that depend on t have bounded variation, therefore the replacement error is again of order $\frac{1}{n}$. We obtain

$$\int_0^1\phi_1(v)^2\cdot\sum_{k_1,k_2,l_1,l_2,k_3}\hat{\psi}_1(k_1)\hat{\psi}_2(k_2)\hat{\psi}_3(l_1)\hat{\psi}_4(l_2)c(v,k_3)c(v,k_3+k_1-l_1-k_2+l_2)\,\mathrm{d}v$$
$$=\ 2\pi\int_0^1\phi_1(v)^2\cdot\int_{-\pi}^{\pi}f(v,\lambda)^2\cdot\psi_1(\lambda)\psi_2(-\lambda)\psi_3(-\lambda)\psi_4(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}v.$$

The second term in (2.5.17) can be dealt with in the same way.
Cumulant term: Using the representation $X_{t,n}=\sum_{j\in\mathbb{Z}}a_{t,n}(t-j)\varepsilon_j$, we get

$$\frac{1}{n}\sum_{k_1,k_2,l_1,l_2,s,t}\Psi(s,t,k,l)\kappa(X_{t-k_1,n},X_{t-k_2,n},X_{s-l_1,n},X_{s-l_2,n})$$
$$=\ \frac{\kappa_4}{n}\sum_{k_1,k_2,l_1,l_2,s,t}\Psi(s,t,k,l)\cdot r\left(\frac{t-k_1}{n}\right)r\left(\frac{t-k_2}{n}\right)r\left(\frac{s-l_1}{n}\right)r\left(\frac{s-l_2}{n}\right)\cdot$$
$$\sum_{i\in\mathbb{Z}}a_{t-k_1,n}(t-k_1-i)a_{t-k_2,n}(t-k_2-i)a_{s-l_1,n}(s-l_1-i)a_{s-l_2,n}(s-l_2-i) \tag{2.5.19}$$

Replacing $h(\frac{t-k_1}{n})$ by $h(\frac{t}{n})$ gives the replacement error

$$\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\cdot\sum_t\left|r\left(\frac{t-k_1}{n}\right)-r\left(\frac{t}{n}\right)\right|$$
$$\cdot\sum_i\frac{1}{\chi(t-k_1-i)\chi(t-k_2-i)}\sum_s\frac{1}{\chi(s-l_1-i)\chi(s-l_2-i)}$$
$$\leq\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\cdot\frac{|k_1|}{\chi(k_1-k_2)\chi(l_1-l_2)}\leq\frac{C}{n},$$

All other replacements lead to similar bounds with k_1, k_2, l_1, l_2 in the nominator of the next-to-last line. Now replace $a_{t-k_1,n}(t-k_1-i)$ in (2.5.19) by $a(\frac{t-k_1}{n}, t-k_1-i)$ and after that by $a(\frac{t}{n}, t-k_1-i)$. This leads to the replacement error (part 1)

$$\begin{aligned}
&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\sum_{t,i}\left|a_{t-k_1,n}(t-k_1-i)-a\left(\frac{t-k_1}{n},t-k_1-i\right)\right| \\
&\qquad\cdot\frac{1}{\chi(t-k_2-i)}\sum_s\frac{1}{\chi(s-l_1-i)\chi(s-l_2-i)} \\
\leq\;&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\cdot\left(\sup_i\sum_t\left|a_{t-k_1,n}(t-k_1-i)-a\left(\frac{t-k_1}{n},t-k_1-i\right)\right|\right) \\
&\qquad\cdot\sum_i\frac{1}{\chi(t-k_2-i)}\sum_s\frac{1}{\chi(s-l_1-i)\chi(s-l_2-i)} \\
\leq\;&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\Psi(k,l)\frac{1}{\chi(l_1-l_2)}\leq\frac{C}{n},
\end{aligned}$$

and with $j := t-k_1-i$ and therefore $i = t-k_1-j$ to the replacement error (part 2),

$$\begin{aligned}
&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\sum_{t,i}\left|a\left(\frac{t-k_1}{n},t-k_1-i\right)-a\left(\frac{t}{n},t-k_1-i\right)\right| \\
&\qquad\qquad\times\frac{1}{\chi(t-k_2-i)}\frac{1}{\chi(l_1-l_2)} \\
\leq\;&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\sum_{t,j}\left|a\left(\frac{t-k_1}{n},j\right)-a\left(\frac{t}{n},j\right)\right|\cdot\frac{1}{\chi(k_1-k_2+j)}\frac{1}{\chi(l_1-l_2)} \\
\leq\;&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\sum_j\frac{|k_1|}{\chi(j)}\frac{1}{\chi(k_1-k_2+j)}\frac{1}{\chi(l_1-l_2)} \\
\leq\;&\frac{C}{n}\sum_{k_1,k_2,l_1,l_2}\tilde{\Psi}(k,l)\frac{|k_1|}{\chi(k_1-k_2)\chi(l_1-l_2)}\leq\frac{C}{n}.
\end{aligned}$$

Replacing the other terms in (2.5.19) lead to similar replacement errors. Therefore, we have the new representation for the covariance (2.5.17)

$$\begin{aligned}
&\frac{\kappa_4}{n}\sum_{k_1,k_2,l_1,l_2,s,t}\Psi(s,t,k,l)\cdot r\left(\frac{t}{n}\right)^2 r\left(\frac{s}{n}\right)^2 \\
&\quad\cdot\sum_i a\left(\frac{t}{n},t-k_1-i\right)a\left(\frac{t}{n},t-k_2-i\right)a\left(\frac{s}{n},s-l_1-i\right)a\left(\frac{s}{n},s-l_2-i\right)
\end{aligned}\tag{2.5.20}$$

Now we replace $a(\frac{s}{n}, s-l_1-i)$ by $a(\frac{t}{n}, s-l_1-i)$. The replacement error can be written by (define $d := s - t$ such that $s = t + d$ and $j := i - t$):

$$\begin{aligned}
& \frac{C}{n} \sum_{k_1,k_2,l_1,l_2,s,t} \Psi(s,t,k,l) \sum_{s,t,j} \left| a\left(\frac{t+d}{n}, d - l_1 - j\right) - a\left(\frac{t}{n}, d - l_1 - j\right) \right| \\
& \qquad \cdot \frac{1}{\chi(k_1 - j)\chi(k_2 - j)\chi(d - l_2 - j)} \\
\leq \; & \frac{C}{n} \sum_{k_1,k_2,l_1,l_2,s,t} \Psi(s,t,k,l) \sum_{d,j} \frac{|d|}{\chi(d - l_1 - j)\chi(d - l_2 - j)\chi(k_1 - j)\chi(k_2 - j)} \\
\leq \; & \frac{C}{n} \sum_{k_1,k_2,l_1,l_2,s,t} \Psi(s,t,k,l) \sum_{d,j} \frac{|d - l_1 - j| + |j - k_1| + |k_1| + |l_1|}{\chi(d - l_1 - j)\chi(d - l_2 - j)\chi(k_1 - j)\chi(k_2 - j)} \\
\leq \; & \frac{C}{n} \sum_{k_1,k_2,l_1,l_2,s,t} \Psi(s,t,k,l) \left(1 + \frac{|k_1| + |l_1|}{\chi(k_1 - k_2)\chi(l_1 - l_2)}\right) \leq \frac{C}{n}.
\end{aligned}$$

Replacing $r(\frac{s}{n})$ by $r(\frac{t}{n})$ and $\Psi(s,t,k,l)$ by $\Psi(t,t,k,l)$ and so on gives similar replacement errors (use the same substitutions). By first summing over s and then over i, we obtain with $c(u,k) = \sum_{j \in \mathbb{Z}} a(u, j+k) a(u,j)$:

$$\begin{aligned}
& \frac{\kappa_4}{n} \sum_{k_1,k_2,l_1,l_2,s,t} \Psi(t,t,k,l) \cdot r\left(\frac{t}{n}\right)^4 \\
& \qquad \cdot \sum_i a\left(\frac{t}{n}, t - k_1 - i\right) a\left(\frac{t}{n}, t - k_2 - i\right) a\left(\frac{t}{n}, s - l_1 - i\right) a\left(\frac{t}{n}, s - l_2 - i\right) \\
= \; & \frac{\kappa_4}{n} \sum_{k_1,k_2,l_1,l_2,t} \Psi(t,t,k,l) \cdot r\left(\frac{t}{n}\right)^4 \cdot c\left(\frac{t}{n}, l_1 - l_2\right) c\left(\frac{t}{n}, k_1 - k_2\right)
\end{aligned}$$

Replacing the sum by an integral (replacement error $\frac{C}{n}$ as before) gives:

$$\begin{aligned}
& \kappa_4 \int_0^1 \phi_1(v)\phi_2(v) \Big(\sum_{k_1,k_2} \hat{\psi}_1(k_1)\hat{\psi}_2(k_2) c(v, k_1 - k_2) \Big) \cdot \Big(\sum_{l_1,l_2} \hat{\psi}_3(l_1)\hat{\psi}_4(l_2) c(v, l_1 - l_2) \Big) \, \mathrm{d}v \\
= \; & \kappa_4 \int_0^1 \phi_1(v)\phi_2(v) \Big(\int_{-\pi}^{\pi} f(v,\lambda)\psi_1(\lambda)\psi_2(-\lambda) \, \mathrm{d}\lambda \Big) \cdot \Big(\int_{-\pi}^{\pi} f(v,\lambda)\psi_3(\lambda)\psi_3(-\lambda) \, \mathrm{d}\lambda \Big) \, \mathrm{d}v.
\end{aligned}$$

□

We are now able to formulate the main result of this section:

Proposition 2.5.5 (The bias-variance decomposition of $\nabla L_{n,b}(u, \theta_0(u))$)**.** *Let Assumptions 2.1.1, 2.2.1 and 2.2.2 hold. Then for each $b \in B_n$, there exists a decomposition*

$$\mathbb{E} \left| \nabla_\theta L_{n,b}(u, \theta_0(u)) \right|^2_{I(\theta_0(u))^{-1}} = v(u,b) + |B(u,b)|^2_{I(\theta_0(u))^{-1}}, \tag{2.5.21}$$

where

$$v(u,h) \;=\; \frac{\mu_K}{nb}\left[d+\kappa_4\left|\frac{1}{4\pi}\int_{-\pi}^{\pi}\nabla_\theta \log f_{\theta_0(u)}(\lambda)\;d\lambda\right|^2_{I(\theta_0(u))^{-1}}\right]+o((nb)^{-1}) \tag{2.5.22}$$

$$B(u,h) \;=\; \frac{1}{4\pi}\frac{1}{b}\int_0^1 K\left(\frac{v-u}{b}\right)\cdot\int_{-\pi}^{\pi}\Big(f(v,\lambda)-f(u,\lambda)\Big)\cdot\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\;d\lambda\;dv + O((nb)^{-1}) \tag{2.5.23}$$

uniformly in $u \in [b/2, 1-b/2]$*,* $b \in B_n$ *with* $\mu_K := \int K(x)^2\;dx$*.*
Furthermore it holds that

$$|B(u,b)|^2_{I(\theta_0(u))^{-1}} \leq Cb^{2\beta} + o((nb)^{-1})$$

uniformly in $u \in [b/2, 1-b/2]$*,* $b \in B_n$*. Moreover, if* θ_0 *is twice continuously differentiable (i.e.* $\beta \geq 2$*), we have with* $d_K := \int x^2 K(x)\;dx$*:*

$$|B(u,b)|^2_{I(\theta_0(u))^{-1}} \;=\; \frac{b^4}{4}d_K^2\left|\frac{1}{4\pi}\int_{-\pi}^{\pi}\partial_u^2 f(u,\lambda)\cdot\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\;d\lambda\right|^2_{I(\theta_0(u))^{-1}} + o(b^4)+o((nb)^{-1}). \tag{2.5.24}$$

uniformly in $u \in [b, 1-b/2]$*,* $b \in B_n$*.*

Proof of Proposition 2.5.5: Obviously, we have a decomposition of the form (2.5.21) with

$$\begin{aligned} B(u,b) &:= \mathbb{E}[\nabla_\theta L_{n,b}(u,\theta_0(u))],\\ v(u,b) &= \mathbb{E}\left|\nabla_\theta L_{n,b}(u,\theta_0(u)) - \mathbb{E}[\nabla_\theta L_{n,b}(u,\theta_0(u))]\right|^2_{I(\theta_0(u))^{-1}}. \end{aligned}$$

We use Lemma 2.5.4 component-wise with $\phi_2(v) = \phi_1(v) = K\left(\frac{v-u}{b}\right)$ and

$$\psi_3(\lambda) = \psi_1(\lambda) = A_{\theta_0(u)}(\lambda)^{-1}, \quad \psi_2(\lambda) = \partial_{\theta_i}\big(A_{\theta_0(u)}(\lambda)^{-1}\big), \quad \psi_4(\lambda) = \partial_{\theta_j}\big(A_{\theta_0(u)}(\lambda)^{-1}\big). \tag{2.5.25}$$

Note that $V(\phi_i) \leq L_K$, where L_K is the Lipschitz constant of the kernel function K, and $|\phi_i|_\infty \leq |K|_\infty := \sup_{u\in[0,1]}|K(u)|$. Furthermore, $|k|\cdot|\hat{\psi}_i(k)|$ and $\sum_{k\in\mathbb{Z}}|\hat{\psi}_i(k)|$ are uniformly bounded in θ by Assumption 2.2.1. Note that $\int_0^1 K_b(u-v)\;\mathrm{d}v = 1$ as long as $u \in [b/2, 1-b/2]$ since K has bounded support $[-\frac{1}{2},\frac{1}{2}]$. Thus, we obtain for the bias term: $B(u,b) = T_1(u,b) + T_2(u,b)$, where

$$\begin{aligned} T_1(u,b) &= \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n} K_b\Big(\frac{t}{n}-u\Big)\cdot\nabla_\theta(-\log(\gamma_{\theta_0(u)}(0)^2))\\ &= -\frac{1}{4\pi}\int_{-\pi}^{\pi} f_{\theta_0(u)}(\lambda)\cdot\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\;\mathrm{d}\lambda + O((nb)^{-1}) \end{aligned}$$

uniformly in $u \in [b/2, 1-b/2]$, $b \in B_n$, and

$$\begin{aligned} T_2(u,b) &= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} K_b\Big(\frac{t}{n}-u\Big)\cdot\left(\sum_{k=0}^{t-1}\gamma_\theta(k)X_{t-k,n}\right)\cdot\left(\sum_{l=0}^{t-1}\nabla_\theta\gamma_\theta(k)X_{t-k,n}\right)\right] \\ &= \frac{1}{b}\cdot\mathbb{E}[L_n(\phi_1,\psi_1,\psi_2)] \\ &= \frac{1}{4\pi}\int_0^1 K_b(v-u)\int_{-\pi}^{\pi} f(v,\lambda)\cdot\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\ \mathrm{d}\lambda\ \mathrm{d}v + O((nb)^{-1}). \end{aligned}$$

uniformly in $u \in [b/2, 1-b/2]$, $b \in B_n$. Therefore, $B(u,b)$ has the form (2.5.23). The estimation error for the bias term follows with the usual Taylor arguments from nonparametric statistics: Since $f(u,\lambda) = f_{\theta_0(u)}(\lambda)$, we have that $u \mapsto f(u,\lambda)$ is l_β-times differentiable. Here, we assume that $\beta > 1$, for the case $0 < \beta \leq 1$ the proof is easier. By a Taylor expansion of $f(v,\lambda)$, we obtain

$$\begin{aligned} &T_1(u,b) + T_2(u,b) \\ = &\ \frac{1}{4\pi}\int_0^1 K_b(v-u)\ \mathrm{d}v\cdot\int_{-\pi}^{\pi}\sum_{k=1}^{l_\beta}\partial_u^k f(u,\lambda)\cdot(v-u)^k\cdot\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\ \mathrm{d}\lambda \\ &+\frac{1}{4\pi}\int_0^1 K_b(v-u)\cdot\int_{-\pi}^{\pi}\int_0^1\big(\partial_u^{l_\beta}f(u+s(v-u),\lambda)-\partial_u^{l_\beta}f(u,\lambda)\big)\cdot\frac{s^{l_\beta-1}}{(l_\beta-1)!}\ \mathrm{d}s \\ &\qquad\times(v-u)^{l_\beta}\cdot\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\ \mathrm{d}\lambda\ \mathrm{d}v + O((nb)^{-1}) \quad (2.5.26) \end{aligned}$$

The first term in (2.5.26) is zero since K is of order l_β, thus $\int_0^1 K_b(v-u)(v-u)^k\ \mathrm{d}v = 0$ for $k = 1, ..., \beta_l$. By Faa di Bruno's rule, we have

$$\partial_u^{l_\beta} f_{\theta_0(u)}(\lambda) = \sum_{\sigma\in\Pi}\nabla_\theta^{|\sigma|}f_{\theta_0(u)}(\lambda)\big[\big(\partial_u^{|D|}\theta_0(u)\big)_{D\in\sigma}\big],$$

where Π is the set of all partitions of the set $\{1, ..., l_\beta\}$, $|\sigma|$ denotes the number of elements of the partition σ and $|D|$ denotes the number of elements of block D in partition σ. Here, for $A \in \mathbb{R}^{d^p}$ and vectors $v_1, ..., v_p \in \mathbb{R}^d$, we define $A\big[(v_j)_{j=1,...,p}\big] := \sum_{i_1,...,i_p=1}^{d} A_{i_1,...,i_p} v_{i_1}\cdot\ldots\cdot v_{i_p}$. We will not go into detail of this formula. Note that for $\sigma = \{\{1\}, ..., \{l_\beta\}\}$, we obtain the summand $\nabla_\theta^{l_\beta} f_{\theta_0(u)}(\lambda)\big[(\partial_u\theta_0(u))_{i=1,...,l_\beta}\big]$ with the highest derivative of $\theta \mapsto f_\theta$ is obtained. By assumption, $\theta \mapsto f_\theta(\lambda)$ is $(l_\beta+1)$-times continuously differentiable. Since $\partial_\theta^{l_\beta+1} f_\theta(\lambda)$ is continuous in both components and $\theta_0 \in \Sigma(\beta, L)$, we have $|\nabla_\theta^{l_\beta} f_{\theta_0(u+s(v-u))}(\lambda)\big[(\partial_u\theta_0(u+s(v-u)))_{i=1,...,l_\beta}\big] - \nabla_\theta^{l_\beta} f_{\theta_0(u)}(\lambda)\big[(\partial_u\theta_0(u))_{i=1,...,l_\beta}\big]| \leq \tilde{C}\cdot|v-u|^{\beta-l_\beta}$ with some $\tilde{C} > 0$. The other summands in $\partial_u^{l_\beta} f_{\theta_0(u)}(\lambda)$ can be analyzed in a similar way. Therefore, we obtain that the second term in (2.5.26) is bounded in $|\cdot|_2$-norm by

$$\frac{C}{4\pi(l_\beta-1)!}\cdot\int_0^1 K_b(v-u)\cdot\int_{-\pi}^{\pi}|v-u|^\beta\cdot|\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})|_2\ \mathrm{d}\lambda \leq C\cdot b^\beta + O((nb)^{-1}).$$

with some $C > 0$. This leads to $|B(u,b)|^2_{I(\theta_0(u))^{-1}} \le |B(u,b)|_2^2 |I(\theta_0(u))|_{spec} \le Cb^{2\beta} + o((nb)^{-1})$. The special case for $\beta = 2$ is easily obtained from (2.5.26).

For the variance term we obtain

$$\begin{aligned} v(u,b) &= \mathbb{E}\,|\nabla_\theta L_{n,b}(u,\theta_0(u)) - \mathbb{E}[\nabla_\theta L_{n,b}(u,\theta_0(u))]|^2_{I(\theta_0(u))^{-1}} \\ &= \sum_{i,j=1}^{d} \left[I(\theta_0(u))^{-1}\right]_{ij} \cdot \mathrm{Cov}\left(\partial_{\theta_i} L_{n,b}(u,\theta_0(u)), \partial_{\theta_j} L_{n,b}(u,\theta_0(u))\right). \end{aligned}$$

An application of Lemma 2.5.4 yields

$$\begin{aligned} &\mathrm{Cov}\left(\sqrt{nb}\partial_{\theta_i} L_{n,b}(u,\theta_0(u)), \sqrt{nb}\partial_{\theta_j} L_{n,b}(u,\theta_0(u))\right) \\ =\ &\frac{1}{b}\mathrm{Cov}\left(\sqrt{n}L_n(\phi_1,\psi_1,\psi_2), \sqrt{n}L_n(\phi_2,\psi_3,\psi_4)\right) \\ =\ &\frac{1}{b}\cdot\frac{1}{4\pi}\int_0^1 K\left(\frac{v-u}{b}\right)^2 \cdot \int_{-\pi}^{\pi} \frac{f(v,\lambda)^2}{f_{\theta_0(u)}(\lambda)^2}\partial_{\theta_i}\log f_{\theta_0(u)}(\lambda)\cdot\partial_{\theta_j}\log f_{\theta_0(u)}(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}v \\ &\quad + \frac{\kappa_4}{(4\pi)^2}\cdot\int_0^1 K\left(\frac{v-u}{b}\right)^2\cdot\left(\int_{-\pi}^{\pi} f(v,\lambda)\partial_{\theta_i}(f_{\theta_0(u)}(\lambda)^{-1})\,\mathrm{d}\lambda\right) \\ &\qquad\qquad \cdot\left(\int_{-\pi}^{\pi} f(v,\lambda)\partial_{\theta_j}(f_{\theta_0(u)}(\lambda)^{-1})\,\mathrm{d}\lambda\right)\,\mathrm{d}v + R_n^{(1)} \end{aligned}$$

where $R_n^{(1)} = O((nb)^{-1})$ uniformly in $u \in [0,1]$, $b \in B_n$. Because $\theta_0 \in \Sigma(\beta, L)$, we can replace $f(v,\lambda) = f_{\theta_0(v)}(\lambda)$ by $f_{\theta_0(u)}(\lambda)$ with replacement error $R_n^{(2)} = O(b^\beta)$ uniformly in $u \in [b/2, 1-b/2]$, $b \in B_n$ (see the calculations regarding the bias above). Therefore, we have using $x'Ax = \mathrm{tr}(Axx')$ for matrices A and vectors x:

$$\begin{aligned} v(u,b) &= \frac{\mu_K}{nb}\cdot\left[\frac{1}{4\pi}\int_{-\pi}^{\pi}(\nabla_\theta\log f_{\theta_0(u)}(\lambda))'\cdot I(\theta_0(u))^{-1}\cdot(\nabla_\theta\log f_{\theta_0(u)}(\lambda))\,\mathrm{d}\lambda\right. \\ &\qquad \left. + \frac{\kappa_4}{(4\pi)^2}\left|\int_{-\pi}^{\pi} f(u,\lambda)\nabla_\theta(f_{\theta_0(u)}(\lambda)^{-1})\,\mathrm{d}\lambda\right|^2_{I(\theta_0(u))^{-1}}\right] + \frac{1}{nb}\left(R_n^{(1)} + R_n^{(2)}\right) \\ &= \frac{\mu_K}{nb}\left[d + \kappa_4\left|\frac{1}{4\pi}\int_{-\pi}^{\pi}\nabla_\theta\log f_{\theta_0(u)}(\lambda)\,\mathrm{d}\lambda\right|^2_{I(\theta_0(u))^{-1}}\right] + \frac{1}{nb}\left(R_n^{(1)} + R_n^{(2)}\right). \end{aligned}$$

□

Corollary 2.5.6 (MISE representation, integrated and summed Bias). *Let Assumptions 2.1.1, 2.2.1 and 2.2.2 hold. Then we have uniformly in $b \in B_n$:*

$$d_M^*(\hat\theta_b,\theta_0) = \mathbb{E}[d_I^*(\hat\theta_b,\theta_0)] = \frac{V_0}{nb} + B^2(b) + o((nb)^{-1}), \tag{2.5.27}$$

$$\mathbb{E}[d_A^*(\hat\theta_b,\theta_0)] = \frac{V_0}{nb} + B^2(b) + o((nb)^{-1}). \tag{2.5.28}$$

where V_0 is defined in (2.2.4) and the integrated bias is

$$B^2(b) \quad := \quad \int_0^1 |B(u,b)|^2_{I(\theta_0(u))^{-1}} w(u) \ du.$$

Furthermore, with the discrete summed bias

$$B^2_{dis}(b) := \frac{1}{n} \sum_{t=1}^{n} \left| B\left(\frac{t}{n}, b\right) \right|^2_{I(\theta_0(t/n))^{-1}} w(t/n),$$

we have uniformly in $b \in B_n$,

$$B^2(b) - B^2_{dis}(b) = o((nb)^{-1}). \tag{2.5.29}$$

Furthermore it holds with a constant $c_0 > 0$:

$$B^2(b) \geq c_0 \int_0^1 |B(u,b)|^2 w(u) \ du, \qquad B^2_{dis}(b) \geq \frac{c_0}{n} \sum_{t=1}^{n} \left| B\left(\frac{t}{n}, b\right) \right|^2 w(t/n). \tag{2.5.30}$$

Proof of Corollary 2.5.6: (2.5.27), (2.5.28) follow from Proposition 2.5.5, where we need the bounded variation of θ_0 and K to approximate the sums by integrals in (2.5.28) and (2.5.29). The estimation (2.5.30) follows from the assumption that the smallest eigenvalue of $I(\theta)$ is uniformly bounded from below by some $\frac{1}{c_0}$, so that $x'I(\theta)^{-1}x \geq x'xc_0$. $\square$

2.5.4 Uniform convergence results and moment inequalities for the local likelihood $L_{n,b}(u,\theta)$ and the maximum likelihood estimator $\hat{\theta}_b(u)$

In this section we show the uniform convergence of quadratic forms of the locally stationary process $X_{t,n}$ towards their expectations. We give convergence rates and prove uniform consistency (w.r.t. u and b) of the maximum likelihood estimator $\hat{\theta}_b(u)$ towards $\theta_0(u)$.

Proposition 2.5.7 (Moment inequality)**.** *Let Assumption 2.1.1 hold. Let $\psi_1, \psi_2 : [-\pi, \pi] \to \mathbb{C}$ be functions which fulfill for some constant $C_1 > 0$:*

$$\sum_{k \in \mathbb{Z}} |\hat{\psi}_i(k)| \leq C_1, \quad i = 1, 2.$$

Let $\phi_1 : [0,1] \to \mathbb{R}$. Then it holds that (see (2.5.11) for the definition of L_n), $p > 2$:

$$\|L_n(\phi_1, \psi_1, \psi_2) - \mathbb{E}[L_n(\phi_1, \psi_1, \psi_2)]\|_p \leq \frac{\tilde{C}_p^{(1)} C_1^2}{n} \cdot \left(\sum_{t=1}^{n} \phi_1 \left(\frac{t}{n}\right)^2 \right)^{1/2},$$

where $\tilde{C}_p^{(1)}$ is the constant from Lemma 2.5.12 which does not depend on the functions ψ_1, ψ_2, ϕ_1.

Proof of Proposition 2.5.7: With the data taper $r(x) = \mathbb{1}_{(0,1]}(x)$, we have

$$\begin{aligned} L_n(\phi_1, \psi_1, \psi_2) &= \frac{1}{n} \sum_{k_1,k_2=0}^{\infty} \hat{\psi}_1(k_1)\hat{\psi}_2(k_2) \sum_{t=1}^{n} \phi_1\left(\frac{t}{n}\right) \cdot X_{t-k_1,n}^{(r)} X_{t-k_2,n}^{(r)} \\ &= \frac{1}{n} \sum_{k_1,k_2=0}^{\infty} \hat{\psi}_1(k_1)\hat{\psi}_2(k_2) \sum_{s=1}^{n} \phi_1\left(\frac{s+\min(k_1,k_2)}{n}\right) \cdot X_{s,n}^{(r)} X_{s-|k_2-k_1|,n}^{(r)}, \end{aligned}$$

therefore we get from the triangle inequality and Lemma 2.5.12 below:

$$\begin{aligned} & \|L_n(\phi_1,\psi_1,\psi_2) - \mathbb{E}[L_n(\phi_1,\psi_1,\psi_2)]\|_p \\ \leq\ & \frac{1}{n} \sum_{k_1,k_2=0}^{\infty} |\hat{\psi}_1(k_1)| \cdot |\hat{\psi}_2(k_2)| \\ & \quad \cdot \left\| \sum_{s=1}^{n-\min(k,l)} \phi_1\left(\frac{s+\min(k_1,k_2)}{n}\right) \left(X_{s,n}^{(r)} X_{s-|k_2-k_1|,n}^{(r)} - \mathbb{E}[X_{s,n}^{(r)} X_{s-|k_2-k_1|,n}^{(r)}] \right) \right\|_p \\ \leq\ & \frac{\tilde{C}_p^{(1)}}{n} \sum_{k_1,k_2=0}^{\infty} |\hat{\psi}_1(k_1)| \cdot |\hat{\psi}_2(k_2)| \\ & \qquad \times \left(\sum_{s=1}^{n-\min(k_1,k_2)} \phi_1\left(\frac{s+\min(k_1,k_2)}{n}\right)^2 r\left(\frac{s}{n}\right)^2 r\left(\frac{s-|k_1-k_2|}{n}\right)^2 \right)^{1/2} \\ \leq\ & \frac{\tilde{C}_p^{(1)}}{n} \Big(\sum_{k_1=0}^{\infty} |\hat{\psi}_1(k_1)| \Big) \cdot \Big(\sum_{k_2=0}^{\infty} |\hat{\psi}_2(k_2)| \Big) \cdot \Big(\sum_{s=1}^{n} \phi_1\left(\frac{s}{n}\right)^2 \Big)^{1/2} \\ \leq\ & \frac{\tilde{C}_p^{(1)} C_1^2}{n} \left(\sum_{s=1}^{n} \phi_1\left(\frac{s}{n}\right)^2 \right)^{1/2}. \end{aligned}$$

□

Proposition 2.5.8 (Uniform convergence of likelihoods and its derivatives)**.** *Let Assumptions 2.1.1, 2.2.1 and 2.2.2 hold. For $k = 0,1,2,3,4$, we have for all $0 < \alpha < \frac{1}{2}$:*

$$\sup_{b\in B_n} \sup_{u\in[0,1]} \sup_{\theta\in\Theta} (nb)^{\frac{1}{2}-\alpha} \Big| \nabla_\theta^k L_{n,b}(u,\theta) - \mathbb{E}[\nabla_\theta^k L_{n,b}(u,\theta)] \Big| \to 0. \tag{2.5.31}$$

Moreover, for $u \in supp(w)$ and n large enough, we have

$$\begin{aligned} \mathbb{E}[L_{n,b}(u,\theta)] &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f_\theta(\lambda) + \frac{f(u,\lambda)}{f_\theta(\lambda)} \right\} d\lambda + O((nb)^{-1} + b^\beta), \\ \mathbb{E}[\nabla_\theta L_{n,b}(u,\theta)] &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{f(u,\lambda) - f_\theta(\lambda)}{f_\theta(\lambda)} \cdot \nabla_\theta \log f_\theta(\lambda)\ d\lambda + O((nb)^{-1} + b^\beta), \\ \mathbb{E}[\nabla_\theta^2 L_{n,b}(u,\theta)] &= \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{f(u,\lambda) - f_\theta(\lambda)}{f_\theta(\lambda)} \cdot \nabla_\theta^2 \log f_\theta(\lambda)\ d\lambda + I(\theta) + O((nb)^{-1} + b^\beta). \end{aligned}$$

Proof. In this proof, we will use C as a generic constant that may change from line to line but is independent of θ, u, b, n. The term $\nabla_\theta^k L_{n,b}(u,\theta)$ splits (up to determistic terms) into summands of the form

$$L_n(\phi_1, \psi_1, \psi_2)$$

with $\phi(v) = \phi_{h,u}(v) = \frac{1}{b} K\left(\frac{v-u}{b}\right)$ and, for example, $\psi_1 = \psi_{1,\theta} = \psi_2 = \psi_{2,\theta} = \frac{1}{A_\theta}$ for $L_{n,b}$. We will show the stated convergence for $L_n(\phi_1, \psi_1, \psi_2)$, then the assertion (2.5.31) follows. We define

$$f(\xi) := L_n(\phi_1, \psi_1, \psi_2) - \mathbb{E}[L_n(\phi_1, \psi_1, \psi_2)],$$

where $\xi = (b, u, \theta) \in \Xi_n := B_n \times [0,1] \times \Theta$. For each $r > 0$, we can find a space Ξ_n' with $\#\Xi_n' < c_q n^q$ such that the compact space Ξ_n is approximated in the following way: for each $\xi = (b, u, \theta) \in \Xi_n$ there is a $\xi' = (b', u', \theta') \in \Xi_n'$ such that $|\xi - \xi'|_1 \leq c_r n^{-r}$. Then we have for $0 < \alpha < \frac{1}{2}$,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\xi \in \Xi_n} (nb)^{\frac{1}{2}-\alpha} |f(\xi)| > \varepsilon\Big) \\
\leq\ & \mathbb{P}\Big(\sup_{\xi' \in \Xi_n'} (nb)^{\frac{1}{2}-\alpha} |f(\xi')| > \frac{\varepsilon}{2}\Big) + \mathbb{P}\Big(\sup_{\xi \in \Xi_n, \xi' \in \Xi_n', |\xi - \xi'|_1 \leq c_r n^{-r}} (nb)^{\frac{1}{2}-\alpha} |f(\xi) - f(\xi')| > \frac{\varepsilon}{2}\Big) \\
=:\ & I_n + II_n.
\end{aligned}$$

Our goal now is to bound I_n, II_n by absolutely summable sequences in n to apply Borel-Cantelli's lemma.

Because of Assumption 2.2.1 (see Lemma 2.5.1), the sums $\sum_{k=0}^\infty |\hat{\psi}_i(k)| \leq C_1$, $\sum_{k=0}^\infty |\nabla \hat{\psi}_i(k)| \leq C_1$ $(i = 1, 2)$ are uniformly bounded in θ. From Proposition 2.5.7 we obtain

$$\begin{aligned}
\|f(\xi)\|_p &= \|L_n(\phi_1, \psi_1, \psi_2) - \mathbb{E} L_n(\phi_1, \psi_1, \psi_2)\|_p \leq \frac{C_p''}{nb} \left(\sum_{t=1}^n K\left(\frac{\frac{t}{n} - u}{b}\right)^2 \right)^{1/2} \\
&\leq C(nb)^{-1/2}.
\end{aligned}$$

We conclude that

$$\begin{aligned}
\sup_{\xi \in \Xi_n} \mathbb{P}\Big((nb)^{\frac{1}{2}-\alpha} |f(\xi)| > \varepsilon/2\Big) &\leq \sup_{\xi \in \Xi_n} \frac{(nb)^{\frac{1}{2}-\alpha} \|f(\xi)\|_p^p}{(\varepsilon/2)^p} \leq \left(\frac{C}{\varepsilon/2}\right)^p \cdot \sup_{b \in B_n} \left[(nb)^{-\alpha p}\right] \\
&\leq C \cdot n^{-\alpha \delta p},
\end{aligned}$$

and thus for p large enough,

$$I_n \leq \#\Xi_n' \cdot \sup_{\xi \in \Xi_n} \mathbb{P}\Big((nb)^{\frac{1}{2}-\alpha} |f(\xi)| > \varepsilon/2\Big) \leq C \cdot n^{q - \alpha \delta p}$$

is bounded by an absolutely summable sequence in n.
Let us introduce some notation to simplify the proof for the second term II_n. Define the Toeplitz matrix

$$T_n(\psi)_{jk} := \int_{-\pi}^{\pi} \psi(\lambda) e^{i\lambda(j-k)}\, \mathrm{d}\lambda,$$

an set $G_\theta := T_n(\psi_1)$, $H_\theta := T_n(\psi_2)$. Put

$$D_{u,b} = \frac{1}{b}\mathrm{diag}\Big(K\Big(\frac{\frac{t}{n}-u}{b}\Big) : t = 1, ..., n\Big).$$

Note that with $\underline{X} := (X_{1,n}, ..., X_{n,n})'$ we have

$$L_n(\phi_1, \psi_1, \psi_2) = \underline{X}'G_\theta' D_{u,b} H_\theta \underline{X}.$$

Now we prove that such functions form a Lipschitz class with respect to $\Xi_n = B_n \times [0,1] \times \Theta$. The well-known inequalities $|x'Ax| \le x'x \cdot |A|_{spec}$ and $|AB|_{spec} \le |A|_{spec}|B|_{spec}$ give for $\xi = (b, u, \theta)$, $\xi' = (b', u', \theta') \in \Xi_n$:

$$\begin{aligned} & |\underline{X}'A_\theta' D_{u,b} B_\theta \underline{X} - \underline{X}'G_{\theta'}' D_{u',b'} H_{\theta'} \underline{X}| \\ \le\ & \underline{X}'\underline{X} \cdot \Big[|G_\theta - G_{\theta'}|_2 |D_{u,b}|_2 |H_\theta|_{spec} + |D_{u,b} - D_{u',b'}|_{spec}|G_{\theta'}|_{spec}|H_\theta|_{spec} \\ & \quad + |H_\theta - H_{\theta'}|_{spec}|G_{\theta'}|_{spec}|D_{u',b'}|_{spec}\Big] \end{aligned}$$

Now we give estimates for the terms appearing above. For Toeplitz matrices it holds that $|T_n(\psi)|_{spec} \le \sum_{k\in\mathbb{Z}} |\hat{\psi}(k)|$. Uniformly in $(b, u, \theta) \in \Xi_n$, we therefore have with some constant $C_1 > 0$,

$$|D_{u,b}|_{spec} \le \frac{K(0)}{\underline{b}}, \quad |G_\theta|_{spec}, |H_\theta|_{spec} \le C_1.$$

Furthermore, with some intermediate value $\bar{\theta} \in \Theta$ and some constant $C_2 > 0$,

$$|G_\theta - G_{\theta'}|_{spec} = |T_n(\nabla_\theta \psi_{1,\bar{\theta}} \cdot (\theta - \theta'))|_{spec} \le d \cdot C_2 \cdot |\theta - \theta'|_1,$$

the same holds for H_θ. Finally, note that

$$|D_{u,b} - D_{u',b'}|_{spec} \le \sup_{v\in[0,1]} \left| \frac{1}{b}K\left(\frac{v-u}{b}\right) - \frac{1}{b'}K\left(\frac{v'-u}{b'}\right)\right| \le \frac{C}{\underline{b}^3}\Big(|u-u'| + |b-b'|\Big).$$

For the expectation $\mathbb{E}L_n(\phi_1, \psi_1, \psi_2)$ of $L_n(\phi_1, \psi_1, \psi_2)$, we can use the same bounds as used above. We have shown that (keep in mind that $\underline{b} \ge c_0 n^{\delta-1}$, see Assumption 2.2.2)

$$|f(\xi) - f(\xi')| \le C(n) \cdot \Big(\underline{X}'\underline{X} + \mathbb{E}[\underline{X}'\underline{X}]\Big) \cdot |\xi - \xi'|_1,$$

where the deterministic $C(n)$ grows only polynomially fast in n. Choose r large enough such that $C(n)n^{-r} = o(n^{-(1+\gamma)})$, then we have

$$\begin{aligned} II_n &\le \mathbb{P}\left(c_r C(n) n^{-r}\Big(\underline{X}'\underline{X} + \mathbb{E}[\underline{X}'\underline{X}]\Big) > \frac{\varepsilon}{2}\right) \\ &\le C\frac{\|\underline{X}'\underline{X} - \mathbb{E}[\underline{X}'\underline{X}]\|_2^2}{n^{2(1+\gamma)}} + \mathbb{P}\left(Cn^{-(1+\gamma)}\mathbb{E}[\underline{X}'\underline{X}] > \frac{\varepsilon}{2}\right) \\ &\le \frac{C}{n^{1+2\gamma}} + \mathbb{P}\left(Cn^{-(1+\gamma)}\mathbb{E}[\underline{X}'\underline{X}] > \frac{\varepsilon}{2}\right). \end{aligned}$$

which is absolutely summable again (note that $\frac{1}{n}\underline{X}'\underline{X} = L_n(1,1,1)$, from that we get the estimation $\|\underline{X}'\underline{X} - \mathbb{E}[\underline{X}'\underline{X}]\|_2^2 \leq Cn$ from Proposition 2.5.7 and $\frac{1}{n}\mathbb{E}[\underline{X}'\underline{X}] = \int_0^1 \int_{-\pi}^{\pi} f(v,\lambda)\, \mathrm{d}\lambda\, \mathrm{d}v + O(n^{-1})$ from Lemma 2.5.4).

The second part of the assertion follows from Lemma 2.5.4 and the Hoelder continuity of θ_0 which allows us to replace $f(v,\lambda)$ by $f(u,\lambda)$ with replacement error $O(b^\beta)$ uniformly in b, u, θ. □

Lemma 2.5.9 (Uniform convergence of the leave-one-out likelihood)**.** *For* $k = 0, 1, 2, 3, 4$ *we have for all* $0 < \alpha < 1$ *almost surely*

$$\sup_{s=1,\dots,n} \sup_{b \in B_n} \sup_{u \in [0,1]} \sup_{\theta \in \Theta} (nb)^{1-\alpha} \Big| \nabla_\theta^k L_{n,b,-s}(u,\theta) - \nabla_\theta^k L_{n,b}(u,\theta) \Big| \to 0.$$

Proof of Lemma 2.5.9: Because the structure of $\nabla_\theta^k L_{n,b}(u,\theta)$ is the same for $k = 0, 1, 2, 3, 4$, we only look at the case $k = 0$. Here, for $s = 1, ..., n$ we have

$$L_{n,b}(u,\theta) - L_{n,b,-s}(u,\theta) = \frac{1}{nb} K\left(\frac{\frac{s}{n} - u}{b}\right) \ell_{s,n}(\theta).$$

Since K is bounded, we have with $\phi_1(k) = \phi_{1,s}(k) = \mathbb{1}_{\{k \leq s\}} \frac{1}{\chi(s-k)}$ and using the results from Proposition 2.5.1 and (2.5.8):

$$(nb) \cdot |L_{n,b}(u,\theta) - L_{n,b,-s}(u,\theta)| \leq C\Big(1 + \sum_{k=1}^{s} \frac{X_{k,n}^2}{\chi(s-k)}\Big) = C\Big(1 + L_n(\phi_{1,s}, 1, 1)\Big).$$

As a consequence,

$$\sup_{s,b,u,\theta} (nb)^{1-\alpha} |L_{n,b}(u,\theta) - L_{n,b,-s}(u,\theta)| \leq \frac{C}{n^{\delta\alpha}} + \frac{C}{n^{\delta\alpha}} \sup_{s=1,\dots,n} |L_n(\phi_{1,s}, 1, 1) - \mathbb{E} L_n(\phi_{1,s}, 1, 1)|$$

Application of Proposition 2.5.7 gives

$$\|L_n(\phi_{1,s}, 1, 1) - \mathbb{E}[L_n(\phi_{1,s}, 1, 1)]\|_p \leq C_p'' \left(\sum_{k=1}^{s} \frac{1}{\chi(s-k)^2}\right)^{1/2} \leq C,$$

Thus for n large enough,

$$\begin{aligned} & \mathbb{P}(\sup_{s,b,u,\theta} (nb)^{1-\alpha} |L_{n,b}(u,\theta) - L_{n,b,-s}(u,\theta)| > \varepsilon) \\ \leq \;& n \cdot \sup_{s=1,\dots,n} \mathbb{P}\left(\frac{C}{n^{\delta\alpha}} |L_n(\phi_{1,s}, 1, 1) - \mathbb{E} L_n(\phi_{1,s}, 1, 1)| > \varepsilon/2\right) \\ \leq \;& C n^{1-\delta\alpha p}, \end{aligned}$$

which is absolutely summable for p large enough and thus yields the result with Borel-Cantelli's lemma. □

Theorem 2.5.10 (Uniform strong consistency of the maximum likelihood estimator)**.** *Let Assumptions 2.1.1, 2.2.1 and 2.2.2 hold. Define*

$$L(u,\theta) \ := \ \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big\{\log f_\theta(\lambda) + \frac{f(u,\lambda)}{f_\theta(\lambda)}\Big\}\ d\lambda.$$

Then it holds that

$$\sup_{b\in B_n}\sup_{u\in supp(w)}\sup_{\theta\in\Theta}\Big|L_{n,b}(u,\theta) - L(u,\theta)\Big| \to 0 \quad a.s.$$

and

$$\sup_{b\in B_n}\sup_{u\in supp(w)}\Big|\hat\theta_b(u) - \theta_0(u)\Big| \to 0 \quad a.s. \tag{2.5.32}$$

Proof. Use Proposition 2.5.8 for the uniform convergence of $L_{n,b}$. The identifiability condition in Assumption 2.2.1 implies that $L(u,\theta)$ attains its local minimum at $\theta = \theta_0(u)$ since

$$\log\frac{f_\theta(\lambda)}{f(u,\lambda)} + \frac{f(u,\lambda)}{f_\theta(\lambda)} - 1 \geq \Big(1 - \frac{f(u,\lambda)}{f_\theta(\lambda)}\Big) + \frac{f(u,\lambda)}{f_\theta(\lambda)} - 1 = 0$$

with equality if and only if $\theta = \theta_0(u)$ (here, we used the inequality $\log(x^{-1}) \geq 1 - x$). Standard arguments provide the uniform convergence of $\hat\theta_b$. For more details, we refer to Chapter 3 and the proof of Theorem 3.3.2. □

2.5.5 Bounds for moments of sums, quadratic and cubic forms of covariates

The inequalities derived in this section are needed to prove the moment inequalities for the local likelihoods and its derivatives in Section 2.5.4. The proofs of the following two lemmata mimic ideas from Subba Rao (2010), Lemma 4.2 therein.

Lemma 2.5.11. *Let Assumption 2.1.1 hold. Define $\eta_{t,k} := X_{t,n}X_{t-k,n} - \mathbb{E}[X_{t,n}X_{t-k,n}]$ let $\mathcal{F}_t := \sigma(\varepsilon_s : s \leq t)$ be the σ-algebra generated from ε_s, $s \leq t$. For integers $t, j_1, j_2, j_3, i_1, i_2 \geq 0$ and $k, l, m \geq 0$, define*

$$\begin{aligned}
M_{j_1,k}(t-j_1) \ &:= \ \mathbb{E}[\eta_{t,k}|\mathcal{F}_{t-j_1}] - \mathbb{E}[\eta_{t,k}|\mathcal{F}_{t-j_1-1}],\\
A_{j_1,j_2,i}(t-j_1-i_1) \ &:= \ \mathbb{E}[M_{j_1,k}(t-j_1)M_{j_2,l}(t-j_1)|\mathcal{F}_{t-j_1-i_1}]\\
&\qquad -\mathbb{E}[M_{j_1,k}(t-j_1)M_{j_2,l}(t-j_1)|\mathcal{F}_{t-j_1-i_1-1}],\\
B_{j_1,j_2,j_3,i_1,i_2}(t-j_3-i_2) \ &:= \ \mathbb{E}[A_{j_1,j_2,i_1}(t-j_3)\cdot M_{j_3,m}(t-j_3)|\ \mathcal{F}_{t-j_3-i_2}]\\
&\qquad -\mathbb{E}[A_{j_1,j_2,i_1}(t-j_3)\cdot M_{j_3,m}(t-j_3)|\ \mathcal{F}_{s-j_3-i_2-1}].
\end{aligned}$$

Define the absolutely summable sequence $\psi_k(j) := \frac{1}{\chi(j)} + \mathbb{1}_{\{j\geq k\}}\frac{1}{\chi(j-k)}$. Fix some $p > 2$. Then there exist constants $C_p^{(i)} > 0$ ($i = 1,2,3$). dependent only on p, the moments of

ε_0 *and the constant in (2.1.2) such that:*

$$\begin{aligned}
\|M_{j_1,k}(t-j_1)\|_p &\leq C_p^{(1)} \cdot \psi_k(j_1), \\
\|A_{j_1,j_2,i_1}(t-j_1-i_1)\|_p &\leq C_p^{(2)} \cdot \psi_k(j_1)\psi_l(j_2) \cdot \Big(\psi_k(j_1+i_1) + \psi_l(j_2+i_1)\Big), \\
\|B_{j_1,j_2,j_3,i_1,i_2}(t-j_3-i_2)\|_p &\leq C_p^{(3)} \cdot \psi_k(j_1)\psi_l(j_2)\psi_m(j_3) \cdot \Big(\psi_k(j_1+i_1) + \psi_l(j_2+i_1)\Big) \\
&\quad \cdot\Big(\psi_k(j_1+i_1+i_2) + \psi_l(j_2+i_1+i_2) + \psi_m(j_3+i_2)\Big).
\end{aligned}$$

Proof of Lemma 2.5.11: Let $g(j)$ be a generic sequence (maybe dependent on t, n and the indices j_1, j_2, j_3, i_1, i_2) with the property $|g(j)| \leq \frac{C}{\chi(j)}$ uniformly in t, n and the indices. Define a second generic sequence $h_k(j) := \mathbb{1}_{\{j \geq k\}} g(j-k)$. Furthermore, let $\mathcal{F}_{>t}$ denote a (generic) random variable with expectation 0 which is $\sigma(\varepsilon_{t+s} : s > 0)$-measurable. Then we have

$$X_{t,n} = \mathcal{F}_{>t-j_1} + \sum_{j=j_1}^{\infty} g(j)\varepsilon_{t-j}, \quad X_{t-k,n} = \mathcal{F}_{>t-j_1} + \sum_{j=j_1}^{\infty} h_k(j)\varepsilon_{t-j}.$$

It follows

$$\mathbb{E}[\eta_{t,k}|\mathcal{F}_{t-j_1}] = \left(\sum_{j\geq j_1} g(j)\varepsilon_{t-j}\right)\left(\sum_{j'\geq j_1} h_k(j')\varepsilon_{t-j'}\right) - \mathbb{E}[\text{this term}],$$

thus with the definition $g_k(j_1, j) := g(j_1)h_k(j) + h_k(j_1)g(j)$, we have

$$M_{j_1,k}(t-j_1) = \varepsilon_{t-j_1} \sum_{j\geq j_1+1} g_k(j_1, j)\varepsilon_{t-j} + g(j_1)h_k(j_1) \cdot (\varepsilon_{t-j_1}^2 - 1). \tag{2.5.33}$$

Recall the definition of $\psi_k(j)$. Note that

$$|g_k(j, j')| = |g(j)h_k(j') + h_k(j)g(j')| \leq \psi_k(j)\psi_k(j').$$

Applying Rio (2009), Theorem 2.1 therein and the Hoelder inequality yields

$$\begin{aligned}
\|M_{j_1,k}(t-j_1)\|_p &\leq (2p-1)^{1/2}\|\varepsilon_0\|_{2p}^2 \cdot \left(\sum_{j\geq j_1+1} g_k(j_1, j)^2\right)^{1/2} \\
&\quad + |g(j_1)h_k(j_1)| \cdot (\|\varepsilon_0\|_{2p}^2 + 1) \\
&\leq \psi_k(j_1) \cdot C^2 \cdot \Big\{\|\varepsilon_0\|_{2p}^2 \cdot \Big[2\tilde{C} \cdot (2p-1)^{1/2} + 1\Big] + 1\Big\} =: \psi_k(j_1) \cdot C_p^{(1)}.
\end{aligned}$$

In the last line we used $\left(\sum_j \psi_k(j)^2\right)^{1/2} \leq 2\left(\sum_j \frac{1}{\chi(j)^2}\right)^{1/2} =: \tilde{C}$. Using the formula (2.5.33) with $t + j_2 - j_1$ instead of t yields with $j^*(j) := j - j_1 + j_2$

$$M_{j_2,l}(t-j_1) = \varepsilon_{t-j_1} \sum_{j\geq j_1+1} g_l(j_2, j^*)\varepsilon_{t-j} + g(j_2)h_l(j_2) \cdot (\varepsilon_{t-j_1}^2 - 1).$$

Straightforward calculations reveal

$$\begin{aligned}
& A_{j_1,j_2,i_1}(t-j_1-i_1) \\
= \; & \varepsilon_{t-j_1-i_1} \sum_{j \geq j_1+i_1+1} \Big[g_k(j_1,j) g_l(j_2,j_2+i_1) + g_k(j_1,j_1+i_1) g_l(j_2,j^*) \Big] \varepsilon_{t-j} \\
& +\mathbb{E}[\varepsilon_0^3] \Big[g(j_1) h_k(j_1) g_l(j_2,j_2+i_1) + g(j_2) h_k(j_2) g_k(j_1,j_1+i_1) \Big] \varepsilon_{t-j_1-i} \\
& +(\varepsilon_{t-j_1-i}^2 - 1) g_k(j_1,j_1+i_1) g_l(j_2,j_2+i_1). \qquad (2.5.34)
\end{aligned}$$

Using

$$\begin{aligned}
& \Big| g_k(j_1,j) g_l(j_2,j_2+i_1) + g_k(j_1,j_1+i_1) g_l(j_2,j^*) \Big| \\
\leq \; & \psi_k(j_1)\psi_l(j_2) \Big[\psi_k(j)\psi_l(j_2+i_1) + \psi_k(j_1+i_1)\psi_l(j^*) \Big],
\end{aligned}$$

Rio (2009) (Theorem 2.1 therein), the Hoelder inequality, and $x \cdot (y-1) \leq 2(y-1) \leq y + (y-2) \leq y \Leftrightarrow xy \leq x+y$ for $x, y \leq 2$ we obtain for the last term:

$$\begin{aligned}
\|A_{j_1,j_2,i_1}(t-j_1-i_1)\|_p \;\; \leq \;\; & \psi_k(j_1)\psi_l(j_2) \Big[\psi_k(j_1+i_1) + \psi_l(j_2+i_1) \Big] \\
& \cdot C^4 \cdot \Big\{ \|\varepsilon_0\|_{2p}^2 \Big\{ (2p-1)^{1/2} 2^{3/2} \tilde{C} + |\mathbb{E}\varepsilon_0^3| + 1 \Big\} + 1 \Big\} \\
=: \;\; & \psi_k(j_1)\psi_l(j_2) \Big[\psi_k(j_1+i_1) + \psi_l(j_2+i_1) \Big] \cdot C_p^{(2)}.
\end{aligned}$$

For the third inequality we will only look at the first term, the other terms can be handled similar. First, we have from (2.5.34) with $t - j_3 + j_1 + i_1$ instead of t:

$$A_{j_1,j_2,i_1}(t-j_3) = \varepsilon_{t-j_3} \sum_{j \geq j_3+1} g_k(j_1, j-j_3+j_1+i_1) g_l(j_2,j_2+i_1) \varepsilon_{t-j} + \text{more terms},$$

thus for $i_2 \geq 1$, we have

$$\begin{aligned}
& \mathbb{E}[A_{j_1,j_2,i_1}(t-j_3) M_{j_3}(t-j_3) | \mathcal{F}_{t-j_3-i_2}] - \mathbb{E}[A_{j_1,j_2,i_1}(t-j_3) M_{j_3}(t-j_3)] \\
= \; & \sum_{j,j' \geq j_3+i_2} g_k(j_1, j-j_3+j_1+i_1) g_l(j_2,j_2+i_1) g_m(j_3,j') \varepsilon_{t-j}\varepsilon_{t-j'} - \mathbb{E}[\text{this term}] \\
& \qquad + \text{more terms},
\end{aligned}$$

and so

$$\begin{aligned}
& B_{j_1,j_2,j_3,i_1,i_2}(t-j_3-i_2) \\
= \; & \varepsilon_{t-j_3-i_2} \sum_{j \geq j_3+i_2} \Big[g_k(j_1, j-j_3+j_1+i_1) g_l(j_2,j_2+i_1) g_m(j_3,j_3+i_2) \\
& \quad + g_k(j_1,j_1+i_1+i_2) g_l(j_2,j_2+i_1) g_m(j_3,j) \Big] \varepsilon_{t-j} + \text{more terms},
\end{aligned}$$

thus applying $\|\cdot\|_p$ and using

$$\begin{aligned}
& |g_k(j_1, j - j_3 + j_1 + i_1) g_l(j_2, j_2 + i_1) g_m(j_3, j_3 + i_2)| \\
\leq\ & \psi_k(j_1)\psi_l(j_2)\psi_m(j_3)\psi_k(j - j_3 + j_1 + i_1)\psi_l(j_2 + i_1)\psi_m(j_3 + i_2), \\
& |g_k(j_1, j_1 + i_1 + i_2) g_l(j_2, j_2 + i_1) g_m(j_3 + i_2)| \\
\leq\ & \psi_k(j_1)\psi_l(j_2)\psi_m(j_3)\psi_k(j_1 + i_1 + i_2)\psi_l(j_2 + i_1)\psi_m(j),
\end{aligned}$$

we get the bound as asserted.

□

Lemma 2.5.12. *Let the definitions and assumptions of Lemma 2.5.11 hold. Let* $b_{t,k}, b_{s,t,k,l}, b_{s,t,\tau,k,l,m}$ *be deterministic constants (maybe dependent on* n*). Then we have with constants* $\tilde{C}_p^{(i)}$ *only dependent on* p *and* $C_p^{(i)}$, $i = 1, 2, 3$, *(see Lemma 2.5.11):*

$$\begin{aligned}
\left\| \sum_{t=1}^{n} b_{t,k} \cdot \eta_{t,k} \right\|_p &\leq \tilde{C}_p^{(1)} \cdot \left(\sum_{t=1}^{n} |b_{t,k}|^2 \right)^{1/2}, \\
\left\| \sum_{s,t=1}^{n} b_{s,t,k,l} \cdot \Big(\eta_{s,k}\eta_{t,l} - \mathbb{E}[\eta_{s,k}\eta_{t,l}] \Big) \right\|_p &\leq \tilde{C}_p^{(2)} \cdot \left(\sum_{s,t=1}^{n} |b_{s,t,k,l}|^2 \right)^{1/2}, \\
\left\| \sum_{s,t,\tau=1}^{n} b_{s,t,\tau,k,l,m} \cdot \Big(\eta_{s,k}\eta_{t,l}\eta_{\tau,m} - \mathbb{E}[\eta_{s,k}\eta_{t,l}\eta_{\tau,m}] \Big) \right\|_p &\leq \tilde{C}_p^{(3)} \cdot \left(\sum_{s,t,\tau=1}^{n} |b_{s,t,\tau,k,l,m}|^2 \right)^{1/2}.
\end{aligned}$$

Proof of Lemma 2.5.12: We start by showing the first inequality. From Lemma 2.5.11, we have $\|M_{j_1,k}(t - j_1)\|_p \leq C_p^{(1)} \psi_k(j_1)$ and therefore

$$\eta_{t,k} = \sum_{j_1=0}^{\infty} M_{j_1,k}(t - j_1) \quad a.s. \tag{2.5.35}$$

Note that $(M_{j_1,k}(t - j_1))_t$ are martingale differences w.r.t. $(\mathcal{F}_{t-j_1})_t$, so we can use Rio (2009), Theorem 2.1 therein, to get

$$\begin{aligned}
\left\| \sum_{t=1}^{n} b_{t,k}\eta_{t,k} \right\|_p &\leq \sum_{j_1=0}^{\infty} \left\| \sum_{t=1}^{n} b_{t,k} M_{j_1,k}(t - j_1) \right\|_p \\
&\leq (p-1)^{1/2} \sum_{j_1=0}^{\infty} \left(\sum_{t=1}^{n} |b_{t,k}|^2 \, \|M_{j_1,k}(t - j_1)\|_p^2 \right)^{1/2} \\
&\leq (p-1)^{1/2} C_p^{(1)} \sum_{j_1=0}^{\infty} \psi_k(j_1) \left(\sum_{t=1}^{n} |b_{t,k}|^2 \right)^{1/2} \leq \tilde{C}_p^{(1)} \left(\sum_{t=1}^{n} |b_{t,k}|^2 \right)^{1/2},
\end{aligned}$$

where $\tilde{C}_p^{(1)} := 2(p-1)^{1/2} C_p^{(1)} \tilde{C}_2$, and $\tilde{C}_2 := \sum_{j \in \mathbb{Z}} \frac{1}{\chi(j)}$.

To show the second inequality, we again use the representation (2.5.35) to get the upper bound

$$
\begin{aligned}
& \Big\| \sum_{s,t=1}^{n} b_{s,t,k,l} \Big(\eta_{s,k}\eta_{t,l} - \mathbb{E}[\eta_{s,k}\eta_{t,l}] \Big) \Big\|_p \\
= \; & \Big\| \sum_{j_1,j_2=0}^{\infty} \sum_{s,t=1}^{n} b_{s,t,k,l} \Big(M_{j_1,k}(s-j_1) M_{j_2,l}(t-j_2) - \mathbb{E}[M_{j_1,k}(s-j_1) M_{j_2,l}(t-j_2)] \Big) \Big\|_p \\
\leq \; & I + II + III_1 + III_2,
\end{aligned}
$$

where

$$
\begin{aligned}
I &= \sum_{j_1,j_2=0}^{\infty} \Big\| \sum_{s=1}^{n} \sum_{t<s-j_1+j_2} b_{s,t,k,l} M_{j_1,k}(s-j_1) M_{j_2,l}^{(2)}(t-j_2) \Big\|_p, \\
III_1 &= \Big\| \sum_{j_1,j_2=0, j_1 \geq j_2}^{\infty} \sum_{s=1, s \geq j_1-j_2}^{n} b_{s,s-j_1+j_2,k,l} \\
& \qquad \Big(M_{j_1,k}(s-j_1) M_{j_2,l}(s-j_1) - \mathbb{E}[M_{j_1,k}(s-j_1) M_{j_2,l}(s-j_1)] \Big) \Big\|_p,
\end{aligned}
$$

and II has the same form as I with reversed roles for s,t. III_1, III_2 are obtained by splitting the case $s-j_1 = t-j_2$ in the two subcases $t \leq s$ and $t > s$, thus III_2 has a similar form like III_1. We first discuss I. Note that

$$
A_s^{(j_1,j_2)} := M_{j_1,k}(s-j_1) \cdot \sum_{t<s-j_1+j_2} b_{s,t,k,l} M_{j_2,l}(t-j_2)
$$

is a martingale difference sequence w.r.t. $(\mathcal{F}_{s-j_1})_s$. Using again Rio (2009), Theorem 2.1, the Cauchy Schwarz inequality, and the result from Lemma 2.5.11, we get

$$
\begin{aligned}
& \Big\| \sum_{s=1}^{n} A_s^{(j_1,j_2)} \Big\|_p \\
\leq \; & (p-1)^{1/2} \Big(\sum_{s=1}^{n} \|A_s^{(j_1,j_2)}\|_p^2 \Big)^{\frac{1}{2}} \\
\leq \; & (p-1)^{1/2} \Big(\sum_{s=1}^{n} \|M_{j_1,k}(s-j_1)\|_{2p}^2 \cdot \Big\| \sum_{t<s-j_1+j_2} b_{s,t,k,l} M_{j_2,l}(t-j_2) \Big\|_{2p}^2 \Big)^{1/2} \\
\leq \; & (p-1)^{1/2} (2p-1)^{1/2} \Big(\sum_{s=1}^{n} \|M_{j_1,k}(s-j_1)\|_{2p}^2 \cdot \sum_{t<s-j_1+j_2} |b_{s,t,k,l}|^2 \cdot \|M_{j_2,l}(t-j_2)\|_{2p}^2 \Big)^{1/2} \\
\leq \; & (p-1)^{1/2} (2p-1)^{1/2} (C_{2p}^{(1)})^2 \Big(\sum_{s,t=1}^{n} |b_{s,t,k,l}|^2 \Big) \cdot \psi_k(j_1) \psi_l(j_2).
\end{aligned}
$$

so that

$$I = \sum_{j_1,j_2=0}^{\infty} \Big\| \sum_{s=1}^{n} A_s^{(j_1,j_2)} \Big\|_p \leq D_p \cdot \Big(\sum_{s,t=1}^{n} |b_{s,t,k,l}|^2 \Big),$$

where $D_p := 4\tilde{C}_2^2(p-1)^{1/2}(2p-1)^{1/2}(C_{2p}^{(1)})^2$. The argumentation for II is the same. For III_1, we use an upper bound for $\|A_{j_1,j_2,i}(s-j_1-i)\|_p$ from Lemma 2.5.11 to obtain

$$\Big(M_{j_1,k}(s-j_1)M_{j_2,l}(s-j_1) - \mathbb{E}[M_{j_1,k}(s-j_1)M_{j_2,l}(s-j_1)] \Big) = \sum_{i=0}^{\infty} A_{j_1,j_2,i}(s-j_1-i) \quad a.s.$$

where $(A_{j_1,j_2,i}(s-j_1-i))_t$ are martingale differences w.r.t. $(\mathcal{F}_{s-j_1-i})_s$. So we get with the same methods as before

$$\begin{aligned}
III_1 &= \Big\| \sum_{j_1,j_2=0,j_1-j_2\geq 0}^{\infty} \sum_{s=1,s\geq j_1-j_2}^{b} b_{s,s-j_1+j_2,k,l} \sum_{i=0}^{\infty} A_{j_1,j_2,i}(s-j_1-i) \Big\|_p \\
&\leq \sum_{j_1,j_2=0,j_1-j_2\geq 0}^{\infty} \sum_{i=0}^{\infty} \Big\| \sum_{s=1,s\geq j_1-j_2}^{n} b_{s,s-j_1+j_2,k,l} A_{j_1,j_2,i}(s-j_1-i) \Big\|_p \\
&\leq (p-1)^{1/2} \sum_{j_1,j_2=0,j_1-j_2\geq 0}^{\infty} \sum_{i=0}^{\infty} \Big(\sum_{s=1,s\geq j_1-j_2}^{n} |b_{s,s-j_1+j_2,k,l}|^2 \|A_{j_1,j_2,i}(s-j_1-i)\|_p^2 \Big)^{1/2} \\
&\leq (p-1)^{1/2} C_p^{(2)} \sum_{i,j_1,j_2=0,j_1-j_2\geq 0}^{\infty} \psi_k(j_1)\psi_l(j_2) \cdot \\
&\qquad \Big(\psi_k(j_1+i) + \psi_l(j_2+i) \Big) \cdot \Big(\sum_{s=1,s\geq j_1-j_2}^{n} |b_{s,s-j_1+j_2,k,l}|^2 \Big)^{1/2} \\
&\leq E_p \sup_{d=0,\dots,n-1} \Big(\sum_{s=d+1}^{n} |b_{s,s-d,k,l}|^2 \Big)^{1/2} \leq E_p \Big(\sum_{d=0}^{n-1} \sum_{s=d+1}^{n} |b_{s,s-d,k,l}|^2 \Big)^{1/2} \\
&\leq E_p \Big(\sum_{s,t=1}^{n} |b_{s,t,k,l}|^2 \Big)^{1/2},
\end{aligned}$$

where $E_p := 16(p-1)^{1/2}C_p^{(2)}\tilde{C}_2^3$. The same argumentation can be done for III_2. In total, we proved the second inequality with $\tilde{C}_p^{(2)} := D_p + E_p$. The proof of the third inequality follows the same lines, but further case distinctions are needed. □

2.5.6 Proofs of the statements in Section 2.2

Here, we give the proofs of the Lemmas and Corollaries from Section 2.2. We start with the proof of Lemma 2.2.9 which is maybe the most important part. For this proof, we need a more detailed analysis of the behavior of $\ell_{t,n}(\theta)$ and its derivatives if the true parameter $\theta = \theta_0(t/n)$ is plugged in. This results are formulated in the following lemma.

Lemma 2.5.13 (Detailed analysis of $\ell_{t,n}(\theta)$)**.** *Suppose that Assumption 2.2.1 holds. Then there exists a decomposition*

$$\nabla \ell_{t,n}(\theta_0(t/n)) = MDS_t^{(1)} + R_t^{(1)} \tag{2.5.36}$$

with a martingale difference sequence $MDS_t^{(1)} := \varepsilon_t \nabla d_{t,n}(\theta_0(t/n)) - \nabla\gamma_{\theta_0(t/n)}(0) a_{t,n}(0)$ *with respect to* $\mathcal{F}_t := \sigma(\varepsilon_s : s \le t)$ *(the* σ*-algebra generated from* ε_s, $s \le t$*), and some random variable* $R_t^{(1)}$*. Furthermore, there exists a constant* $\tilde{C} > 0$ *independent of* n *such that*

$$\sum_{t=1}^{n} |\mathbb{E}\nabla \ell_{t,n}(\theta_0(t/n))|_1 \le \tilde{C}, \qquad \sum_{t=1}^{n} |\nabla^2 \ell_{t,n}(\theta_0(t/n)) - I(\theta_0(t/n))|_1 \le \tilde{C}. \tag{2.5.37}$$

Proof of Lemma 2.5.13: In this proof, C is a generic constant independent of n which may change its value from line to line. Note that

$$\begin{aligned}
\nabla \ell_{t,n}(\theta) &= d_{t,n}(\theta) \cdot \nabla d_{t,n}(\theta) - \frac{\nabla\gamma_\theta(0)}{\gamma_\theta(0)}, \\
\nabla^2 \ell_{t,n}(\theta) &= \big(\nabla d_{t,n}(\theta) \cdot \nabla d_{t,n}(\theta)' + \frac{\nabla\gamma_\theta(0) \cdot \nabla\gamma_\theta(0)'}{\gamma_\theta(0)^2}\big) + \big(d_{t,n}(\theta) \cdot \nabla^2 d_{t,n}(\theta) - \frac{\nabla^2\gamma_\theta(0)}{\gamma_\theta(0)}\big).
\end{aligned}$$

Note that by Proposition 2.5.2, we have $\varepsilon_t = \sum_{k=0}^{\infty} \gamma_{\theta_0(u)}(k)\tilde{X}_{t-k}(\theta_0(u))$. Using this representation for $u = t/n$ gives

$$d_{t,n}(\theta_0(t/n)) = \varepsilon_t + R_t,$$

where $R_t := -\sum_{k=t}^{\infty} \gamma_{\theta_0(t/n)}(k) X_{t-k,n} + \sum_{k=0}^{\infty} \gamma_{\theta_0(t/n)}(k) \cdot \Big(X_{t-k,n} - \tilde{X}_{t-k}\big(\theta_0(\tfrac{t}{n})\big)\Big)$. Furthermore, since $\gamma_\theta(0) = a_\theta(0)^{-1}$, we have that

$$\nabla d_{t,n}(\theta) = \frac{\nabla\gamma_\theta(0)}{\gamma_\theta(0)}\varepsilon_t + \nabla\gamma_\theta(0)\big(a_{t,n}(0) - a_\theta(0)\big)\varepsilon_t + R_t^{(2)},$$

where $R_t^{(2)} = \nabla d_{t,n}(\theta) - \frac{\nabla\gamma_\theta(0)}{\gamma_\theta(0)}\varepsilon_t - \nabla\gamma_\theta(0)(a_{t,n}(0) - a_\theta(0)) \in \mathcal{F}_{t-1}$. This shows that there is a decomposition of the form (2.5.36) into a martingale difference sequence $\mathrm{MDS}_t^{(1)} := \big(\varepsilon_t^2 - 1\big)\frac{\nabla\gamma_{\theta_0(t/n)}(0)}{\gamma_{\theta_0(t/n)}(0)} + \varepsilon_t R_t^{(2)} = \varepsilon_t \nabla d_{t,n}(\theta_0(t/n)) - \nabla\gamma_{\theta_0(t/n)}(0) a_{t,n}(0)$ and $R_t^{(1)} := \nabla\gamma_{\theta_0(t/n)}(0)(a_{t,n}(0) - a(t/n,0))\varepsilon_t^2 + R_t \cdot \nabla d_{t,n}(\theta_0(t/n))$. Since $\mathrm{Cov}(X_{t-k,n}, X_{t-l,n}) \le \frac{C}{\chi(k-l)}$,

$$|\mathrm{Cov}(X_{t-k,n} - \tilde{X}_{t-k}(\theta_0(t/n)), X_{t-l,n})| \le C \sum_{k_1=0}^{\infty} \big|a_{t-k,n}(k_1) - a(t/n,k_1)\big| \frac{1}{\chi(k+k_1-l)}$$

and $\sum_{t=1}^{n} |a_{t-k,n}(k_1) - a\big(\frac{t-k}{n}, k_1\big)| + \sum_{t=1}^{n} \big|a\big(\frac{t-k}{n}, k_1\big) - a(t/n,k_1)\big| \le C(1 + \frac{|k|}{\chi(k_1)})$ by the results of Lemma 2.5.1, we conclude that component-wise:

$$\begin{aligned}
&\sum_{t=1}^{n} |\mathbb{E}\nabla \ell_{t,n}(\theta_0(t/n))| \\
\leq\ & C \sum_{t=1}^{n} \sum_{k=t}^{\infty} \sum_{l=0}^{\infty} \frac{1}{\chi(k)\chi(l)} \mathrm{Cov}(X_{t-k,n}, X_{t-l,n}) \\
&\quad + C \sum_{t=1}^{n} \sum_{k,l=0}^{\infty} \frac{1}{\chi(k)\chi(l)} \mathrm{Cov}(X_{t-k,n} - \tilde{X}_{t-k}\big(\theta_0(t/n)\big), X_{t-l,n}) \\
&\quad + C \cdot \sum_{t=1}^{n} |a_{t,n}(0) - a(t/n, 0)| \\
\leq\ & C \cdot \sum_{t=1}^{n} \frac{1}{\chi(t)} \sum_{k,l=0}^{\infty} \frac{1}{\chi(l)\chi(k-l)} + C \sum_{k_1,k,l=0}^{\infty} \frac{|k|}{\chi(k)\chi(l)\chi(k+k_1-l)\chi(k_1)} + C \leq C.
\end{aligned}$$

With exactly the same arguments we can show that

$$\sum_{t=1}^{n} \big| d_{t,n}(\theta_0(t/n)) \cdot \nabla^2 d_{t,n}(\theta_0(t/n)) - \frac{\nabla^2 \gamma_{\theta_0(t/n)}(0)}{\gamma_{\theta_0(t/n)}(0)} \big| \leq C.$$

In addition, similar calculations as above show that

$$\sum_{t=1}^{n} |\nabla d_{t,n}(\theta_0(t/n)) \cdot \nabla d_{t,n}(\theta_0(t/n))' - \sum_{l,k=0}^{\infty} \nabla \gamma_{\theta_0(t/n)}(k) \nabla \gamma_{\theta_0(t/n)}(l)' c(t/n, k-l)| \leq C.$$

Since

$$\begin{aligned}
&\sum_{l,k=0}^{\infty} \nabla \gamma_{\theta_0(t/n)}(k) \nabla \gamma_{\theta_0(t/n)}(l)' c(t/n, k-l) \\
=\ & \int_{-\pi}^{\pi} \nabla (A_{\theta_0(t/n)}(-\lambda)^{-1}) \cdot \nabla (A_{\theta_0(t/n)}(-\lambda)^{-1}) \cdot f(t/n, \lambda) \ \mathrm{d}\lambda,
\end{aligned}$$

we obtain the second result in (2.5.37). □

Proof of Lemma 2.2.9. In this proof we will use C as a generic constant which may change its value from line to line but does not depend on b, n, u. To keep the notation simple, let us use the abbreviations $\hat{\theta} := \hat{\theta}_{b,-t}(u)$, $\theta_0 := \theta_0(u)$ and $L(\cdot) := L_{n,b,-t}(u, \cdot)$. The general idea of the proof is to use Taylor expansions in the nominator of (2.2.8) to separate the expression into terms where we can use Proposition 2.5.8, Lemma 2.5.9, Theorem 2.5.10 and the continuity of $I(\cdot)$ to show that these terms are of lower order than $d_M^*(\hat{\theta}_b, \theta_0)$ and into terms where we have to calculate expectations and use continuity arguments to show the same. First use a third-order Taylor expansion to write

$$\ell_{t,n}(\hat{\theta}) - \ell_{t,n}(\theta_0) = \nabla \ell_{t,n}(\theta_0) \cdot (\hat{\theta} - \theta_0) + |\hat{\theta} - \theta_0|^2_{\nabla^2 \ell_{t,n}(\theta_0)} + |\hat{\theta} - \theta_0|^3_{\nabla^3 \ell_{t,n}(\bar{\theta}_1)}, \quad (2.5.38)$$

where $\bar{\theta}_1(u) \in \Theta$ is some intermediate value with $|\bar{\theta}_1 - \theta_0|_2 \leq |\hat{\theta} - \theta_0|_2$. Using (2.5.38), we obtain the decomposition

$$\begin{aligned} & CV(b) - \frac{1}{n}\sum_{t=1}^{n} \ell_{t,n}\Big(\theta_0\Big(\frac{t}{n}\Big)\Big)w(t/n) - \overline{d_A}(\hat{\theta}_b, \theta_0) \\ = \ & \frac{1}{n}\sum_{t=1}^{n} \nabla \ell_{t,n}(\theta_0(t/n)) \cdot (\hat{\theta}_{b,-t}(t/n) - \theta_0(t/n))w(t/n) \\ & + \frac{1}{n}\sum_{t=1}^{n} |\hat{\theta}_{b,-t}(t/n) - \theta_0(t/n)|^2_{\nabla^2 \ell_{t,n}(\theta_0(t/n) - I(\theta_0(t/n))} w(t/n) \\ & + \frac{1}{n}\sum_{t=1}^{n} |\hat{\theta}_{b,-t}(t/n) - \theta_0(t/n)|^3_{\nabla^3 \ell_{t,n}(\bar{\theta}_1(t/n))} w(t/n). \qquad (2.5.39) \end{aligned}$$

In view of Corollary 2.5.6, it is enough to show that each term of (2.5.39) is almost surely of order $o((nb)^{-1} + B^2(b))$ or $o((nb)^{-1/2}B(b))$, respectively. We will discuss the three terms in (2.5.39) separately.

Third term in (2.5.39): Note that $\theta_0(u)$ is in the interior of Θ for all $u \in \mathrm{supp}(w)$. Because of Theorem 2.5.10 it follows that $\sup_{b \in B_n} \sup_{u \in \mathrm{supp}(w)} |\nabla L_{n,b}(u, \hat{\theta}_b(u))|_2 = 0$ for n large enough. Using a second-order Taylor argument, we obtain

$$\hat{\theta} - \theta_0 = -\big(\nabla^2 L(\bar{\theta}_2)\big)^{-1} \cdot \nabla L(\theta_0). \qquad (2.5.40)$$

with some intermediate value $\bar{\theta}_2 \in \Theta$ which fulfills $|\bar{\theta}_2 - \theta_0| \leq |\hat{\theta} - \theta_0|$. We know from Proposition 2.5.8, Lemma 2.5.9, Theorem 2.5.10 and the continuity of $I(\cdot)$ that $\nabla^2 L_{n,b,-t}(u, \bar{\theta}_2(u)) \to I(\theta_0(u))$ uniformly in $b \in B_n$, $u \in \mathrm{supp}(w)$. Together with (2.5.40), for n large enough, the third term in (2.5.39) is bounded by

$$\begin{aligned} & \frac{C}{n}\sum_{t=1}^{n} |\nabla^3 \ell_{t,n}(\bar{\theta}_1(t/n))|_1 \cdot |\nabla L_{n,b,-t}(t/n, \theta_0(t/n))|^3 w(t/n) \\ \leq \ & C \cdot \sup_{u \in \mathrm{supp}(w)} \sup_{t=1,\dots,n} |\nabla L_{n,b,-t}(u, \theta_0(u)) - B(u,b)|^3 \cdot \frac{1}{n}\sum_{t=1}^{n} |\nabla^3 \ell_{t,n}(\bar{\theta}_1(t/n))|_1 \\ & + \frac{C}{n}\sum_{t=1}^{n} |\nabla^3 \ell_{t,n}(\bar{\theta}_1(t/n))|_1 \cdot |B(t/n,b)|^3 w(t/n). \qquad (2.5.41) \end{aligned}$$

It is easily seen that

$$\frac{1}{n}\sum_{t=1}^{n} |\nabla^3 \ell_{t,n}(\bar{\theta}_1(t/n))|_1 \leq C\left(1 + \frac{1}{n}\sum_{t=1}^{n} X_{t,n}^2\right)$$

is bounded a.s. (see the results of Proposition 2.5.7 and use a Borel-Cantelli argument), thus the first term in (2.5.41) is of order $O((nh)^{3/2-\alpha})$ with arbitrary $\alpha > 0$ (see Proposition 2.5.8 and Lemma 2.5.9). Define $Z_t := \sum_{k=0}^{t-1} \frac{X_{t-k}^2}{\chi(k)}$. Z_t fulfils

$$|\nabla^3 \ell_{t,n}(\bar{\theta}_1(t/n))|_1 \leq C(1 + Z_t),$$

and we can estimate the second term in (2.5.41) by

$$\frac{C}{n}\sum_{t=1}^{n}(Z_t - \mathbb{E}Z_t)\cdot|B(t/n,b)|^3 w(t/n) + \frac{C}{n}\sum_{t=1}^{n}|B(t/n,b)|^3 w(t/n) \tag{2.5.42}$$

The second term in (2.5.42) is of order $o(B^2(b))$. Define the abbreviation $g_b(l/n) := \sum_{t=l}^{n}\frac{1}{\chi(t-l)}|B(t/n,b)|^3 w(t/n)$, then the first term in (2.5.42) has the representation

$$\frac{1}{n}\sum_{t=1}^{n}(Z_t - \mathbb{E}Z_t)\cdot|B(t/n,b)|^3 w(t/n) = L_n(g_b,1,1) - \mathbb{E}[L_n(g_b,1,1)]$$

and it is easy to see that this is of order $O(n^{-1/2}B(b))$, use the notation and the same method for the proof as done for (2.5.48) below.

Second term in (2.5.39): The second term can be written as

$$\frac{1}{n}\sum_{t=1}^{n}|I(\theta_0(t/n))^{-1}\nabla L_{n,b,-t}(t/n,\theta_0(t/n))|^2_{\nabla^2\ell_{t,n}(\theta_0(t/n))-I(\theta_0(t/n))} + R_n \tag{2.5.43}$$

where

$$\begin{aligned}|R_n| \quad &\leq \quad \frac{C}{n}\sum_{t=1}^{n}|\nabla^2\ell_{t,n}(\theta_0(t/n)) - I(\theta_0(t/n))|_1\cdot|\nabla L_{n,b,-t}(t/n,\theta_0(t/n))|^2 \\ &\qquad\cdot\left|[\nabla^2 L_{n,b,-t}(t/n,\bar\theta_1(t/n))]^{-1} - I(\theta_0(t/n))^{-1}\right|_2.\end{aligned}$$

Again, with Proposition 2.5.8, Lemma 2.5.9, Theorem 2.5.10 and the continuity of $I(\cdot)$ we can show that this term is of order $o((nb)^{-1} + B^2(b))$ with the same methods used for the third term of (2.5.39).

First term in (2.5.39): Using a fourth-order taylor argument, we have

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\nabla\ell_{t,n}(\theta_0(t/n))\cdot(\hat\theta-\theta_0) \\ =\quad &-\frac{1}{n}\sum_{t=1}^{n}\nabla\ell_{t,n}(\theta_0(t/n))\cdot\nabla^2 L(\theta_0)^{-1}\cdot\left[\nabla L(\theta_0) + |\hat\theta-\theta_0|^2_{\nabla^3 L(\theta_0)} + |\hat\theta-\theta_0|^3_{\nabla^4 L(\bar\theta_3)}\right]\end{aligned} \tag{2.5.44}$$

with some intermediate value $\bar\theta_3(u)\in\Theta$ with $|\bar\theta_3-\theta_0|\leq|\hat\theta-\theta_0|$. The last term in (2.5.44) can be bounded via

$$\begin{aligned}&\frac{1}{n}\left|\sum_{t=1}^{n}\nabla\ell_{t,n}(\theta_0(t/n))\cdot(\nabla^2 L_{n,b,-t}(t/n,\theta_0(t/n)))^{-1}\cdot|\hat\theta_{b,-t}(t/n)-\theta_0(t/n)|^3_{\nabla^4 L(\bar\theta_3(t/n))}\right| \\ \leq\quad &\frac{1}{n}\sum_{t=1}^{n}|\nabla\ell_{t,n}(\theta_0(t/n))|_1\cdot|\hat\theta_{b,-t}(t/n)-\theta_0(t/n)|^3\cdot|\nabla^4 L_{n,b,-t}(t/n,\bar\theta_3(t/n))|_2 \\ &\qquad\cdot|(\nabla^2 L_{n,b,-t}(t/n,\theta_0(t/n)))^{-1}|_2.\end{aligned}$$

Again, this can be handled like the third term in (2.5.39).
The second term in (2.5.44) can be handled as follows: First replace $\nabla^3 L(\theta_0)$ by $\mathbb{E}[\nabla^3 L(\theta_0)]$, then $\hat{\theta} - \theta_0$ by $I(\theta_0)^{-1}\nabla L(\theta_0)$ and after that $\nabla^2 L(\theta_0)^{-1}$ by $I(\theta_0)^{-1}$. The replacement errors again can be handled like the third term in (2.5.39), only

$$\frac{1}{n}\sum_{t=1}^{n} \nabla \ell_{t,n}(\theta_0(t/n)) \cdot I(\theta_0(t/n))^{-1} \cdot |I(\theta_0(t/n))^{-1} \nabla L_{n,b,-t}(t/n, \theta_0(t/n))|^2_{\mathbb{E}[\nabla^3 L_{n,b,-t}(t/n,\theta_0(t/n))]} \tag{2.5.45}$$

ist left. For the first term in (2.5.44) use the expansion

$$[\nabla^2 L(\theta_0)]^{-1} = [\nabla^2 L(\theta_0)]^{-1} \cdot \Big[\mathbb{E}[\nabla^2 L(\theta_0)] - \nabla^2 L(\theta_0)\Big] \cdot \mathbb{E}[\nabla^2 L(\theta_0)]^{-1} + \mathbb{E}[\nabla^2 L(\theta_0)]^{-1}$$

and we get the two terms

$$\frac{1}{n}\sum_{t=1}^{n} \nabla \ell_{t,n}(\theta_0(t/n)) \cdot \mathbb{E}[\nabla^2 L_{n,b,-t}(t/n, \theta_0(t/n))]^{-1} \cdot \nabla L_{n,b,-t}(t/n, \theta_0(t/n)) \tag{2.5.46}$$

and

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} \nabla \ell_{t,n}(\theta_0(t/n)) \cdot \Big[\mathbb{E}[\nabla^2 L_{n,b,-t}(t/n, \theta_0(t/n))] - \nabla^2 L_{n,b,-t}(t/n, \theta_0(t/n))\Big] \\ \cdot \mathbb{E}[\nabla^2 L_{n,b,-t}(t/n, \theta_0(t/n))]^{-1} \cdot \nabla L_{n,b,-t}(t/n, \theta_0(t/n)),\end{aligned} \tag{2.5.47}$$

where we replaced $[\nabla^2 L(\theta_0)]^{-1}$ by $\mathbb{E}[\nabla^2 L(\theta_0)]^{-1}$ in the last term (replacement error handled as before).

In the last step we have to show that (2.5.43), (2.5.45), (2.5.46) and (2.5.47) fulfil $\sup_{b\in B_n} \left|\frac{\text{term}}{d_M^*(\hat{\theta}_b, \theta_0)}\right| \to 0$. We will do this in a little bit more abstract way. All terms we have to discuss are finite sums (i.e., not more than $C = C(d)$ terms, where d is the dimension of the parameter space) of the form

$$\frac{1}{n}\sum_{t=1}^{n} w(t/n) \cdot f_t \cdot x_t^{(1)} \cdot x_t^{(2)} \cdot x_t^{(3)}, \tag{2.5.48}$$

where f_t is deterministic and bounded uniformly in t, n, b, and $x_t^{(i)}$ are random variables. More precisely, we have with $j_1, j_2, j_3, j_4 \in \{1, ..., d\}$:

$$\begin{aligned}
x_t^{(1)} &= \partial_{j_1}\ell_{t,n}(\theta_0(t/n)), \quad x_t^{(2)} = \partial_{j_2} L_{n,b,-t}(t/n, \theta_0(t/n)), \quad x_t^{(3)} = 1, && (2.5.49)\\
x_t^{(1)} &= 1, \quad x_t^{(2)} = \partial_{j_1} L_{n,b}(t/n, \theta_0(t/n)), \quad x_t^{(3)} = \partial_{j_2} L_{n,b}(t/n, \theta_0(t/n)), && (2.5.50)\\
x_t^{(1)} &= (\partial_{j_1}\partial_{j_2}\ell_{t,n}(\theta_0(t/n)) - I(\theta_0(t/n))_{j_1,j_2}), \quad x_t^{(2)} = \partial_{j_3} L_{n,b,-t}(t/n, \theta_0(t/n)), \\
&\qquad x_t^{(3)} = \partial_{j_4} L_{n,b,-t}(t/n, \theta_0(t/n)), && (2.5.51)\\
x_t^{(1)} &= \partial_{j_1}\ell_{t,n}(\theta_0(t/n)), \quad x_t^{(2)} = \partial_{j_2} L_{n,b,-t}(t/n, \theta_0(t/n)), \\
&\qquad x_t^{(3)} = \partial_{j_3} L_{n,b,-t}(t/n, \theta_0(t/n)), && (2.5.52)\\
x_t^{(1)} &= \partial_{j_1}\ell_{t,n}(\theta_0(t/n)), \quad x_t^{(2)} = \partial_{j_2} L_{n,b,-t}(t/n, \theta_0(t/n)), \\
&\qquad x_t^{(3)} = \partial_{j_3}\partial_{j_4} L_{n,b,-t}(t/n, \theta_0(t/n)) - \mathbb{E}[\partial_{j_3}\partial_{j_4} L_{n,b,-t}(t/n, \theta_0(t/n))]. && (2.5.53)
\end{aligned}$$

Here, (2.5.50) does not appear in the calculations above but in the proof of Lemma 2.2.6. Since the structure is the same as the terms which appear above, we discuss it here. Note that for this case, the deterministic term in (2.5.54) must not be discussed since the exspecations were already subtracted before.

Our goal is to show that the sum (2.5.48) has order $o(d_M^*(\hat{\theta}_b, \theta_0)) = o((nb)^{-1} + B^2(b))$ uniformly in $b \in B_n$. This is done with the same technique used in the proof of Proposition 2.5.8. First, we will show that each term of the form 2.5.48 fulfils $\|\text{term}\|_p \leq Cn^{-\tau p}$ with some $\tau > 0$ independent of p. In the second step we will show that the random terms have a Hoelder continuity property with respect to $b \in B_n$. We use the decomposition (where $j, k \in \{1,2,3\} \backslash \{i\}$ are the two indices which are not i):

$$\begin{aligned} x_t^{(1)} x_t^{(2)} x_t^{(3)} &= (x_t^{(1)} - \mathbb{E}[x_t^{(1)}])(x_t^{(2)} - \mathbb{E}[x_t^{(2)}])(x_t^{(3)} - \mathbb{E}[x_t^{(3)}]) \\ &\quad + \sum_{i=1}^{3} \mathbb{E}[x_t^{(i)}](x_t^{(j)} - \mathbb{E}[x_t^{(j)}])(x_t^{(k)} - \mathbb{E}[x_t^{(k)}]) \\ &\quad + \sum_{i=1}^{3} \mathbb{E}[x_t^{(j)}]\mathbb{E}[x_t^{(k)}](x_t^{(i)} - \mathbb{E}[x_t^{(i)}]) \\ &\quad + \mathbb{E}[x_t^{(1)}]\mathbb{E}[x_t^{(2)}]\mathbb{E}[x_t^{(3)}]. \end{aligned} \tag{2.5.54}$$

Deterministic term in (2.5.54): It is easy to see that $\mathbb{E}[x_t^{(2)}], \mathbb{E}[x_t^{(3)}]$ are bounded uniformly in t, b, n; moreover we have that $\sum_{t=1}^{n} |\mathbb{E}[x_t^{(1)}]|$ is bounded in b, n since by Lemma 2.5.13 (note again that we do not have to consider $x_t^{(1)} = 1$). Thus,

$$\frac{1}{n} \sum_{t=1}^{n} w(t/n) f_t \mathbb{E}[x_t^{(1)}]\mathbb{E}[x_t^{(2)}]\mathbb{E}[x_t^{(3)}] = O(n^{-1}).$$

Terms with one random variable in (2.5.54): we observe that these terms have the form

$$F_n := \frac{1}{n} \sum_{k_1,k_2=0}^{\infty} \sum_{t=1}^{n} g_t(k_1,k_2) \cdot r\Big(\frac{t-k_1}{n}\Big) r\Big(\frac{t-k_2}{n}\Big) \cdot \eta_{t-\min(k_1,k_2),|k_1-k_2|},$$

where $g_t = g_t(k_1,k_2)$ is deterministic (defined later), $r(x) = \mathbb{1}_{(0,1]}(x)$ is a data taper and η_t is defined in Lemma 2.5.11. Therefore, by Lemma 2.5.12,

$$\|F_n\|_p \;\leq\; \frac{C}{n} \sum_{k_1,k_2=0}^{\infty} \Big(\sum_{t=1}^{n} g_t(k_1,k_2)^2 \Big)^{1/2}.$$

If $i = 1$, then g_t is of the form

$$g_t = w(t/n) f_t \mathbb{E}[x_t^{(j)}]\mathbb{E}[x_t^{(k)}] \hat{\phi}_1(t/n, k_1) \hat{\phi}_2(t/n, k_2) \tag{2.5.55}$$

where at least one of the expectations has the form $B_j(t/n, b) + O((nb)^{-1})$, and the other expectation is bounded. Therefore, in this case we have with Lemma 2.5.12

$$\|F_n\|_p \leq \frac{C}{n} \Big(\sum_{t=1}^{n} w(t/n) |B(t/n,b)|^2 \Big)^{1/2} + C\frac{b^{1/2}}{nb} \leq \frac{1}{n^{1/2}} B(b) + C\frac{b^{1/2}}{nb}. \tag{2.5.56}$$

If $i \neq 1$, then g_t is of the form

$$g_t = \sum_{s=1}^{n} w(s/n) f_s \mathbb{E}[x_s^{(1)}]\mathbb{E}[x_s^{(j)}] \frac{1}{nb} K\left(\frac{t-s}{nb}\right) \hat{\phi}_1(s/n, k_1)\hat{\phi}_2(s/n, k_2),$$

where $\mathbb{E}[x_s^{(j)}]$ is bounded. If we are not in the case (2.5.50), we conclude with Lemma 2.5.12, that

$$\|F_n\|_p \leq \frac{C}{n(nb)} \left(\sum_{t=1}^{n} \left(\sum_{s=1}^{n} |\mathbb{E}[x_s^{(1)}]| \right)^2 \right)^{1/2} \leq \frac{C}{\sqrt{n}} \frac{1}{nb}.$$

In the case (2.5.50), we use that $x_s^{(1)} = 1$ and $\mathbb{E}[x_s^{(j)}] = B_j(s/n, b)$, thus with the Cauchy Schwarz inequality we have

$$\begin{aligned}
\|F_n\|_p &\leq \frac{C}{n(nb)} \left(\sum_{t=1}^{n} \left(\sum_{s=1}^{n} w(s/n) B_j(s/n, b) K\left(\frac{t-s}{nb}\right) \right)^2 \right)^{1/2} \\
&\leq \frac{C}{n(nb)} \Big(\sum_{t=1}^{n} \sum_{s=1}^{n} w(s/n) |B(s/n, b)|^2 \Big| K\Big(\frac{t-s}{nb}\Big) \Big| \cdot \sum_{s=1}^{n} \Big| K\Big(\frac{t-s}{nb}\Big) \Big| \Big)^{1/2} \\
&\leq \frac{C}{n^{1/2}} B^2(b).
\end{aligned}$$

Terms with two random variables in (2.5.54): These terms have the form

$$\begin{aligned}
F_n &:= \frac{1}{n} \sum_{k_1,k_2,l_1,l_2=0}^{\infty} \sum_{s,t=1}^{n} g_{s,t}(k_1, k_2, l_1, l_2) \cdot r(\frac{t-k_1}{n}) r(\frac{t-k_2}{n}) r(\frac{s-l_1}{n}) r(\frac{s-l_2}{n}) \cdot \\
&\quad \Big(\eta_{t-\min(k_1,k_2),|k_1-k_2|} \eta_{s-\min(l_1,l_2),|l_1-l_2|} \Big),
\end{aligned}$$

therefore with Lemma 2.5.12,

$$\|F_n - \mathbb{E}F_n\|_p \leq \frac{C}{n} \sum_{k_1,k_2,l_1,l_2=0}^{\infty} \Big(\sum_{s,t=1}^{n} g_{s,t}(k_1, k_2, l_1, l_2)^2 \Big)^{1/2}.$$

If $i = 1$, we have

$$\begin{aligned}
g_{s,t} &= \frac{1}{(nb)^2} \sum_{u=1}^{n} \mathbb{E}[x_u^{(1)}] w(u/n) f_u \Big| K\Big(\frac{t-u}{nb}\Big) K\Big(\frac{s-u}{nb}\Big) \Big| \\
&\quad \cdot |\hat{\phi}_1(u/n, k_1)| \cdot |\hat{\phi}_2(u/n, k_2)| \cdot |\hat{\psi}_1(u/n, l_1)| \cdot |\hat{\psi}_2(u/n, l_2)|, \quad (2.5.57)
\end{aligned}$$

therefore with the Cauchy-Schwarz inequality

$$\begin{aligned}
\|F_n - \mathbb{E}F_n\|_p &\leq \frac{C}{n(nb)^2}\Big(\sum_{s,t=1}^{n}\Big(\sum_{u=1}^{n}\Big|K\Big(\frac{t-u}{nb}\Big)K\Big(\frac{s-u}{nb}\Big)\Big|\Big)^2\Big)^{1/2} \\
&\leq \frac{C}{n(nb)^2}\Big(\sum_{s,t=1,|s-t|\leq 2nb}^{n}\sum_{u=1}^{n}K^2\Big(\frac{t-u}{nb}\Big)\cdot\sum_{u=1}^{n}K^2\Big(\frac{s-u}{nb}\Big)\Big)^{1/2} \\
&\leq C\frac{b^{1/2}}{nb}. \qquad (2.5.58)
\end{aligned}$$

In the case that $i \neq 1$, we have

$$g_{s,t} = \frac{1}{nb}K\left(\frac{t-s}{nb}\right)w(t/n)\mathbb{E}[x_t^{(i)}]f_t|\hat{\phi}_1(t/n,k_1)|\cdot|\hat{\phi}_2(t/n,k_2)|\cdot|\hat{\psi}_1(t/n,l_1)|\cdot|\hat{\psi}_2(t/n,l_2)|,$$

thus

$$\|F_n - \mathbb{E}F_n\|_p \leq \frac{C}{n(nb)}\Big(\sum_{s,t=1}^{n}K\Big(\frac{t-s}{nb}\Big)^2\Big)^{1/2} \leq \frac{Cb^{1/2}}{nb}.$$

The discussion of the expectation

$$\mathbb{E}F_n = \frac{1}{n}\sum_{t=1}^{n}w(t/n)f_t\mathbb{E}[x_t^{(i)}]\cdot\mathbb{E}\Big[(x_t^{(j)} - \mathbb{E}x_t^{(j)})\cdot(x_t^{(k)} - \mathbb{E}x_t^{(k)})\Big].$$

is left. Using Assumption 2.1.1 and Proposition 5.4 in Dahlhaus and Polonik (2009), we have

$$\sum_{s=1}^{n}|\mathrm{Cov}(X_{t-k_1,n}X_{t-k_2,n}, X_{s-l_1,n}X_{s-l_2,n})| \leq C.$$

Choose $\phi_j(u,\lambda), \psi_j(u,\lambda)$ from $A_{\theta_0(u)}(\lambda), \partial_l A_{\theta_0(u)}(\lambda),\ \partial_k\partial_l A_{\theta_0(u)}(\lambda)$ and choose $g_{s_1,s_2} = g_{s_1,s_2}(t)$ bounded such that the first equality in the following derivation holds. Then we have

$$\begin{aligned}
&\Big|\mathbb{E}\Big[(x_t^{(j)} - \mathbb{E}x_t^{(j)})\cdot(x_t^{(k)} - \mathbb{E}x_t^{(k)})\Big]\Big| \\
= &\ \Big|\mathbb{E}\Big[\sum_{s_1,s_2=1}^{n} g_{s_1,s_2}\Big(\Big(\sum_{k_1=0}^{s_1-1}\hat{\phi}_1(s_1/n,k_1)X_{s_1-k_1,n}\Big)\Big(\sum_{k_2=0}^{s_1-1}\hat{\phi}_2(s_1/n,k_2)X_{s_1-k_2,n}\Big) - \mathbb{E}[...]\Big) \\
&\quad\cdot\Big(\Big(\sum_{l_1=0}^{s_2-1}\hat{\psi}_1(s_2/n,l_1)X_{s_2-l_1,n}\Big)\Big(\sum_{l_2=0}^{s_2-1}\hat{\psi}_2(s_2/n,l_2)X_{s_2-l_2,n}\Big) - \mathbb{E}[...]\Big)\Big]\Big| \\
\leq &\ C\sum_{k_1,k_2,l_1,l_2=0}^{\infty}|\tilde{\phi}_1(k_1)\tilde{\phi}_2(k_2)\tilde{\psi}_1(l_1)\tilde{\psi}_2(l_2)| \\
&\qquad\times\sum_{s_1,s_2=1}^{n}|g_{s_1,s_2}|\cdot|\mathrm{Cov}(X_{s_1-k_1,n}X_{s_1-k_2,n}, X_{s_2-l_1,n}X_{s_2-l_2,n})| \\
\leq &\ C\sum_{s_2=1}^{n}\sup_{s_1}|g_{s_1,s_2}|.
\end{aligned}$$

Note that $g_{s_1,s_2}(t)$ is bounded. Thus,

$$|\mathbb{E}F_n| \leq \frac{C}{n} \sum_{t=1}^{n} |\mathbb{E}x_t^{(i)}| \cdot \sum_{s_2=1}^{n} \sup_{s_1} |g_{s_1,s_2}(t)|. \tag{2.5.59}$$

If $i = 1$, we have $\sum_{t=1}^{n} |\mathbb{E}x_t^{(i)}| \leq C$, thus (2.5.59) is of order $O(n^{-1})$.

If $i \neq 1$ and $x_t^{(i)}$ is no deterministic term, we have $\sup_{t,h} |\mathbb{E}x_t^{(i)}| = o(1)$ and $g_{s_1,s_2}(t) = \mathbb{1}_{\{s_2=t\}} \cdot \frac{1}{nb} \cdot K\big(\frac{s_1-t}{nb}\big)$, thus $|\mathbb{E}F_n|$ is of order $o((nb)^{-1})$.

If $i \neq 1$ and $x_t^{(i)}$ is deterministic, we are in the case (2.5.49). Here a more precise analysis is needed. In fact, this expectation is the reason why we had to choose the likelihoods and cross validation functionals as projection error-type terms. First note that in this case the expectation can be bounded by (here $k, l \in \{1, ..., d\}$ are arbitrary indices):

$$|\mathbb{E}F_n| \leq \frac{C}{n(nb)} \sum_{s,t=1, s\neq t}^{n} \Big|K\Big(\frac{t-s}{nb}\Big)\Big| |\mathrm{Cov}(\partial_k \ell_{t,n}(\theta_0(t/n)), \partial_l \ell_{s,n}(\theta_0(t/n)))| \tag{2.5.60}$$

By Lemma 2.5.13, we have a decomposition $\partial_k \ell_{t,n}(\theta_0(t/n)) = \mathrm{MDS}_t^{(1)} + R_t^{(1)}$. Similarly to the result in Lemma 2.5.13, we have another decomposition

$$\partial_l \ell_{s,n}(\theta_0(t/n)) = \mathrm{MDS}_{s,t}^{(2)} + R_{s,t}^{(2)},$$

where $\mathrm{MDS}_{s,t}^{(2)} := \varepsilon_s \partial_l d_{s,n}(\theta_0(t/n)) - \partial_l \gamma_{\theta_0(t/n)}(0) a_{s,n}(0)$, and

$$\begin{aligned} R_{s,t}^{(2)} \;\; := \;\; & \partial_l \gamma_{\theta_0(t/n)}(0)(a_{s,n}(0) - a(t/n, 0))\varepsilon_s^2 - \sum_{k=s}^{\infty} \gamma_{\theta_0(t/n)}(k) X_{s-k,n} \cdot \partial_l d_{s,n}(\theta_0(t/n)) \\ & + \sum_{k=0}^{\infty} \gamma_{\theta_0(t/n)}(k) \big(X_{s-k,n} - \tilde{X}_{s-k}(\theta_0(t/n))\big) \cdot \partial_l d_{s,n}(\theta_0(t/n)). \end{aligned}$$

Thus, (2.5.60) can be bounded by

$$\frac{C}{n(nb)} \sum_{s,t=1}^{n} K\Big(\frac{t-s}{nb}\Big) \Big[|\mathrm{Cov}(\mathrm{MDS}_t^{(1)}, R_{s,t}^{(2)})| + |\mathrm{Cov}(R_t^{(1)}, \mathrm{MDS}_{s,t}^{(2)})| + |\mathrm{Cov}(R_t^{(1)}, R_{s,t}^{(2)})|\Big]. \tag{2.5.61}$$

We will only discuss the first summand, the other terms can be handled with the same arguments. We replace ε_t by $\varepsilon_t = \sum_{k=0}^{\infty} \gamma_{\theta_0(t/n)}(k) \tilde{X}_{t-k}(t/n)$ to avoid case distinctions. According to the definition of $R_{s,t}^{(2)}$, the first summand of (2.5.61) can be bounded by three terms T_1, T_2, T_3, T_4. For the first bound we obtain

$$\begin{aligned}
|T_1| &\le \frac{C}{n(nb)} \sum_{s,t=1}^{n} \sum_{k_1,k_2,l_2=0}^{\infty} \sum_{l_1=s}^{\infty} \frac{|\mathrm{Cov}(\tilde{X}_{t-k_1}(t/n) X_{t-k_2,n}, X_{s-l_1,n} X_{s-l_2,n})|}{\chi(k_1)\chi(k_2)\chi(l_1)\chi(l_2)} \\
&\le \frac{C}{n(nb)} \sum_{s=1}^{n} \frac{1}{\chi(s)} \sum_{k_1,k_2,l_2=0}^{\infty} \frac{1}{\chi(k_1)\chi(k_2)\chi(l_2)} \\
&\quad \times \sum_{t=1}^{n} \sum_{l_1=0}^{\infty} |\mathrm{Cov}(\tilde{X}_{t-k_1}(t/n) X_{t-k_2,n}, X_{s-l_1,n} X_{s-l_2,n})| \\
&\le \frac{C}{n(nb)},
\end{aligned}$$

because $\sum_{t=1}^{n} \sum_{l_1=0}^{\infty} |\mathrm{Cov}(\tilde{X}_{t-k_1}(t/n) X_{t-k_2,n}, X_{s-l_1,n} X_{s-l_2,n})| \le C$ uniformly in s, k_1, k_2, l_2. The second bound has the form

$$|T_2| \le \frac{C}{n(nb)} \sum_{s,t=1}^{n} \sum_{k_1,k_2,l_1,l_2=0}^{\infty} \frac{|\mathrm{Cov}(\tilde{X}_{t-k_1}(t/n) X_{t-k_2,n}, (X_{s-l_1,n} - \tilde{X}_{s-l_1}(\frac{s}{n})) X_{s-l_2,n})|}{\chi(k_1)\chi(k_2)\chi(l_1)\chi(l_2)}$$

Using $\sup_k \sum_{s=1}^{n} |a_{s,n}(k) - a(s/n, k)| \le C$ (see Assumption 2.1.1) and $V(a(\cdot,k)) \le \frac{C}{\chi(k)}$, we get the bound $|T_2| \le \frac{C}{n(nb)}$. The third bound reads

$$\begin{aligned}
|T_3| &\le \frac{C}{n(nb)} \sum_{s,t=1}^{n} \Big| K\Big(\frac{t-s}{nb}\Big)\Big| \sum_{k_1,k_2,l_1,l_2=0}^{\infty} \frac{1}{\chi(k_1)\chi(k_2)\chi(l_1)\chi(l_2)} \\
&\quad \cdot |\mathrm{Cov}(\tilde{X}_{t-k_1}(t/n) X_{t-k_2,n}, (\tilde{X}_{s-l_1,n}(\frac{s}{n}) - \tilde{X}_{s-l_1,n}(\frac{t}{n})) X_{s-l_2,n})|.
\end{aligned}$$

The Hoelder continuity of θ_0 allows us to write $|K(\frac{t-s}{nb})| \sup_k \left|a(\frac{s}{n}, k) - a(\frac{t}{n}, k)\right| \le C b^{\beta}$. This gives $|T_3| \le C\frac{b^{\beta}}{nb}$. For the last term, note that

$$\begin{aligned}
|T_4| &\le \frac{C}{n(nb)} \sum_{s,t=1}^{n} K\Big(\frac{t-s}{nb}\Big) \sum_{k_1,k_2,l_1,l_2=0}^{\infty} \frac{1}{\chi(k_1)\chi(k_2)\chi(l_1)\chi(l_2)} \\
&\quad \cdot |\mathrm{Cov}(\tilde{X}_{t-k_1}(t/n) X_{t-k_2,n}, (a_{s,n}(0) - a(t/n, 0)) \tilde{X}_{s-l_1}(t/n) \tilde{X}_{s-l_2}(t/n))|.
\end{aligned}$$

Using the same arguments as for T_2, T_3 (namely, $\sum_{s=1}^{n} |a_{s,n}(0) - a(s/n, 0)| \le C$ and $|K(\frac{t-s}{nb})| \cdot |a(s/n, 0) - a(t/n, 0)| \le C b^{\beta}$) yield $|T_4| \le \frac{C}{nb}\big(n^{-1} + b^{\beta}\big)$.

Terms with three random variables in (2.5.54): Here, we have

$$\begin{aligned}
F_n &= \frac{1}{n} \sum_{k_1,k_2,l_1,l_2,m_1,m_2=0}^{\infty} \sum_{s,t,\tau=1}^{n} g_{s,t,\tau}(k_1, k_2, l_1, l_2, m_1, m_2) \\
&\quad \cdot r\big(\frac{t-k_1}{n}\big) r\big(\frac{t-k_2}{n}\big) r\big(\frac{s-l_1}{n}\big) r\big(\frac{s-l_2}{n}\big) r\big(\frac{\tau-m_1}{n}\big) r\big(\frac{\tau-m_2}{n}\big) \\
&\quad \cdot \eta_{t-\min(k_1,k_2),|k_1-k_2|} \eta_{s-\min(l_1,l_2),|l_1-l_2|} \eta_{\tau-\min(m_1,m_2),|m_1-m_2|},
\end{aligned}$$

and from Lemma 2.5.12 we obtain

$$\|F_n - \mathbb{E}F_n\|_p \leq \frac{C_p}{n}\left(\sum_{s,t,\tau=1}^{n} g_{s,t,\tau}^2\right)^{1/2}.$$

In all cases, we have:

$$g_{s,t,\tau} = w(t/n) f_t \frac{1}{(nb)^2} K\left(\frac{t-s}{nb}\right) K\left(\frac{t-\tau}{nb}\right).$$

Thus,

$$\begin{aligned}\frac{1}{n}\left(\sum_{s,t,\tau=1}^{n} g_{s,t,\tau}^2\right)^{1/2} &\leq \frac{C}{n(nb)^2}\left(\sum_{s,t,\tau} K\left(\frac{t-s}{nb}\right)^2 K\left(\frac{t-\tau}{nb}\right)^2\right)^{1/2} \leq \frac{C(n(nb)^2)^{1/2}}{n(nb)^2} \\ &\leq \frac{C}{nb}\frac{1}{\sqrt{n}}.\end{aligned}$$

At last, we have to discuss the expectation $\mathbb{E}F_n$, i.e.

$$\frac{1}{n}\sum_{t=1}^{n} w(t/n) f_t \mathbb{E}[(x_t^{(1)} - \mathbb{E}x_t^{(1)})(x_t^{(2)} - \mathbb{E}x_t^{(2)})(x_t^{(3)} - \mathbb{E}x_t^{(3)})]. \tag{2.5.62}$$

We will not go into details here, but let us mention that all terms are $o((nb)^{-1})$.

Hoelder continuity property: Note that if $x_t^{(i)}$ depends on b, then it has the form $F(b) := \nabla^k L_{n,b,-t}(t/n, \theta_0(t/n))$ or $F(b) := \nabla^k L_{n,b}(t/n, \theta_0(t/n))$. Similar as in the proof of Proposition 2.5.8 it can be shown that F has a Hoelder continuity property,

$$|F(b) - F(b')| \leq |b - b'| \cdot C(n) \cdot \Big[\underline{X}'\underline{X} + 1\Big],$$

where $\underline{X} := (X_{1,n}, ..., X_{n,n})'$ and $C(n)$ grows only polynomially fast in n, but may change from line to line in the following. Because of (2.5.2) it can be shown that

$$\sup_{b\in B_n} |F(b)| \leq C(n)\Big(1 + \underline{X}'\underline{X}\Big),$$

Looking at (2.5.48) as a function of b,

$$G(b) := \frac{1}{n}\sum_{t=1}^{n} w(t/n) \cdot f_t \cdot x_t^{(1)} \cdot x_t^{(2)} \cdot x_t^{(3)},$$

we obtain

$$|G(b) - G(b')| \leq |b - b'| C(n)\Big[\underline{X}'\underline{X} + 1\Big]^2 \cdot \sum_{t=1}^{n} |x_t^{(1)}| \leq |b - b'| C(n)\Big[\underline{X}'\underline{X} + 1\Big]^3.$$

□

Proof of Corollary 2.2.7. $\theta_0(u)$ is in the interior of Θ. Thus, for n large enough, by uniform consistency (Theorem 2.5.10), $\hat{\theta}_b(u)$ lies in the interior of Θ, too. A standard maximum likelihood expansion yields

$$\begin{aligned}\hat{\theta}_b(u) - \theta_0(u) &= -(\nabla_\theta^2 L_{n,b}(u, \bar{\theta}(u)))^{-1} \cdot \nabla_\theta L_{n,b}(u, \theta_0(u)), \\ &= \big(I_d + R_b(u)\big) \cdot v_b(u),\end{aligned}$$

where $I_d \in \mathbb{R}^{d\times d}$ is the identity matrix,

$$\begin{aligned}v_b(u) &:= -I(\theta_0(u))^{-1} \cdot \nabla_\theta L_{n,b}(u, \theta_0(u)), \\ R_b(u) &:= (\nabla_\theta^2 L_{n,b}(u, \bar{\theta}(u)))^{-1} I(\theta_0(u)) - I_d\end{aligned}$$

and $|\bar{\theta}(u) - \theta_0(u)|_2 \le |\hat{\theta}_b(u) - \theta_0(u)|_2$ where $\bar{\theta}(u) \in \Theta$ is some intermediate value. Using the elementary formula $x'Ax \le |x|_2^2 |A|_{spec}$, we have

$$\begin{aligned}&\Big| |(I_d + R(u))v(u)|^2_{I(\theta_0(u))} - |v(u)|^2_{I(\theta_0(u))} \Big| \\ \le\ & 2 \cdot |\langle v(u), I(\theta_0(u)) R_b(u) v_b(u)\rangle| + |R_b(u) v_b(u)|^2_{I(\theta_0(u))} \\ \le\ & \Big(2|I(\theta_0(u))R_b(u)|_2 + |R_b(u)' I(\theta_0(u)) R_b(u)|_2\Big) \cdot |v_b(u)|^2. \qquad (2.5.63)\end{aligned}$$

Because of Proposition 2.5.8 and (2.5.32) we have

$$\sup_{b\in B_n} \sup_{u\in \mathrm{supp}(w)} \big|\nabla_\theta^2 L_{n,b}(u, \bar{\theta}(u)) - I(\theta_0(u))\big| \to 0,$$

thus

$$\sup_{b\in B_n} \sup_{u\in \mathrm{supp}(w)} |R_b(u)| \to 0. \qquad (2.5.64)$$

According to Assumption 2.2.1, let $c_0 > 0$ be the value which bounds all eigenvalues from $I(\theta_0(u))$ from below. Using the representations

$$\begin{aligned}d_I^*(\hat{\theta}_b, \theta_0) &= \int_0^1 |v_b(u)|^2_{I(\theta_0(u))} w(u)\ \mathrm{d}u, \\ d_I(\hat{\theta}_b, \theta_0) &= \int_0^1 |(I_d + R_b(u)) \cdot v_b(u)|^2_{I(\theta_0(u))} w(u)\ \mathrm{d}u,\end{aligned}$$

we conclude with (2.5.63), (2.5.64):

$$\begin{aligned}&\sup_{b\in B_n} \left| \frac{d_I(\hat{\theta}_b, \theta_0) - d_I^*(\hat{\theta}_b, \theta_0)}{d_I^*(\hat{\theta}_b, \theta_0)} \right| \\ \le\ & \frac{1}{c_0} \sup_{b\in B_n} \frac{\int_0^1 \big| |(I_d + R_b(u))v_b(u)|^2_{I(\theta_0(u))} - |v_b(u)|^2_{I(\theta_0(u))}\big| \cdot w(u)\ \mathrm{d}u}{\int_0^1 |v_b(u)|^2 \cdot w(u)\ \mathrm{d}u} \to 0 \quad (n \to \infty).\end{aligned}$$

Using the shortcuts $d_I = d_I(\hat{\theta}_b, \theta_0)$ (similarly for d_I^*, d_M^*), we have

$$\frac{d_I - d_M^*}{d_M^*} = \frac{d_I - d_I^*}{d_I^*} \cdot \left(\frac{d_I^* - d_M^*}{d_M^*} + 1\right) + \frac{d_I^* - d_M^*}{d_M^*},$$

hence, the assertion follows. The proof for d_A is the same by using sums instead of integrals. $\square$

Proof of Lemma 2.2.8. We define

$$\overline{d}_A^*(\hat{\theta}_b, \theta_0) := \frac{1}{n} \sum_{t=1}^{n} \left\| \nabla L_{n,b,-t}\Big(\frac{t}{n}, \theta_0\Big(\frac{t}{n}\Big)\Big) \right\|^2_{I(\theta_0(t/n))^{-1}}.$$

We have to show that

$$\sup_{b \in B_n} \left| \frac{\overline{d}_A^*(\hat{\theta}_b, \theta_0) - d_A^*(\hat{\theta}_b, \theta_0)}{d_M^*(\hat{\theta}_b, \theta_0)} \right| \to 0, \tag{2.5.65}$$

then it follows immediately from Lemma 2.2.6:

$$\sup_{b \in B_n} \left| \frac{\overline{d}_A^*(\hat{\theta}_b, \theta_0) - d_M^*(\hat{\theta}_b, \theta_0)}{d_M^*(\hat{\theta}_b, \theta_0)} \right| \to 0. \tag{2.5.66}$$

Using the same techniques as in the proof of Corollary 2.2.7 and using Lemma 2.5.9, it can be shown

$$\sup_{b \in B_n} \left| \frac{\overline{d}_A(\hat{\theta}_b, \theta_0) - \overline{d}_A^*(\hat{\theta}_b, \theta_0)}{\overline{d}_A^*(\hat{\theta}_b, \theta_0)} \right| \to 0, \tag{2.5.67}$$

and we can conclude from (2.5.66), (2.5.67) like in the proof of Corollary 2.2.7 that

$$\sup_{b \in B_n} \left| \frac{\overline{d}_A(\hat{\theta}_b, \theta_0) - d_M^*(\hat{\theta}_b, \theta_0)}{d_M^*(\hat{\theta}_b, \theta_0)} \right| \to 0.$$

We now show (2.5.65). We have

$$\nabla L_{n,b}(t/n, \theta_0(t/n)) - \nabla L_{n,b,-t}(t/n, \theta_0(t/n)) = \frac{K(0)}{nb} \nabla \ell_{t,n}(\theta_0(t/n)),$$

thus using the Cauchy Schwarz inequality we obtain

$$\begin{aligned}
& |d_A^*(\hat{\theta}_b, \theta_0) - \overline{d}_A^*(\hat{\theta}_b, \theta_0)| && (2.5.68)\\
\leq \; & \frac{2K(0)}{nb} \cdot \frac{1}{n} \sum_{t=1}^{n} |\langle \nabla \ell_{t,n}(\theta_0(t/n)), I(\theta_0(t/n))^{-1} \nabla L_{n,b}(t/n, \theta_0(t/n)) \rangle| \\
& + \frac{K(0)^2}{(nb)^2} \cdot \frac{1}{n} \sum_{t=1}^{n} |\nabla \ell_{t,n}(\theta_0(t/n))|^2_{I(\theta_0(t/n))^{-1}} \\
\leq \; & \frac{2K(0)}{nb} Z_t^{1/2} \cdot d_A^*(\hat{\theta}_b, \theta_0)^{1/2} + \frac{K(0)^2}{(nb)^2} \cdot Z_t, && (2.5.69)
\end{aligned}$$

where

$$Z_t := \frac{1}{n} \sum_{t=1}^{n} |\nabla \ell_{t,n}(\theta_0(t/n))|^2_{I(\theta_0(t/n))^{-1}}.$$

Define $x_t := \nabla \ell_{t,n}(\theta_0(t/n))$, $A_t := I(\theta_0(t/n))^{-1}$. By writing

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}|x_t|_{A_t}^2 &= \frac{1}{n}\sum_{s,t=1}^{n}\mathbb{1}_{\{s=t\}}\Big(\langle x_t-\mathbb{E}x_t, A_t(x_s-\mathbb{E}x_s)\rangle - \mathbb{E}\langle x_t-\mathbb{E}x_t, A_t(x_s-\mathbb{E}x_s)\rangle\Big)\\ &\quad +\frac{2}{n}\sum_{t=1}^{n}\langle x_t-\mathbb{E}[x_t], A_t\mathbb{E}[x_t]\rangle + \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}|x_t|_{A_t}^2\end{aligned}$$

and using Lemma 2.5.12 and the Markov inequality, it is easy to show that the first and second term converge to zero almost surely. A straightforward calculation shows that the third (deterministic) term is bounded. In total, Z_t is bounded almost surely. (2.5.65) now follows from (2.5.69),

$$d_A^* = d_M^* \cdot \Big(1 + \frac{d_A^* - d_M^*}{d_M^*}\Big)$$

and Lemma 2.2.6. □

Proof of Lemma 2.2.6. We use the decomposition

$$\begin{aligned}&|\nabla_\theta L_{n,b}(u,\theta_0(u))|_{I(\theta_0(u))^{-1}}^2 - \mathbb{E}|\nabla_\theta L_{n,b}(u,\theta_0(u))|_{I(\theta_0(u))^{-1}}^2\\ =\;& \Big(|\nabla_\theta L_{n,b}(u,\theta_0(u)) - \mathbb{E}\nabla_\theta L_{n,b}(u,\theta_0(u))|_{I(\theta_0(u))^{-1}}^2\\ &\quad -\mathbb{E}|\nabla_\theta L_{n,b}(u,\theta_0(u)) - \mathbb{E}\nabla_\theta L_{n,b}(u,\theta_0(u))|_{I(\theta_0(u))^{-1}}^2\Big)\\ &\quad +2\left\langle \nabla_\theta L_{n,b}(u,\theta_0(u)) - \mathbb{E}\nabla_\theta L_{n,b}(u,\theta_0(u)), I(\theta_0(u))^{-1}\mathbb{E}\nabla_\theta L_{n,b}(u,\theta_0(u))\right\rangle.\end{aligned} \tag{2.5.70}$$

This shows that the proof of the uniform convergence

$$\sup_{b\in B_n}\Big|\frac{d_A^*(\hat\theta_b,\theta_0) - \mathbb{E}[d_A^*(\hat\theta_b,\theta_0)]}{d_M^*(\hat\theta_b,\theta_0)}\Big| \to 0 \tag{2.5.71}$$

is already covered by the proof of the almost sure convergence of (2.5.50) in the second part of the proof of Lemma 2.2.9. Note that the form of the decomposition (2.5.70) implies that there is no need to discuss any convergences of expectations in Lemma 2.2.9. Similar argumentations with integrals instead of sums lead to

$$\sup_{b\in B_n}\Big|\frac{d_I^*(\hat\theta_b,\theta_0) - d_M^*(\hat\theta_b,\theta_0)}{d_M^*(\hat\theta_b,\theta_0)}\Big| \to 0. \tag{2.5.72}$$

Because of Corollary 2.5.6, we have

$$\sup_{b\in B_n}\Big|\frac{d_M^*(\hat\theta_b,\theta_0) - \mathbb{E}[d_A^*(\hat\theta_b,\theta_0)]}{d_M^*(\hat\theta_b,\theta_0)}\Big| \to 0. \tag{2.5.73}$$

The assertion follows from (2.5.71), (2.5.72) and (2.5.73). □

Chapter 3

An approximation theory for recursively defined locally stationary processes

In this chapter we do a first step towards a general asymptotic theory for nonlinear locally stationary processes. In the literature, most of the asymptotic results are obtained if the explicit structure of the time series model is known, such as in the case of tvAR models (cf. Dahlhaus and Giraitis (1998)), linear models (cf. Dahlhaus and Polonik (2009)), the tvARCH case (cf. Dahlhaus and Subba Rao (2006)) and random coefficient models (cf. Subba Rao (2006)). To prove their results, the authors of these papers heavily use the structure of these models.

We consider a quite general Markov-structured non-stationary process $X_{t,n}$, $t = 1, ..., n$. With this model we cover many well-known locally stationary processes (especially the models mentioned above) which are obtained by replacing the constant parameters by time-dependent parameter curves evaluated at t/n.

To formulate our results, we will use the functional dependence measure introduced in Wu (2005). Some recent publications which are using this framework also allow for locally stationary processes: Karmakar and Wu (2016) deal with strong approximations, Zhou and Wu (2009) discuss quantile regression and Liu, Xiao and Wu (2013) obtain inequalities for tail probabilities (which they claim are also valid for nonstationary models). Up to now, standard results as laws of large numbers and central limit theorems for general nonstationary processes have not been proved yet. Furthermore, most of the statements in those publications are given in a normalized way meaning that the expectation of the underlying process is 0. A more comprehensive study of the asymptotic behavior of the expectation of locally stationary processes is usually missing.

In section 3.1, we show that $X_{t,n}$ can be approximated by some stationary process $\tilde{X}_t(u)$ as long as $|t/n-u| \ll 1$ and $n^{-1} \ll 1$. We prove that under reasonable conditions $X_{t,n}$ has a Taylor-like expansion into $\tilde{X}_t(u)$ and so-called derivative processes which form the key quantities in the following derivations. Derivative processes were already defined for specific models as in the case of tvAR (cf. Dahlhaus (2011)) and tvARCH (cf.

Dahlhaus and Subba Rao (2006)). However, the existence of these processes and their properties are completely unclear if an explicit representation of $X_{t,n}$ is not available.

In section 3.2, we use derivative processes to prove expansions of expectations, covariance functions and the distribution function of $X_{t,n}$. Furthermore, we present laws of large numbers and a central limit theorem which hold under minimal moment assumptions on the process. The proofs are based on a partition of the sum over $X_{t,n}$ into sums over smaller ranges of t where $X_{t,n}$ can be approximated by stationary processes $\tilde{X}_t(u)$ by exploiting their smoothness properties. We then make use of the asymptotic theory for sums of stationary sequences. It should be noted that the theorems we derive in this section do not use the special Markovian structure of the process but only expansion and dependence properties which are results from section 3.1. This means that as long similar results as presented in section 3.1 can be obtained for a process, all the theorems are applicable in principle.

In section 3.3, we apply the previous results to nonparametric maximum likelihood estimation of parameter curves in locally stationary processes. The general framework for this chapter was already given in Dahlhaus (2011). Some concluding remarks are given in section 3.4. Some proofs are postponed to section 3.5.

3.1 Stationary approximations and derivative processes

For some fixed natural number $p > 0$, an i.i.d. sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ of real-valued random variables and a function $G : \mathbb{R} \times \mathbb{R}^p \times [0,1] \to \mathbb{R}, (\varepsilon, x, u) \mapsto G_\varepsilon(x,u)$, consider the process $X_{t,n}$ defined by the recursion

$$X_{t,n} = G_{\varepsilon_t}\big(X_{t-1,n}, ..., X_{t-p,n}, \frac{t}{n} \vee 0\big), \quad t \le n, \tag{3.1.1}$$

where $a \vee b = \max\{a,b\}$. We assume that the process is observed at $t = 1, ..., n$ meaning that in the above model the time is rescaled to the unit interval due to $\frac{t}{n} \in [0,1]$. At a fixed time point $u \in [0,1]$, we define the *stationary approximation* as the stationary process $\tilde{X}_t(u)$, $t \in \mathbb{Z}$ given by the recursion

$$\tilde{X}_t(u) = G_{\varepsilon_t}\big(\tilde{X}_{t-1}(u), ..., \tilde{X}_{t-p}(u), u\big), \quad t \in \mathbb{Z}. \tag{3.1.2}$$

The notion of local stationarity now means that for each $u \in (0,1)$ and sufficiently small $\delta > 0$ the processes $X_{t,n}$ and $\tilde{X}_t(u)$ are close to each other for $\frac{t}{n} \in [u-\delta, u+\delta]$ (see Proposition 3.1.5 below). It is obvious that this requires smoothness assumptions on the function G specified below in Assumption 3.1.2. The stationary approximation can fruitfully be used to derive mathematical results on the process $X_{t,n}$ (cf. sections 3.2 and 3.3). More mathematical tools and a deeper understanding are provided by the derivative processes $\partial_u \tilde{X}_t(u)$, $\partial_u^2 \tilde{X}_t(u)$, etc. They reflect the slope and curvature of the nonstationary process at time u - for example we have the Taylor expansion

$$X_{t,n} \approx \tilde{X}_t\big(\frac{t}{n}\big) \approx \tilde{X}_t(u_0) + \Big(\frac{t}{n} - u_0\Big)\partial_u \tilde{X}_t(u_0) + \frac{1}{2}\Big(\frac{t}{n} - u_0\Big)^2 \partial_u^2 \tilde{X}_t(u_0)$$

under suitable regularity assumptions. Usually the derivative processes are also stationary and even more ergodic making the a powerful tool for proofs. These properties are proved for the general class of processes of the form (3.1.1) below in a series of theorems. Furthermore, we establish that the uniform functional dependence measure decays geometrically which is a key property for proving asymptotic results. Before we start we give some examples for models which fulfill (3.1.1). These include in particular several classical parametric time series models where the constant parameters have been replaced by time-dependent parameter curves.

Example 3.1.1. *(i) the tvAR(p) process (cf. Dahlhaus and Giraitis (1998)): Given parameter curves $a_i, \sigma : [0,1] \to \mathbb{R}$ $(i = 1, ..., p)$,*

$$X_{t,n} = a_1\big(\frac{t}{n}\big)X_{t-1,n} + ... + a_p\big(\frac{t}{n}\big)X_{t-p,n} + \sigma\big(\frac{t}{n}\big)\varepsilon_t.$$

(ii) the tvARCH(p) process (cf. Dahlhaus and Subba Rao (2006)): Given parameter curves $a_i : [0,1] \to \mathbb{R}$ $(i = 0, ..., p)$,

$$X_{t,n} = \Big(a_0\big(\frac{t}{n}\big) + a_1\big(\frac{t}{n}\big)X_{t-1,n}^2 + ... + a_p\big(\frac{t}{n}\big)X_{t-p,n}^2\Big)^{1/2}\varepsilon_t.$$

(iii) the tvTAR(1) process (cf. Zhou and Wu (2009)): Given parameter curves $a_1, a_2 : [0,1] \to \mathbb{R}$, define

$$X_{t,n} = a_1\big(\frac{t}{n}\big)X_{t-1,n}^+ + a_2\big(\frac{t}{n}\big)X_{t-1,n}^- + \varepsilon_t,$$

where $x^+ := \max\{x, 0\}$ and $x^- := \max\{-x, 0\}$.

(iv) the time-varying random coefficient model (cf. Subba Rao (2006)): With some parameter functions $a_i(\cdot)$, $i = 0, ..., p$,

$$X_{t,n} = a_0(\varepsilon_t) + a_1(\varepsilon_t)X_{t-1,n} + ... + a_p(\varepsilon_t)X_{t-p,n}.$$

Recall that for $q > 0$ we denote the weighted ℓ^q-norm by $|x|_{w,q} = \big(\sum_{i=1}^p w_i|x_i|^q\big)^{1/q}$ and for real-valued random variables Z we use $\|Z\|_q = (\mathbb{E}|Z|^q)^{1/q} < \infty$. Recall the definition of the shift process $\mathcal{F}_t = (\varepsilon_t, \varepsilon_{t-1}, ...)$, the uniform functional dependence measure $\delta_q^Y(k)$ and the projection operator $P_j\cdot = \mathbb{E}[\cdot|\mathcal{F}_j] - \mathbb{E}[\cdot|\mathcal{F}_{j-1}]$ from the preliminaries of this thesis. The subsequent theorems contain results on the geometric decay of this functional dependence measure which will be used in sections 3.2 and 3.3 to provide asymptotic results like uniform laws of large numbers and central limit theorems.

We will use $\partial_1 G_\varepsilon(y,u)$, $\partial_2 G_\varepsilon(y,u)$ to denote the derivatives of $G_\varepsilon(y,u)$ with respect to y, u, respectively.

We work with the following set of assumptions:

Assumption 3.1.2. *There exists $q > 0$, $\chi = (\chi_1, ..., \chi_p) \in \mathbb{R}_{\geq 0}^p$ with $|\chi|_1 < 1$ and $y_0 \in \mathbb{R}^p$ such that:*

(i) $\sup_{u\in[0,1]} \|G_{\varepsilon_0}(y_0,u)\|_q < \infty$, *and*

$$\sup_{u\in[0,1]} \sup_{y\neq y'} \frac{\|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y',u)\|_q}{|y-y'|_{\chi,q'}} \leq 1. \tag{3.1.3}$$

(ii) $(y,u) \mapsto G_\varepsilon(y,u)$ *is continuous for all* ε, $\| \sup_{u\in[0,1]} |G_{\varepsilon_0}(y_0,u)|\, \|_q < \infty$, *and*

$$\Big\| \sup_{u\in[0,1]} \sup_{y\neq y'} \frac{|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y',u)|}{|y-y'|_{\chi,q'}} \Big\|_q \leq 1. \tag{3.1.4}$$

(iii) $(y,u) \mapsto G_\varepsilon(y,u)$ *is continuously differentiable for all* ε,
$\| \sup_{u\in[0,1]} |\partial_2 G_{\varepsilon_0}(y_0,u)|\, \|_q < \infty$, *and*

$$C_i := \Big\| \sup_{u\in[0,1]} \sup_{y\neq y'} \frac{|\partial_i G_{\varepsilon_0}(y,u) - \partial_i G_{\varepsilon_0}(y',u)|}{|y-y'|_{1,q'}} \Big\|_q < \infty \tag{3.1.5}$$

(iv) For some $0 < \alpha \leq 1$, *it holds that*

$$C := \sup_{u\in[0,1]} \|C(\tilde{Y}_t(u))\|_q < \infty, \quad \text{where} \quad C(y) := \sup_{u\neq u'} \frac{\|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y,u')\|_q}{|u-u'|^\alpha} \tag{3.1.6}$$

Discussion: Note that (i) - (iii) impose increasingly strong smoothness assumptions on the recursion function $G_\varepsilon(y,u)$. While (i) - (iii) are directly verifiable, (iv) includes conditions on the stationary approximation $\tilde{X}_t(u)$. Note that the upcoming theorems also state properties of $\tilde{X}_t(u)$. Their results can be used to verify this assumption.

3.1.1 Existence and uniqueness of $X_{t,n}$ and $\tilde{X}_t(u)$

We now establish existence and uniqueness of $X_{t,n}$ and $\tilde{X}_t(u)$ under mild contraction conditions.

Proposition 3.1.3. *(i) Existence of a stationary approximation: Suppose that Assumption 3.1.2(i) holds. Then for all* $u \in [0,1]$, *the recursion (3.1.2) has a unique stationary and ergodic solution* $\tilde{X}_t(u) = H(u,\mathcal{F}_t)$ *and we have*

$$\sup_{u\in[0,1]} \delta_q^{\tilde{X}(u)}(k) \leq C\rho^k, \qquad \sup_{u\in[0,1]} \|\tilde{X}_0(u)\|_q < \infty$$

with some $C > 0$ *and* $0 < \rho < 1$.
(ii) Existence of the nonstationary process: Under the above conditions, there exists an a.s. unique solution of (3.1.1) with a representation $X_{t,n} = H_{t,n}(\mathcal{F}_t)$ *and* $\sup_{n\in\mathbb{N}} \sup_{t=1,\dots,n} \|X_{t,n}\|_q < \infty$. *Furthermore, it holds that*

$$\sup_{n\in\mathbb{N}} \delta_q^{X_{\cdot,n}}(k) \leq C\rho^k$$

with some $C > 0$ *and* $0 < \rho < 1$.

The proof of (i) for fixed $u \in [0,1]$ is similar to the proof in Shao and Wu (2007), Theorem 5.1. Since we state the results uniformly in $u \in [0,1]$, we will give the proof in the appendix for completeness. Since the definition of $X_{t,n}$ and $\tilde{X}_t(0)$ coincide for $t \leq 0$, existence and uniqueness of $X_{t,n}$ follow from the existence and uniqueness of $\tilde{X}_t(0)$. Therefore, (ii) is an immediate corollary of (i).

3.1.2 A uniform L^q-approximation

We now prove that $X_{t,n}$ can be approximated by the stationary process $\tilde{X}_t(u)$ uniformly in a L^q-sense. We will use the shortcuts

$$Y_{t-1,n} := (X_{t-1,n}, ..., X_{t-p,n})', \qquad \tilde{Y}_{t-1}(u) := (\tilde{X}_{t-1}(u), ..., \tilde{X}_{t-p}(u))'$$

to keep the notation of the recursion equations simple.

Lemma 3.1.4. *Suppose that Assumption 3.1.2(i),(iv) hold. Then,*

$$\sup_{u \neq u'} \frac{\|\tilde{X}_t(u) - \tilde{X}_t(u')\|_q}{|u-u'|^{\alpha}} \leq \frac{C}{(1-|\chi|_1)^{1/q'}}. \tag{3.1.7}$$

Furthermore, we have:

$$\sup_{t=1,...,n} \|X_{t,n} - \tilde{X}_t(t/n)\|_q \leq Cp^{\alpha} \left(\frac{|\chi|_1}{(1-|\chi|_1)^2} \right)^{1/q'} \cdot n^{-\alpha}. \tag{3.1.8}$$

Note that the approximation error in (3.1.8) cannot be avoided - cf. Dahlhaus (2011), (49), for the tvAR(1) case (with a different error due to different assumptions). The following approximation result is now obtained as a corollary.

Proposition 3.1.5. *Under the assumptions of Lemma 3.1.4, we have the following strong approximation of $X_{t,n}$ uniformly for all $t = 1, ..., n$:*

$$\|X_{t,n} - \tilde{X}_t(u)\|_q \leq \frac{C}{(1-|\chi|_1)^{1/q'}} \Big| u - \frac{t}{n} \Big|^{\alpha} + Cp^{\alpha} \Big(\frac{|\chi|_1}{(1-|\chi|_1)^2} \Big)^{1/q'} \cdot n^{-\alpha} = O\Big(\Big| u - \frac{t}{n} \Big|^{\alpha} + n^{-\alpha} \Big).$$

3.1.3 Existence of continuous modifications and derivative processes

Proposition 3.1.3 gives the almost sure uniqueness of $\tilde{X}_t(u)$ for each $u \in [0,1]$, but not continuity of $u \mapsto \tilde{X}_t(u)$ since this involves uncountably many points $u \in [0,1]$. In order to guarantee the existence of a continuous or even differentiable modification $\hat{X}_t(u)$ of $\tilde{X}_t(u)$ we have to impose stronger conditions on the recursion function G in (3.1.1) ($\hat{X}_t(u)$ is a modification of $\tilde{X}_t(u)$ if for all $u \in [0,1]$, $\hat{X}_t(u) = \tilde{X}_t(u)$ a.s.). A natural way would be to apply the Kolmogorov-Chentzov theorem, but this theorem contains a tradeoff in its conditions between moment assumptions and smoothness of the process which usually leads to either strong moment or smoothness assumptions which may not be useful in practice. Furthermore it does not use the specific structure of the process which is known and we could not give a bound for moments of $\sup_{u \in [0,1]} |\hat{X}_t(u)|$. We therefore use a different approach.

Theorem 3.1.6 (Existence of a continuous modification)**.** *Suppose that Assumption 3.1.2(ii) holds. Then for each $t \in \mathbb{Z}$, there exists a continuous modification $(\hat{X}_t(u))_{u\in[0,1]}$ of $(\tilde{X}_t(u))_{u\in[0,1]}$ from Proposition 3.1.3 with* $\sup_{u\in[0,1]} |\hat{X}_t(u)| \in L^q$.

Because most of the recursively defined stationary models have (component-wise) Lipschitz continuous recursion functions G, condition (3.1.4) is fulfilled for these models with an appropriate parameter space. The supremum taken over $u \in [0,1]$ in (3.1.4) however seems to restrict the parameter space. If for example ε_t has a distribution with mean 0 and variance 1, and $G_\varepsilon(x,u) = a(u)x + b(u)x \cdot \varepsilon$ with continuous functions $a(u)$, $b(u)$, then (3.1.4) for $q = 2$ reads $\| \sup_{u\in[0,1]} |a(u) + b(u)\varepsilon_t| \|_2 < 1$. The following result implies that this condition can be relaxed to $\sup_{u\in[0,1]} \|a(u) + b(u)\varepsilon_t\|_2 < 1$ under certain assumptions.

Proposition 3.1.7. *In the situation of Theorem 3.1.6, instead of (3.1.4) assume that $x \mapsto G_\varepsilon(x,u)$ is differentiable for all ε, u and that for all $u_0 \in [0,1]$,*

$$\limsup_{\delta\to 0} \Big\| \sup_{|u-u_0|\le\delta} \sup_x |\partial_1 G_{\varepsilon_0}(x,u) - \partial_1 G_{\varepsilon_0}(x,u_0)|_1 \Big\|_q = 0$$

and

$$\sup_{u\in[0,1]} \Big\| \sup_{y\neq y'} \frac{|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y',u)|}{|y-y'|_{\chi,q'}} \Big\|_q \le 1.$$

Then the results of Theorem 3.1.6 are still valid.

In the following we will assume that $(y,u) \mapsto G_\varepsilon(y,u)$ is differentiable in both components. For the moment, assume that there exists a modification of the process $(\tilde{X}_t(u))_{u\in[0,1]}$ with differentiable paths (denote the modification by $\tilde{X}_t(u)$ again) and denote the derivative by $D_u\tilde{X}_t(u)$. Define $D_u\tilde{Y}_{t-1}(u) := (D_u\tilde{X}_{t-1}(u), ..., D_u\tilde{X}_{t-p}(u))'$. Then the following recursion equation for $D_u\tilde{X}_t(u)$, obtained by differentiating (3.1.2) should hold:

$$D_u\tilde{X}_t(u) = \langle \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u), D_u\tilde{Y}_{t-1}(u)\rangle + \partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u), \tag{3.1.9}$$

The first part of the next theorem shows that given the existence of the process $\tilde{X}_t(u)$ from Theorem 3.1.3, the recursion (3.1.9) has a solution $D_u\tilde{X}_t(u)$; the second part shows that $\tilde{X}_t(u)$ is differentiable with respect to u and that the derivative coincides with $D_u\tilde{X}_t(u)$.

Theorem 3.1.8. *Suppose that Assumptions 3.1.2(ii),(iii) hold. Then the following statements hold.*

(i) Existence of the first derivative process: For all $u \in [0,1]$, the recursion (3.1.9) has a unique stationary and ergodic solution $D_u\tilde{X}_t(u) = \tilde{H}(u, \mathcal{F}_t)$ and it holds that

$$\delta_q^{D_u\tilde{X}(u)}(k) \le C\rho^k, \qquad \sup_{u\in[0,1]} \|D_u\tilde{X}_t(u)\|_q < \infty$$

with some $C > 0$, $0 < \rho < 1$.

(ii) Differentiability:

(a) There exists a continuously differentiable modification $(\hat{X}_t(u))_{u\in[0,1]}$ of the process $(\tilde{X}_t(u))_{u\in[0,1]}$ from Proposition 3.1.3 where $\partial_u \hat{X}_t(u)$ is a modification of $D_u \tilde{X}_t(u)$,

(b) $\sup_{u\in[0,1]} |\partial_u \hat{X}_t(u)| \in L^q$.

As it can be seen in the proof, the statements of Theorem 3.1.8(i) can be obtained under milder conditions. More precisely, one has to suppose that Assumption 3.1.2(i) holds and that for all ε, the mapping $(y,u) \mapsto G_\varepsilon(y,u)$ is differentiable with

$$\sup_{u\in[0,1]} \|C(\tilde{Y}_{t-1}(u))\|_q < \infty, \quad \text{where} \quad C(y) := \sup_{u\in[0,1]} \|\partial_2 G_{\varepsilon_0}(y,u)\|_q, \tag{3.1.10}$$

The results of Theorem 3.1.8 allow us to Taylor expand $\tilde{X}_t(t/n)$ around $\tilde{X}_t(u)$:

Corollary 3.1.9 (Taylor expansion of $\tilde{X}_t(t/n)$)**.** *Suppose that Assumptions 3.1.2(ii),(iii) hold. Then we have for all t,n and $u \in [0,1]$:*

$$\tilde{X}_t\big(\frac{t}{n}\big) = \tilde{X}_t(u) + \big(\frac{t}{n} - u\big) \cdot \partial_u \tilde{X}_t(u) + R_{t,n} \quad a.s.,$$

where $R_{t,n} = \big(\frac{t}{n} - u\big)\Big\{\partial_u \tilde{X}_t(\bar{u}_{t,n}) - \partial_u \tilde{X}_t(u)\Big\}$ and $\bar{u}_{t,n}$ is a random variable with $|\bar{u}_{t,n} - u| \le |\frac{t}{n} - u|$. If $|\frac{t}{n} - u| = o(1)$, it holds that $R_{t,n} = o(|\frac{t}{n} - u|)$.

Under suitable conditions, similar results hold for higher order derivatives of $\hat{X}_t(u)$. For some models it is possible to obtain explicit expressions for the corresponding derivative processes.

Example 3.1.10 (Explicit representations for derivative processes)**.**

(i) The tvAR(p) process $X_{t,n} = \sum_{j=1}^p a_j\big(\frac{t}{n}\big) X_{t-j,n} + \varepsilon_t$ has the corresponding stationary approximation $\tilde{X}_t(u) = \sum_{j=1}^p a_j(u)\tilde{X}_{t-j}(u) + \varepsilon_t$ which has an explicit representation $\tilde{X}_t(u) = \sum_{j=0}^\infty \psi_j(u) \cdot \varepsilon_{t-j}$ with differentiable ψ_j $(j = 0,1,2,...)$. It is easy to see that $\partial_u \tilde{X}_t(u) = \sum_{j=0}^\infty \partial_u \psi_j(u) \cdot \varepsilon_{t-j}$ is the a.s. uniquely determined derivative process.

(ii) For tvARCH(p) processes, explicit expressions for the derivative processes were obtained in Dahlhaus and Subba Rao (2006).

In the following we will write $\tilde{X}_t(u)$ even if we want to refer to the differentiable modification to keep the notation simple. Since all our results only involve finitely (or at most countably) many observations, this will not cause any problems.

3.1.4 Higher order derivative processes

If $\tilde{X}_t(u)$ has a twice continuously differentiable modification and $(y,u) \mapsto G_\varepsilon(y,u)$ is twice continuously differentiable, then the following recursion equation for $\partial_u^2 \tilde{X}_t(u)$ should hold:

$$\begin{aligned}\partial_u^2 \tilde{X}_t(u) &= \langle \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u),u), \partial_u^2 \tilde{Y}_{t-1}(u)\rangle + \langle \partial_1^2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u),u)\partial_u \tilde{Y}_{t-1}(u), \partial_u \tilde{Y}_{t-1}(u)\rangle \\ &\quad +2\langle \partial_1\partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u),u), \partial_u \tilde{Y}_{t-1}(u)\rangle + \partial_2^2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u),u). \end{aligned} \tag{3.1.11}$$

Using the same techniques as in Theorem 3.1.8, one can find similar conditions as in Assumption 3.1.2 such that a second (or even higher) order derivative process $\partial_u^2 \tilde{X}_t(u)$ exists. These results can be used in situations where a higher order Taylor expansion is necessary, see Chapter 4 of this thesis.

In most of the practical situations one would expect that the processes $\tilde{X}_t(u)$, $\partial_u \tilde{X}_t(u)$ and $\partial_u^2 \tilde{X}_t(u)$ allow for the same moments, i.e. if one of the processes is in L^q then the other processes fulfill this, too. In (3.1.11) however there seems to occur an imbalance because of the term $\langle \partial_1^2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u),u)\partial_u \tilde{Y}_{t-1}(u), \partial_u \tilde{Y}_{t-1}(u)\rangle$ which seems to have a q-th moment only in the case that $\partial_u \tilde{X}_t(u)$ has a $2q$-th moment. Following the proof techniques of Theorem 3.1.8 this would lead to the fact that $\partial_u^2 \tilde{X}_t(u)$ only has a q-th moment under conditions on G_{ε_0} and its derivatives (similar to (3.1.5)) which involve $2q$-th moments. It can be seen in special cases where an explicit representation of the process is available (for example tvAR(p), Dahlhaus (2011) or tvARCH(p), Dahlhaus and Subba Rao (2006)) that $2q$-th moments are not necessary in general. We conjecture that the reason for this lies in the behaviour of $\partial_1^2 G$ which in these cases satisfies that $|\langle \partial_1^2 G_\varepsilon(y,u), y\rangle|_1$ is still bounded uniformly in y, u.

The formalization of this is beyond the scope of this chapter. We will close this section by presenting a result on the Hoelder continuity of the first derivative process which already contains the higher moment assumption discussed above.

Lemma 3.1.11 (Hoelder property of the first derivative process). *Suppose that Assumption 3.1.2(ii),(iii) hold. Additionally assume that for some* $1 \geq \alpha_2 > 0$ *and* $i = 1,2$ *it holds component-wise:*

$$D_i := \sup_u \|D_i(\tilde{Y}_t(u))\|_q < \infty, \quad D_i(y) := \sup_{u \neq u'} \frac{\|\partial_i G_{\varepsilon_0}(y,u) - \partial_i G_{\varepsilon_0}(y,u')\|_q}{|u-u'|^{\alpha_2}} \tag{3.1.12}$$

Then

$$\sup_{u\neq u'} \frac{\|\partial_u \tilde{X}_t(u) - \partial_u \tilde{X}_t(u')\|_{q/2}}{|u-u'|^{\alpha_2}} \leq C.$$

with some constant $C > 0$.

3.1.5 A simulation study

To quantify the quality of the approximations given in Lemma 3.1.4 and Corollary 3.1.9, we consider the tvARCH(1) model

$$X_{t,n} := \Big(a_0 + a_1\big(\frac{t}{n}\big)X_{t-1,n}^2\Big)^{1/2}\varepsilon_t$$

with $a_0 := 0.2$, $a_1(u) = 0.95u^2$ and $\varepsilon_0 \sim N(0,1)$. Note that if t/n tends to 1, the values of $X_{t,n}$ are more dependent to each other than for smaller values of t/n. We generated realizations of $X_{t,n}$, $\tilde{X}_t(\frac{t}{n})$ with $n = 500$ (see Figure 3.1(a),(b) for a realization of $X_{t,n}$ and $X_{t,n} - \tilde{X}_t(\frac{t}{n})$). In Figure 3.1(c) we have the plotted empirical 5%- and 95%-quantile curves of the difference $X_{t,n} - \tilde{X}_t(\frac{t}{n})$ for $N = 1000$ replications. It can be seen that with stronger dependence, the quality of the approximation $X_{t,n} \approx \tilde{X}_t(\frac{t}{n})$ gets worse. Secondly we consider the approximation quality of $\tilde{X}_t(t/n)$ by $\tilde{X}_t(u)$ and $\tilde{X}_t(u) + (\frac{t}{n} - u)\partial_u \tilde{X}_t(u)$, respectively. Since these approximations are only working locally (for $|t/n - u| \ll 1$), we compare them by dividing the whole time line $t = 1, ..., n$ into subsets $(u_i - b, u_i + b]$, where $b = 50$ and $u_i = (2i-1)b$ for $i = 1, ..., 5$. In Figure 3.1(d) empirical 5%- and 95%-quantile curves obtained from $N = 1000$ replications for the differences $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i)$ and $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i) - (\frac{t}{n} - u_i)\partial_u \tilde{X}_t(u_i)$ (where $t \in (u_i - b, u_i + b]$) are depicted, respectively. We emphasize that the improvement of the (pointwise) approximation $\tilde{X}_t(\frac{t}{n})$ by taking into account the derivative process is remarkable. However, both approximations again get worse if the dependence of $X_{t,n}$ to earlier values increases.

3.2 Asymptotic properties of functionals of $X_{t,n}$

3.2.1 Mean expansions

To get results for a wide range of interesting functionals, we define the following class $\mathcal{H}_r(\alpha, M, C)$ of real-valued functions which have a Hoelder property where the Hoelder constant may depend at most polynomially on the location.

Definition 3.2.1 (The classes $\mathcal{H}_r(\beta, M, C)$ and $\mathcal{L}_r(M, C)$)**.** *We say that a function $g : \mathbb{R}^r \to \mathbb{R}$ is in the class $\mathcal{H}_r(\beta, M, C)$ if $M \geq 0$, $1 \geq \beta > 0$ and it holds that*

$$\sup_{y \neq y'} \frac{|g(y) - g(y')|}{|y - y'|_1^\beta \cdot (1 + |y|_1^M + |y'|_1^M)} \leq C. \tag{3.2.1}$$

If $\beta = 1$, we say that $g \in \mathcal{L}_r(M, C)$.

Let us abbreviate $Z_{t,n} := (X_{t,n}, ..., X_{t-r+1,n})$ and $\tilde{Z}_t(u) := (\tilde{X}_t(u), ..., \tilde{X}_{t-r+1}(u))$. An immediate consequence of the existence of a continuously differentiable modification of $(\tilde{X}_t(u))_{u \in [0,1]}$ is an expansion of the corresponding mean $\mathbb{E}g(Z_{t,n})$ and the corresponding stationary version $\mathbb{E}g(\tilde{Z}_t(u))$:

Proposition 3.2.2. *Assume that $g \in \mathcal{L}_r(M, C)$. Suppose that Assumption 3.1.2(i),(iv) are fulfilled for some $1 \geq \alpha > 0$ and $q = M+1$. Then we have uniformly for $t = 1, ..., n$:*

$$\mathbb{E}g(Z_{t,n}) = \mathbb{E}g\big(\tilde{Z}_t\big(\frac{t}{n}\big)\big) + O(n^{-\alpha}) = \mathbb{E}g(\tilde{Z}_t(u)) + O\big(n^{-\alpha} + \big|\frac{t}{n} - u\big|^\alpha\big). \tag{3.2.2}$$

If additionally Assumption 3.1.2(ii),(iii) are fulfilled and g is continuously differentiable, then $\mu(g, \cdot)$ is continuously differentiable with derivative

$$\partial_u \mathbb{E}g(\tilde{Z}_t(u)) = \sum_{j=1}^r \mathbb{E}[\partial_j g(\tilde{X}_t(u), ..., \tilde{X}_{t-r+1}(u)) \cdot \partial_u \tilde{X}_{t-j+1}(u)]. \tag{3.2.3}$$

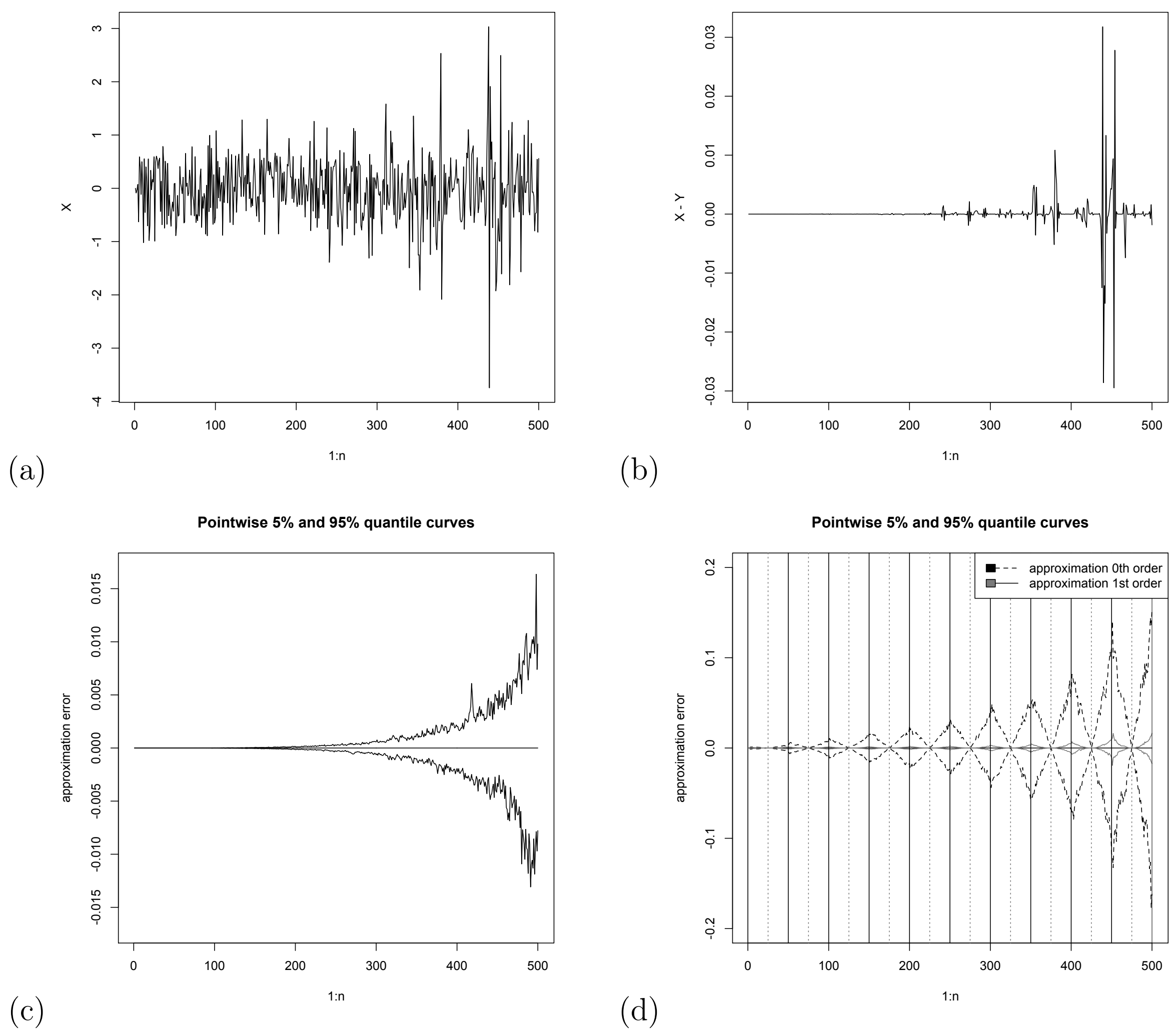

Figure 3.1: Top: (a) Realization of one $X_{t,n}$, $t = 1, ..., n$. (b) Difference $X_{t,n} - \tilde{X}_t(\frac{t}{n})$ for one realization. Bottom: (c) empirical 5%- and 95%-quantile curves of $X_{t,n} - \tilde{X}_t(\frac{t}{n})$ for $N = 1000$ replications. (d) Dashed and Solid: empirical 5%- and 95%-quantile curves of $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i)$ and $\tilde{X}_t(\frac{t}{n}) - \tilde{X}_t(u_i) - (\frac{t}{n} - u_i)\partial_u \tilde{X}_t(u_i)$ for $t \in (u_i - b, u_i + b]$ (black vertical lines) and $N = 1000$ replications, respectively. Here, $b = 50$ and $u_i = (2i-1)b$ (dotted vertical lines), $i = 1, ..., 5$.

The proof of (3.2.2) is immediate from the Hoelder-type property (3.2.1) of g and the results from Lemma 3.1.4. The second statement (3.2.3) follows from the expansion

$$g(\tilde{Z}_t(v)) = g(\tilde{Z}_t(u)) + (v-u)\partial_u g(\tilde{Z}_t(u)) + \int_u^v \left\{ \partial_u g(\tilde{Z}_t(s)) - \partial_u g(\tilde{Z}_t(u)) \right\} \mathrm{d}s$$

which holds almost surely since g is continuously differentiable and $\| \sup_u |\partial_u g(\tilde{Z}_t(u))| \|_1 < \infty$.

The result of Proposition 3.2.2 enables us to get expansions of the mean, the covariance and the distribution function of $X_{t,n}$. Suppose in the following that Assumption 3.1.2(ii),(iii) holds for $q = M+1$.

Corollary 3.2.3 (Mean expansion, $M = 0$)**.** *Choosing* $g : \mathbb{R} \to \mathbb{R}, g(y) = y$ *yields*

$$\mathbb{E}X_{t,n} = \mathbb{E}\tilde{X}_t(t/n) + O(n^{-1}),$$

where $\mu(u) := \mathbb{E}\tilde{X}_0(u)$ *is continuously differentiable with derivative* $\partial_u \mu(u) = \mathbb{E}\partial_u \tilde{X}_0(u)$.

Corollary 3.2.4 (Covariance expansion, $M = 1$)**.** *Define* $\gamma(u,r) := Cov(\tilde{X}_t(u), \tilde{X}_{t-r}(u))$. *Choosing* $g : \mathbb{R}^{r+1} \to \mathbb{R}, g(y) = y_1 y_{r+1}$ *and using the results from Corollary 3.2.3, we obtain uniformly for* $t = 1, ..., n$*:*

$$\gamma_{t,n}(r) := Cov(X_{t,n}, X_{t-r,n}) = \gamma(\frac{t}{n}, r) + O(n^{-1}) \tag{3.2.4}$$

and $\gamma(u,r)$ *is continuously differentiable with derivative*

$$\partial_u \gamma(u,r) = Cov(\partial_u \tilde{X}_0(u), \tilde{X}_r(u)) + Cov(\tilde{X}_0(u), \partial_u \tilde{X}_r(u)).$$

Similar expansions can be derived for higher-order cumulants. The expansion (3.2.4) is only valid for fixed $r \geq 0$. To give expansions of the Wigner-Ville spectrum (cf. Martin und Flandrin (1985)), one has to analyze the expression more carefully:

Corollary 3.2.5 (Expansion of the Wigner-Ville spectrum, $M = 1$)**.** *The function* $f_n(u,\lambda) := \sum_{r\in\mathbb{Z}} Cov(X_{\lfloor un-\frac{r}{2}\rfloor,n}, X_{\lfloor un+\frac{r}{2}\rfloor,n})e^{i\lambda r}$ *is called the Wigner-Ville spectrum of the process* $X_{t,n}$ *(here,* $\lfloor a\rfloor := \max\{k \in \mathbb{Z} : k \leq a\}$*). Define the time-varying spectral density* $f(u,\lambda) := \sum_{r\in\mathbb{Z}} \gamma(u,r)e^{i\lambda r}$ *(cf. Dahlhaus and Polonik (2009)). We have uniformly in* $u \in [0,1]$, $\lambda \in [0, 2\pi)$*:*

$$f_n(u,\lambda) = f(u,\lambda) + O(\log(n)^2 n^{-1}). \tag{3.2.5}$$

Furthermore, $u \mapsto f(u,\lambda)$ *is differentiable with derivative*

$$\partial_u f(u,\lambda) = \sum_{r\in\mathbb{Z}} \partial_u \gamma(u,r)e^{i\lambda r}.$$

Proof of Corollary 3.2.5: Let us use the abbreviation $t_r(u) := \lfloor un - \frac{r}{2}\rfloor$. Note that $\lfloor un + \frac{r}{2}\rfloor = t_r(u) + r$. We have for $r \geq 0$:

$$\mathrm{Cov}(X_{t_r(u),n}, X_{t_r(u)+r,n}) = \sum_{k=0}^{\infty} \mathbb{E}\big[P_{t_r(u)-k}X_{t_r(u),n}P_{t_r(u)-k}X_{t_r(u)+r,n}\big].$$

Define $\delta(r) := \max\{\sup_{n\in\mathbb{N}} \delta_2^{X_{\cdot,n}}(r), \sup_u \delta_2^{X(u)}(r)\} \leq C'\rho^r$ with some $C' > 0, 0 < \rho < 1$. By Lemma 3.1.4, we have with some constant C independent of r, n, t:

$$\begin{aligned}
& \big|\mathbb{E}\big[P_{t_r(u)-k}X_{t_r(u)+r,n}P_{t_r(u)-k}\big(X_{t_r(u),n} - \tilde{X}_{t_r(u)}(u)\big)\big]\big| \\
\leq\ & \big\|P_{t_r(u)-k}X_{t_r(u)+r,n}\big\|_2 \big\|P_{t_r(u)-k}\big(X_{t_r(u),n} - \tilde{X}_{t_r(u)}(u)\big)\big\|_2 \\
\leq\ & \delta(r+k)\min\{2\delta(k), C\Big(n^{-1} + \Big|\frac{t_r(u)}{n} - u\Big|\Big)\} \leq \delta(r+k)\min\{2\delta(k), C\frac{r+2}{2n}\},
\end{aligned}$$

and

$$\begin{aligned}
& \big|\mathbb{E}\big[P_{t_r(u)-k}\big(X_{t_r(u)+r,n} - \tilde{X}_{t_r(u)+r}(u)\big)P_{t_r(u)-k}\tilde{X}_{t_r(u)}(u)\big]\big| \\
\leq\ & \big\|P_{t_r(u)}\big(X_{t_r(u)+r,n} - \tilde{X}_{t_r(u)+r}(u)\big)\big\|_2 \big\|P_{t_r(u)-k}\tilde{X}_{t_r(u)}(u)\big\|_2 \\
\leq\ & \delta(k)\min\{2\delta(r+k), C\Big(n^{-1} + \Big|\frac{t_r(u)+r}{n} - u\Big|\Big)\} \leq \delta(k)\min\{2\delta(r+k), C\frac{r+2}{2n}\}.
\end{aligned}$$

The bounds for $r < 0$ are similar. Since $|e^{i\lambda r}| \leq 1$ and

$$\sum_{r\geq 0}\big(\rho^r \wedge \frac{r}{n}\big) = \sum_{r=0}^{\lfloor \log(n^{-1})/\log(\rho)\rfloor} \frac{r}{n} + \sum_{r=\lfloor \log(n^{-1})/\log(\rho)\rfloor+1}^{\infty} \rho^r = O(\log(n)^2 n^{-1} + n^{-1}),$$

we obtain (3.2.5). Note that

$$\begin{aligned}
|\mathrm{Cov}(\tilde{X}_0(u), \partial_u \tilde{X}_r(u))| &\leq \sum_{k=r}^{\infty} |\mathbb{E}[P_{r-k}\tilde{X}_0(u)\cdot P_{r-k}\partial_u\tilde{X}_r(u)]| \\
&\leq \sum_{k=r}^{\infty} \|P_{r-k}\tilde{X}_0(u)\|_2 \|P_{r-k}\partial_u\tilde{X}_r(u)\|_2 \\
&\leq \sum_{k=r}^{\infty} \delta_2^{\tilde{X}(u)}(r-k)\cdot\delta_2^{\partial_u\tilde{X}(u)}(k).
\end{aligned}$$

Similar arguments can be used to bound $\mathrm{Cov}(\tilde{X}_r(u), \partial_u\tilde{X}_0(u))$. By the results of Corollary 3.2.4, Proposition 3.1.3 and Theorem 3.1.8(i), we have $\sum_{r\in\mathbb{Z}}\sup_u|\partial_u\gamma(u,r)| < \infty$. This enables us to swap differentiation and summation leading to differentiability of $u \mapsto f(u,\lambda)$ with derivative

$$\partial_u f(u,\lambda) = \sum_{r\in\mathbb{Z}} \partial_u\gamma(u,r)e^{i\lambda r}.$$

$\square$

As a last application of Proposition 3.2.2, we present an expansion of the distribution function of $X_{t,n}$ which may also be used to approximate quantiles of such nonstationary processes.

Example 3.2.6 (Expansion of the distribution function, $M = 0$)**.** *Assume that ε has a Lipschitz continuous distribution and continuously differentiable function F_ε with Lipschitz constant L_ε and derivative f_ε. Assume that $(\varepsilon, y, u) \mapsto G_\varepsilon(y,u)$ is continuously differentiable and that the derivative $\partial_\varepsilon G_\varepsilon(y,u) \geq \delta_G > 0$ is uniformly bounded from below by some positive constant $\delta_G > 0$. This assumption guarantees that the variance of the innovation in a step of the recursion cannot be arbitrarily small. By the inverse function theorem we know that there exists a continuously differentiable inverse*

$x \mapsto H(x,y,u)$ of $\varepsilon \mapsto G_\varepsilon(y,u)$.
Finally, assume that for all $x \in \mathbb{R}$, the expressions

$$C(x) := \sup_{u\in[0,1]} \sup_{y\neq y'} \frac{|H(x,y,u) - H(x,y',u)|}{|y-y'|_1},$$

are finite. In this situation it holds that the distribution function of $X_{t,n}$,

$$F_{X_{t,n}}(x) = \mathbb{E}\big[\mathbb{P}(G_{\varepsilon_t}(Y_{t-1,n},t/n) \le x|\mathcal{F}_{t-1})\big] = \mathbb{E}\big[F_\varepsilon(H(x,Y_{t-1,n},t/n))\big]$$

can be approximated by the distribution function $F_{\tilde{X}_t(u)}(x) := \mathbb{P}(\tilde{X}_t(u) \le x)$ by

$$\begin{aligned} |F_{X_{t,n}}(x) - F_{\tilde{X}_t(t/n)}(x)| &\le L_\varepsilon \|H(x,Y_{t-1,n},t/n) - H(x,\tilde{Y}_{t-1}(t/n),t/n)\|_1 \\ &\le L_\varepsilon C_1(x) \sum_{j=1}^p \|X_{t-j-1,n} - \tilde{X}_{t-j-1}(t/n)\|_1 \le \frac{C}{n}\cdot L_\varepsilon \cdot C_1(x) \end{aligned}$$

with some constant C independent of x,t,n (cf. Lemma 3.1.4). Furthermore $u \mapsto F_{\tilde{X}_t(u)}(x)$ is differentiable with derivative

$$\begin{aligned} \partial_u F_{\tilde{X}_t(u)}(x) &= \mathbb{E}\Big[f_\varepsilon(H(x,\tilde{Y}_{t-1}(u),u)) \\ &\qquad \times\big(\langle \partial_2 H(x,\tilde{Y}_{t-1}(u),u), \partial_u \tilde{Y}_{t-1}(u)\rangle + \partial_3 H(x,\tilde{Y}_{t-1}(u),u)\big)\Big]. \end{aligned}$$

Another important application of the results from Section 3.1 is the expansion of functionals of $X_{t,n}$ in sums with a weighting kernel $K : [-\frac{1}{2},\frac{1}{2}] \to \mathbb{R}$ of bounded variation satisfying $\int K \,\mathrm{d}x = 1$. The results of the following Proposition can be used to obtain bias expansions for nonparametric estimators (see Section 3.3). Define $K_b(x) := \frac{1}{b}K(\frac{x}{b})$ with some bandwidth $b = b_n \to 0$ satisfying $nb \to \infty$.

Proposition 3.2.7 (Bias expansion)**.** *Assume that $g \in \mathcal{H}_r(\beta,M,C)$. If Assumption 3.1.2(i),(iv) are fulfilled with $q = M+1$ and some $1 \ge \alpha > 0$, we have uniformly in $u \in [0,1]$:*

$$\frac{1}{n}\sum_{t=1}^n K_b\big(\frac{t}{n}-u\big)\big(\mathbb{E}g(Z_{t,n}) - \mathbb{E}g(\tilde{Z}_t(t/n))\big) = O(n^{-\alpha\beta}), \tag{3.2.6}$$

and uniformly in $u \in [\frac{b}{2}, 1-\frac{b}{2}]$:

$$\frac{1}{n}\sum_{t=1}^n K_b\big(\frac{t}{n}-u\big)\mathbb{E}g(\tilde{Z}_t(t/n) - \mathbb{E}g(\tilde{Z}_0(u)) = O(b^{\alpha\beta}) + O\big((nb)^{-1}\big), \tag{3.2.7}$$

If additionally K is symmetric, $g \in \mathcal{L}_r(M,C)$ is continuously differentiable and Assumption 3.1.2(ii),(iii) holds with $q = M+1$, then (3.2.6) and (3.2.7) are valid with $\alpha = \beta = 1$ and we have uniformly in $u \in [\frac{b}{2}, 1-\frac{b}{2}]$

$$\frac{1}{n}\sum_{t=1}^n K_b\big(\frac{t}{n}-u\big)\mathbb{E}g(\tilde{Z}_t(t/n)) - \mathbb{E}g(\tilde{Z}_0(u)) = o(b) + O\big((nb)^{-1}\big), \tag{3.2.8}$$

The proof of (3.2.6) and (3.2.7) is immediate from Proposition 3.2.2 and the fact that K has bounded variation and $\int K \,\mathrm{d}x = 1$. To prove (3.2.8), note that

$$g(\tilde{Z}_t(t/n)) = g(\tilde{Z}_t(u)) + \big(\frac{t}{n} - u\big) \cdot \partial_u g(\tilde{Z}_t(u)) + \int_u^{t/n} \big\{\partial_u g(\tilde{Z}_t(s)) - \partial_u g(\tilde{Z}_t(u))\big\} \,\mathrm{d}s.$$

Furthermore, as long as $|\frac{t}{n} - u| \leq b$, we have

$$\big|\mathbb{E}\int_u^{t/n} \big\{\partial_u g(\tilde{Z}_t(s)) - \partial_u g(\tilde{Z}_t(u))\big\} \,\mathrm{d}s\big| \leq b \cdot \sup_{|u-s|\leq b} \|\partial_u g(\tilde{Z}_t(s)) - \partial_u g(\tilde{Z}_t(u))\|_1 = o(b)$$

since $u \mapsto \partial_u g(\tilde{Z}_t(u))$ is continuous and $\| \sup_u |\partial_u g(\tilde{Z}_t(u))|\|_1 < \infty$. Finally, because K has bounded variation and is symmetric,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^n K_b\big(\frac{t}{n} - u\big)\big(\mathbb{E}g(\tilde{Z}_t(t/n)) - \mathbb{E}g(\tilde{Z}_t(u))\big) \\
=\;& \mathbb{E}[\partial_u g(\tilde{Z}_t(u))] \cdot \frac{1}{n}\sum_{t=1}^n K_b\big(\frac{t}{n} - u\big) \cdot \big(\frac{t}{n} - u\big) + o(b) = O(n^{-1}) + o(b).
\end{aligned}$$

Remark 3.2.8. *Note that in the situation of Proposition 3.2.7, derivative processes were used to get $o(b)$ instead of $O(b)$ in (3.2.7). Even smaller rates can be obtained by using the results of Lemma 3.1.11 and/or higher order derivative processes together with higher order kernels.*
If we assume that $u \mapsto \tilde{X}_t(u)$ has a twice continuously differentiable modification and g is twice continuously differentiable, we obtain a bias decomposition whose structure is well-known from nonparametric statistics:

$$\begin{aligned}
\frac{1}{nb}\sum_{t=1}^n K\Big(\frac{t/n - u}{b}\Big)\mathbb{E}g(Z_{t,n}) - \mathbb{E}g(\tilde{Z}_t(u)) \;&=\; \int x^2 K(x)\ dx \cdot \mathbb{E}[\partial_u^2 g(\tilde{Z}_t(u))] \cdot b^2 \\
&\qquad + o(b^2) + O\big((nb)^{-1}\big).
\end{aligned}$$

3.2.2 A weak local and global law of large numbers

The smoothness of $X_{t,n}$ in the time direction can be used to obtain laws of large numbers by only assuming the existence of the first moment of $X_{t,n}$. The key step of the proof is to split the sum over $X_{t,n}$ into sums over smaller ranges of t where $X_{t,n}$ can be approximated by stationary processes. We will also provide results for localized sums. For this, we will assume that $K : [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}$ is a function of bounded variation with $\int K \,\mathrm{d}x = 1$.

Let us first cite a Lemma from Dahlhaus and Subba Rao (2006) (Lemma A.1 and A.2) which can be easily generalized to convergence in L^1:

Lemma 3.2.9. *Assume that (Y_t) is a stationary and ergodic process with $\mathbb{E}|Y_1| < \infty$. Let $b = b_n \to 0$ such that $nb_n \to \infty$. Then the following convergence holds in L^1:*

$$\frac{1}{nb}\sum_{t=1}^n K\Big(\frac{t/n - u}{b}\Big)Y_t \to \mathbb{E}Y_1.$$

Proposition 3.2.10 (Weak law - global and local version). *Suppose that Assumption 3.1.2(i),(iv) holds with some $q \geq 1$ and $1 \geq \alpha > 0$.*

(i) If $q \geq 1$, we have

$$\frac{1}{n}\sum_{t=1}^{n} X_{t,n} \to \int_0^1 \mathbb{E}\tilde{X}_0(u)\ du$$

in L^1 as $n \to \infty$. For each $u \in [0,1]$ it holds that

$$\frac{1}{nb}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) \cdot X_{t,n} \to \mathbb{E}\tilde{X}_0(u)$$

in L^1 as $n \to \infty$, $nb \to \infty$ and $b = b_n \to 0$.

(ii) If $q > 1$, then

$$\Big\| \sup_{u\in[0,1]} \Big|\frac{1}{nb}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) \cdot \big(X_{t,n} - \mathbb{E}X_{t,n}\big)\Big|\Big\|_q \leq \frac{B_K q}{(q-1)^2}\Delta_{q,0}^{X_{\cdot,n}} \cdot n^{1/q-1}b^{-1}.$$

and thus by Markov's inequality, for all $x > 0$:

$$\mathbb{P}\Big(\sup_{u\in[0,1]} \Big|\frac{1}{nb}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big)\cdot\big(X_{t,n}-\mathbb{E}X_{t,n}\big)\Big| > x\Big) \leq \frac{1}{x^q}\Big(\frac{B_K q}{(q-1)^2}\Delta_{0,q}^{X_{\cdot,n}}\Big)^q \cdot n^{1-q}b^{-q}.$$

If $q > 2$, then there exist constants C_1, C_2 not depending on n, b such that for all $x > 0$:

$$\begin{aligned}\mathbb{P}\Big(\sup_{u\in[0,1]} \Big|\frac{1}{nb}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) \cdot \big(X_{t,n} - \mathbb{E}X_{t,n}\big)\Big| > x\Big)&\\ \leq \frac{2C_1(B_K\Delta_{0,q}^{X_{\cdot,n}})^q n^{1-q}b^{-q}}{x^q} + 8G_{1-2/q}\Big(\frac{C_2 n^{1/2} b}{B_K\Delta_{0,q}^{X_{\cdot,n}}}\Big)&,\end{aligned}$$

with positive constants C_1, C_2 not depending on n, b and $G_q(y) := \sum_{j=1}^{\infty} e^{-j^q y^2}$ a Gaussian-like tail function.

Remark 3.2.11. *(i) For $q > 1$ and $b = o(n^{1-\frac{1}{q}})$, the results of Proposition 3.2.10(ii) and Proposition 3.2.7 can be used to obtain uniform convergence of the mean estimator $\hat{\mu}_b(u) := \frac{1}{nb}\sum_{t=1}^{n} K\big(\frac{t/n-u}{b}\big)X_{t,n}$ towards $\mu(u) := \mathbb{E}\tilde{X}_0(u)$ in the sense that*

$$\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]} |\hat{\mu}_b(u) - \mu(u)| \xrightarrow{\mathbb{P}} 0.$$

(ii) The results of Proposition 3.2.10(i) remain true if $X_{t,n}$ is replaced by $g(Z_{t,n})$ with some $g \in \mathcal{H}_r(\beta, M, C)$ and the assumptions are fulfilled for $q \geq M + \beta$. The

reason is that $g(\tilde{Z}_t(u))$ is still stationary and, by Hoelder's inequality applied to the conjugated pair $(\frac{M+\beta}{\beta}, \frac{M+\beta}{M})$,

$$\|g(Z_{t,n}) - g(\tilde{Z}_t(u))\|_1 \leq C \cdot \Big(\sum_{j=1}^{d} \|X_{t-j,n} - \tilde{X}_{t-j}(u)\|_{M+\beta} \Big)^{\beta}.$$

with some constant $C > 0$ independent of t, n, u. Thus the key steps of the proof of Proposition 3.2.10(i) carry over to this situation. Similar generalizations are possible for Proposition 3.2.10(ii) since $\sup_{n\in\mathbb{N}} \delta_q^{g(Z_{\cdot,n})}(k) \leq C\big(\sup_{n\in\mathbb{N}} \delta_{q(M+\beta)}^{X_{\cdot,n}}(k)\big)^{\beta}$ with some constant $C > 0$ independent of k, n.

3.2.3 A Central limit theorem

We provide local and global central limit theorems which may be useful in particular to find asymptotic distributions of (nonparametric) estimators of locally stationary processes, see section 3.3. It should be noted that the results of Theorem 3.2.12 can be generalized to functionals $g(X_{t,n})$ of $X_{t,n}$ since the proofs do not use the specific structure (3.1.1) of $X_{t,n}$.

Theorem 3.2.12 (Central limit theorem - global version)**.** *Suppose that Assumption 3.1.2(ii),(iv) holds with some $q \geq 2$. Define $S_n := \sum_{t=1}^{n}(X_{t,n} - \mathbb{E}X_{t,n})$.*

(i) If $q \geq 2$, then we have the following invariance principle:

$$\{S_{\lfloor nu \rfloor}/\sqrt{n}, 0 \leq u \leq 1\} \Rightarrow \Big\{ \int_0^u \sigma(v)\ dB(v), 0 \leq u \leq 1 \Big\},$$

where $B(v)$ is a standard-Brownian motion and the long-run variance $\sigma^2(v)$ is given by

$$\sigma^2(v) = \Big\| \sum_{i=0}^{\infty} P_0 \tilde{X}_i(v) \Big\|_2^2 = \sum_{k\in\mathbb{Z}} Cov(\tilde{X}_0(v), \tilde{X}_k(v)).$$

(ii) Strong approximation: If $q > 2$, then there exists a probability space $(\Omega_c, \mathcal{A}_c, \mathbb{P}_c)$ on which we can define random variables X_i^c with the partial sum process $S_n^c := \sum_{i=1}^{n} X_i^c$ and a Gaussian process $G_i^c = \sum_{j=1}^{i} Y_j^c$ with Y_j^c being independent Gaussian random variables with mean 0 such that $(S_i^c)_{i=1,\ldots,n} \overset{d}{=} (S_i)_{i=1,\ldots,n}$ and

$$\max_{1\leq i\leq n} |S_i^c - G_i^c| = o_p(n^{1/q}) \quad \text{in} \quad (\Omega_c, \mathcal{A}_c, \mathbb{P}_c).$$

Theorem 3.2.13 (Central limit theorem - local version)**.** *Assume that $g \in \mathcal{L}_r(M, C)$ for some $M \geq 0$. Suppose that Assumption 3.1.2(ii),(iv) hold with some $1 \geq \alpha > 0$, $q = 2(M+1)$. Then, provided that $\sqrt{nb}n^{-\alpha} \to 0$, $b \to 0$ and $nb \to \infty$:*

$$W_{n,b} \ := \ \frac{1}{\sqrt{nb}} \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) \cdot \Big(g(Z_{t,n}) - \mathbb{E}g(Z_{t,n})\Big) \xrightarrow{d} N(0, \sigma^2(u))$$

with $\sigma^2(u) := \big\| \sum_{l=0}^{\infty} P_0 g(\tilde{Z}_l(u))\big\|_2^2$*, and*
(i) We have the following bias decomposition uniformly in $u \in [\frac{b}{2}, 1-\frac{b}{2}]$*:*

$$\frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big)\mathbb{E}g(Z_{t,n}) - \mathbb{E}g(\tilde{Z}_0(u)) = O(\sqrt{nb^{1+2\alpha}}) + O\big((nb)^{-1/2}\big).$$

(ii) If additionally, g is continuously differentiable and Assumption 3.1.2(iii) is fulfilled with $q = M+1$*, then uniformly in* $u \in [\frac{b}{2}, 1-\frac{b}{2}]$*,*

$$\frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big)\mathbb{E}g(Z_{t,n}) - \mathbb{E}g(\tilde{Z}_0(u)) = o(\sqrt{nb^3}) + O\big((nb)^{-1/2}\big).$$

3.3 Application to Maximum Likelihood estimation

Many recursively defined locally stationary processes $X_{t,n}$ in (3.1.1) are obtained by replacing the constant parameters $\theta \in \Theta \subset \mathbb{R}^d$ of a recursively defined stationary model

$$X_t(\theta) = \tilde{G}_{\varepsilon_t}\big(X_{t-1}(\theta), ..., X_{t-p}(\theta), \theta\big), \quad t = 1, ..., n$$

by time dependent parameter curves $\theta_0 : [0,1] \to \Theta$ evaluated at the rescaled time $\frac{t}{n}$, see Example 3.1.1. In this section, let us assume that $G_\varepsilon(y,u) := \tilde{G}_\varepsilon(y, \theta_0(u))$, so $X_{t,n}$ obeys the recursion

$$X_{t,n} = \tilde{G}_{\varepsilon_t}\big(X_{t-1,n}, ..., X_{t-p,n}, \theta_0(\frac{t}{n})\big), \quad t = 1, ..., n.$$

Note that there is a strong connection between the stationary approximation $\tilde{X}_t(u)$ of $X_{t,n}$ and the original stationary process due to $\tilde{X}_t(u) = X_t(\theta_0(u))$. Our goal is to obtain estimators for $\theta_0(\cdot)$ based on $X_{t,n}$, $t = 1, ..., n$ with a quasi maximum likelihood approach.

Suppose for the moment that $\varepsilon \mapsto G_\varepsilon(y,\theta)$ is continuously differentiable for all ε, y, u and that the derivative $\partial_\varepsilon \tilde{G}_\varepsilon(y,\theta) \geq \delta_G > 0$ is bounded uniformly from below with some constant $\delta_G > 0$. This ensures that the new innovation ε_t has an impact on the value of $X_{t,n}$ which is not too small. Under these conditions, there exists a continuously differentiable inverse $x \mapsto H(x,y,\theta)$ of $\varepsilon \mapsto G_\varepsilon(y,\theta)$ (see also Example 3.2.6).
Suppose that ε_0 has a continuous density f_ε. The negative conditional log likelihood of $X_t(\theta) = x$ given $(X_{t-1}(\theta), ..., X_{t-p}(\theta)) = y$ is then

$$\ell(x,y,\theta) = -\log f_\varepsilon(H(x,y,\theta)) - \log \partial_x H(x,y,\theta). \qquad (3.3.1)$$

In the following derivations, we do not make use of the specific structure of ℓ. This means especially that we allow for model misspecifications due to a false density f_ε. Many authors prefer the case of a Gaussian density $f_\varepsilon(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ because then a minimizer θ of ℓ can be interpreted as a minimum (quadratic) distance estimator (see Dahlhaus and Giraitis (1998) in the tvAR case, Dahlhaus and Subba Rao (2006) in the tvARCH case).

Based on this we define $\ell_{t,n}(\theta) := \ell(X_{t,n}, Y_{t-1,n}, \theta)$. Given a bandwidth $b \in (0,1)$ and kernel function $K : [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}$ with $\int K(x)\ \mathrm{d}x = 1$, $K_b(x) := \frac{1}{b}K(\frac{x}{b})$ we define the local log conditional likelihood

$$L_{n,b}(u,\theta) := \frac{1}{n}\sum_{t=p+1}^{n} K_b\Big(\frac{t}{n} - u\Big) \cdot \ell_{t,n}(\theta).$$

For $u \in [0,1]$, the estimator of $\theta_0(u)$ is defined via

$$\hat{\theta}_b(u) := \arg\min_{\theta\in\Theta} L_{n,b}(u,\theta). \tag{3.3.2}$$

We will now discuss conditions such that $\hat{\theta}_b(\cdot)$ is consistent and asymptotically normal. A convenient way to formulate these results is to make a structural assumption on ℓ: We suppose that ℓ is Hoelder continuous in its first two components with at most polynomially increasing Hoelder constant. To make this more precise, we introduce the following class of functions:

Definition 3.3.1 (The class $\tilde{\mathcal{H}}_p(\beta, M, C)$)**.** *We say that a function $g : \mathbb{R}^{p+1} \times \Theta \to \mathbb{R}$ is in the class $\tilde{\mathcal{H}}_p(\beta, M, C)$ with $C = (C_z, C_\theta)$ and constants $C_z, C_\theta \geq 0$ and $M \geq 0$, $1 \geq \beta > 0$ if for all $z \in \mathbb{R}^{p+1}, \theta \in \Theta$ it holds that $g(\cdot, \theta) \in \mathcal{H}_{p+1}(\beta, M, C_z)$ and $g(z, \cdot) \in \mathcal{H}_d\big(1, 0, C_\theta(1 + |z|_1^{M+\beta})\big)$.*

It turns out in Theorem 3.3.2 that the (pointwise) consistency of $\hat{\theta}_b$ can be obtained by posing conditions on the likelihood of the corresponding stationary process which is defined via $L(u,\theta) := \mathbb{E}[\tilde{\ell}_t(u,\theta)]$ with $\tilde{\ell}_t(u,\theta) := \ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta)$. Especially if ℓ is taken to be of the form (3.3.1) with f_ε the standard Gaussian density, the properties of $L(u,\theta)$ are usually well-known from the maximum likelihood theory of the stationary process $X_t(\theta)$ and therefore are easy to verify (see also Example 3.3.5).

Theorem 3.3.2 (Pointwise and uniform consistency of $\hat{\theta}_b$)**.** *Assume that $\ell \in \tilde{\mathcal{H}}_p(\beta, M, C)$ for some $M \geq 0$, $1 \geq \beta > 0$. Suppose that Assumption 3.1.2(i),(iv) holds with some $1 \geq \alpha > 0$ and $q = M + \beta$.*
Furthermore suppose that for all $u \in [0,1]$, $\theta_0(u) \in int(\Theta)$ is the unique minimizer of $L(u,\theta)$ over $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^d$ is a compact set. Then:

(i) For all $u \in (0,1)$ with $b \to 0$ and $bn \to \infty$:

$$\hat{\theta}_b(u) \xrightarrow{\mathbb{P}} \theta_0(u).$$

(ii) If additionally $q > M + \beta$ and $b = o(n^{1-\frac{M+\beta}{q}})$ and $\theta_0(\cdot)$ is continuous, we have

$$\sup_{u\in[\frac{b}{2}, 1-\frac{b}{2}]} |\hat{\theta}_b(u) - \theta_0(u)| \xrightarrow{\mathbb{P}} 0.$$

Remark 3.3.3. *Note that in nearly all cases, the conditions of Assumption 3.1.2(iv) assumed in Theorem 3.3.2 implicitly impose a Hoelder continuity condition on* $\theta_0(\cdot)$.

Proof of Theorem 3.3.2. (i) For fixed $u \in [0,1]$ and $\theta \in \Theta$, note that $\ell(\cdot,\cdot,\theta) \in \mathcal{H}_{p+1}(\beta, M, C_z)$. Application of Proposition 3.2.10(i) (see also Remark 3.2.11(ii)) leads to

$$L_{n,b}(u,\theta) = \frac{1}{n}\sum_{t=1}^{n} K_b\Big(\frac{t}{n} - u\Big) \cdot \ell(X_{t,n}, Y_{t-1,n}, \theta) \xrightarrow{\mathbb{P}} \mathbb{E}\ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta) = L(u,\theta).$$

Define $m := M + \beta$, $m' := m \wedge 1$. The function $\theta \mapsto L(u,\theta)$ is continuous since

$$\begin{aligned} |L(u,\theta) - L(u,\theta')| &\leq \|\ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta) - \ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta')\|_1 \\ &\leq C_\theta \cdot |\theta - \theta'|_1 \cdot \big(1 + \Big(\sum_{j=0}^{p} \|\tilde{X}_t(u)\|_m^{m'}\Big)^{\frac{m}{m'}}\big). \end{aligned}$$

It remains to show stochastic equicontinuity of $L_{n,h}(u,\theta)$: Define $h : \mathbb{R}^{p+1} \to \mathbb{R}$, $h(z) = C_\theta(1 + |z|_1^m)$. Fix $\eta > 0$. We have

$$|L_{n,b}(u,\theta) - L_{n,b}(u,\theta')| \leq |\theta - \theta'|_1 \cdot \frac{1}{n}\sum_{t=1}^{n} \Big|K_b\Big(\frac{t}{n} - u\Big)\Big| \cdot h(X_{t,n}, Y_{t-1,n}).$$

Obviously, $h \in \mathcal{H}_{p+1}(m', (m \vee 1) - 1, C)$ with some constant $C > 0$. Application of Proposition 3.2.10(i) to $K/\int K \, \mathrm{d}x$ and h (see also Remark 3.2.11(ii)) yields for all $u \in (0,1)$:

$$\frac{1}{n}\sum_{t=1}^{n} \Big|K_b\Big(\frac{t}{n} - u\Big)\Big| \cdot h(X_{t,n}, Y_{t-1,n}) \xrightarrow{\mathbb{P}} \int |K| \, \mathrm{d}x \cdot \mathbb{E}h(\tilde{X}_t(u), \tilde{Y}_{t-1}(u)) =: c(u). \quad (3.3.3)$$

Choosing $\delta = \frac{\eta}{2c(u)}$ yields

$$\begin{aligned} &\mathbb{P}\Big(\sup_{|\theta-\theta'|_1 \leq \delta} |L_{n,b}(u,\theta) - L_{n,b}(u,\theta')| > \eta\Big) \\ \leq\ &\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{t=1}^{n} \Big|K_b\Big(\frac{t}{n} - u\Big)\Big| \cdot h(X_{t,n}, Y_{t-1,n}) - c(u)\Big| > c(u)\Big) \to 0 \quad (n \to \infty). \end{aligned}$$

This gives $\sup_{\theta\in\Theta} |L_{n,b}(u,\theta) - L(u,\theta)| \xrightarrow{\mathbb{P}} 0$. By standard arguments (cf. Van der Vaart (2009), Theorem 5.7), the proof is complete.
To prove (ii), we apply Proposition 3.2.10(ii) with $\tilde{q} = \frac{q}{M+\beta} > 1$ (see also Remark 3.2.11(ii)) to obtain for each $\theta \in \Theta$ that

$$\sup_{u\in[0,1]} \big|L_{n,b}(u,\theta) - \mathbb{E}L_{n,b}(u,\theta)\big| = O_p(n^{\frac{M+\beta}{q}-1} b^{-1}).$$

By Proposition 3.2.7 we have $\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}|\mathbb{E}L_{n,b}(u,\theta)-L(u,\theta)| = O(b^{\alpha\beta})+O((nb)^{-1})$, which yields

$$\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}\big|L_{n,b}(u,\theta)-L(u,\theta)\big| \xrightarrow{\mathbb{P}} 0.$$

Similarly we can strengthen (3.3.3) to

$$\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}\Big|\frac{1}{n}\sum_{t=1}^{n}\Big|K_b\Big(\frac{t}{n}-u\Big)\Big|\cdot h(X_{t,n},Y_{t-1,n})-c(u)\Big| \xrightarrow{\mathbb{P}} 0.$$

Now define $c := \inf_u c(u)$. Choosing $\delta = \frac{\eta}{2c}$ yields

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}\sup_{|\theta-\theta'|_1\le\delta}|L_{n,b}(u,\theta)-L_{n,b}(u,\theta')|>\eta\Big)\\
\le\ &\mathbb{P}\Big(\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}\Big|\frac{1}{n}\sum_{t=1}^{n}\Big|K_b\Big(\frac{t}{n}-u\Big)\Big|\cdot h(X_{t,n},Y_{t-1,n})-c(u)\Big|>c\Big)\to 0 \quad (n\to\infty).
\end{aligned}$$

So we have seen that $\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}\sup_{\theta\in\Theta}|L_{n,b}(u,\theta)-L(u,\theta)| \xrightarrow{\mathbb{P}} 0$. Standard arguments give the result (see also the appendix). $\square$

We now provide a central limit theorem for $\hat{\theta}_b$ including a bias decomposition. Let ∇ denote the derivative with respect to θ.

Theorem 3.3.4 (A central limit theorem for $\hat{\theta}_b$). *Additionally to Theorem 3.3.2, suppose that*

- $\nabla\ell \in \tilde{\mathcal{H}}_p(1,M',C')$ *for some* $M'\ge 0$, $\nabla^2\ell\in\tilde{\mathcal{H}}_p(\beta'',M'',C'')$ *for some* $M''\ge 0$, $1\ge\beta''>0$,
- *Assumption 3.1.2(i),(iv) is fulfilled with* $q=\max\{2(M'+1),M''+\beta''\}$ *and some* $1\ge\alpha'>0$, *Assumption 3.1.2(ii) is fulfilled with* $q=2(M'+1)$.

Assume that the model is correct in the weak sense that $\mathbb{E}[\nabla\tilde{\ell}(u,\theta_0(u))|\mathcal{F}_{t-1}]=0$, *i.e.* $\nabla\tilde{\ell}_t(u,\theta_0(u))$ *is a martingale difference sequence with respect to* $(\mathcal{F}_t)$. *Then we have for* $b\to 0$, $nb\to\infty$ *and* $nb^{1+2\alpha'}=o(1)$*:*

$$\sqrt{nb}(\hat{\theta}_b(u)-\theta_0(u)) \xrightarrow{d} N\big(0,V(u)^{-1}I(u)V(u)^{-1}\big), \tag{3.3.4}$$

where $I(u) := \mathbb{E}[\nabla\tilde{\ell}_t(u,\theta_0(u))\nabla\tilde{\ell}_t(u,\theta_0(u))']$ *and* $V(u) := \nabla^2 L(u,\theta_0(u))$ *is assumed to be positive definite.*
If additionally $\nabla\ell$ *is continuously differentiable,* K *is symmetric and Assumption 3.1.2 (iii) is fulfilled for* $q=M'+1$, *the result (3.3.4) remains true if* $nb^3=O(1)$.

Proof of Theorem 3.3.4: The conditions on $\nabla^2\ell$ imply that $u \mapsto \nabla^2 L(u,\theta) = \mathbb{E}[\nabla^2\tilde{\ell}_t(u,\theta)]$ is continuous. Note that by Theorem 3.2.13, we have

$$\begin{aligned}
&\sqrt{nb}\nabla L_{n,b}(u,\theta_0(u)) \\
= \;& \frac{1}{\sqrt{nb}}\sum_{t=p+1}^{n} K\Big(\frac{t/n-u}{b}\Big)\Big(\nabla\ell(X_{t,n},Y_{t-1,n},\theta_0(u)) - \mathbb{E}\nabla\ell(X_{t,n},Y_{t-1,n},\theta_0(u))\Big) \\
\xrightarrow{d} \;& N(0,\sigma^2(u)),
\end{aligned}$$

where $\sigma^2(u) = \Big\| \sum_{l=0}^{\infty} P_0\nabla\tilde{\ell}_t(u,\theta_0(u))\Big\|_2^2 = I(u)$ by the martingale difference property. Furthermore Theorem 3.2.13 gives that

$$\begin{aligned}
&\frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big)\mathbb{E}\nabla\ell(X_{t,n},Y_{t-1,n},\theta_0(u)) \\
= \;& \frac{1}{\sqrt{nb}}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big)\big(\mathbb{E}\nabla\ell(X_{t,n},Y_{t-1,n},\theta_0(u)) - \mathbb{E}\nabla\ell(\tilde{X}_t(u),\tilde{Y}_{t-1}(u),\theta_0(u))\big)
\end{aligned}$$

is $O(\sqrt{nb^{1+2\alpha'}}) + O((nb)^{-1/2})$ or $o(\sqrt{nb^3}) + O((nb)^{-1/2})$ dependent on the assumptions. Since $\nabla^2\ell$ fulfills the same assumptions as ℓ in Theorem 3.3.2, we can mimic its proof and obtain

$$\sup_{\theta\in\Theta} |\nabla^2 L_{n,b}(u,\theta) - \nabla^2 L(u,\theta)| \xrightarrow{\mathbb{P}} 0.$$

By continuity of $\theta \mapsto \nabla^2 L(u,\theta)$, we obtain for each sequence $\tilde{\theta}_n \xrightarrow{\mathbb{P}} \theta_0(u)$ that

$$\begin{aligned}
&|\nabla^2 L_{n,b}(u,\tilde{\theta}_n) - \nabla^2 L(u,\theta_0(u))| \\
\leq \;& |\nabla^2 L_{n,b}(u,\tilde{\theta}_n) - \nabla^2 L(u,\tilde{\theta}_n)| + |\nabla^2 L(u,\tilde{\theta}_n) - \nabla^2 L(u,\theta_0(u))| \xrightarrow{\mathbb{P}} 0.
\end{aligned}$$

Standard arguments now give the result. □

An important special case is the case of Gaussian conditional likelihoods combined with autoregressive models. Specific examples for these are given in Example 3.1.1.

Example 3.3.5 (Autoregressive models)**.** *In this example we discuss the model $\tilde{G}_\varepsilon(y,\theta) = \mu(y,\theta) + \sigma(y,\theta)\varepsilon$, where $\mu,\sigma : \mathbb{R}^p \times \Theta \to \mathbb{R}$ satisfy*

$$\sup_\theta \sup_{y\neq y'} \frac{|\mu(y,\theta)-\mu(y',\theta)|}{|y-y'|_{\chi,1}} + \sup_\theta \sup_{y\neq y'} \frac{|\sigma(y,\theta)-\sigma(y',\theta)|}{|y-y'|_{\chi,1}}\|\varepsilon_0\|_2 \leq 1 \tag{3.3.5}$$

with some $\chi \in \mathbb{R}^p_{\geq 0}$ with $|\chi|_1 < 1$. Assume that $\mathbb{E}\varepsilon_0 = 0$ and $\mathbb{E}\varepsilon_0^2 = 1$ and that $\theta_0(\cdot) \in \Sigma(\alpha, L)$, i.e. θ_0 is Hoelder-continuous with exponent α. Then Assumption 3.1.2(ii) is fulfilled with $q = 2$.

If we choose f_ε to be the standard Gaussian density, we obtain from (3.3.1):

$$\ell(x,y,\theta) = \frac{1}{2}\Big(\frac{x-\mu(y,\theta)}{\sigma(y,\theta)}\Big)^2 - \frac{1}{2}\log\sigma^2(y,\theta) + const. \tag{3.3.6}$$

Furthermore assume that

$$\sup_y \sup_{\theta\neq\theta'} \frac{|\mu(y,\theta)-\mu(y,\theta')|}{|\theta-\theta'|_1\cdot(1+|y|_1)} < \infty, \qquad \sup_y \sup_{\theta\neq\theta'} \frac{|\sigma(y,\theta)-\sigma(y,\theta')|}{|\theta-\theta'|_1\cdot(1+|y|_1)} < \infty. \tag{3.3.7}$$

Let $\sigma(\cdot)\geq\delta_\sigma$ *be uniformly bounded from below with some* $\delta_\sigma>0$*. Then* $\ell\in\tilde{\mathcal{H}}_p(1,1,C)$ *with some* $C>0$*, and Assumption 3.1.2(i),(iv) is fulfilled with* $q=2$ *and* α *from above.*

Fix $u\in[0,1]$*. Suppose that*

$$\mu(\tilde{Y}_{t-1}(u),\theta)=\mu(\tilde{Y}_{t-1}(u),\theta_0(u)) \quad \text{and} \quad \sigma(\tilde{Y}_{t-1}(u),\theta)=\sigma(\tilde{Y}_{t-1}(u),\theta_0(u)) \quad a.s.$$

implies $\theta=\theta_0(u)$*. Then* $\theta\mapsto L(u,\theta)$ *has a unique minimum in* $\theta=\theta_0(u)$ *since* $\log(x)\leq x-1$ *if and only if* $x=1$ *and* $x^2\geq 0$ *if and only if* $x=0$ *and, omitting the argument* $\tilde{Y}_{t-1}(u)$*,*

$$2\big(L(u,\theta)-L(u,\theta_0(u))\big)=\mathbb{E}\Big(\frac{\mu(\theta)-\mu(\theta_0(u))}{\sigma(\theta)}\Big)^2+\mathbb{E}\Big[\log\frac{\sigma(\theta)^2}{\sigma(\theta_0(u))^2}-1+\frac{\sigma(\theta_0(u))^2}{\sigma(\theta)^2}\Big]\geq 0.$$

If additionally Θ *is compact and* $\theta_0(u)\in int(\Theta)$*, the assumptions of Theorem 3.3.2 are fulfilled and we obtain for* $\hat{\theta}_b$ *defined by (3.3.2):*

$$\hat{\theta}_b(u)\xrightarrow{\mathbb{P}}\theta_0(u).$$

We now will show asymptotic normality of $\hat{\theta}_b$*. To keep the presentation simple, we will assume* $\sigma(\cdot,\cdot)\equiv 1$*,* $\mathbb{E}\varepsilon_0^4<\infty$ *and replace* $\mathbb{E}\varepsilon_0^2=1$ *by* $\mathbb{E}\varepsilon_0^2=\sigma_0^2>0$*. Note that Assumption (3.1.2)(ii) is fulfilled with* $q=4$*. Then, omitting the arguments of* μ*, we have*

$$\nabla\ell=-(x-\mu)\nabla\mu, \qquad \nabla^2\ell=\nabla\mu\cdot\nabla\mu'-(x-\mu)\nabla^2\mu.$$

Then $\mathbb{E}[\nabla\ell(\tilde{X}_t(u),\tilde{Y}_{t-1}(u),\theta_0(u))|\mathcal{F}_{t-1}]=0$ *and* $I(u)=\mathbb{E}[\nabla\ell\cdot\nabla\ell']=\sigma_0^2\mathbb{E}[\nabla\mu\cdot\nabla\mu']=\sigma_0^2V(u)$ *with* $V(u):=\nabla^2L(u,\theta_0(u))$*. If additionally*

$$\sup_\theta \sup_{y\neq y'} \frac{|\nabla\mu(y,\theta)-\nabla\mu(y',\theta)|_1}{|y-y'|_1} < \infty, \qquad \sup_y \sup_{\theta\neq\theta'} \frac{|\nabla\mu(y,\theta)-\nabla\mu(y,\theta')|_1}{|\theta-\theta'|_1(1+|y|_1)} < \infty \tag{3.3.8}$$

and similar assumptions are fulfilled for $\nabla^2\mu$*, then we have* $\nabla\ell,\nabla^2\ell\in\tilde{\mathcal{H}}_p(1,1,C')$ *with some* $C'>0$*. This shows that all conditions of the first part of Theorem 3.3.4 are fulfilled and we obtain for* $b\to 0$*,* $nb\to\infty$ *and* $nb^3=o(1)$*:*

$$\sqrt{nb}\big(\hat{\theta}_b(u)-\theta_0(u)\big)\xrightarrow{d}N\big(0,\sigma_0^2\cdot V(u)^{-1}\big). \tag{3.3.9}$$

If additionally, $\mu,\nabla\mu$ *and* θ_0 *are continuously differentiable and,*

$$\sup_\theta \sup_{y\neq y'} \frac{|\partial_i\mu(y,\theta)-\partial_i\mu(y',\theta)|_1}{|y-y'|_1} < \infty, \quad (i=1,2), \tag{3.3.10}$$

then $\nabla\ell$ *is continuously differentiable and Assumption 3.1.2(iii) is fulfilled with* $q=2$*. So all conditions of the second part of Theorem 3.3.4 are fulfilled and we obtain (3.3.9) even if* $nb^3=O(1)$*.*

We close this section by using the results of Example 3.3.5 in a more specific example of the tvExpAR(1) process which is a locally stationary version of the ExpAR(1) process discussed in Jones (1978). Up to now, there is no asymptotic theory available for the parameter estimator in this model; we show that our theory immediately provides consistency and asymptotic normality of the corresponding maximum likelihood estimator.

Example 3.3.6 (Maximum likelihood estimation in the tvExpAR(1) process)**.** *Assume that there exists $\theta_0 : [0,1] \to \Theta$ (where the image of θ_0 is in the interior of Θ) with $\Theta := \{\theta \in \mathbb{R} : 0 \le \theta \le \rho\}$ and some fixed $\rho > 0$, $0 < |a_0| < 1$ such that*

$$X_{t,n} = a_0 \exp\Big(-\theta_0\big(\frac{t}{n}\big) X_{t-1,n}^2 \Big) X_{t-1,n} + \varepsilon_t, \quad t = 1, ..., n.$$

Assume that $\mathbb{E}\varepsilon_0 = 1$, $\mathbb{E}\varepsilon_0^2 = \sigma_0^2 > 0$ and $\mathbb{E}\varepsilon_0^4 < \infty$. It is easily seen that this model fulfills the smoothness assumptions (3.3.5), (3.3.7), (3.3.8) and (3.3.10) with $\mu(y,\theta) := a_0 \exp(-\theta y^2) y$ and $\sigma(\cdot,\cdot) \equiv 1$. Let $\tilde{X}_t(u)$ denote the corresponding stationary approximation of $X_{t,n}$. Identifiability of θ is obtained due to

$$\mathbb{E}[(\mu(\tilde{X}_t(u),\theta) - \mu(\tilde{X}_t(u),\theta'))^2] \ge a_0^2 \mathbb{E}[\exp(-2\rho \tilde{X}_0(u)^2)\tilde{X}_0(u)^6] \cdot |\theta - \theta'|^2,$$

since $\mathbb{E}[\exp(-2\rho\tilde{X}_t(u)^2)\tilde{X}_t(u)^6] = 0$ would imply $\tilde{X}_t(u) = 0$ a.s. which is a contradiction to $\mathbb{E}[\tilde{X}_t(u)^2] \ge \sigma_0^2$ which follows from the recursion of $\tilde{X}_t(u)$. Let $\hat{\theta}_b(u)$ defined by (3.3.2) based on the likelihood (3.3.6). We obtain for $b \to 0$, $bn \to \infty$:

$$\hat{\theta}_b(u) \xrightarrow{\mathbb{P}} \theta_0(u),$$

and for $nb^3 = O(1)$:

$$\sqrt{nb}\big(\hat{\theta}_b(u) - \theta_0(u)\big) \xrightarrow{d} N(0, \sigma_0^2 V(u)^{-1}),$$

where $V(u) = a_0^2 \mathbb{E}[\exp(-2\theta_0(u)\tilde{X}_0(u)^2)\tilde{X}_0(u)^6]$.

3.4 Concluding Remarks

In this chapter, we made a first step to derive a general asymptotic theory for nonstationary processes $X_{t,n}$. We introduced derivative processes which have shown to be a powerful tool to show mean expansions of functionals of $X_{t,n}$. We could see in Figure 3.1 that the pointwise approximation of $X_{t,n}$ by the Taylor expansion of $\tilde{X}_t(t/n)$ around some time point $u \in [0,1]$ with derivative processes has very low variance as long as $|t/n - u| \ll 1$, $n^{-1} \ll 1$ and the dependence of the process is small. This also motivates to use these expansions in other fields of statistics which are well-studied for stationary processes.

We formulated laws of large numbers and central limit theorems for such processes under minimal moment assumptions by using the smoothness of the approximating stationary process. We applied the results to nonparametric maximum likelihood estimation and formulated easy verifiable conditions which are applicable to a wide range of well-known locally stationary processes.

3.5 Lemmas and Proofs

3.5.1 Proofs of section 3.1

Here, we prove the results from section 3.1. The following lemma from Duflo (1997), Lemma 6.2.10 therein will be used frequently to verify the geometric decay of the difference of recursively defined processes:

Lemma 3.5.1. *Assume that $p > 0$ is a positive natural number, $\chi \in \mathbb{R}^p_{\geq 0}$ with $|\chi|_1 < 1$ and that there are sequences of real-valued nonnegative numbers $(z_s)_{s>-p}$, $(\mu_s)_{s>0}$ which fulfill for all $s = 1, 2, ...$:*

$$z_s \leq \sum_{i=1}^{p} \chi_i z_{s-i} + \mu_s. \tag{3.5.1}$$

Then there exist constants $\lambda_0 \in (0,1)$, $C_\lambda > 0$ only depending on χ, p such that for all $s = 1, 2, ...$:

$$z_s \leq C_\lambda \Big(\lambda_0^s \cdot |(z_0, ..., z_{-p+1})|_1 + \sum_{i=0}^{s-1} \lambda_0^i \mu_{s-i} \Big).$$

Sometimes we will apply the lemma for $s = 0, 1, 2, ...$ instead of $s = 1, 2, 3, ...$.

For the following proofs, recall that $Y_{t-1,n} := (X_{t-1,n}, ..., X_{t-p,n})$ and $\tilde{Y}_{t-1}(u) = (\tilde{X}_{t-1}(u), ..., \tilde{X}_{t-p}(u))$. For $y \in \mathbb{R}^p$, we will use the abbreviation $G_{\varepsilon,u}(y) := G_\varepsilon(y, u)$. Define the random map $R_{\varepsilon,u}(y) := (G_{\varepsilon,u}(y), y_1, ..., y_{p-1})$. Let $X_{n,u}(y)$ be the first element of the vector $H_{n,u}(y) := R_{\varepsilon_0,u} \circ R_{\varepsilon_{-1},u} \circ ... \circ R_{\varepsilon_{-n},u}(y)$, where $n = 0, 1, 2, ...$ For consistency of the following argumentations, define $X_{n,u}(y) := y_{-n}$ for $n = -1, ..., -p$. Note that $H_{n,u}(y)_j = X_{n-j+1,u}(y)$ (in distribution) for $j = 1, ..., p$. Let $J_{n,u}(y)$ be defined similarly to $H_{n,u}(y)$ but based on $\varepsilon_{-1}, ..., \varepsilon_{-n-1}$ instead of $\varepsilon_0, ..., \varepsilon_{-n}$. Note that $X_{n,u}(y) = G_{\varepsilon_0,u}(J_{n-1,u}(y))$ and that $J_{n-1,u}(y) = H_{n-1,u}(y) = (X_{n-1,u}(y), ..., X_{n-p,u}(y))'$ holds in distribution.

Proof of Proposition 3.1.3. (i) Note that $(|a| + |b|)^{q'} \leq |a|^{q'} + |b|^{q'}$ since $0 < q' \leq 1$. By (3.1.3), we obtain

$$\begin{aligned}
& \|X_{n,u}(y) - X_{n,u}(y')\|_q^{q'} \\
\leq\ & \|G_{\varepsilon_0,u}(J_{n-1,u}(y)) - G_{\varepsilon_0,u}(J_{n-1,u}(y'))\|_q^{q'} \\
\leq\ & \mathbb{E}\big[\mathbb{E}\big[|G_{\varepsilon_0,u}(J_{n-1,u}(y)) - G_{\varepsilon_0,u}(J_{n-1,u}(y'))|^q \big| \mathcal{F}_{-1}\big]\big]^{q'/q} \\
\leq\ & \mathbb{E}\big[|J_{n-1,u}(y) - J_{n-1,u}(y')|_{\chi,q'}^q\big]^{q'/q} \\
\leq\ & \mathbb{E}\Big[\Big(\sum_{j=1}^{p} \chi_j |X_{n-j,u}(y) - X_{n-j,u}(y')|^{q'}\Big)^{q/q'}\Big]^{q'/q} \\
=\ & \Big\| \sum_{j=1}^{p} \chi_j |X_{n-j,u}(y) - X_{n-j,u}(y')|^{q'} \Big\|_{q/q'} \\
\leq\ & \sum_{j=1}^{p} \chi_j \big\| |X_{n-j,u}(y) - X_{n-j,u}(y')|^{q'} \big\|_{q/q'} = \sum_{j=1}^{p} \chi_j \big\| X_{n-j,u}(y) - X_{n-j,u}(y') \big\|_q^{q'}.
\end{aligned}$$

By Lemma 3.5.1, we have with some $C_\lambda > 0, \lambda_0 \in (0,1)$ independent of $u \in [0,1]$ that for all $n \in \mathbb{N}$:

$$\|X_{n,u}(y) - X_{n,u}(y')\|_q^{q'} \le C_\lambda \lambda_0^{n+1} \cdot |y - y'|_1^{q'}. \tag{3.5.2}$$

Applying (3.5.2) to $y = y_0$ and $y' = R_{\varepsilon_{-n-1},u}(y_0)$, we obtain

$$\begin{aligned}\Big\| \sum_{n=0}^{\infty} |X_{n,u}(y_0) - X_{n+1,u}(y_0)| \Big\|_q^{q'} &\le \sum_{n=0}^{\infty} \|X_{n,u}(y_0) - X_{n+1,u}(y_0)\|_q^{q'} \\ &\le C_\lambda \sum_{n=0}^{\infty} \lambda_0^{n+1} \cdot \| |y_0 - R_{\varepsilon_{-n-1},u}(y_0)|_1 \|_q^{q'} < \infty.\end{aligned}$$

By the Markov inequality, this shows that $(X_{n,u}(y_0))_{n\in\mathbb{N}}$ is a Cauchy sequence a.s. and thus has an almost sure limit $\tilde{X}_0(u)$ (say). Furthermore, we have

$$\|X_{n,u}(y_0)\|_q^{q'} \le |y_0|_1^{q'} + \sum_{k=0}^{n-1} \|X_{k+1,u}(y_0) - X_{k,u}(y_0)\|_q^{q'} \le |y_0|_1^{q'} + \frac{C_\lambda \lambda_0}{1-\lambda_0} \| |y_0 - R_{\varepsilon_{-n-1},u}(y_0)|_1 \|_q^{q'}.$$

By Fatou's lemma,

$$\sup_{u\in[0,1]} \|\tilde{X}_0(u)\|_q^{q'} \le \sup_{u\in[0,1]} \liminf_{n\to\infty} \|X_{n,u}(y_0)\|_q^{q'} < \infty,$$

since $\sup_{u\in[0,1]} \|G_{\varepsilon_0}(y_0,u)\|_q < \infty$ by assumption.
Since $\tilde{X}_0(u)$ is $\mathcal{F}_0$-measurable, we can write $\tilde{X}_0(u) = H(u, \mathcal{F}_0)$ for some measurable function H. By (3.5.2), $X_{n,u}(y)$ converges almost surely to the same limit $\tilde{X}_0(u)$ for arbitrary $y \in \mathbb{R}^p$. This shows a.s. uniqueness and we can express $\tilde{X}_t(u) = H(u, \mathcal{F}_t)$. Because $\tilde{X}_t(u)$ obeys (3.1.1), we have for $X_t^{*0}(u) = H(u, \mathcal{F}_t^{*0})$ by (3.1.3):

$$\|\tilde{X}_t(u) - \tilde{X}_t^{*0}(u)\|_q^{q'} \le \sum_{j=1}^{p} \chi_j \|\tilde{X}_{t-j}(u) - \tilde{X}_{t-j}^{*0}(u)\|_q^{q'}$$

By Lemma 3.5.1, we conclude $\|\tilde{X}_t(u) - \tilde{X}_t^{*0}(u)\|_q^{q'} \le 2pC_\lambda \lambda_0^t \|\tilde{X}_0(u)\|_q^{q'}$.

(ii) Because $X_{0,n} = \tilde{X}_0(0)$ by means of (3.1.1), the existence and uniqueness statement is obvious from Proposition 3.1.3. From (3.1.3) and the triangle inequality, we obtain

$$\begin{aligned}\|X_{t,n}\|_q^{q'} &\le \sum_{j=1}^{p} \chi_j \|X_{t-j,n} - y_{0j}\|_q^{q'} + \big\|G_{\varepsilon_0}\big(y_0, \frac{t}{n}\big)\big\|_q^{q'} \\ &\le \sum_{j=1}^{p} \chi_j \|X_{t-j,n}\|_q^{q'} + |y_0|_1^{q'} + \sup_{u\in[0,1]} \|G_{\varepsilon_0}(y_0,u)\|_q^{q'}.\end{aligned}$$

Since $\|X_{s,n}\|_q^{q'} = \|\tilde{X}_0(0)\|_q^{q'}$ for $s \le 0$, Lemma 3.5.1 implies $\|X_{t,n}\|_q^{q'} \le C_\lambda p \lambda_0^t \|\tilde{X}_0(0)\|_q^{q'} + (1-\lambda_0)^{-1}\big(|y_0|_1^{q'} + \sup_{u\in[0,1]} \|G_{\varepsilon_0}(y_0,u)\|_q^{q'}\big)$ for all $t = 1, ..., n$, which provides that

$\sup_{n\in\mathbb{N}}\sup_{t=1,...,n}\|X_{t,n}\|_q^{q'} < \infty$. Note that for arbitrary $t \geq 0$, $k \geq 0$, we have by (3.1.3):

$$\|X_{t,n} - X_{t,n}^{*(t-k)}\|_q^{q'} \leq \sum_{j=1}^{p} \chi_j \mathbb{E}|X_{t-j,n} - X_{t-j,n}^{*(t-k)}|^{q'} \leq \sum_{j=1}^{p} \chi_j \|X_{t-j,n} - X_{t-j,n}^{*(t-k)}\|_q^{q'}.$$

Note that $z_s := \|X_{s+(t-k),n} - X_{s+(t-k),n}^{*(t-k)}\|_q^{q'} = 0$ for $s < 0$,
and $z_0 \leq 2\sup_{n\in\mathbb{N}}\sup_{t=1,...,n}\|X_{t,n}\|_q^{q'}$. Lemma 3.5.1 implies $\|X_{t,n} - X_{t,n}^{*(t-k)}\|_q^{q'} = z_k \leq 2C_\lambda\lambda_0^k\sup_{n\in\mathbb{N}}\sup_{t=1,...,n}\|X_{t,n}\|_q^{q'}$. □

Proof of Lemma 3.1.4: The first inequality (3.1.7) is a consequence of

$$\begin{aligned}
& \|\tilde{X}_t(u) - \tilde{X}_t(u')\|_q^{q'} \\
\leq\; & \|G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u) - G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u')\|_q^{q'} + \|G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u') - G_{\varepsilon_t}(\tilde{Y}_{t-1}(u'), u')\|_q^{q'} \\
\leq\; & \|C(\tilde{Y}_{t-1}(u))\|_q^{q'}|u-u'|^{\alpha q'} + \sum_{j=1}^{k}\chi_j\|\tilde{X}_{t-j}(u) - \tilde{X}_{t-j}(u')\|_q^{q'} \\
\leq\; & C^{q'}|u-u'|^{\alpha q'} + |\chi|_1 \cdot \|\tilde{X}_t(u) - \tilde{X}_t(u')\|_q^{q'}.
\end{aligned}$$

For the second inequality, note that we have for all $s = 1, ..., n$:

$$\begin{aligned}
& \left\|X_{s,n} - \tilde{X}_s\left(\frac{s}{n}\right)\right\|_q^{q'} = \left\|G_{\varepsilon_t}\left(Y_{s-1,n}, \frac{s}{n}\right) - G_{\varepsilon_t}\left(\tilde{Y}_{s-1}\left(\frac{s}{n}\right), \frac{s}{n}\right)\right\|_q^{q'} \\
\leq\; & \sum_{i=1}^{p}\chi_i \cdot \left\|X_{s-i,n} - \tilde{X}_{s-i}\left(\frac{s}{n}\right)\right\|_q^{q'} \\
\leq\; & \sum_{i=1}^{p}\chi_i \cdot \left\|X_{s-i,n} - \tilde{X}_{s-i}\left(\frac{s-i}{n}\vee 0\right)\right\|_q^{q'} + \sum_{i=1}^{p}\chi_i \cdot \left\|\tilde{X}_{s-i}\left(\frac{s-i}{n}\vee 0\right) - \tilde{X}_{s-i}\left(\frac{s}{n}\right)\right\|_q^{q'} \\
\leq\; & \sum_{i=1}^{p}\chi_i \cdot \left\|X_{s-i,n} - \tilde{X}_{s-i}\left(\frac{s-i}{n}\vee 0\right)\right\|_q^{q'} + C^{q'}p^{\alpha q'}\frac{|\chi|_1}{1-|\chi|_1}\cdot n^{-\alpha q'}.
\end{aligned}$$

Define $z_s := \|X_{s,n} - \tilde{X}_s(\frac{s}{n}\vee 0)\|_q^{q'}$. Note that $z_s = 0$ for $s \leq 0$ and define $\mu := C^{q'}p^{\alpha q'}\frac{|\chi|_1}{1-|\chi|_1}\cdot n^{-\alpha q'}$. In this special case we can calculate the constants from Lemma 3.5.1 directly, since $z_{s-i_1-...-i_s} = 0$ for $i_1, ..., i_s \in \{1, ..., p\}$:

$$z_s \leq \sum_{i_1=1}^{p}\chi_{i_1}z_{s-i_1} + \mu \leq \sum_{i_1,i_2=1}^{p}\chi_{i_1}\chi_{i_2}z_{s-i_1-i_2} + \mu(1+|\chi|_1) \leq ... \leq \mu(1+|\chi|_1+...+|\chi|_1^{s-1}),$$

which yields $z_s \leq \frac{\mu}{1-|\chi|_1}$ and thus

$$\sup_{s=1,...,n}\left\|X_{s,n} - \tilde{X}_s\left(\frac{s}{n}\right)\right\|_q^{q'} \leq C^{q'}p^{\alpha q'}\frac{|\chi|_1}{(1-|\chi|_1)^2}n^{-\alpha q'}.$$

□

Proof of Theorem 3.1.6. With out loss of generality, we prove the statement for $t = 0$. Because of the continuity of G, the process $(X_{n,u}(y_0))_{u\in[0,1]}$ is continuous and thus a random element of the normed space $(C[0,1], |\cdot|_\infty)$ where $|\cdot|_\infty$ denotes the supremum norm on $[0,1]$. With condition (3.1.4) we obtain for two functions $u \mapsto y(u), y(u)'$:

$$\Big\| \sup_{u\in[0,1]} |X_{n,u}(y) - X_{n,u}(y')| \Big\|_q^{q'} \le \sum_{j=1}^p \chi_j \cdot \Big\| \sup_{u\in[0,1]} |X_{n-j,u}(y) - X_{n-j,u}(y')| \Big\|_q^{q'}.$$

Lemma 3.5.1 implies

$$\| \sup_{u\in[0,1]} |X_{n,u}(y) - X_{n,u}(y')| \|_q^{q'} \le C_\lambda \lambda_0^{n+1} \sup_{u\in[0,1]} |y - y'|_1^{q'}. \tag{3.5.3}$$

Taking $y(u) = y_0$, $y'(u) = R_{\varepsilon_{-n-1},u}(y_0)$, we conclude

$$\Big\| \sup_{u\in[0,1]} |X_{n+1,u}(y_0) - X_{n,u}(y_0)| \Big\|_q^{q'} \le C_\lambda \lambda_0^{n+1} \| \sup_{u\in[0,1]} |y_0 - R_{\varepsilon_0}(y_0, u)|_1 \|_q^{q'}. \tag{3.5.4}$$

This implies that the sequence $(X_{n,u}(y_0))_{u\in[0,1]}$, $n \in \mathbb{N}$ of elements of $C[0,1]$ is a Cauchy sequence in $(C[0,1], |\cdot|_\infty)$ almost surely. Since this space is complete, there exists a continuous limit $\hat{X}_0 = (\hat{X}_0(u))_{u\in[0,1]}$. It was already shown in the proof of Proposition 3.1.3 that $X_{n,u}(y_0) \to \tilde{X}_0(u)$ a.s. for fixed $u \in [0,1]$. This implies that $\hat{X}_0$ is a continuous modification of $(\tilde{X}_0(u))_{u\in[0,1]}$. By (3.5.4), we have

$$\begin{aligned}\Big\| \sup_{u\in[0,1]} |X_{n,u}(y_0)| \Big\|_q^{q'} &\le \sum_{k=0}^{n-1} \Big\| \sup_{u\in[0,1]} |X_{k,u}(y_0) - X_{k+1,u}(y_0)| \Big\|_q^{q'} + |y_0|_1^{q'} \\ \le \quad & \frac{C_\lambda \lambda_0}{1-\lambda_0} \| \sup_{u\in[0,1]} |y_0 - R_{\varepsilon_0}(y_0, u)|_1 \|_q^{q'} + |y_0|_1^{q'} =: D^{q'}.\end{aligned}$$

Because for $M \in \mathbb{N}$, $M \wedge \sup_{u\in[0,1]} |\cdot|$ is a bounded and continuous functional, we obtain $\big\| M \wedge \sup_{u\in[0,1]} |\hat{X}_0(u)| \big\|_q \le D$ and by the monotone convergence theorem, $\sup_{u\in[0,1]} |\hat{X}_t(u)| \in L^q$. □

Proof of Proposition 3.1.7. For fixed $u_0 \in [0,1]$, the fundamental theorem of calculus gives

$$\begin{aligned}&G_{\varepsilon_0}(y, u) - G_{\varepsilon_0}(y', u) \\ = \quad & \int_0^1 \langle \partial_1 G_{\varepsilon_0}(y' + s\cdot(y - y'), u) - \partial_1 G_{\varepsilon_0}(y' + s\cdot(y-y'), u_0), y - y' \rangle \, \mathrm{d}s \\ & + \big(G_{\varepsilon_0}(y, u_0) - G_{\varepsilon_0}(y', u_0)\big).\end{aligned}$$

The first term is bounded in absolute value by $\sup_x |\partial_1 G_{\varepsilon_0}(x, u) - \partial_1 G_{\varepsilon_0}(x, u_0)|_1 \cdot |y - y'|_\infty$. Since $|\chi|_1 < 1$, we can assume w.l.o.g. that $\chi_j > 0$ for all $j = 1, ..., p$ (if for instance $\chi_1 = 0$, one can define $\chi' := \chi + (1 - |\chi|_1/2, 0, ..., 0)$ which still fulfills $|\chi'|_1 < 1$).

Now choose $\beta > 1$ such that $\beta|\chi|_1 < 1$, and define $\chi' := \delta\chi$. We have $|y-y'|_\infty \leq \frac{1}{\min(\chi')}|y-y'|_{\chi',q'}$. For $\delta > 0$ small enough, we have

$$\begin{aligned}
&\Big\| \sup_{|u-u_0|\leq\delta} \sup_{x\neq y} \frac{|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y',u)|}{|y-y'|_{\chi',q'}} \Big\|_q^{q'} \\
\leq\ & \frac{1}{\min(\chi')^{q'}} \Big\| \sup_{|u-u_0|\leq\delta} \sup_x |\partial_1 G_{\varepsilon_0}(x,u) - \partial_1 G_{\varepsilon_0}(x,u_0)|_1 \Big\|_q^{q'} \\
&\qquad + \frac{1}{\beta^{q'}} \sup_{|u-u_0|\leq\delta} \Big\| \sup_{x\neq y} \frac{|G_{\varepsilon_0}(y,u) - G_{\varepsilon_0}(y',u)|}{|y-y'|_{\chi,q'}} \Big\|_q^{q'} < 1.
\end{aligned}$$

Partitioning of $[0,1]$ into (overlapping) closed intervals $I_1, ..., I_K$ of at most length δ and applying Theorem 3.1.6 on each of these intervals I_k, $k = 1, ..., K$ provides the existence of a continuous modification of $(\hat{X}_t^{(k)}(u))_{u\in I_k}$ on each of these subintervals with $\sup_{u\in I_k} |\hat{X}_t^{(k)}(u)| \in L^q$. For fixed $k, k' \in \{1, ..., K\}$ with $I_k \cap I_{k'} \neq \emptyset$ the continuous processes $(\hat{X}_t^{(k)}(u))_{u\in I_k}$, $(\hat{X}_t^{(k')}(u))_{u\in I_{k'}}$ are a.s. equal on $I_k \cap I_{k'}$ which ensures continuity of a process $(\hat{X}_t(u))_{u\in[0,1]}$ which is assembled from $(\hat{X}_t^{(k)}(u))_{u\in I_k}$, $k = 1, ..., K$. □

Proof of Theorem 3.1.8: (i) Note that Assumption 3.1.2(ii),(iii) imply 3.1.2(i) and (3.1.10). We will only use these conditions for the following proof. Since the process $\tilde{X}_t(u)$ is already known to exist, we will define a new recursion function. For $y \in \mathbb{R}^p$, define the random map $\hat{G}_t(y,u) := \langle \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u), y\rangle + \partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u)$ and $\hat{R}_{t,u}(y) := (\hat{G}_t(y,u), y_1, ..., y_{p-1})$, and let $DX_{t,n,u}(y)$ be the first element of $\hat{R}_{t,u} \circ \hat{R}_{t-1,u} \circ ... \circ \hat{R}_{t-n,u}(y)$ for $n \in \mathbb{N}$. For $y, y' \in \mathbb{R}^p$, (3.1.3) and Fatou's lemma imply

$$\begin{aligned}
&\|\hat{R}_{t,u}(y) - \hat{R}_{t,u}(y')\|_q = \|\langle \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u), y - y'\rangle\|_q \\
\leq\ & \liminf_{h\to 0} \frac{\|G_{\varepsilon_t}(\tilde{Y}_{t-1}(u) + h(y-y'), u) - G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u)\|_q}{h} \\
\leq\ & \liminf_{h\to 0} \frac{\|G_{\varepsilon_t}(\tilde{Y}_{t-1}(u) + h(y-y'), u) - G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u)\|_q}{|h(y-y')|_{\chi,q'}} \cdot |y-y'|_{\chi,q'} \\
\leq\ & |y-y'|_{\chi,q'}.
\end{aligned} \tag{3.5.5}$$

Similar to the proof of Proposition 3.1.3, we obtain $C_\lambda > 0, \lambda_0 \in (0,1)$ with

$$\|DX_{t,n,u}(y) - DX_{t,n,u}(y')\|_q^{q'} \leq C_\lambda \cdot \lambda_0^{n+1} |y-y'|^{q'}.$$

Applying this to $y = y_0$ and $y' = \hat{R}_{t-n-1,u}(y_0)$ we obtain

$$\Big\| \sum_{n=0}^\infty |DX_{t,n,u}(y_0) - DX_{t,n+1,u}(y_0)| \Big\|_q^{q'} \leq C_\lambda \sum_{n=0}^\infty \lambda_0^{n+1} \cdot \| |y_0 - \hat{R}_{t-n-1,u}(y_0)|_1 \|_q^{q'}$$

which is finite by (3.1.10) and (3.5.5). This implies that $DX_{0,n,u}(y_0)$ converges a.s. to some limit $D\tilde{X}_0(u)$, say. Because $\tilde{X}_k(u) \in \mathcal{F}_k$ $(k \in \mathbb{Z})$, it is obvious that $D\tilde{X}_0(u)$ is $\mathcal{F}_0$-measurable and therefore has a representation $D\tilde{X}_0(u) = \hat{H}(u, \mathcal{F}_0)$. The rest of the proof is the same as in Proposition 3.1.3(i).

(ii) Because of the continuous differentiability of G, the process $(X_{n,u}(y_0))_{u\in[0,1]}$ is a random element of $(C^1[0,1], |\cdot|_{C^1})$, where $\|f\|_{C^1} = |f|_\infty + |f'|_\infty$ and $|\cdot|_\infty$ denotes the supremum norm on $[0,1]$. Define $\tilde{q} := q/2$ and $\tilde{q}' := \min(\tilde{q}, 1)$. Because of $X_{n,u}(y) = G_{\varepsilon_0,u}(J_{n-1,u}(y))$, we have for two differentiable functions $u \mapsto y_1(u), y_2(u) \in \mathbb{R}^p$:

$$\partial_u X_{n,u}(y_1) = \langle \partial_1 G_{\varepsilon_0}(J_{n-1,u}(y_1), u), \partial_u J_{n-1,u}(y_1)\rangle + \partial_2 G_{\varepsilon_0}(J_{n-1,u}(y_1), u).$$

This shows (use similar techniques as in (3.5.5)):

$$\begin{aligned}
&\big\| \sup_{u\in[0,1]} |\partial_u X_{n,u}(y_1)|\big\|_q^{q'}\\
\le\ & \sum_{j=1}^p \chi_j \big\| \sup_u |\partial_u X_{n-j,u}(y_1)|\big\|_q^{q'} + \big\| \sup_u |\partial_2 G_{\varepsilon_0,u}(J_{n-1,u}(y_1))|\big\|_q^{q'}\\
\le\ & \sum_{j=1}^p \chi_j \big\| \sup_u |\partial_u X_{n-j-1,u}(y_1)|\big\|_q^{q'} + C_2 \sum_{j=1}^p \big\| \sup_u |X_{n-j,u}(y_1)|\big\|_q^{q'} + \big\| \sup_u |\partial_2 G_{\varepsilon_0}(0,u)|\big\|_q^{q'}.
\end{aligned}$$

The third term is finite by assumption, and in the proof of Theorem 3.1.6 it was shown that $\|\sup_u |X_{n,u}(y_1)|\|_q^{q'} \le D(y_1)^{q'}$ for all $n \in \mathbb{N}$. Since $|\chi|_1 < 1$, Lemma 3.5.1 implies for all $n \in \mathbb{N}$:

$$\begin{aligned}
&\big\| \sup_{u\in[0,1]} |\partial_u X_{n,u}(y_1)|\big\|_q^{q'}\\
\le\ & C_\lambda\big(\lambda_0^{n+1} |\partial_u y_1|_1^{q'} + (1-\lambda_0)^{-1}\big(C_2 p D(y_1)^{q'} + \big\| \sup_u |\partial_2 G_{\varepsilon_0}(0,u)|\big\|_q^{q'}\big) =: E(y_1)^{q'}.
\end{aligned} \tag{3.5.6}$$

Using the triangle inequality, we obtain

$$\begin{aligned}
&\big\| \sup_{u\in[0,1]} \big|\partial_u X_{n,u}(y_1) - \partial_u X_{n,u}(y_2)\big|\big\|_{\tilde{q}}^{\tilde{q}'}\\
\le\ & \big\| \sup_{u\in[0,1]} \big|\langle \partial_1 G_{\varepsilon_0,u}(J_{n-1,u}(y_1), u) - \partial_1 G_{\varepsilon_0,u}(J_{n-1,u}(y_2), u), \partial_u J_{n-1,u}(y_1)\rangle\big|\big\|_{\tilde{q}}^{\tilde{q}'}\\
&\quad + \big\| \sup_{u\in[0,1]} \big|\langle \partial_1 G_{\varepsilon_0,u}(J_{n-1,u}(y_2), u), \partial_u J_{n-1,u}(y_1) - \partial_u J_{n-1,u}(y_2)\rangle\big|\big\|_{\tilde{q}}^{\tilde{q}'}\\
&\quad + \big\| \sup_{u\in[0,1]} \big|\partial_2 G_{\varepsilon_0,u}(J_{n-1,u}(y_1), u) - \partial_2 G_{\varepsilon_0,u}(J_{n-1,u}(y_2), u)\big|\big\|_{\tilde{q}}^{\tilde{q}'} =: A_1 + A_2 + A_3.
\end{aligned}$$

Condition (3.1.5) and the result (3.5.3) from the proof of Theorem 3.1.6 (use $\tilde{C}_\lambda$, $\tilde{\lambda}_0$ for the result therein) implies

$$\begin{aligned}
A_3 &\le C_2 \cdot \big\| \sup_{u\in[0,1]} |J_{n-1,u}(y_1) - J_{n-1,u}(y_2)|_1\big\|_{\tilde{q}}^{\tilde{q}'}\\
&\le C_2 \cdot \Big(\sum_{j=1}^p \big\| \sup_{u\in[0,1]} |X_{n-j-1,u}(y_1) - X_{n-j-1,u}(y_2)|\big\|_q^{q'}\Big)^{\tilde{q}'/q'}\\
&\le C_2\big(\tilde{C}_\lambda p \tilde{\lambda}_0^{n-p}\big)^{\tilde{q}'/q'} \sup_u |y_1 - y_2|_1^{\tilde{q}'}.
\end{aligned}$$

A similar technique as in (3.5.5) gives

$$A_2 \le \sum_{j=1}^{p} \chi_j \big\| \sup_{u\in[0,1]} |\partial_u X_{n-j,u}(y_1) - \partial_u X_{n-j,u}(y_2)| \big\|_{\tilde q}^{\tilde q'}.$$

By Cauchy Schwarz' inequality, we have

$$\begin{aligned}
A_1 \quad &\le \quad \sum_{j=1}^{p} \Big\| \sup_{u\in[0,1]} \big| \big(\partial_1 G_{\varepsilon_0}(J_{n-1,u}(y_1),u) - \partial_1 G_{\varepsilon_0}(J_{n-1,u}(y_2),u)\big)_j \big| \cdot \big| \partial_u J_{n-1,u}(y_1)_j \big| \Big\|_{\tilde q}^{\tilde q'} \\
&\le \quad C_1 \sum_{j=1}^{p} \Big\| \sup_{u\in[0,1]} \big| J_{n-1,u}(y_1) - J_{n-1,u}(y_2) \big|_1 \cdot \big| \partial_u J_{n-1,u}(y_1)_j \big| \Big\|_{\tilde q}^{\tilde q'} \\
&\le \quad C_1 \sum_{j=1}^{p} \Big(\sum_{i=1}^{p} \Big\| \sup_{u\in[0,1]} \big| X_{n-i-1,u}(y_1) - X_{n-i-1,u}(y_2) \big| \Big\|_q^{q'} \Big)^{\tilde q'/q'} \\
&\qquad\qquad \times \Big\| \sup_{u\in[0,1]} \big| \partial_u X_{n-j-1,u}(y_1) \big| \Big\|_q^{\tilde q'} \\
&\le \quad C_1 p E(y_1)^{\tilde q'} \big(\tilde C_\lambda p \lambda_0^{n-p} \big)^{\tilde q'/q'} \sup_u |y_1 - y_2|_1^{\tilde q'}
\end{aligned}$$

Finally we have shown that exists a constant $C(y_1) > 0$ such that

$$\begin{aligned}
&\Big\| \sup_{u\in[0,1]} \big| \partial_u X_{n,u}(y_2) - \partial_u X_{n,u}(y_1) \big| \Big\|_{\tilde q}^{\tilde q'} \\
\le \quad &\sum_{j=1}^{p} \chi_j \Big\| \sup_{u\in[0,1]} \big| \partial_u X_{n-j,u}(y_2) - \partial_u X_{n-j,u}(y_1) \big| \Big\|_{\tilde q}^{\tilde q'} + C(y_1) \tilde\lambda_0^n \sup_u |y_1 - y_2|_1^{\tilde q'}.
\end{aligned}$$

Lemma 3.5.1 implies that there exist constants $C_\lambda > 0$, $\lambda_0 \in (0,1)$ such that for $n \in \mathbb{N}$:

$$\begin{aligned}
&\Big\| \sup_{u\in[0,1]} \big| \partial_u X_{n,u}(y') - \partial_u X_{n,u}(y) \big| \Big\|_{\tilde q}^{\tilde q'} \\
\le \quad &C_\lambda \big(\lambda_0^{n+1} \sup_u |\partial_u y_1 - \partial_u y_2|_1^{\tilde q'} + C(y_1) \sum_{i=0}^{n} \lambda_0^i \tilde\lambda_0^{n-i} \big) \sup_u |y_1 - y_2|_1^{\tilde q'}.
\end{aligned}$$

Put $y_1(u) \equiv y_0$, $y_2(u) = R_{\varepsilon_0}(y_0,u)$.
Using $\| \sup_u |\partial_u y_1 - \partial_u y_2|_1 \|_{\tilde q} = \| \sup_u |\partial_2 G_{\varepsilon_0}(y_0,u)| \|_q < \infty$ and $\| \sup_u |y_1 - y_2|_1 \|_{\tilde q} \le \| \sup_u |y_0 - R_{\varepsilon_0}(y_0,u)|_1 \|_{\tilde q} < \infty$ by assumption, we obtain that for all $n \in \mathbb{N}$:

$$\big\| \sup_{u\in[0,1]} |\partial_u X_{n+1,u}(y_0) - \partial_u X_{n,u}(y_0)| \big\|_{\tilde q}^{\tilde q} \le \hat C_\lambda(y_0) \hat\lambda_0^n \tag{3.5.7}$$

with $0 < \hat\lambda_0 := \max(\lambda_0, \tilde\lambda_0) < 1$ and some constant $\hat C_\lambda(y_0) > 0$. Together with the result (3.5.4), we obtain that the sequence $(X_{n,u}(y_0))_{u\in[0,1]}$, $n \in \mathbb{N}$ of elements of $C^1[0,1]$ is a Cauchy sequence in $(C^1[0,1], |\cdot|_{C^1})$ almost surely. Since this space is complete, there exists a continuously differentiable limit $\hat{X}_0 = (\hat{X}_0(u))_{u\in[0,1]}$. Because $\hat{X}_0$ is $\mathcal{F}_0$-measurable, there exists a measurable function $\{u \mapsto \hat{H}(u,\cdot)\}$ on $\mathbb{R}^{\mathbb{N}}$ such that $u \mapsto \hat{H}(u,z)$ is continuously differentiable for all $z \in \mathbb{R}^{\mathbb{N}}$. We may define $\partial_u\hat{X}_t(u) := \hat{H}(u,\mathcal{F}_t)$ for arbitrary $t \in \mathbb{Z}$. The process $X_{t,n,u}(y)$ defined similarly as $X_{n,u}(y)$ but with $\varepsilon_0, ..., \varepsilon_{-n}$ replaced by $\varepsilon_t, ..., \varepsilon_{t-n}$ has the same distributional properties as $X_{n,u}(y)$ and therefore $X_{t,n,u}(y) \to \hat{H}(u,\mathcal{F}_t)$ a.s. and $\partial_u X_{t,n,u}(y) \to \partial_u\hat{H}(u,\mathcal{F}_t)$ a.s. Since

$$X_{t,n,u}(y) = G_{\varepsilon_t}(X_{t-1,n-1,u}(y), u)$$

and

$$\partial_u X_{t,n,u}(y) = \langle \partial_1 G_{\varepsilon_t}(X_{t-1,n-1,u}(y), u), \partial_u X_{t-1,n-1,u}(y)\rangle + \partial_2 G_{\varepsilon_t}(X_{t-1,n-1,u}(y), u)$$

we obtain for $n \to \infty$ that $\hat{X}_t(u)$ fulfills (3.1.1) and $\partial_u\hat{X}_t(u)$ fulfills (3.1.9) a.s. for all $t \in \mathbb{Z}$. Since (3.1.1) and (3.1.9) only allow for a.s. unique solutions, we conclude that $(\hat{X}_t(u))_{u\in[0,1]}$ is a continuously differentiable modification of $(\tilde{X}_t(u))_{u\in[0,1]}$ and $(\partial_u\hat{X}_t(u))_{u\in[0,1]}$ is a continuous modification of $(D\tilde{X}_t(u))_{u\in[0,1]}$.
The uniform convergence $\sup_u |\partial_u X_{n,u}(y_0) - \partial_u\hat{X}_0(u)| \to 0$ together with Fatou's lemma and (3.5.6) implies $\sup_u |\partial_u\hat{X}_0(u)| \in L^q$. □

Proof of Lemma 3.1.11. Define $\tilde{q} := q/2$ and $\tilde{q}' := \tilde{q}/2$. Because $\partial_u\tilde{X}_t(u)$ obeys (3.1.9), we have with the Cauchy Schwarz inequality:

$$\begin{aligned} &\|\partial_u\tilde{X}_t(u) - \partial_u\tilde{X}_t(u')\|_{\tilde{q}}^{\tilde{q}'} \\ \le\ & \sum_{j=1}^{p} \big\|\big(\partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u) - \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u'), u')\big)_j\big\|_q^{\tilde{q}'} \cdot \|\partial_u\tilde{X}_{t-j}(u)\|_q^{\tilde{q}'} \\ &+\big\|\langle \partial_1 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u'), u'), \partial_u\tilde{X}_{t-1}(u) - \partial_u\tilde{X}_{t-1}(u')\rangle\big\|_{\tilde{q}}^{\tilde{q}'} \\ &+\|\partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u) - \partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u'), u')\|_{\tilde{q}}^{\tilde{q}'}. \end{aligned} \tag{3.5.8}$$

(3.1.5) and (3.1.12) give

$$\|\partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), u) - \partial_2 G_{\varepsilon_t}(\tilde{Y}_{t-1}(u'), u')\|_{\tilde{q}}^{\tilde{q}'} \le C_2^{\tilde{q}'} p^{\tilde{q}'/q'} \cdot \|\tilde{X}_t(u) - \tilde{X}_t(u')\|_q^{\tilde{q}'} + D_2^{\tilde{q}'}|u-u'|^{\alpha_2\tilde{q}'}.$$

Similar results are obtained for the first term in (3.5.8). Note that $\|\sup_u |\partial_u\tilde{X}_t(u)|\|_q \le M$ with some $M > 0$ by Theorem 3.1.8. The conditions of Lemma 3.1.4 are fulfilled for $\alpha = 1$, alternatively it can be seen directly that

$$\|\tilde{X}_t(u) - \tilde{X}_t(u')\|_q = \Big\|\int_0^1 |\partial_u\tilde{X}_t(u' + (u-u')s)\ \mathrm{d}s\Big\|_q |u-u'| \le \big\|\sup_v |\partial_u\tilde{X}_t(v)|\big\|_q |u-u'|.$$

A similar technique as in (3.5.5) now implies

$$\begin{aligned} \|\partial_u\tilde{X}_t(u) - \partial_u\tilde{X}_t(u')\|_q^{\tilde{q}'} \ \le\ & |\chi|_1 \|\partial_u\tilde{X}_t(u) - \partial_u\tilde{X}_t(u')\|_q^{\tilde{q}'} \\ &+pM^{\tilde{q}'}\big(C_1^{\tilde{q}'} p^{\tilde{q}'/q'} \cdot M|u-u'|^{\tilde{q}'} + D_1^{\tilde{q}'}|u-u'|^{\alpha_2\tilde{q}'}\big) \\ &+\big(C_2^{\tilde{q}'} p^{\tilde{q}'/q'} \cdot M|u-u'|^{\tilde{q}'} + D_2^{\tilde{q}'}|u-u'|^{\alpha_2\tilde{q}'}\big), \end{aligned}$$

which gives the result since $|\chi|_1 < 1$. □

3.5.2 Proofs of section 3.2

Proof of Proposition 3.2.10. (i) For $K \in \mathbb{N}$ and $k = 1, ..., 2^K$ define intervals of indices $I_{k,K,n} := \{t : t/n \in (\frac{k-1}{2^K}, \frac{k}{2^K}]\}$ such that $\bigcup_{K=1}^{2^K} I_{k,K,n} = \{1, ..., n\}$. For fixed $K \in \mathbb{N}$, we have

$$\begin{aligned}
&\Big\|\frac{1}{n}\sum_{t=1}^{n} X_{t,n} - \frac{1}{2^K}\sum_{k=1}^{2^K}\frac{1}{|I_{k,K,n}|}\sum_{t\in I_{k,K,n}} X_{t,n}\Big\|_1 \\
\leq\ & \Big\|\sum_{k=1}^{2^K}\Big(\frac{|I_{k,K,n}|}{n} - \frac{1}{2^K}\Big)\cdot\frac{1}{|I_{k,K,n}|}\sum_{t\in I_{k,K,n}} X_{t,n}\Big\|_1 \\
\leq\ & \sum_{k=1}^{2^K}\Big|\frac{|I_{k,K,n}|}{n} - \frac{1}{2^K}\Big|\cdot\sup_{t=1,\dots,n}\|X_{t,n}\|_1 \leq \frac{2^K}{n}\cdot\sup_{t=1,\dots,n}\|X_{t,n}\|_1
\end{aligned}$$

and

$$\begin{aligned}
&\Big\|\frac{1}{2^K}\sum_{k=1}^{2^K}\frac{1}{|I_{k,K,n}|}\sum_{t\in I_{k,K,n}} X_{t,n} - \frac{1}{2^K}\sum_{k=1}^{2^K}\frac{1}{|I_{k,K,n}|}\sum_{t\in I_{k,K,n}}\tilde{X}_t\Big(\frac{k}{2^K}\Big)\Big\|_1 \\
\leq\ & \sup_{t=1,\dots,n}\|X_{t,n} - \tilde{X}_t(t/n)\|_1 + \sup_{|u-v|\leq 2^{-K}}\|\tilde{X}_t(u) - \tilde{X}_t(v)\|_1
\end{aligned}$$

Note that for fixed K, by the ergodic theorem for stationary sequences we have for $n \to \infty$:

$$E(K,n) := \frac{1}{2^K}\sum_{k=1}^{2^K}\frac{1}{|I_{k,K,n}|}\sum_{t\in I_{k,K,n}}\tilde{X}_t\Big(\frac{k}{2^K}\Big) \to \frac{1}{2^K}\sum_{k=1}^{2^K}\mathbb{E}\tilde{X}_0\Big(\frac{k}{2^K}\Big) =: E(K)$$

a.s. and in L^1. By the continuity of $[0,1] \to \mathbb{R}, u \mapsto \mathbb{E}\tilde{X}_0(u)$, we have

$$\mathbb{E}(K) := \frac{1}{2^K}\sum_{k=1}^{2^K}\mathbb{E}\tilde{X}_0\Big(\frac{k}{2^K}\Big) \to \int_0^1 \mathbb{E}\tilde{X}_0(u)\,\mathrm{d}u =: E \quad (K \to \infty).$$

Finally,

$$\begin{aligned}
&\Big\|\frac{1}{n}\sum_{t=1}^{n} X_{t,n} - E\Big\|_1 \\
\leq\ & \frac{2^K}{n}\cdot\sup_{t=1,\dots,n}\|X_{t,n}\|_1 + \sup_{t=1,\dots,n}\|X_{t,n} - \tilde{X}_t(t/n)\|_1 + \sup_{|u-v|\leq 2^{-K}}\|\tilde{X}_t(u) - \tilde{X}_t(v)\|_1 \\
&+\|E(K,n) - E(K)\|_1 + |E(K) - E|.
\end{aligned}$$

Thus for all $K \in \mathbb{N}$:

$$\limsup_{n\to\infty}\Big\|\frac{1}{n}\sum_{t=1}^{n} X_{t,n} - E\Big\|_1 \leq \sup_{|u-v|\leq 2^{-K}}\|\tilde{X}_t(u) - \tilde{X}_t(v)\|_1 + |E(K) - E|.$$

The limit $K \to \infty$ gives the result.
For proving the local weak law of large numbers, first note that

$$\begin{aligned}
&\Big\| \frac{1}{nb} \sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) \cdot \big(X_{t,n} - \tilde{X}_t(u)\big) \Big\|_1 \\
\leq\ & |K|_\infty \Big(\sup_{t=1,\ldots,n} \|X_{t,n} - \tilde{X}_t(t/n)\|_1 + \sup_{|u-v|\leq h/2} \|\tilde{X}_t(u) - \tilde{X}_t(v)\|_1 \Big) \to 0.
\end{aligned}$$

This shows that it is enough to consider the convergence of the sum with the corresponding stationary sequence. From Lemma 3.2.9, we have that $\frac{1}{nb}\sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) \cdot \tilde{X}_t(u) \to \mathbb{E}\tilde{X}_t(u)$ holds in L^1, which finishes the proof.
(ii) Define $S_n(u) := \sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big)\big(X_{t,n} - \mathbb{E}X_{t,n}\big)$ and $S_{k,n} := \sum_{t=1}^{k} X_{t,n}$. By partial summation, we have

$$S_n(u) = \sum_{t=1}^{n-1} \Big[K\Big(\frac{t/n-u}{b}\Big) - K\Big(\frac{(t+1)/n-u}{b}\Big)\Big] \cdot S_{t,n} + K\Big(\frac{1-u}{b}\Big) S_{n,n}.$$

Since K is of bounded variation, we have $\sum_{t=1}^{n-1} \Big|K\Big(\frac{t/n-u}{b}\Big) - K\Big(\frac{(t+1)/n-u}{b}\Big)\Big| \leq B_K$ and thus

$$|S_n(u)| \leq B_K \cdot \sup_{t=1,\ldots,n} |S_{t,n}|. \tag{3.5.9}$$

First assume $1 < q \leq 2$. By using the decomposition $X_{t,n} - \mathbb{E}X_{t,n} = \sum_{l=0}^{\infty} P_{t-l}X_{t,n}$ and applying Doob's maximal inequality, Burkholder's inequality and the elementary inequality $(|a_1| + |a_2|)^{q/2} \leq |a_1|^{q/2} + |a_2|^{q/2}$, we obtain

$$\begin{aligned}
\| \sup_{t=1,\ldots,n} |S_{t,n}| \|_q \ &\leq\ \sum_{l=0}^{\infty} \Big\| \sup_{t=1,\ldots,n} \Big| \sum_{s=1}^{t} P_{s-l}X_{s,n} \Big| \Big\|_q \\
&\leq\ \sum_{l=0}^{\infty} \frac{q}{q-1} \Big\| \sum_{s=1}^{n} P_{s-l}X_{s,n} \Big\|_q \leq \sum_{l=0}^{\infty} \frac{q}{(q-1)^2} \Big(\mathbb{E}\Big(\sum_{s=1}^{n} (P_{s-l}X_{s,n})^2 \Big)^{q/2} \Big)^{1/q} \\
&\leq\ \frac{q}{(q-1)^2} \sum_{l=0}^{\infty} \Big(\sum_{s=1}^{n} \|P_{s-l}X_{s,n}\|_q^q \Big)^{1/q} \\
&\leq\ \frac{q}{(q-1)^2} \cdot n^{1/q} \cdot \sum_{l=0}^{\infty} \delta_q^{X_{\cdot,n}}(l).
\end{aligned}$$

which shows that

$$\Big\| \sup_{u\in[0,1]} |(nb)^{-1} S_n(u)| \Big\|_q \leq \frac{B_K q}{(q-1)^2} \Delta_{0,q}^{X_{\cdot,n}} \cdot n^{1/q-1} b^{-1}.$$

Note that in our case, $\Delta_{m,q}^{X_{\cdot,n}} = O(r^m)$ with some $0 < r < 1$. If $q > 2$, we use a Nagaev-type inequality from Liu, Xiao and Wu (2013), Theorem 2(ii) which also holds in our situations as the authors point out in their section 4. Applying this theorem to $S_{t,n}$ and $-S_{t,n}$, we have for all $x > 0$:

$$\mathbb{P}\big(\sup_{t=1,\ldots,n}|S_{t,n}| > x\big) \leq \frac{2C_1(\Delta_{0,q}^{X_{\cdot,n}})^q n}{x^q} + 8G_{1-2/q}\Big(\frac{C_2 x}{\sqrt{n}\Delta_{0,q}^{X_{\cdot,n}}}\Big)$$

with positive constants C_1, C_2 not depending on n. Using (3.5.9), we obtain

$$\begin{aligned}
\mathbb{P}\Big(\sup_{u\in[0,1]}|(nb)^{-1}S_n(u)| > x\Big) &\leq \mathbb{P}\Big(\sup_{t=1,\ldots,n}|S_{t,n}| > \frac{nbx}{B_K}\Big)\\
&\leq \frac{2C_1(B_K\Delta_{0,q}^{X_{\cdot,n}})^q n(nb)^{-q}}{x^q} + 8G_{1-2/q}\Big(\frac{C_2 nb}{\sqrt{n}B_K\Delta_{0,q}^{X_{\cdot,n}}}\Big).
\end{aligned}$$

□

Proof of Proposition 3.2.12. (i) Define $S_{n,L} := \sum_{l=0}^{L-1}\sum_{t=1}^n P_{t-l}X_{t,n}$. Use the abbreviation l.i.m. for $\limsup_{L\to\infty}\limsup_{n\to\infty}$. Because $P_{t-l}X_{t,n} - \mathbb{E}X_{t,n} \to 0$ a.s. and in L^1 for $l\to\infty$, we have by Doob's maximal inequality:

$$\begin{aligned}
&\text{l.i.m.}\Big\|\sup_{u\in[0,1]}|S_{\lfloor nu\rfloor}/\sqrt{n} - S_{\lfloor nu\rfloor,L}/\sqrt{n}|\Big\|_2\\
\leq\ & \text{l.i.m.}\sum_{l=L}^{\infty}\frac{1}{\sqrt{n}}\Big\|\sup_{T=1,\ldots,n}\Big|\sum_{t=1}^{T}P_{t-l}X_{t,n}\Big|\Big\|_2 \leq \text{l.i.m.}\sum_{l=L}^{\infty}\frac{2}{\sqrt{n}}\Big\|\sum_{t=1}^{n}P_{t-l}X_{t,n}\Big\|_2\\
\leq\ & \text{l.i.m.}\sum_{l=L}^{\infty}\frac{2}{\sqrt{n}}\Big(\sum_{t=1}^{n}\|P_{t-l}X_{t,n}\|_2^2\Big)^{1/2} \leq \text{l.i.m. } 2\sum_{l=L}^{\infty}\delta_2^X(l) = 0.
\end{aligned}$$

Now define $\tilde{S}_{n,L} := \sum_{l=0}^{L-1}\sum_{t=1}^n P_{t-l}\tilde{X}_t\big(\frac{t}{n}\big)$. Note that

$$\begin{aligned}
\|P_{t-l}(X_{t,n} - \tilde{X}_t(t/n))\|_2 &\leq \min\Big\{\delta_2^{\tilde{X}(t/n)}(l) + \delta_2^X(l), \sup_{t=1,\ldots,n}\|X_{t,n} - \tilde{X}_t(t/n)\|_2\Big\}\\
&=: \min\{\delta(l), c_n\}.
\end{aligned}$$

By similar arguments as the calculation above, we obtain

$$\begin{aligned}
&\text{l.i.m.}\Big\|\sup_{u\in[0,1]}|S_{\lfloor nu\rfloor,L}/\sqrt{n} - \tilde{S}_{\lfloor nu\rfloor,L}/\sqrt{n}|\Big\|_2\\
\leq\ & \text{l.i.m. } 2\sum_{l=0}^{L-1}\min\{\delta(l), c_n\} \leq \text{l.i.m.}\Big(\sum_{0\leq l\leq c_n^{-1/2}} c_n + \sum_{l > c_n^{-1/2}}\delta(l)\Big)\\
\leq\ & \text{l.i.m.}\Big(c_n^{1/2} + \sum_{l > c_n^{-1/2}}\delta(l)\Big) = 0.
\end{aligned}$$

Now fix $L\in\mathbb{N}$. Define $\hat{S}_{n,L} := \sum_{t=1}^n\Big(\sum_{l=0}^{L-1}P_t\tilde{X}_{t+l}\big(\frac{t+l}{n}\big)\Big)$, where $\tilde{X}_t(u) := \tilde{X}_t(1)$ for $u > 1$. We have

$$|\tilde{S}_{T,L} - \hat{S}_{T,L}| \leq \sum_{l=0}^{L-1}\sum_{t=1}^{l}\big|P_{t-l}\tilde{X}_t\big(\frac{t}{n}\big)\big| + \sum_{l=0}^{L-1}\sum_{t=T-l+1}^{T}\big|P_t\tilde{X}_{t+l}\big(\frac{t+l}{n}\big)\big|$$

Define $M_{t,l} := P_t\tilde{X}_{t+l}\big(\frac{t+l}{n}\big) \overset{d}{=} P_0\tilde{X}_l\big(\frac{t+l}{n}\big) =: M(\frac{t+l}{n})$. We have

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sup_{T=1,\ldots,n}|M_{t,l}| \geq \varepsilon\Big) &\leq n\cdot\sup_{t=1,\ldots,n}\mathbb{P}(|M_{t,l}|\geq\varepsilon\sqrt{n}) \leq \sup_{t=1,\ldots,n}\mathbb{E}[|M_{t,l}|^2\mathbb{1}_{\{|M_{t,l}|\geq\varepsilon\sqrt{n}\}}] \\
&= \sup_{u\in[0,1]}\mathbb{E}[M(u)^2\mathbb{1}_{\{|M(u)|\geq\varepsilon\sqrt{n}\}}] \\
&\leq \mathbb{E}\Big[\big(\sup_u|M(u)|\big)^2\cdot\mathbb{1}_{\{\sup_u|M(u)|\geq\varepsilon\sqrt{n}\}}\Big] \to 0,
\end{aligned}$$

which shows $\frac{1}{\sqrt{n}}\sup_{u\in[0,1]}|\tilde{S}_{\lfloor nu\rfloor,L} - \hat{S}_{\lfloor nu\rfloor,L}| \xrightarrow{\mathbb{P}} 0$.
We now investigate the weak convergence of $\hat{S}_{\lfloor nu\rfloor,L}/\sqrt{n}$ with a martingale central limit theorem from Billingsley, Theorem 18.2. Note that $\sum_{l=0}^{L-1}M_{t,l}/\sqrt{n}$ is a martingale difference sequence with respect to $\mathcal{F}_t$. By elementary operations it can be seen that for each $T = 1,...,n$ and each $\varepsilon > 0$,

$$\sum_{t=1}^{T}\mathbb{E}\Big[\Big(\sum_{l=0}^{L-1}M_{t,l}/\sqrt{n}\Big)^2\mathbb{1}_{\{|\sum_{l=0}^{L-1}M_{t,l}|\geq\varepsilon\sqrt{n}\}}\Big]$$

is bounded by finitely many (dependent on L) terms of the form

$$\frac{1}{n}\sum_{t=1}^{T}\mathbb{E}[M_{t,l}^2\mathbb{1}_{\{|M_{t,l'}|\geq\varepsilon\sqrt{n}\}}],$$

where $l, l' \in \{0,...,L-1\}$. By using similar techniques as above, it can be shown that this converges to 0.
It remains to investigate the behaviour of

$$\sum_{t=1}^{T}\mathbb{E}\Big[\Big(\sum_{l=0}^{L-1}M_{t,l}/\sqrt{n}\Big)^2\Big|\mathcal{F}_{t-1}\Big] = \sum_{l,l'=0}^{L-1}\frac{1}{n}\sum_{t=1}^{T}\mathbb{E}[M_{t,l}M_{t,l'}|\mathcal{F}_{t-1}]$$

for $T = \lfloor sn\rfloor$, $s\in(0,1]$. Define $I_{k,K,T} := \{t : \frac{t}{T}\in(\frac{k-1}{2^K},\frac{k}{2^K}]\}$, then we have for $K\in\mathbb{N}$:

$$\begin{aligned}
&\Big\|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[M_{t,l}M_{t,l'}|\mathcal{F}_{t-1}] - \frac{1}{2^K}\sum_{k=1}^{2^K}\frac{1}{|I_{k,K,T}|}\sum_{t\in I_{k,K,T}}\mathbb{E}[M_{t,l}M_{t,l'}|\mathcal{F}_{t-1}]\Big\|_1 \\
\leq\ &\frac{2^K}{T}\cdot\sup_{t=1,\ldots,n}\sup_{l=0,\ldots,L-1}\|M_{t,l}M_{t,l'}\|_1,
\end{aligned}$$

which is bounded by $\frac{2^K}{T}\sup_u\|\tilde{X}_0(u)\|_2^2$. Furthermore, since $\frac{t}{T}\in I_{k,K,T}\Rightarrow|\frac{t+l}{n}-\frac{k}{2^K}s|\leq 2^{-K}+\frac{L}{n}$, we obtain

$$\begin{aligned}
&\Big\|\frac{1}{2^K}\sum_{k=1}^{2^K}\frac{1}{|I_{k,K,T}|}\sum_{t\in I_{k,K,n}}\Big(\mathbb{E}[M_{t,l}M_{t,l'}|\mathcal{F}_{t-1}] - \mathbb{E}[M_{t,l}(\frac{k}{2^K}s)M_{t,l'}(\frac{k}{2^K}s)|\mathcal{F}_{t-1}]\Big)\Big\|_1 \\
\leq\ &2\big(\sup_{|u-v|\leq 2^{-K}}\|\tilde{X}_0(u)-\tilde{X}_0(v)\|_2 + \sup_{|u-v|\leq Ln^{-1}}\|\tilde{X}_0(u)-\tilde{X}_0(v)\|_2\big)\cdot\sup_u\|\tilde{X}_0(u)\|_2.
\end{aligned}$$

with $M_{t,l}(u) := P_t\tilde{X}_{t+l}(u)$. Since $\mathbb{E}[M_{t,l}(u)M_{t,l'}(u)|\mathcal{F}_{t-1}]$ is ergodic, we have

$$\frac{1}{|I_{k,K,T}|}\sum_{t\in I_{k,K,T}}\mathbb{E}[M_{t,l}(\frac{k}{2^K}s)M_{t,l'}(\frac{k}{2^K}s)|\mathcal{F}_{t-1}]\xrightarrow{\mathbb{P}}\mathbb{E}[M_{0,l}(\frac{k}{2^K}s)M_{0,l'}(\frac{k}{2^K}s)].$$

In total, performing first $n\to\infty$ and afterwards $K\to\infty$, we obtain

$$\begin{aligned}\sum_{l,l'=0}^{L-1}\frac{1}{n}\sum_{t=1}^{\lfloor ns\rfloor}\mathbb{E}[M_{t,l}M_{t,l'}|\mathcal{F}_{t-1}] &\to \sum_{l,l'=0}^{L-1}s\cdot\int_0^1\mathbb{E}[M_{0,l}(xs)M_{0,l'}(xs)]\ \mathrm{d}x\\ &= \int_0^s\Big\|\sum_{l=0}^{L-1}P_0\tilde{X}_l(y)\Big\|_2^2\ \mathrm{d}y.\end{aligned}$$

So we have seen that $\{S_{\lfloor nu\rfloor}/\sqrt{n}, 0\le u\le 1\}\xrightarrow{d}\{\int_0^u\Big\|\sum_{l=0}^{L-1}P_0\tilde{X}_l(v)\Big\|_2\ \mathrm{d}B(v), 0\le u\le 1\}$. By the dominated convergence theorem, $\int_0^u\Big\|\sum_{l=0}^{L-1}P_0\tilde{X}_l(v)\Big\|_2^2\ \mathrm{d}v\to\int_0^u\sigma^2(v)\ \mathrm{d}v$, which completes the proof.
(ii) Since $\Delta_{m,q}=O(r^m)$ with some $0<r<1$, this follows directly from Theorem 2.1 in Karmakar and Wu (2016). □

Proof of Theorem 3.2.13: Define $M_t(u):=g(\tilde{Z}_t(u))$. Note that

$$\Big\|W_{n,b}-\frac{1}{\sqrt{nb}}\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)\cdot M_t(t/n)\Big\|_1\le|K|_\infty\sqrt{nb}\sup_{t=1,\dots,n}\|g(Z_{t,n})-M_t(t/n)\|_1.$$

Since $\|g(Z_{t,n})-M_t(t/n)\|_1\le C\cdot\sup_{t=1,\dots,n}\sum_{j=1}^d\|X_{t-j+1,n}-\tilde{X}_{t-j+1}(t/n)\|_{M+1}\le C'n^{-\alpha}$, the term above is of order $\sqrt{nb}n^{-\alpha}$.
Since $\sum_{k=0}^\infty\sup_u\delta_2^{M(u)}(k)<\infty$, $|K|_\infty<\infty$ and $(K_b(t/n-u)P_{t-l}M_t(t/n))_t$ is a martingale difference sequence with respect to $(\mathcal{F}_{t-l})$, we can use the same technique as in the proof of Theorem 3.2.12 to show that

$$\limsup_{L\to\infty}\limsup_{n\to\infty}\Big\|\frac{1}{\sqrt{nb}}\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)\cdot\Big[(M_t(t/n)-\mathbb{E}M_t(t/n))-\sum_{l=0}^{L-1}P_{t-l}M_t(t/n)\Big]\Big\|_2=0.$$

Now fix $L\in\mathbb{N}$. Since K is Lipschitz continuous and $\sup_t\|M_t((t+l)/n)-M_t(t/n)\|_1\le C'n^{-1}$, it is enough to consider the weak convergence of $\sum_{t=1}^n W_t(t/n)$, where we define $W_t(v):=\sum_{l=0}^{L-1}K\Big(\frac{t/n-u}{b}\Big)P_tM_{t+l}(v)/\sqrt{nb}$. Note that $W_t(t/n)$ is a martingale difference sequence w.r.t. $\mathcal{F}_t$. It holds that

$$\begin{aligned}&\sum_{t=1}^n\|W_t^2(t/n)-W_t^2(u)\|_1\\ \le\ &\sum_{l,l'=0}^{L-1}\frac{1}{nb}\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)^2\|P_tM_{t+l}(t/n)P_tM_{t+l'}(t/n)-P_tM_{t+l}(u)P_tM_{t+l'}(u)\|_1\\ \le\ &2\sum_{l,l'=0}^{L-1}\frac{1}{nb}\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)^2\|M_0(t/n)-M_0(u)\|_2\cdot\sup_u\|M_0(u)\|_2=o(1).\end{aligned}$$

By Lemma 3.2.9,

$$\begin{aligned}\sum_{t=1}^{n}\mathbb{E}[W_t^2(u)|\mathcal{F}_{t-1}] &= \sum_{l,l'=0}^{L-1}\frac{1}{nb}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)^2\mathbb{E}[P_tM_{t+l}(u)P_tM_{t+l'}(u)|\mathcal{F}_{t-1}]\\ &\xrightarrow{\mathbb{P}} \int K^2(x)\ \mathrm{d}x\cdot\Big\|\sum_{l=0}^{L-1}P_0M_l(u)\Big\|_2^2.\end{aligned}$$

Fix $\varepsilon>0$. The sum $\sum_{t=1}^{n}\mathbb{E}[W_t^2(t/n)\mathbb{1}_{\{|W_t(t/n)|\geq\varepsilon\}}]$ is bounded by finitely many (dependent on L) terms of the form

$$\begin{aligned}&\frac{1}{nb}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)^2\mathbb{E}[(P_tM_{t+l}(t/n))^2\mathbb{1}_{\{|K|_\infty|P_tM_{t+l'}(t/n)|\geq\varepsilon\sqrt{nb}\}}]\\ \leq\ & |K|_\infty^2\sup_{u\in[0,1]}\mathbb{E}[(P_0M_l(u))^2\mathbb{1}_{\{|P_0M_{l'}(u)|\geq\varepsilon\sqrt{nb}/|K|_\infty\}}]\\ \leq\ & |K|_\infty^2\mathbb{E}[(\sup_u|P_0M_l(u)|)^2\mathbb{1}_{\{\sup_u|P_0M_{l'}(u)|\geq\varepsilon\sqrt{nb}/|K|_\infty\}}]\end{aligned}$$

which converges to 0 since

$$\|\sup_u|P_0M_l(u)|\|_2\leq 2\|\sup_u|M_l(u)|\|_2\leq C\|\sup_u|\tilde{X}_0(u)|\|_{2(M+\alpha)}^{M+\alpha}<\infty.$$

So we can apply Theorem 18.1. from Billingsley to obtain

$$\sum_{t=1}^{n}W_t(t/n)\xrightarrow{d}\int K^2(x)\ \mathrm{d}x\cdot\Big\|\sum_{l=0}^{L-1}P_0M_l(u)\Big\|_2N(0,1)$$

and thus

$$\frac{1}{\sqrt{nb}}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)\cdot(M_t(t/n)-\mathbb{E}M_t(t/n))\xrightarrow{d}\int K^2(x)\ \mathrm{d}x\cdot\Big\|\sum_{l=0}^{\infty}P_0M_l(u)\Big\|_2N(0,1).$$

It remains to analyse the bias term

$$\frac{1}{\sqrt{nb}}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)\mathbb{E}M_t(t/n).$$

The results (i) and (ii) are immediate from Proposition 3.2.7. □

Proof of Theorem 3.3.2, uniform convergence of $\hat{\theta}_b$: Since a sequence converges in probability to some random variable Z if each subsequence has a further subsequence that converges almost surely towards Z, we may assume w.l.o.g. that

$$\sup_{u\in[\frac{b}{2},1-\frac{b}{2}]}\sup_{\theta\in\Theta}|L_{n,b}(u,\theta)-L(u,\theta)|\to 0\quad a.s. \tag{3.5.10}$$

Since θ_0 is continuous and $\theta_0(u) \in \mathrm{int}(\Theta)$ for all $u \in [0,1]$, the whole curve θ_0 has a positive $|\cdot|_1$-distance $c_{min} := \inf_{u\in[0,1]} \mathrm{dist}(\theta_0(u), \partial\Theta) > 0$ to the boundary $\partial\Theta$ of Θ. Choose $\varepsilon \in (0, c_{min})$ arbitrarily. For each $u \in D_u = D_u(n) := [\frac{b}{2}, 1-\frac{b}{2}]$, define $\Theta(u,\varepsilon) := \{\theta \in \Theta : |\theta - \theta_0(u)|_1 < \varepsilon\}$. Define

$$\theta^*(u) :\in \mathrm{argmin}_{\theta\in\Theta\cap\Theta(u,\varepsilon)^c} L(u,\theta).$$

Here, $\theta^*(u)$ has not to be unique, but we choose one of the possible values. Because $\Theta \cap \Theta(u,\varepsilon)^c$ is compact, there has to exist at least one. Because $\theta_0(u)$ is the unique minimum of $\theta \mapsto L(u,\theta)$ over Θ, there exists $\delta(u) > 0$ such that

$$L(u,\theta^*(u)) - L(u,\theta_0(u)) = \delta(u).$$

It holds that $\delta := \inf_{u\in[0,1]} \delta(u) > 0$. Otherwise, because of the compactness of $[0,1]$, there would exist a sequence $(u_n) \subset [0,1]$ with $u_n \to u^* \in [0,1]$ and $\delta(u_n) \to 0$. By the continuity of L, θ_0 and $u \mapsto \inf_{\theta\in\Theta\cap\Theta(u,\varepsilon)^c} L(u,\theta)$ (use Berge's Maximum theorem and the fact that $u \mapsto \Theta \cap \Theta(u,\varepsilon)^c$ is a continuous set function) this would imply

$$0 \leftarrow \delta(u_n) = \inf_{\theta\in\Theta\cap\Theta(u_n,\varepsilon)^c} L(u_n,\theta) - L(u_n,\theta_0(u_n)) \to \inf_{\theta\in\Theta\cap\Theta(u^*,\varepsilon)^c} L(u^*,\theta) - L(u^*,\theta_0(u^*)),$$

which is a contradiction to the fact that $\theta_0(u^*)$ is the unique minimum of $L(u^*,\theta)$. By (3.5.10) we may choose $N \in \mathbb{N}$ such that for all $n \geq N$, $\sup_{u\in D_u} \sup_{\theta\in\Theta} |L_{n,b}(u,\theta) - L(u,\theta)| < \frac{\delta}{2}$. Now suppose that for some $n \geq N$, $\sup_{u\in D_u} |\hat{\theta}_b(u) - \theta_0(u)|_1 \geq \varepsilon$. Then we have for some $u \in D_u$ that

$$\begin{aligned} L_{n,b}(u,\hat{\theta}_b(u)) &> L(u,\hat{\theta}_b(u)) - \frac{\delta}{2} \geq L(u,\theta^*(u)) - \frac{\delta}{2} \\ &= L(u,\theta_0(u)) + \delta(u) - \frac{\delta}{2} \geq L(u,\theta_0(u)) + \frac{\delta}{2} > L_{n,b}(u,\theta_0(u)), \end{aligned}$$

which is a contradiction to the maximal property of $\hat{\theta}_b(u)$. □

Chapter 4

Local bandwidth selection with a contrast minimization approach

In this chapter we discuss data adaptive local bandwidth selection for quasi maximum likelihood estimators in a very general class of locally stationary time series models. Our theory and assumptions cover recursively defined processes such as tvAR, tvARCH and tvTAR as well as linear processes, for instance the tvARMA process. Let us mention that some minimax results for bandwidth selectors in the very special case of tvAR processes are available: Arkoun and Pergamenchtchikov (2014) consider minimax optimal local bandwidth selection in the case of tvAR(1) processes under the assumption of differentiability of the known true parameter curve. Furthermore, some online adaptive estimation results were obtained by Arkoun (2010) and Giraud, Roueff and Sanchez-Perez (2015).

In Section 4.1, we introduce the model and the quasi maximum likelihood approach. To measure pointwise distances between elements of the finite-dimensional parameter space, we consider the Euclidean norm and a weighted Euclidean norm which can be interpreted as an approximation of the Kullback-Leibler divergence. For both distance measures, the bandwidth selection procedure is done via contrast minimization which is motivated by the general approach of Goldenshluger and Lepski (2011). We state our main results, which are a minimax lower bound if the true curve is in a Hoelder class, and the fact that our bandwidth selector achieves the minimax-optimal rate up to a log factor which usually arises in local procedures.

In Section 4.2 we present the conditions under which the main results hold. We emphasize that nearly all assumptions are stated in terms of a stationary approximation of the observed process, whose properties are usually well-known. Besides standard assumptions from maximum likelihood theory we have to assume that the difference between the quasi maximum likelihood estimator and the true value behaves like a martingale difference sequence so that we can apply a Bernstein inequality for martingale differences from Van de Geer (2000). Furthermore we impose that the limit of the quasi likelihood attains its optimum at the true curve with some known rate so that we obtain results for the rate of the quasi maximum likelihood estimator even if a Taylor expansion is not possible. Dependence assumptions are stated with the functional dependence measure. In the second part of Section 4.2, we present some more specific examples for which the stated assumptions are fulfilled.

In Section 4.3, we analyze the quality of the bandwidth selection procedure for some time series models and compare the two different minimization approaches based on the two different distance measures. Some concluding remarks are drawn in Section 4.4.

All proofs and some more general examples are postponed to the Appendix, Section 4.5. Some results therein may be of independent interest. For instance, we introduce a step-by-step approximation theory for localized empirical processes of locally stationary processes and provide either a deterministic or random bias expansion under the assumption of the existence of derivative processes which were introduced in Chapter 3. Furthermore, we provide exponential inequalities based on martingale decompositions and the decay of the functional dependence measure.

4.1 Introduction and Main Results

The Model. Let $\varepsilon_t, t \in \mathbb{Z}$ be a sequence of i.i.d. random variables and $\mathcal{F}_t := (\varepsilon_t, \varepsilon_{t-1}, ...)$ the shift process. We assume that we observe a Bernoulli shift process

$$X_{t,n} = J_{t,n}(\mathcal{F}_t, \theta_0), \quad t = 1, ..., n \tag{4.1.1}$$

where $J_{t,n}$ is a measurable function which may vary for each $t = 1, ..., n$ and $n \in \mathbb{N}$, and $\theta_0 : [0,1] \to \Theta \subset \mathbb{R}^d$ is an unknown parameter curve. We allow the process to depend on n since we are working in the infill asymptotics framework, assuming that $X_{t,n}$ mainly depends on the rescaled time $\frac{t}{n}$ to obtain a meaningful asymptotic theory. Our aim is to provide minimax-optimal estimators for θ_0 based on observations $X_{t,n}$, $t = 1, ..., n$. To do so, we impose structural assumptions on $X_{t,n}$ by claiming that the process is near to a stationary process $\tilde{X}_t(u)$ ($u \in [0,1]$) as long as $|\frac{t}{n} - u| \ll 1$ and $n^{-1} \ll 1$ (this is made precise in Assumption 4.2.3). We ask

$$\tilde{X}_t(u) = J(\mathcal{F}_t, \theta_0(u)) \tag{4.1.2}$$

to depend on θ_0 and u solely through $\theta_0(u)$, where J is some measurable function. Furthermore, we assume that $\tilde{X}_t(u)$ obeys the recursion

$$\tilde{X}_t(u) = G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), \theta_0(u)), \quad t \in \mathbb{Z}, \tag{4.1.3}$$

where $\tilde{Y}_{t-1}(u) := (\tilde{X}_s(u) : s \le t-1)$ are the past values of the process, $G_\varepsilon(y, \theta)$ a measurable function, where $\varepsilon \in \mathbb{R}$, $y \in \mathbb{R}^{\mathbb{N}}$ and $\theta \in \Theta$. It should be noted that we only pose a structural Markovian assumption on the approximating stationary process which allows us to include a wide range of invertible linear processes in our model which would not obey (4.1.4), see the following Example 4.1.1:

Example 4.1.1. *(i) Recursively defined locally stationary processes $X_{t,n}$ which are obtained by replacing the constant parameters in stationary processes by time-dependent parameter curves θ_0 evaluated at the rescaled time $\frac{t}{n}$ obey (4.1.1). More* *precisely, they obey*

$$X_{t,n} = G_{\varepsilon_t}\big(Y_{t-1,n}, \theta_0\big(\frac{t}{n} \vee 0\big)\big), \quad t \le n \tag{4.1.4}$$

where $Y_{t-1,n} := (X_{t-1,n}, ..., X_{t-p,n})$ *and* $G_\varepsilon(y, \theta)$ *is some measurable function. ? and Zhou and Wu (2009) discussed properties of such processes. Some special cases are:*

(a) *the tvAR(p) process (cf. Dahlhaus and Giraitis (1998), Dahlhaus and Polonik (2009), Dahlhaus (2011)): Given parameter curves* $a_i, \sigma : [0,1] \to \mathbb{R}$ $(i = 1, ..., p)$,

$$X_{t,n} = a_1\big(\frac{t}{n}\big)X_{t-1,n} + ... + a_p\big(\frac{t}{n}\big)X_{t-p,n} + \sigma\big(\frac{t}{n}\big)\varepsilon_t.$$

(b) *the tvARCH(p) process (cf. Dahlhaus and Subba Rao (2006)): Given parameter curves* $a_i : [0,1] \to \mathbb{R}$ $(i = 0, ..., p)$,

$$X_{t,n} = \big(a_0\big(\frac{t}{n}\big) + a_1\big(\frac{t}{n}\big)X_{t-1,n}^2 + ... + a_p\big(\frac{t}{n}\big)X_{t-p,n}^2\big)^{1/2}\varepsilon_t$$

(c) *the tvTAR(1) process (cf. Zhou and Wu (2009)): Given parameter curves* $a_1, a_2 : [0,1] \to \mathbb{R}$,

$$X_{t,n} = a_1\big(\frac{t}{n}\big)X_{t-1,n}^+ + a_2\big(\frac{t}{n}\big)X_{t-1,n}^- + \varepsilon_t,$$

where $x^+ := \max\{x, 0\}$ *and* $x^- := \max\{-x, 0\}$.

(ii) *Linear locally stationary processes (cf. Dahlhaus and Polonik (2009)): For each* $t = 1, ..., n$, $n \in \mathbb{N}$ *assume that there exist coefficients* $a_{t,n}(k)$ *such that*

$$X_{t,n} = \sum_{k=0}^{\infty} a_{t,n}(k)\varepsilon_{t-k}. \tag{4.1.5}$$

Well-known special cases are the tvAR(p) process (see (i)(a)) and the tvMA(p) process: Given parameter curves $a_i : [0,1] \to \mathbb{R}$ $(i = 1, ..., p)$,

$$X_{t,n} = \varepsilon_t + a_1\big(\frac{t}{n}\big)\varepsilon_{t-1} + ... + a_p\big(\frac{t}{n}\big)\varepsilon_{t-p}.$$

(iii) *Nonparametric iid regression: Given* $\theta_0 : [0,1] \to \mathbb{R}$, $X_{t,n} = \theta_0\big(\frac{t}{n}\big) + \varepsilon_t$.

Quasi maximum likelihood approach. The estimation of θ_0 is performed by a nonparametric quasi maximum likelihood method. For this, we assume that some weight function $\ell(x, y, \theta)$ is given (which naturally should mimic the negative log conditional likelihood of $X_{t,n}$ given $Y_{t,n}$), where $x \in \mathbb{R}$ and y, θ as before. For the truncated past vector $Y_{t-1,n}^c := (X_{t-1,n}, ..., X_{1,n}, 0, 0, ...)$, we define

$$\ell_{t,n}(\theta) := \ell(X_{t,n}, Y_{t-1,n}^c, \theta).$$

Fix $u \in [0,1]$. For some kernel function $K : [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}$ with $\int K \, \mathrm{d}x = 1$ and some bandwidth $b \in (0,1]$, we introduce a local likelihood

$$L_{n,b}(u,\theta) := \frac{1}{K_{n,b}(u)} \sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) \cdot \ell_{t,n}(\theta),$$

where $K_{n,b}(u) := \sum_{t=1}^{n} K(\frac{t/n-u}{b})$. An estimator of $\theta_0(u)$ is given by

$$\hat{\theta}_b(u) = \mathrm{argmin}_{\theta \in \Theta} L_{n,b}(u,\theta). \tag{4.1.6}$$

Our theory holds for general weight functions ℓ, but let us emphasize an important special case. Suppose for the moment that $\varepsilon \mapsto G_\varepsilon(y,\theta)$ is continuously differentiable for all ε, y and that the derivative $\partial_\varepsilon G_\varepsilon(y,\theta) \geq \delta_G > 0$ is bounded uniformly from below with some constant $\delta_G > 0$. This ensures that the new innovation ε_t has an impact on the value of $\tilde{X}_t(u)$ which is not too small. Under these conditions, there exists a continuously differentiable inverse $x \mapsto H(x,y,\theta)$ of $\varepsilon \mapsto G(\varepsilon,y,\theta) := G_\varepsilon(y,\theta)$. Suppose that ε_0 has a continuous density f_ε. The negative conditional log likelihood of $\tilde{X}_t(u) = x$ given $\tilde{Y}_{t-1}(u) = y$ then takes the form

$$\ell(x,y,\theta) = -\log f_\varepsilon(H(x,y,\theta)) - \log \partial_x H(x,y,\theta). \tag{4.1.7}$$

In the following derivations, we do not make use of the specific structure of ℓ. This means especially that we allow for model misspecifications due to a false density f_ε. Many authors prefer the case of a Gaussian density $f_\varepsilon(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ because then a minimizer θ of ℓ can be interpreted as a minimum (quadratic) distance estimator (see Dahlhaus and Giraitis (1998) in the tvAR case, Dahlhaus and Subba Rao (2006) in the tvARCH case). See also Example 4.5.9.

Distance measures. In the following, let ∇ denote the derivative with respect to $\theta \in \Theta$. Define $\tilde{\ell}_t(u,\theta) := \ell(\tilde{X}_t(u), \tilde{Y}_{t-1}(u), \theta)$. In **?** it was shown that $\hat{\theta}_b(u)$ is consistent and asymptotically normal for processes $X_{t,n}$ which obey (4.1.4) and fulfill some regularity conditions. More precisely, it holds that

$$\sqrt{nb}\big(\hat{\theta}_b(u) - \theta_0(u) - \mathrm{bias}_n(b)\big) \xrightarrow{d} N\big(0, \textstyle\int K^2 \, \mathrm{d}x \cdot V(u)^{-1} I(u) V(u)^{-1}\big),$$

where $I(u) := \mathbb{E}[\nabla\tilde{\ell}_t(u,\theta_0(u)) \cdot \nabla\tilde{\ell}_t(u,\theta_0(u))']$ is the Fisher information matrix of the stationary process, $V(u) = \mathbb{E}[\nabla^2\tilde{\ell}_t(u,\theta_0(u))]$ and $\mathrm{bias}_n(b) \approx \mathbb{E}[\hat{\theta}_b(u) - \theta_0(u)]$ some bias term. It is immediate that there exists a typical bias-variance decomposition

$$\mathbb{E}|\hat{\theta}_b(u) - \theta_0(u)|_2^2 \approx \mathrm{bias}_n(b) + \int K^2 \, \mathrm{d}x \cdot V(u)^{-1} I(u) V(u)^{-1},$$

where $|\cdot|_2$ denotes the 2-norm in $\mathbb{R}^d$. An important question is what bandwidth $b(u) = b_n(u)$ leads to the optimal rate for the mean squared error and how one can choose this bandwidth adaptively from the data. In this chapter we tackle the problem of local adaptive bandwidth selection (i.e., for each $u \in [0,1]$ an estimator $\hat{b}(u)$ is proposed) which as far as we know was not discussed theoretically in the literature.

Besides the quadratic distance between the estimator $\hat{\theta}_b(u)$ and $\theta_0(u)$, the Kullback-Leibler divergence arises naturally as a distance measure in maximum likelihood theory since $\theta_0(u)$ is the minimizer of the Kullback-Leibler divergence between the model (4.1.3) with parameter θ (instead of $\theta_0(u)$) and (4.1.3). It can be shown that under regularity conditions, the Kullback-Leibler divergence between $\hat{\theta}_b(u)$ and $\theta_0(u)$ is dominated by $|\hat{\theta}_b(u) - \theta_0(u)|^2_{V(u)}$ (see Proposition 4.5.8). To discuss both distances, we introduce a weighted squared distance measure at $u \in [0,1]$, i.e.

$$d_{u,\Xi}(\theta_1, \theta_0) := |\theta_1(u) - \theta_0(u)|^2_{\Xi}, \tag{4.1.8}$$

where $|x|^2_{\Xi} := \langle x, \Xi x\rangle$ for vectors $x \in \mathbb{R}^d$ and positive definite matrices $\Xi \in \mathbb{R}^{d\times d}$. Let Id denote the identity matrix in $\mathbb{R}^{d\times d}$. Then $\mathbb{E}_{\theta_0} d_{u,\mathrm{Id}}(\hat{\theta}_b, \theta_0)$ corresponds to the mean squared error of $\hat{\theta}_b(u)$ and $\mathbb{E}_{\theta_0} d_{u,V(u)}(\hat{\theta}_b, \theta_0)$ corresponds to the Kullback-Leibler divergence of $\hat{\theta}_b(u)$ w.r.t. $\theta_0(u)$.

All distances $d_{u,\Xi}(\theta_1, \theta_0)$ are equivalent since Ξ is positive definite. It is also clear that the minimizers $b_{opt,\Xi}(u)$ of $\mathbb{E}_{\theta_0} d_{u,\Xi}(\hat{\theta}_b, \theta_0)$ coincide if the dimension d of the parameter space Θ equals 1, but they differ in general for $d > 1$. In simulations (see section 4.3) it turns out that there is a significant difference between the corresponding optimizers which justifies to analyze two different model selection procedures. The theoretic behavior of the two optimizers $b_{opt,\mathrm{Id}}$ and $b_{opt,V(u)}$ can be explained as follows: While $b_{opt,\mathrm{Id}}$ leads to estimators $\hat{\theta}_{b_{opt,\mathrm{Id}}}$ which try to fit best to the unknown parameter curve θ_0, $b_{opt,V(u)}$ leads to curves $\hat{\theta}_{b_{opt,V(u)}}$ which ensure that the associated model $X_{t,n}$ is near to the true model and therefore leads to good prediction properties of $\hat{\theta}_{b_{opt,V(u)}}$. For $d > 1$, this difference can be seen in cases where components of θ_0 have different smoothness properties around some $u \in [0,1]$ and $V(u)$ puts a lot of weight on one specific component of θ_0. Then $\hat{\theta}_{b_{opt,V(u)}(u)}(u)$ will try to fit this component best, while $\hat{\theta}_{b_{opt,\mathrm{Id}}(u)}(u)$ will try to fit all components of θ_0 with equal quality. Note that in cases where all components of θ_0 have equal smoothness properties, they all force b to the same optimal value and therefore weighting would not lead to different behavior of the bandwidth selector. This is why we do not expect $\hat{\theta}_{b_{opt,V(u)}(u)}(u)$ and $\hat{\theta}_{b_{opt,\mathrm{Id}}(u)}(u)$ to have a large difference in these cases.

4.1.1 A fully adaptive model selection procedure

A fully adaptive model selection procecure. In the following we will need estimators $\hat{I}_{n,b}(u)$ and $\hat{V}_{n,b}(u)$ of the matrices $I(u)$ and $V(u)$. Their choice is discussed in section 4.1.2. To select the bandwidth b, we propose a contrast minimization method. The general idea of the contrast minimization approach was introduced by Goldenshluger and Lepski (2011). We start by defining a grid

$$B_n = \{a^{-k} : k \in \mathbb{N}\} \cap [\underline{b}_n, 1], \qquad \underline{b}_n = c_b(\Xi) \cdot \frac{\log(n)^{1+2\alpha M}}{n} \tag{4.1.9}$$

of admissible bandwidths, with some constants $c_b(\Xi), \alpha, M$ (independent of n). Here, α is a measure for the exponential decay of the density of ε_0 (if for instance ε_0 is Gaussian, one has $\alpha = \frac{1}{2}$) and M can be interpreted as the minimum degree of a polynomial in (x, y) which is needed to bound the absolute value of the likelihood $\ell(x, y, \theta)$ (for Gaussian likelihoods, one often has $M = 2$, see Example 4.5.9). The precise definitions can be found in Assumptions 4.2.3 and 4.2.2(v) while $c_b(\Xi)$ is given in Theorem 4.1.4. In the following we will assume that Ξ is either the identity Id or $V(u)$. Define the theoretical penalization term

$$P_{n,\Xi}(u,b) := |\log(b)| \cdot F_{n,b}(u)^{-2} \cdot \mathrm{tr}\big(\Xi V(u)^{-1} I(u) V(u)^{-1}\big), \tag{4.1.10}$$

where $F_{n,b}(u) := K_{n,b}(u) \cdot \big(\sum_{t=1}^{n} K(\frac{t/n-u}{b})^2\big)^{-1/2}$. We set $\hat{P}_{n,\Xi}(u,b)$ to be the same as $P_{n,\Xi}(u,b)$ but with $I(u), V(u)$ replaced by their (truncated) estimators $\tilde{I}_{n,b}(u) := \hat{I}_{n,b}(u) \wedge I_m$ and $\tilde{V}_{n,b}(u) := \hat{V}_{n,b}(u) \vee V_0$, where V_0, I_m given in (4.2.2) and the operators $\wedge, \vee$ are generalized minimum or maximum, respectively, of two matrices defined in Lemma 4.5.16. V_0 can be interpreted as the smallest possible value of $V(u)$, similarly for I_m. Furthermore, Ξ is estimated by $\tilde{\Xi}_{n,b} := \tilde{V}_{n,b}(u)$ in the case of $\Xi = V(u)$ and $\tilde{\Xi}_{n,b} :=$ Id in the case of $\Xi =$ Id. Define the penalization term

$$\widehat{\mathrm{pen}}_{n,\Xi}(u,b) := C_P(\Xi)\big\{\hat{P}_{n,\Xi}(u,b) + \sup_{b' \in B_n, b' \geq b} \hat{P}_{n,\Xi}(u,b')\big\}. \tag{4.1.11}$$

with $C_P(\mathrm{Id}) := 256$, $C_P(V(u)) := 288 + 192\sqrt{2} \leq 560$ and

$$Y_{n,\Xi}(u,b) := \max_{b' \in B_n, b' \leq b} \big\{ \max\{d_{u,\tilde{\Xi}_{n,b}}\big(\hat{\theta}_b, \hat{\theta}_{b'}\big), d_{u,\tilde{\Xi}_{n,b'}}\big(\hat{\theta}_b, \hat{\theta}_{b'}\big)\} - \widehat{\mathrm{pen}}_{n,\Xi}(u,b')\big\}_+, \tag{4.1.12}$$

where $\{y\}_+ := \max\{y, 0\}$ for real numbers y. The bandwidth $\hat{b}(u)$ is selected using the rule

$$\hat{b}_\Xi(u) := \arg\min_{b \in B_n} \{Y_{n,\Xi}(u,b) + \widehat{\mathrm{pen}}_{n,\Xi}(u,b)\}.$$

The final estimator is given by $\tilde{\theta}(u) := \hat{\theta}_{\hat{b}_\Xi(u)}(u)$. Note that either (4.1.10) or (4.1.12) simplify in the cases $\Xi =$ Id or $\Xi = V(u)$.

Remark 4.1.2. *(i) The additional minimization with I_m in $\tilde{I}_{n,b}(u)$ is only done to simplify the proof and gives a natural (deterministic) upper bound for the penalization term. Especially in practice it is possible to omit these terms.*

(ii) In some cases, for instance if the recursion structure (4.1.3) is linear in ε_t, it holds that $I(u)V(u)^{-1} = c_\varepsilon \cdot Id$, where $c_\varepsilon > 0$ is some number possibly dependent on characteristics of ε_t (see also Example 4.2.7). In these cases, the theoretical penalization term (4.1.10) and its estimator can be simplified accordingly.

4.1.2 The choice of the estimators of $I(u)$ and $V(u)$

Choice of $\hat{I}_{n,b}(u)$ and $\hat{V}_{n,b}(u)$. A natural choice for an estimator of $I(u)$ and $V(u)$ would be to replace the expectation by its empirical counterparts based on the observations $X_{1,n}, ..., X_{n,n}$, namely $V_{n,b}^\circ(u) = \nabla^2 L_{n,b}(u, \hat{\theta}_b(u))$ and $I_{n,b}^\circ(u) = \frac{1}{K_{n,b}(u)} \sum_{t=1}^{n} K\big(\frac{t/n-u}{b}\big) \cdot$

$\nabla \ell(Y_{t,n}^c, \hat{\theta}_b(u)) \cdot \ell(Y_{t,n}^c, \hat{\theta}_b(u))$. However, in simulations these estimators seem to have a very high variance and thus are unstable. The reason can be explained best by analyzing the estimators in the special case of linear regression

$$X_{t,n} = \theta_0\big(\frac{t}{n}\big) + \varepsilon_t \tag{4.1.13}$$

with some one-dimensional function $\theta_0 : [0,1] \to \mathbb{R}$. Using the Gaussian conditional likelihood (4.1.7) with $f_\varepsilon(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ for estimation, we would obtain

$$V_{n,b}^\circ(u) = 1, \qquad I_{n,b}^\circ(u) = \frac{1}{K_{n,b}(u)}\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big) \cdot \big(X_{t,n} - \hat{\theta}_b(u)\big)^2 \tag{4.1.14}$$

for estimating $V(u) = 1$ and $I(u) = \mathbb{E}\varepsilon_0^2$. This means that the estimator $\hat{I}_{n,b}(u)$ introduced in (4.1.14) would try to estimate properties of the i.i.d. errors ε_t (which are the same for all $t = 1,...,n$ and thus can be seen as 'global' properties) with local estimators. In more complicated models, $V(u)$ and $I(u)$ do not only depend on properties of ε_0 but also on functionals of $\theta_0(u)$.
This motivates to separate estimation of the properties of ε_0 and estimation of $I(u), V(u)$. Similar ideas were presented, for instance, in Lepski, Mammen and Spokoiny (1997) in the linear regression case (4.1.13) by assuming that the variance of the estimators ε_0 is known. We conjecture that most of the properties of ε_0 can be estimated from $X_{1,n},...,X_{n,n}$ with parametric rates if the underlying time series model is known. Besides the standard approach to use a pre-estimator for $\theta_0(u)$ and afterwards using a simple maximum likelihood approach for these unknown properties, some different procedures were introduced, cf. for instance Kreiss and Paparoditis (2015) in the setting of linear locally stationary processes (Example 4.5.10). In the linear regression case (4.1.13) a well-known procedure to estimate the variance of ε_0 with parametric rates is the first difference method $\frac{1}{2n}\sum_{t=1}^n (X_{t,n} - X_{t-1,n})^2$.

The idea of the estimators introduced in the following is that the 'global' information of ε_t is only contained in $X_{t,n}$ which approximately fulfills $X_{t,n} \approx G_{\varepsilon_t}(Y_{t-1,n}, \theta_0(t/n))$ by the Markov Assumption (4.1.3). We now approximate $X_{t,n}$ in $V_{n,b}^\circ, I_{n,b}^\circ$ by replacing $X_{t,n} \approx G_{\varepsilon_t}(Y_{t-1,n}, \theta_0(t/n))$ by its conditional expectation given $Y_{t-1,n}$ which eliminates ε_t from the estimator, but forces us to know some specific properties of ε_0. In many cases (see Examples 4.5.9 and 4.5.10) these properties correspond to the variance or the fourth moment of ε_0. To make this more precise, assume that the quantities

$$\begin{aligned} g_V(y,\theta) &:= \mathbb{E}\big[\nabla^2 \ell(x,y,\theta)\big|_{x=G_{\varepsilon_0}(y,\theta)}\big], \\ g_I(y,\theta) &:= \mathbb{E}\big[\nabla \ell(x,y,\theta)\big|_{x=G_{\varepsilon_0}(y,\theta)} \cdot \nabla \ell(x,y,\theta)'\big|_{x=G_{\varepsilon_0}(y,\theta)}\big], \end{aligned}$$

are known. Define $\hat{V}_{n,b}(u) := V_{n,b}(u, \hat{\theta}_b(u))$, where

$$V_{n,b}(u,\theta) := \frac{1}{K_{n,b}(u)}\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big) g_V\big(Y_{t-1,n}^c, \theta\big),$$

and similarly $\hat{I}_{n,b}(u)$ and $I_{n,b}(u,\theta)$ with g_I instead of g_V.

4.1.3 Main Results

Recall that $\Sigma(\beta, L)$ is the class of Hoelder continuous functions. We first provide a minimax lower bound for estimation of $\theta_0(u)$.

Theorem 4.1.3 (Lower bound)**.** *Fix some $\beta, L > 0$. Suppose that Assumptions 4.2.3, 4.2.2, 4.2.5 and 4.2.6 hold. Then we have with some constant $c(u)$ independent of n:*

$$\inf_{\tilde{\theta} \in \sigma(X_{1,n},...,X_{n,n})} \sup_{\theta_0 \in \Sigma(\beta,L)} \mathbb{E}_{\theta_0} d_{u,\Xi}(\tilde{\theta}, \theta_0) \geq c(u) \cdot n^{-\frac{2\beta}{2\beta+1}},$$

where the infimum is taken over all possible estimators $\tilde{\theta}$ based on $X_{t,n}$, $t = 1, ..., n$ and $\theta_0 \in \Sigma(\beta, L)$ is meant component-wise where β, L are the same for all components.

The following theorems show that $\hat{\theta}_{\hat{b}(u)}(u)$ is minimax optimal for both quadratic and Kullback-Leibler distance up to a factor $\log(n)$ which is natural in local model selection problems.

Theorem 4.1.4 (Upper bound)**.** *Suppose that Assumptions 4.2.3, 4.2.5 and 4.2.2 hold. Define $\beta' := \beta \vee 1$. Then there exists a constant $c_b(\Xi) > 0$ in (4.1.9) and a constant $C(\Xi, u) > 0$ independent of n such that for all $n \geq 3$:*

$$\begin{aligned}
&\sup_{\theta_0 \in \Sigma(\beta,L)} \mathbb{E}_{\theta_0} d_{u,\Xi}(\hat{\theta}_{\hat{b}_\Xi(u)}(u), \theta_0(u)) \\
\leq\ & \inf_{b \in B_n} \Big\{ N_1(\Xi) \sup_{b \in B_n, b' \geq b} P_{n,\Xi}(u, b) \\
&\qquad + N_2(\Xi) \sum_{b' \in B_n, b' \leq b} \mathbb{E}|\tilde{B}_{n,b'}(\Xi^{1/2} V^{-1} \nabla \ell(\cdot, \theta_0(u)), u)|_2^2 + W_{n,\Xi}(b) \Big\} \\
&\quad + C(\Xi, u) \log(n) \cdot (n^{-1} + n^{-2\beta'}),
\end{aligned} \tag{4.1.15}$$

where $W_{n,\Xi}(b)$ contains asymptotically negligible terms, and $N_1(Id) = 5760$, $N_2(Id) = 360$, $N_1(V(u)) = 11664 + 7776\sqrt{2} \leq 22661$, $N_2(V(u)) = 729 + 486\sqrt{2} \leq 1417$.
If additionally Assumption 4.2.4 is fulfilled, it holds that $W_{n,\Xi}(b) \lesssim \log(n) n^{-1} + \mathbb{1}(b^\beta > c_1)$ with some constant $c_1 > 0$ and $\mathbb{E}|\tilde{B}_{n,b'}(\Xi^{1/2} V^{-1} \nabla \ell(\cdot, \theta_0(u)), u)|_2^2 \lesssim (b')^{2\beta} + n^{-1}$. The choice $b \sim n^{-\frac{1}{2\beta+1}}$ in (4.1.15) gives

$$\sup_{\theta_0 \in \Sigma(\beta,L)} \mathbb{E}_{\theta_0} d_{u,\Xi}(\hat{\theta}_{\hat{b}_\Xi(u)}(u), \theta_0(u)) \lesssim \Big(\frac{\log(n)}{n}\Big)^{-\frac{2\beta}{2\beta+1}}.$$

4.2 Assumptions and Examples

Assumptions. Recall the notations from the preliminaries. Recall that $\|Z\|_q := (\mathbb{E}|Z|^q)^{1/q}$ for a real-valued random variable Z, the ℓ^q distance $|x|_q := (\sum_{i=1}^p |x_i|^q)^{1/q}$ and the weighted ℓ^q distance $|x|_{w,q} := (\sum_{i=1}^\infty w_i |x_i|^q)^{1/q}$ for vectors $x \in \mathbb{R}^p$. Recall the definition of the functional dependence measure $\delta_q^Y(k)$ for processes $Y = (Y_t)_{t \in \mathbb{Z}}$ and the projection operator $P_j \cdot := \mathbb{E}[\cdot|\mathcal{F}_j] - \mathbb{E}[\cdot|\mathcal{F}_{j-1}]$, where $\mathcal{F}_t := (\varepsilon_t, \varepsilon_{t-1}, ...)$.

Since our theory covers a wide range of time series models, we need some structural assumptions on the chosen weight function ℓ to obtain exponential inequalities and approximations for empirical means based on ℓ and its derivatives by using properties of $X_{t,n}$. For this, we will impose a Lipschitz-type condition where we allow the Lipschitz constant to depend on the location at most polynomially. The exact condition which also gives a definition of M is stated in Assumption 4.2.2(v).

Definition 4.2.1 (The class $\mathcal{L}(M,\chi,C_z,C_\theta)$)**.** *We say that a function $g : \mathbb{R}^{\mathbb{N}} \times \Theta \to \mathbb{R}^p$ is in the class $\mathcal{L}(M,\chi,C_z,C_\theta)$ if either $g \equiv 0$ (then use $M = -\infty$), $g(\cdot,\theta)$ is constant (then use $M = 0$) or there exists $M \geq 1$, vectors $C_z, C_\theta \in \mathbb{R}^p_{\geq 0}$ and a sequence $\chi \in \mathbb{R}^{\mathbb{N}}_{\geq 0}$ with $\sum_{j=1}^\infty \chi_j < \infty$ such that for all $i = 1,...,p$ it holds that*

$$\sup_{z\neq z'} \frac{|g_i(z,\theta) - g_i(z',\theta)|}{|z-z'|_{\chi,1}\cdot(1+|z|_{\chi,1}^{M-1}+|z'|_{\chi,1}^{M-1})} \leq C_{z,i}, \qquad \sup_{\theta\neq\theta'} \frac{|g_i(z,\theta)-g_i(z,\theta')|}{|\theta-\theta'|_2(1+|z|_{\chi,1}^{M})} \leq C_{\theta,i} \tag{4.2.1}$$

Since we are dealing with nonparametric maximum likelihood type estimators we need to impose assumptions on the smoothness of the true curve θ_0 and the size of Θ which is done in Assumption 4.2.2(i). If ℓ coincides with the true negative log conditional likelihood (4.1.7) with correctly specified density f_ε of ε_0, it is well-known that under regularity conditions, $\nabla\tilde{\ell}_t(u,\theta_0(u))$ is a martingale difference sequence and $V(u) = I(u)$ is positive definite. Since we allow for general ℓ it is possible to deal with misspecifications: Especially in the case that ℓ is taken to be (4.1.7) but f_ε is wrongly chosen as the standard Gaussian density it is easily possible to retain the martingale difference property and the positive definiteness of $V(u)$, see Example 4.5.9. Since these properties of the estimation procedure are crucial for our proofs to apply Bernstein inequalities, we ask for them in Assumption 4.2.2(iii),(iv), where we ask for a slightly stronger assumption on $G_V(u,\theta) := \mathbb{E}[g_V(\tilde{Y}_0(u),\theta)]$ since we use a different technique to estimate $V(u) = G_V(u,\theta_0(u))$. Finally, we need a possibility to determine the convergence rate of $\hat{\theta}_b(u)$ even in the case when it lies on the boundary of Θ. To do so, we use a technique of Van de Geer (2000) which needs assumptions on $L(u,\theta) := \mathbb{E}\tilde{\ell}_t(u,\theta)$.

Assumption 4.2.2 (Likelihood assumptions)**.** *Assume that for some $\beta, L_0 > 0$,*

(i) $\Theta \subset \mathbb{R}^d$ *is compact,* $\delta_\Theta := \inf_{u\in[0,1]}\inf_{\theta\in\partial\Theta}|\theta_0(u)-\theta|_2 > 0$*, and* $\theta_0 \in \Sigma(\beta, L_0)$.

(ii) There exists $C_L(\cdot) > 0$ *s. t. for* $u \in [0,1]$, $\theta \in \Theta$: $L(u,\theta) - L(u,\theta_0(u)) \geq \frac{1}{C_L(u)}|\theta-\theta_0(u)|_2^2$.

(iii) $\nabla\tilde{\ell}_t(u,\theta_0(u))$ *is a martingale difference sequence with respect to* $\mathcal{F}_t$ *in each component.*

(iv) $\inf_{u\in[0,1]}\inf_{\theta\in\Theta}\lambda_{min}(G_V(u,\theta)) > 0$.

(v) Assume that $g \in \{\ell, \nabla\ell, \nabla^2\ell, g_I, g_V\}$ *fulfills* $g \in \mathcal{L}(M,\chi,C_{g,z},C_{g,\theta})$. *We ask* $\rho(t) := \sum_{j=t+1}^\infty \chi_j$ *to fulfill*

$$\sum_{j=1}^\infty j\chi_j < \infty, \qquad \sum_{t=1}^\infty \rho(t) < \infty, \qquad \rho(n) \leq \frac{C_\rho}{n}.$$

Note that all conditions in Assumption 4.2.2 only deal with the stationary process $\tilde{X}_t(u)$ whose properties are usually well-known. In many models, ℓ depends only on finitely many components of y or χ is a geometrically decaying sequence which immediately fulfills the summability conditions in Assumption 4.2.2(v). The conditions on g_I, g_V usually can be obtained from the conditions on $\nabla\ell, \nabla^2\ell$ but in the case of g_I they may lead to a value of M which is larger than necessary. We now define

$$V_0 := \inf_{u\in[0,1]} \inf_{\theta\in\Theta} \lambda_{min}(G_V(u,\theta)) \cdot \mathrm{Id}, \qquad I_m := \sup_{u\in[0,1]} \sup_{\theta\in\Theta} \lambda_{max}(G_I(u,\theta)) \cdot \mathrm{Id}, \tag{4.2.2}$$

where $G_I(u,\theta) := \mathbb{E}[g_I(\tilde{Y}_0(u),\theta)]$.

In the following assumption we present the conditions we need on the observed process $X_{t,n}$. We specify how $X_{t,n}$ has to be approximated by $\tilde{X}_t(u)$. Since the parameter curve θ_0 plays an important role in the time evolution of $X_{t,n}$ it is obvious that the smoothness properties of θ_0 (especially the Hoelder exponent β) appear here. Furthermore, we pose conditions on the dependence structure and the moments of the approximation process $\tilde{X}_t(u)$ to obtain exponential inequalities for empirical processes based on $\tilde{X}_t(u)$. For this, let us define $N_\alpha(q)^q := \Gamma(\alpha q + 2)$ for $q \geq 1$ and $\alpha \geq 0$ to measure the exponential decay of the distribution of $\tilde{X}_t(u)$. The conditions are completely independent of the estimation procedure and thus can be checked separately for the processes of interest. A wide range of linear processes and recursively defined processes (see Example 4.5.9, 4.5.10) satisfy them.

Assumption 4.2.3 (Moment and dependence assumptions). *Assume that there exists some $\alpha \geq 0$ such that for all $q \geq 1$:*

$$\sup_{u\in[0,1]} \|\tilde{X}_0(u)\|_q < \infty, \qquad \sup_{u\in[0,1]} \delta_q^{\tilde{X}(u)}(k) \leq \delta(k) \cdot N_\alpha(q), \tag{4.2.3}$$

where $\delta(k)$ is a sequence and $\xi(t) := \sum_{j=1}^{t} \chi_j \cdot \delta(t-j+1)$ fulfills $\sum_{t=1}^{\infty} \xi(t) < \infty$. Suppose there exist $C_{B,1}, C_{B,2} > 0$ independent of n such that

$$\sup_{t=1,\dots,n} \|X_{t,n} - \tilde{X}_t(t/n)\|_{2M} \leq C_{B,1} \cdot n^{-\beta'}, \qquad \|\tilde{X}_t(u) - \tilde{X}_t(v)\|_{2M} \leq C_{B,2} \cdot |u-v|^{\beta'}, \tag{4.2.4}$$

where $\beta' := \beta \wedge 1$.

In opposite to standard nonparametric regression, in our model we have to subtract a random (instead of a deterministic) bias term from $\hat{\theta}_b(u) - \theta_0(u)$ to obtain a quantity where a Bernstein inequality is applicable. The random bias term involves the process $\tilde{X}_t(t/n)$ for $t = 1, ..., n$ which has to be replaced by $\tilde{X}_t(u)$. To do so, we have to impose differentiability assumptions on $u \mapsto \tilde{X}_t(u)$ which is done in Assumption 4.2.4. While for linear models as in Example 4.1.1(ii) such differentiability is directly inherited from the deterministic coefficients therein, the problem is more involved for recursively defined models in Example 4.1.1(i). A general theory for them was introduced in **?**. Comparable conditions as in Assumption 4.2.4 were used in Dahlhaus and Subba Rao (2006) and Subba Rao (2006) to discuss the bias.

Assumption 4.2.4 (Bias assumptions)**.** *If $\beta > 1$, assume that $\tilde{X}_t(u)$ has an l_β-times differentiable modification which fulfills $C_{\partial^k X} := \sup_u \|\partial_u^k \tilde{X}_t(u)\|_{2M} < \infty$, $\sum_{j=0}^{\infty} \delta_{2M}^{\partial_u^k \tilde{X}(u)}(j) < \infty$ for all $k = 1, ..., l_\beta$ and*

$$\|\partial_u^{l_\beta} \tilde{X}_t(u) - \partial_u^{l_\beta} \tilde{X}_t(v)\|_{2M} \leq C_{B,2} \cdot |u - v|^{\beta - l_\beta}. \tag{4.2.5}$$

Furthermore, for $g \in \{\ell, \nabla\ell, \nabla^2\ell, g_I, g_V\}$ assume that $g(\cdot, \theta)$ is l_β-times partially differentiable and $\partial_{i_1}...\partial_{i_{l_\beta}} g(\cdot, \theta) \in \mathcal{L}(M - l_\beta, \chi, C_z\psi(i_1) \cdot ... \cdot \psi(i_{l_\beta}), \cdot)$ with an absolutely summable sequence $(\psi(k))_{k\in\mathbb{N}}$ and some constant C_z, where the second condition in (4.2.1) does not have to be fulfilled.

We pose some conditions on the kernel which are standard in nonparametric estimation theory.

Assumption 4.2.5 (Kernel assumptions)**.** *Assume that $K : [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}$ is a function of bounded variation with $\int K \ dx = 1$. Assume that K is a kernel of order l_β, i.e. $\int K(x)x^j \ dx = 0$ for $j = 1, ..., l_\beta$. Suppose that there exists $K_0 > 0$ such that for all $b \in B_n$:*

$$\frac{K_{n,b}(u)}{nb} \geq K_0. \tag{4.2.6}$$

Condition (4.2.6) in Assumption 4.2.5 is usually fulfilled if B_n does not contain too small bandwidths. Since K has bounded variation B_K, it holds that $\big|\frac{K_{n,b}(u)}{nb} - \frac{1}{b}\int_0^1 K\left(\frac{v-u}{b}\right) \ \mathrm{d}v\big| \leq \frac{B_K}{nb}$. If we define $K_0 := \frac{1}{2}\min\big\{\int_0^1 K(y) \ \mathrm{d}y, \int_{-1}^0 K(y) \ \mathrm{d}y\big\} > 0$, we have $\frac{1}{b}\int_0^1 K\left(\frac{v-u}{b}\right) \ \mathrm{d}v \geq 2K_0$, which leads to $\frac{K_{n,b}(u)}{nb} \geq K_0$ as long as $b \geq \frac{B_K}{K_0} \cdot \frac{1}{n}$.

To show a minimax lower bound, we need some knowledge of the properties of the conditional likelihood of $X_{t,n}$ given $Y_{t-1,n} := (X_{s,n} : s \leq n)$. Here, we assume that this likelihood does only depend on θ_0 through its values on the discrete grid $\frac{t}{n}$, $t = 1, ..., n$. This general formulation allows to cover both recursively defined time series and linear models (see Example 4.1.1).

Assumption 4.2.6 (Structural assumptions on $X_{t,n}$)**.** *Suppose that the negative log conditional likelihood of $X_{t,n} = x$ given $Y_{t-1,n} = y$ is given by $\tilde{\ell}(x, y, \theta_0(\frac{t-k}{n})_{k\geq 0})$, where $\tilde{\ell} : \mathbb{R} \times \mathbb{R}^{\mathbb{N}} \times \Theta^{\mathbb{N}} \to \mathbb{R}$ is some function which fulfills $\tilde{\ell}(\cdot, (\theta)_{k\geq 0}) = \ell(\cdot, \theta)$ for all $\theta \in \Theta$. Suppose furthermore,*

(i) $(\theta_k)_{k\geq 0} \mapsto \tilde{\ell}(x, y, (\theta_k)_{k\geq 0})$ is partially continuously differentiable. There exist sequences $(C_{\nabla,i}(k))_{k\geq 0}$ with $\sum_{k=0}^{\infty} kC_{\nabla\ell,i}(k) < \infty$ ($i = 1, 2$) such that $(x, y, \theta_k) \mapsto \nabla_{\theta_j}\tilde{\ell}(x, y, (\theta_i)_{i\geq 0})$ is in $\mathcal{L}(M, \chi, C_{\nabla\ell,1}(j), C_{\nabla\ell,1}(j)C_{\nabla\ell,2}(k))$ for all $k, j \in \mathbb{N}$.

(ii) There exists $z_0 \in \mathbb{R}^{\mathbb{N}}$ with $|z_0|_\infty \leq C_{z_0}$ and $\sup_{\theta\in\Theta} |\nabla_{\theta_j}\ell(z_0, (\theta)_{k\geq 0})|_1 \leq C_{\nabla\ell,1}(j)$ ($j \geq 1$).

If the likelihood ℓ was chosen correctly, Assumption 4.2.6 does not impose any new conditions in the case of recursively defined models (4.1.4). For general models, the idea behind the preceding assumption is that $X_{t,n}$ should obey a recursion of the form

$$X_{t,n} = \tilde{G}_{\varepsilon_t}\big(Y_{t-1,n}, \theta_0\big(\frac{t-k}{n} \vee 0\big)_{k\geq 0}\big), \qquad t \leq n$$

with some measurable $\tilde{G}$ satisfying $\tilde{G}_\varepsilon(\cdot,(\theta)_{j\geq 0}) = G_\varepsilon(\cdot,\theta)$ with G from (4.1.3) and $\tilde{\ell}$ is similarly constructed as in (4.1.7). For linear models (4.1.5), one often has $a_{t,n}(k) = A_k\big(\theta_0\big(\frac{t-j}{n}\vee 0\big)_{j\geq 0}\big)$ with measurable functions $A_k : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$. For instance, this is the case for tvARMA models (see Example 4.2.10).

Examples. In Chapter 3 of this thesis, conditions on a general $G_\varepsilon(y,\theta)$ were discussed such that the conditions of Assumption 4.2.3 are fulfilled. In the next two examples we consider a more special case of a recursion function depending only on finitely many past values and being linear in ε together with a Gaussian likelihood ℓ of the form (4.1.7). We therefore use the notation $Y_{t-1,n} = (X_{t-1,n},...,X_{t-p,n})$ here. We start with models which have a constant conditional variance $\mathbb{E}\varepsilon_0^2$ (which has to be known or pre-estimated). It can easily be seen that the following examples cover tvAR- and tvTAR models. They are an immediate consequence of the more general Lemma 4.5.9 in the appendix, the proofs are therefore omitted. Depending on whether the conditional variance is assumed to be time-varying or not, it is imposed by the use of g_V and g_I to know / pre-estimate either $\mathbb{E}\varepsilon_0^2$ or $\mathbb{E}\varepsilon_0^4$.

Example 4.2.7 (Constant conditional variance)**.** *Assume that there exists $m : \mathbb{R}^p \to \mathbb{R}^d$ such that*

$$X_{t,n} = \big\langle m(Y_{t-1,n}), \theta_0(t/n)\big\rangle + \varepsilon_t, \quad t = 1,...,n. \tag{4.2.7}$$

Suppose that $\mathbb{E}\varepsilon_0 = 0$, $\sigma^2 := \mathbb{E}\varepsilon_0^2$ is known, and

(a) $\sup_{y\neq y'}\frac{|m_i(y)-m_i(y')|}{|y-y'|_{\chi_i,1}} \leq 1$ *with some $\chi_i \in \mathbb{R}^p_{\geq 0}$ $(i = 1,...,d)$,*

(b) the Lebesgue density $f_{|\varepsilon_0|}$ of $|\varepsilon_0|$ fulfills $f_{|\varepsilon_0|}(x) \leq C_f\exp(-x^{1/\alpha})$ for some $\alpha, C_f > 0$,

(c) $m_1(\tilde{Y}_0(u)),...,m_d(\tilde{Y}_0(u))$ are linearly independent in L^2.

Define $\Theta := \big\{\theta \in \mathbb{R}^d : \sum_{i=1}^d\sum_{j=1}^p |\theta_i|\chi_{i,j} \leq \rho\big\}$ with some $0 < \rho < 1$. Assume that $\theta_0 \in \Sigma(\beta, L)$. Then Assumptions 4.2.2, 4.2.3 and 4.2.6 are fulfilled for the Gaussian likelihood (4.1.7) with $M = 2$ and $G_\varepsilon(y,\theta) = \langle m(y),\theta\rangle + \varepsilon$.
In that case, it holds that

$$g_I(y,\theta) = \sigma^2 m(y)m(y)' = \sigma^2 g_V(y,\theta), \qquad I(u) = \sigma^2\mathbb{E}m(\tilde{Y}_0(u))m(\tilde{Y}_0(u))' = \sigma^2 V(u).$$

Note that condition (c) is immediately clear in the tvAR case, since then we have $m_i(y) = y_i$ and thus $\sum_{i=1}^p \xi_i m_i(\tilde{X}_{-i+1}(u)) = 0$ for some $\xi_1,...,\xi_p \in \mathbb{R}$ inductively implies $\xi_j\mathbb{E}[\varepsilon_0^2] = \xi_j\mathbb{E}m_{j+1}(\tilde{X}_0(u))\varepsilon_0 = \sum_{i=1}^p \xi_i\mathbb{E}[m_i(\tilde{X}_{-i+1}(u))\varepsilon_{-j+1}] = 0$ for $j = 1,...,p$.

Example 4.2.8 (Deterministic time-varying conditional variance)**.** *Assume that there exists $m : \mathbb{R}^p \to \mathbb{R}^{d-1}$, $(\tilde{\theta}_0, \sigma_0) : [0,1] \to \mathbb{R}^{d-1}\times[\sigma_{min},\sigma_{max}]$ with $\sigma_{max} > \sigma_{min} > 0$ such that*

$$X_{t,n} = \big\langle m(Y_{t-1,n}), \tilde{\theta}_0(t/n)\big\rangle + \sigma_0(t/n)\varepsilon_t, \quad t = 1,...,n. \tag{4.2.8}$$

*Suppose that $\mathbb{E}\varepsilon_0 = 0$, $\mathbb{E}\varepsilon_0^2 = 1$, $\mu_4 := \mathbb{E}\varepsilon_0^4$ is known and that conditions (a),(b),(c) from Example 4.2.7 hold accordingly. Define $\Theta := \big\{\theta = (\tilde{\theta},\sigma) \in \mathbb{R}^d : \sum_{i=1}^{d-1}\sum_{j=1}^p |\theta_i|\chi_{i,j} \leq$

$\rho\} \times [\sigma_{min}, \sigma_{max}]$ *with some* $0 < \rho < 1$. *Assume that* $\theta_0 = (\tilde{\theta}_0, \sigma_0) \in \Sigma(\beta, L)$. *Then Assumptions 4.2.2, 4.2.3 and 4.2.6 are fulfilled for the Gaussian likelihood (4.1.7) with* $M = 2$ *and* $G_\varepsilon(y, \theta) = \langle m(y), \tilde{\theta}\rangle + \sigma\varepsilon$.
In that case, it holds that

$$g_I(y,\theta) = \frac{1}{\sigma^2}\begin{pmatrix} m(y)m(y)' & 0 \\ 0 & \mu_4 - 1\end{pmatrix}, \quad g_V(y,\theta) = \frac{1}{\sigma^2}\begin{pmatrix} m(y)m(y)' & 0 \\ 0 & 2\end{pmatrix}.$$

In both Examples 4.2.7 and 4.2.8, the conditions of Assumption 4.2.4 are fulfilled under suitable conditions on the differentiability of m, see the results of **?**.

In the case that the conditional variance is random, we have to assume that the noise ε_0 is a.s. bounded. We conjecture that this condition can be relaxed if $\beta \geq \beta_0 > 0$ with some known $\beta_0 > 0$ since then one does not need exponential inequalities to bound empirical processes of $X_{t,n}$. Furthermore, we have to guarantee that the conditional variance is uniformly bounded from below. A prominent example for such models are tvARCH processes.

Example 4.2.9 (Constant conditional variance)**.** *Assume that there exists* $m : \mathbb{R}^p \to \mathbb{R}^d_{\geq 0}$ *such that*

$$X_{t,n} = \sqrt{\langle m(Y_{t-1,n}), \theta_0(t/n)\rangle} \cdot \varepsilon_t, \quad t = 1, ..., n. \tag{4.2.9}$$

Suppose that $\mathbb{E}\varepsilon_0 = 0$, $\mathbb{E}\varepsilon_0^2 = 1$, $\mu_4 := \mathbb{E}\varepsilon_0^4$ *is known, and*

(a) $\sup_{y \neq y'} \frac{|\sqrt{m_i(y)} - \sqrt{m_i(y')}|}{|y-y'|_{\chi_i,1}} \leq 1$ *with some* $\chi_i \in \mathbb{R}^p_{\geq 0}$ *(*$i = 1, ..., d$*). There exists* $m_0 > 0$ *such that* $m_1(y) \geq m_0$ *for all* $y \in \mathbb{R}^p$.

(b) $|\varepsilon_0| \leq C_\varepsilon$ *a.s., put* $\alpha = 0$.

(c) $m_1(\tilde{Y}_0(u)), ..., m_d(\tilde{Y}_0(u))$ *are linearly independent in* $L^2(\Omega, \mathbb{R})$.

Define $\Theta := \big\{\theta \in \mathbb{R}^d_{\geq 0} : \sum_{i=1}^d \sum_{j=1}^p \sqrt{\theta_i}\chi_{i,j} \leq \rho_{max} C_\varepsilon^{-1}, \theta_i \geq \rho_{min}\big\}$ *with some* $0 < \rho_{max} < 1, \rho_{min} > 0$. *Assume that* $\theta_0 \in \Sigma(\beta, L)$. *Then Assumptions 4.2.2, 4.2.3 and 4.2.6 are fulfilled for the Gaussian likelihood (4.1.7) with* $M = 3$ *and* $G_\varepsilon(y,\theta) = \sqrt{\langle m(y), \theta\rangle}\varepsilon$.
In that case, it holds that

$$g_I(y,\theta) = \frac{\mu_4 - 1}{4}\frac{m(y)m(y)'}{\langle m(y), \theta\rangle^2} = \frac{\mu_4 - 1}{2} g_V(y,\theta).$$

In spectral time series analysis, linear locally stationary processes play an important role. Here, we discuss the conditions that have to be imposed on a model of the form (4.1.5) introduced by Dahlhaus and Polonik (2009) such that the main assumptions in our theorems are fulfilled. A very general formulation can be found in Lemma 4.5.10 in the appendix. Here, we only consider the prominent example of the tvARMA process.

Example 4.2.10 (tvARMA(r,s) processes). *Assume that there are functions $a_j, b_k, \sigma : [0,1] \to \mathbb{R}$ ($j = 0, ..., r$, $k = 0, ..., s$) such that $a_0 \equiv b_0 \equiv 1$, $a_j(u) = a_j(0), b_k(u) = b_k(0)$ for $u < 0$ and*

$$\sum_{j=0}^{r} a_j\big(\frac{t}{n}\big)X_{t-j,n} = \sum_{k=0}^{s} b_k\big(\frac{t}{n}\big)\sigma\big(\frac{t-k}{n}\big)\varepsilon_{t-k}, \quad t = 1, ..., n$$

Define $p_\theta(w) := \sum_{j=0}^{r} a_j w^j$ and $q_\theta(w) := \sum_{k=0}^{s} b_k w^k$, $\tilde{\Theta}$ a convex closed subset of

$$\{\theta = (a_1, ..., a_r, b_1, ..., b_s, \sigma) \in \mathbb{R}^{r+s} \times [\sigma_{min}, \sigma_{max}] : \\ p_\theta(w) \neq 0, q_\theta(w) \neq 0 \text{ for } 0 < |w| \leq 1 + \rho\},$$

and

$$\Theta := \{\theta \in \tilde{\Theta} : \text{the zeros of } p_\theta \text{ and } q_\theta \text{ differ by at least } \rho_2\}.$$

with some $\rho, \rho_2 > 0$ and $\sigma_{max} > \sigma_{min} > 0$. Assume that $\theta_0 = (a_1, ..., a_r, b_1, ..., b_s, \sigma) \in \Sigma(\beta, L)$, $\mathbb{E}\varepsilon_0 = 0$, $\mathbb{E}\varepsilon_0^2 = 1$ and $\mathbb{E}\varepsilon_0^4$ is known. Assume that condition (b) from Example 4.2.7 holds.
Then Assumptions 4.2.2, 4.2.3, 4.2.4 and 4.2.6 are fulfilled with $M = 2$. In that case, it holds that

$$g_I(y,\theta) = \frac{1}{\sigma^2}\begin{pmatrix} \mu(y,\theta)\mu(y,\theta)' & 0 \\ 0 & \mu_4 - 1 \end{pmatrix}, \quad g_V(y,\theta) = \frac{1}{\sigma^2}\begin{pmatrix} \mu(y,\theta)\mu(y,\theta)' & 0 \\ 0 & 2 \end{pmatrix},$$

where

$$\mu(y,\theta) = \Big(\big(\sum_{k=1}^{\infty}(B(\theta)^k)_{11}y_{i+k}\big)_{i=1,...,r}, \big(-\sum_{k=1}^{\infty}\sum_{l=1}^{k}(B(\theta)^{l-1})_{11}(B(\theta)^{k-l})_{i1}\sum_{j=0}^{s}a_j y_{j+k}\big)_{i=1,...,s}\Big)',$$

$B(\theta) := \begin{pmatrix} (-b_i)_{i=1,...,s-1} & -b_s \\ Id_{s-1} & 0 \end{pmatrix}$ *and $a_0 := 1$.*

Remark 4.2.11 (The choice of $\tilde{\Theta}$ in Example 4.2.10). *Note that a convex superset $\tilde{\Theta}$ is only needed since we are working in a non-asymptotic framework. A suitable choice of $\tilde{\Theta}$ is given by*

$$\tilde{\Theta} := \{\theta = (a_1, ..., a_r, b_1, ..., b_s, \sigma) \in \mathbb{R}^{r+s} \times [\sigma_{min}, \sigma_{max}] : \sum_{j=1}^{r}|a_j| \leq 1 - \rho_3, \sum_{j=1}^{s}|b_j| \leq 1 - \rho_4\}$$

with $\rho_3 := 1 - (1-\rho)^{-(r+1)} > 0, \rho_4 := 1 - (1-\rho)^{-(s+1)} > 0$. The proof is an easy consequence of the maximum principle from complex analysis: If $\sum_{j=1}^{r}|a_j| < 1 + \rho_3$, we have for $w = (1+\rho)e^{i\lambda}$ with arbitrary $\lambda \in [-\pi, \pi]$: $\big|\sum_{j=1}^{r} a_j w^j\big| \leq \sum_{j=1}^{r}|a_j|(1+\rho)^j \leq \sum_{j=1}^{r}|a_j|(1+\rho)^p \leq (1-\rho_3)(1+\rho)^p < 1$. This shows that $\sum_{j=1}^{r} a_j w^j$ does not attain 1 for $|w| \leq 1 + \rho$ by the maximum principle, hence $p_\theta(w)$ cannot have zeros for $|w| \leq 1 + \rho$. A similar argumentation leads to the same result for q_θ.

4.3 A simulation study

4.3.1 Differences between the minimizers of $b \mapsto d_{u,\Xi}(\hat{\theta}_b, \theta_0)$

Here we briefly discuss the differences between the bandwidth selectors $\hat{b}_{opt,\Xi}(u)$ which minimize $b \mapsto d_{u,\Xi}(\hat{\theta}_b, \theta_0)$ for $\Xi \in \{\mathrm{Id}, V(u)\}$. For this, we simulated the tvAR(1) model $X_{t,n} = a_0\big(\frac{t}{n}\big)X_{t-1,n} + \sigma_0\big(\frac{t}{n}\big)\varepsilon_t$, where $\theta_0 = (a_0, \sigma_0)'$ are step functions (see Figure 4.3.1) and $\varepsilon_t \sim N(0,1)$ standard Gaussian distributed random variables. We chose $H_n = \{1.5^{-k} : k \in \mathbb{N}_0\} \cap [\frac{1}{n}, 1]$ and the Epanechnikov kernel $K(x) = \frac{3}{2}(1-(2x)^2)\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(x)$ for estimation. We assume that the property $\mathbb{E}\varepsilon_0^4 = 3$ of the errors is known which then leads to full knowledge of g_V and g_I. We chose $n = 500$ to be the length of the observed time series and repeated the simulation $N = 1000$ times. At each time point $u \in \{\frac{t}{n} : t = 1, ..., n\}$ we determined $\hat{b}_{opt,\Xi}(u)$. The 5% and 95% quantile curves of the corresponding estimators $\hat{\theta}_{\hat{b}_{opt,\Xi}(u)}(u)$ are plotted in Figure 4.3.1. Furthermore we have visualized the chosen bandwidths for the two procedures and the ratio of $\frac{I_{11}(u)}{I_{22}(u)}$, where $I(u) = V(u) = \mathrm{diag}(\frac{1}{1-a_0(u)^2}, \frac{2}{\sigma_0(u)^2})$ is a diagonal matrix.

It can be seen that the differences of the two estimators $\hat{\theta}_{\hat{b}_{opt,\Xi}(u)}(u)$ ($\Xi \in \{\mathrm{Id}, V(u)\}$) are larger if the ratio $\frac{I_{11}(u)}{I_{22}(u)}$ is far away from 1. For $u \le 0.5$, the Kullback-Leibler-type distance $d_{u,V(u)}(\hat{\theta}_b, \theta_0)$ puts a lot of weight to the second component σ_0 and thus, the estimator of σ_0 associated with $\hat{b}_{opt,V(u)}(u)$ leads to a more precise estimation of σ_0 than the estimator associated with $\hat{b}_{opt,\mathrm{Id}}(u)$. The behavior is mirrored for the first component a_0. In the case $u \in [0.5, 0.75]$ the ratio $\frac{I_{11}(u)}{I_{22}(u)} \approx 1$ and thus the estimators behave nearly the same. For $u \ge 0.75$, the ratio is greater than 1 and thus the behavior is mirrored to the case $u \le 0.5$. It should be noted that in this example, the difference of the two bandwidth selectors is not very large as long as the ratio $\frac{I_{11}(u)}{I_{22}(u)} \in [\frac{1}{5}, 5]$. Only for very large or very small values of this ratio or significant differences in the smoothness properties of the two components of θ_0 we observe significant differences in the choice of $\hat{b}_{opt,\Xi}(u)$ and $\hat{\theta}_{\hat{b}_{opt,\Xi}(u)}(u)$. In simulations, no method conquered the other in view of stability in the situation that B_n contains very small bandwidths. For this reason we will only consider $\Xi = \mathrm{Id}$ in the following simulations.

4.3.2 The estimation procedure

We discuss the quality of our procedure in four different models, the tvTAR(1), the tvAR(1), tvMA(1) and the tvARCH(1) model. In all four models we generate $N = 1000$ replications of a time series of length $n = 1000$. Since the constant $C_P(\mathrm{Id}) = 2^8$ from (4.1.11) which is used in the proof usually leads to too conservative estimators, one has to find meaningful values of $C_P(\mathrm{Id})$ which usually depend on the chosen time series model and the parameter space Θ. The same holds for the constant $c_b(\mathrm{Id})$ in (4.1.9). In practise, one has to find good values for $C_P(\mathrm{Id})$ and $c_b(\mathrm{Id})$ with training data before applying the algorithm to the test data set. A good starting point seems to be to define $c_b(\mathrm{Id}) \approx 1$ and $C_P(\mathrm{Id}) \approx 1$. Here, we analyze the following models:

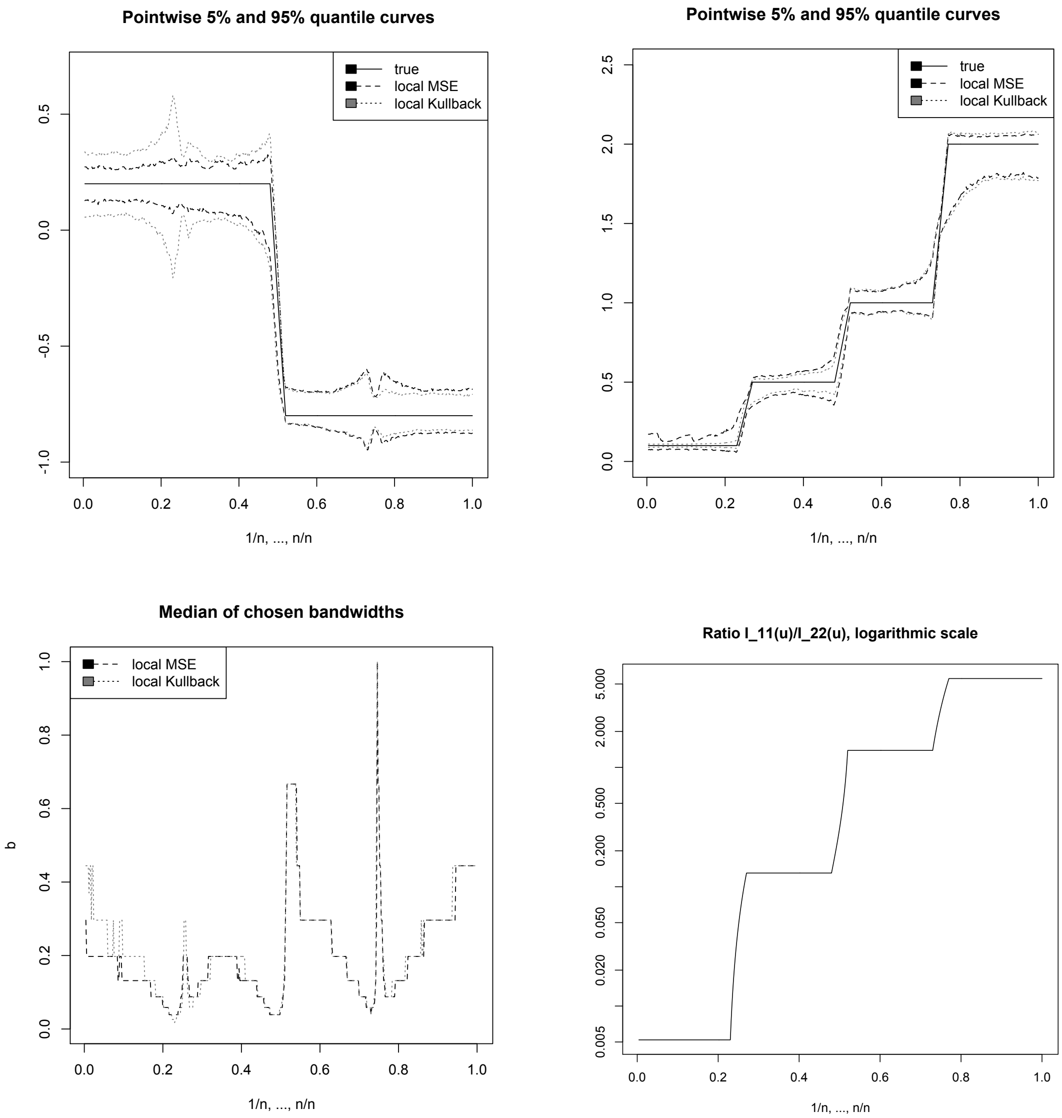

Figure 4.1: Discussion of the two distance measures $d_{u,\Xi}(\hat{\theta}_b, \theta_0)$, $\Xi \in \{\mathrm{Id}, V(u)\}$. Top: Left/Right: Solid lines are true curves a_0 (left) and σ_0 (right). 5% and 95% quantile curves the two components of the estimator $\hat{\theta}_{\hat{b}_{opt,\Xi}(u)}(u)$. Bottom: Left: Median of the chosen bandwidths $\hat{b}_{opt,\Xi}(u)$. Right: The ratio $\frac{I_{11}(u)}{I_{22}(u)}$ in logarithmic scale.

- tvTAR(1): $X_{t,n} = a\left(\frac{t}{n}\right)X_{t-1,n}^{+} + b\left(\frac{t}{n}\right)X_{t-1,n}^{-} + \varepsilon_t$, $\theta = (a, b)$
- tvAR(1): $X_{t,n} = a\left(\frac{t}{n}\right)X_{t-1,n} + \sigma\left(\frac{t}{n}\right)\varepsilon_t$, $\theta = (a, \sigma)$,
- tvMA(1): $X_{t,n} = \sigma\left(\frac{t}{n}\right)\varepsilon_t + a\left(\frac{t}{n}\right)\sigma\left(\frac{t-1}{n}\right)\varepsilon_{t-1}$, $\theta = (a, \sigma)$,
- tvARCH(1): $X_{t,n} = \left(a\left(\frac{t}{n}\right) + b\left(\frac{t}{n}\right)X_{t-1,n}^2\right)^{1/2}\varepsilon_t$, $\theta = (a, b)$.

In each case, we assume $\varepsilon_t \sim N(0,1)$ with known second and fourth moment $\mathbb{E}\varepsilon_0^2 = 1$, $\mathbb{E}\varepsilon_0^4 = 3$. In the grid, we used $a = 1.5$. The following constants $C_P(\text{Id})$, $c_b(\text{Id})$ were chosen for the simulations:

model	tvTAR	tvAR	tvMA	tvARCH
$C_P(\text{Id})$	1.5	0.6	0.7	2.0
$c_b(\text{Id})$	2.0	1.0	2.0	2.0

The simulation results are given in Figures 4.2 and 4.3. The true curve of the first (left) and second (right) component of the estimators of θ_0 is plotted together with the 5%- and 95% quantile curves of the estimator $\hat{\theta}_{\hat{b}_{\text{Id}}(u)}(u)$, the (unknown) optimal local estimator $\hat{\theta}_{\hat{b}_{opt,\text{Id}}(u)}(u)$ and an (unknown) optimal global estimator $\hat{\theta}_{\hat{b}_{opt,\text{Id}}}(u)$ which is chosen by minimizing the averaged squared error $b \mapsto \frac{1}{n}\sum_{t=1}^{n} |\hat{\theta}_b(t/n) - \theta_0(t/n)|_2^2$. Furthermore, the pointwise median of the local bandwidth $\hat{b}_{opt,\text{Id}}(u)$ chosen is shown. It turns out that a good choice of $c_b(\text{Id})$ is crucial to obtain a stable procedure. If B_n contains too small elements, the bandwidth selector tends to choose them occasionally which leads to 'artefacts' in the estimator $\hat{\theta}_{\hat{b}_{\text{Id}}(u)}(u)$. The reason for this can be seen in the proof of Theorem 4.1.4: Bandwidths $b < c_b(\text{Id})\frac{\log(n)^2}{n}$ do not longer guarantee that rare events occur with negligible probability. The selection routine is relatively insensitive to the choice of $C_P(\text{Id})$. In Figures 4.2, 4.3 one can see that for the defined step functions, our method $\hat{b}_{\text{Id}}(u)$ outperforms the estimators associated to the global optimal bandwidth selector $\hat{b}_{opt}$ and works reasonably well compared with the local optimal choice $\hat{b}_{\text{Id},opt}(u)$. In general, the quality of the bandwidth selector $\hat{b}_{\text{Id}}(u)$ depends on the quality of the corresponding quasi-maximum likelihood estimator. Especially in ARCH(1) models, the parameter estimators obtained by the maximum likelihood approach have a very high variance which compromises bandwidth selection.

4.4 Concluding remarks

In this chapter, we proposed a data adaptive bandwidth selection procedure for parameter curves in locally stationary processes. We proved that the bandwidth selector is minimax optimal over Hoelder classes up to a log factor which is common in local procedures. As seen in the simulations, the method is applicable to a wide range of popular time series.

The quality of the selection routine depends strongly on the quality of the corresponding quasi maximum likelihood estimator. Therefore, the method works better in tvAR models than in tvARCH models (where the maximum likelihood estimators have a very high variance).

We conjecture that a generalization to multivariate time series is straightforward. Moreover, it is not hard to allow for a partially known parameter curve, i.e. $\tilde{X}_t(u) = G_{\varepsilon_t}(\tilde{Y}_{t-1}(u), \theta_0(u), u)$ in (4.1.3) can depend on u not only through θ_0. To guarantee the same results in this case, a modification of the bias expansions Lemma 4.5.6 and 4.5.7 is necessary. To reduce technicality, we omitted the details in this paper. It should be

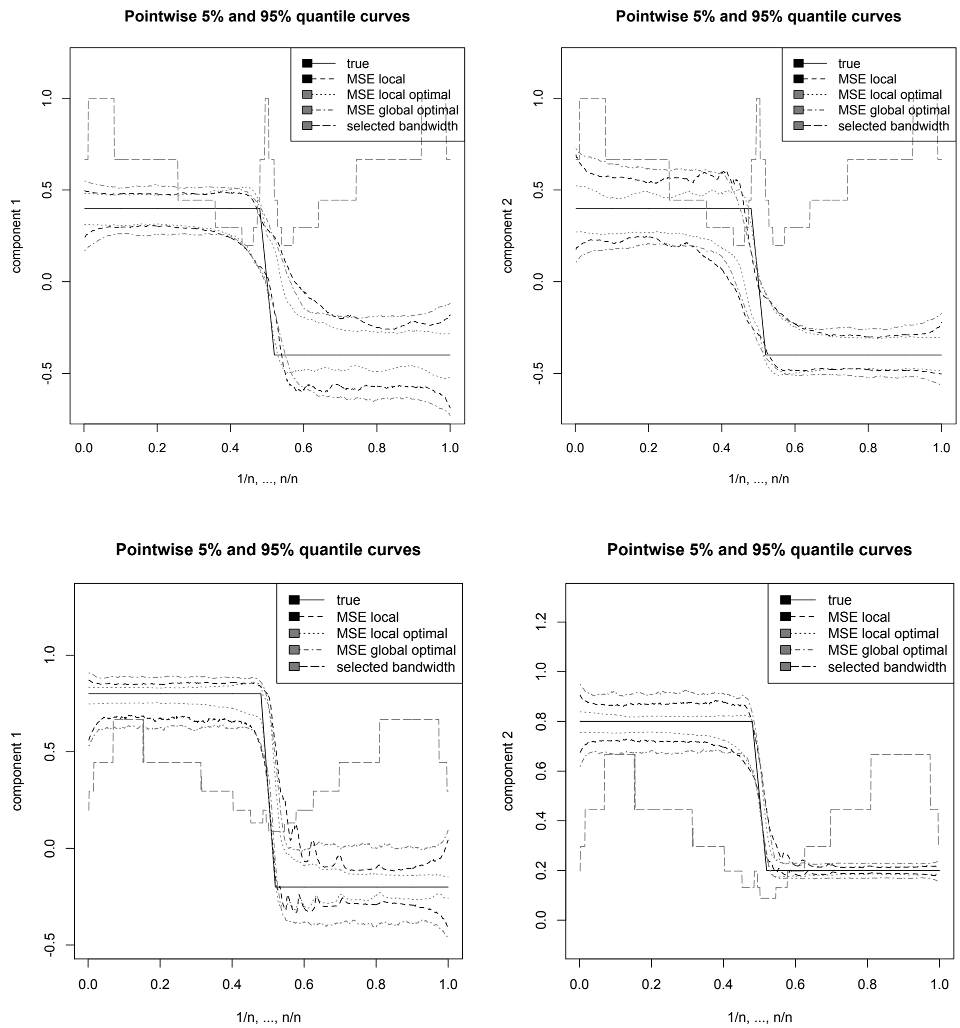

Figure 4.2: 1st row: tvTAR(1) model, 2nd row: tvAR(1).

also possible to relax the differentiability assumption of the model in θ by convoluting the likelihood with a twice differentiable function.

4.5 Lemmas and Proofs

4.5.1 Stationary approximation and exponential inequalities

In this section, we To shorten some expressions, let us introduce the following notations. Since M, χ are fixed in this chapter, we do not mark the dependency on this quantities. Let $C_X := \sum_{k=0}^{\infty} \delta(k) + \sup_{u \in [0,1]} |\mathbb{E}\tilde{X}_0(u)| < \infty$. We then have

$$\|\tilde{X}_t(u)\|_q = \|\tilde{X}_t(u) - \mathbb{E}\tilde{X}_t(u)\|_q + |\mathbb{E}\tilde{X}_t(u)| \leq C_X N_\alpha(q). \tag{4.5.1}$$

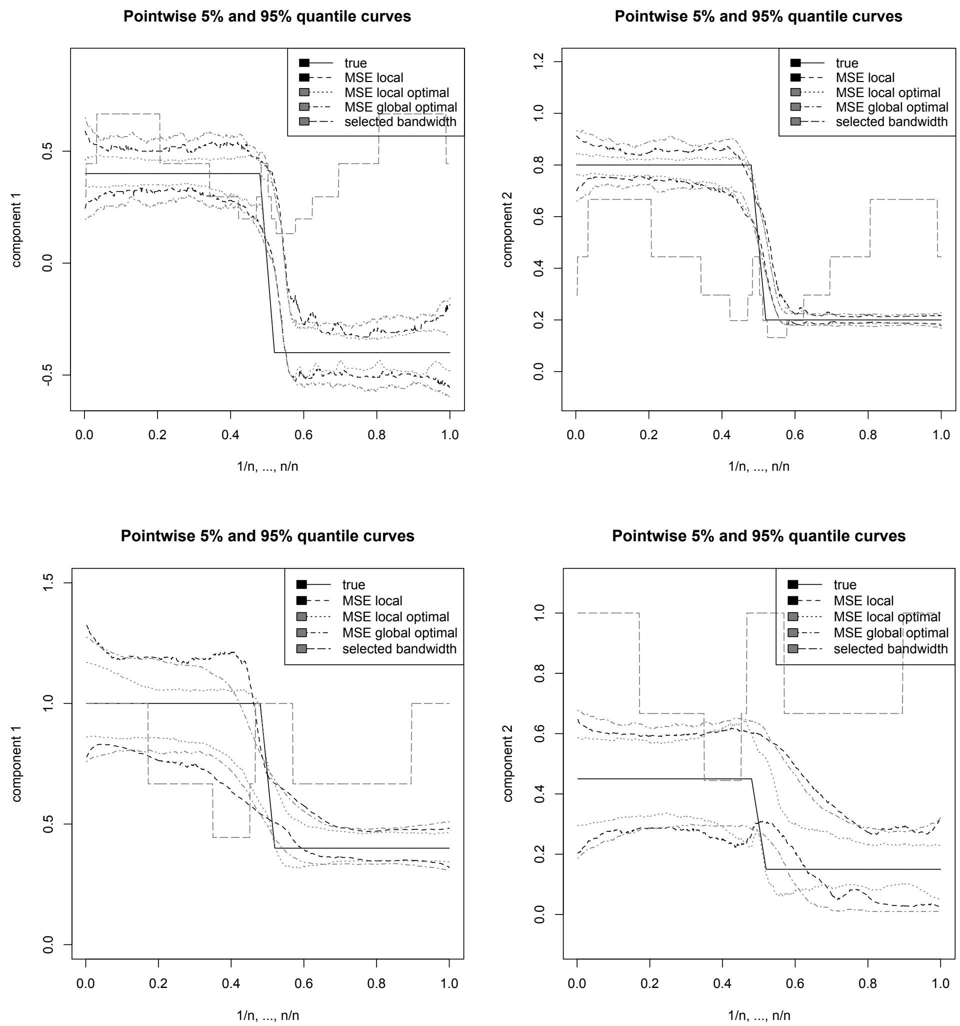

Figure 4.3: 1st row: tvMA(1), 2nd row: tvARCH(1).

Furthermore, set $E_{X,1} := 1 + 2|\chi|_1^{M-1} C_X^{M-1}$ and $E_{X,2} := C_X |\chi|_1 (1 + |\chi|_1^{M-1} C_X^{M-1})$ and $\Theta_{max} := \sup_{\theta \in \Theta} |\theta|_2$.

Lemma 4.5.1 (The stationary approximation)**.** *Let* $g \in \mathcal{L}(M, \chi, C_z, C_\theta)$*. Define*

$$S_{n,b}(g(\cdot,\theta), u) := \frac{1}{K_{n,b}(u)} \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) \cdot \big\{ g(Y_{t,n}^c, \theta) - g(\tilde{Y}_t(t/n)^c, \theta) \big\}.$$

Suppose that Assumption 4.2.3 and 4.2.5 hold. Assume that $\sum_{j=1}^{\infty} j \chi_j < \infty$*. Then there exists a constant* $C_S > 0$ *not depending on* n, b *such that:*

$$\big\| \sup_{\theta \in \Theta} |S_{n,b}(g(\cdot,\theta), u)| \big\|_2 \le C_S \cdot n^{-\beta'}.$$

Proof of Lemma 4.5.1: First note that for all $t = 1, ..., n$ it holds that:

$$\begin{aligned}\|X_{t,n}\|_{2M} &\leq \|X_{t,n} - \tilde{X}_t(t/n)\|_{2M} + \|\tilde{X}_t(t/n)\|_{2M} \leq C_{B,1} n^{-\beta'} + C_X N_\alpha(2M) \\ &\leq C_{B,1} + C_X N_\alpha(2M).\end{aligned}$$

Because $g \in \mathcal{L}(M, \chi, C)$, we have component-wise by Hoelder's inequality:

$$\begin{aligned}&\big\| \sup_{\theta\in\Theta} |g(Y_{t,n}^c, \theta) - g(\tilde{Y}_t(t/n)^c, \theta)|\big\|_2 \\ \leq\ & C_z \textstyle\sum_{j=1}^t \chi_j \big\|X_{t-j+1,n} - \tilde{X}_{t-j+1}\big(\frac{t}{n}\big)\big\|_{2M} \\ &\qquad \cdot \Big(1 + \big(\textstyle\sum_{j=1}^t \chi_j \|X_{t-j+1,n}\|_{2M}\Big)^{M-1} + \Big(\textstyle\sum_{j=1}^t \chi_j \big\|\tilde{X}_{t-j+1}\big(\frac{t}{n}\big)\big\|_{2M}\Big)^{M-1}\Big).\end{aligned} \tag{4.5.2}$$

By Assumption 4.2.3 it holds that $\|X_{t-j+1,n} - \tilde{X}_{t-j+1}(\frac{t-j+1}{n})\|_{2M} \leq C_{B,1} n^{-\beta'}$ and thus $\|X_{t-j+1}\|_{2M} \leq C_{B,1} + C_X N_\alpha(2M)$. Furthermore,

$$\begin{aligned}&\big\|X_{t-j+1,n} - \tilde{X}_{t-j+1}\big(\frac{t}{n}\big)\big\|_{2M} \\ \leq\ & \big\|X_{t-j+1,n} - \tilde{X}_{t-j+1}\big(\frac{t-j+1}{n}\big)\big\|_{2M} + \big\|\tilde{X}_{t-j+1}\big(\frac{t-j+1}{n}\big) - \tilde{X}_{t-j+1}\big(\frac{t}{n}\big)\big\|_{2M} \\ \leq\ & C_{B,1} n^{-\beta'} + C_{B,2} N_\alpha(2M)\big(\frac{j+1}{n}\big)^{\beta'} \leq (C_{B,1} + C_{B,2} N_\alpha(2M)) n^{-\beta'} \cdot (j+1).\end{aligned}$$

We conclude that (4.5.2) is bounded by

$$C_z \sum_{j=1}^\infty (j+1)\chi_j \cdot (C_{B,1} + C_{B,2} N_\alpha(2M))(1 + 2(|\chi|_1 (C_{B,1} + C_X N_\alpha(2M)))^{M-1}) n^{-\beta'}.$$

□

Lemma 4.5.2 (The crop approximation)**.** *Let $g \in \mathcal{L}(M, \chi, C_z, C_\theta)$. Define*

$$C_{n,b}(g(\cdot, \theta), u) := \frac{1}{K_{n,b}(u)} \sum_{t=1}^n K\Big(\frac{t/n - u}{b}\Big) \cdot \big\{g(\tilde{Y}_t(t/n)^c, \theta) - g(\tilde{Y}_t(t/n), \theta)\big\}.$$

Assume that $\rho(t) := \sum_{j=t+1}^\infty \chi_j < \infty$ and $\sum_{t=1}^\infty \rho(t) < \infty$, $\rho(n) \leq \frac{C_\rho}{n}$. Suppose that Assumption 4.2.3 and 4.2.5 holds. Then, for each $u \in (0, 1]$ there exists a constant $C_C(u) > 0$ not depending on n, b such that

$$\big\| \sup_{\theta\in\Theta} |C_{n,b}(g(\cdot, \theta), u)|\big\|_2 \leq \frac{C_C(u)}{n}.$$

Proof of Lemma 4.5.2: Because $g \in \mathcal{L}(M, \chi, C_z, C_\theta)$, we have component-wise for all $u \in [0, 1]$ by Hoelder's inequality:

$$\begin{aligned}&\big\| \sup_{\theta\in\Theta} |g(\tilde{Y}_t(u)^c, \theta) - g(\tilde{Y}_t(u), \theta)|\big\|_2 \\ \leq\ & C_z \textstyle\sum_{j=t+1}^\infty \chi_j \|\tilde{X}_{t-j+1}(u)\|_{2M} \cdot \big(1 + 2\big(\textstyle\sum_{j=1}^\infty \chi_j \|\tilde{X}_{t-j+1}(u)\|_{2M}\big)^{M-1}\big) \\ \leq\ & C_z \big(\textstyle\sum_{j=t+1}^\infty \chi_j\big) \cdot C_X N_\alpha(2M)\big(1 + 2(|\chi|_1 C_X N_\alpha(2M))^{M-1}\big) =: \rho(t) \cdot D_X.\end{aligned}$$

We conclude for $u \in (0,1]$:

$$\big\| \sup_{\theta\in\Theta} |C_{n,b}(g,u,\theta)| \big\|_2 \leq \frac{D_X}{K_{n,b}(u)} \sum_{t=1}^{n} \Big| K\Big(\frac{t/n-u}{b}\Big)\Big| \rho(t). \tag{4.5.3}$$

There are two cases: If $b \leq u$, then the term in the sum over t is only different from 0 if $\frac{t}{n} \geq u - \frac{b}{2} \geq \frac{u}{2}$ which implies $t \geq \frac{u}{2} \cdot n$. Then $\rho(t) \leq \frac{C_\rho}{t} \leq \frac{2C_\rho}{u} \cdot \frac{1}{n}$ and thus (4.5.3) can be bounded by

$$\frac{D_X}{K_{n,b}(u)} \sum_{t=1}^{n} \Big| K\Big(\frac{t/n-u}{b}\Big)\Big| \cdot \frac{2C_\rho}{u} \cdot \frac{1}{n} \leq D_X C_\rho \cdot \frac{2}{u} \cdot \frac{1}{n}.$$

In the case $b \geq u$, we have $\sum_{t=1}^{n} |K\big(\frac{t/n-u}{b}\big)| \rho(t) \leq \|K\|_\infty \sum_{t=1}^{\infty} \rho(t)$ and, by Assumption 4.2.5, $K_{n,b}(u) \geq c_0 \cdot (nb) \geq c_0 \cdot n \cdot \frac{u}{2}$, thus (4.5.3) is bounded by

$$D_X \cdot \frac{2}{u} \cdot \frac{|K|_\infty}{c_0} \cdot \sum_{t=1}^{\infty} \rho(t) \cdot \frac{1}{n}.$$

Thus, the assertion holds with $C_C(u) := \frac{2D_X}{u} \cdot \max\{C_\rho, \frac{|K|_\infty}{c_0} \sum_{t=1}^{\infty} \rho(t)\}$. □

Lemma 4.5.3 (Exponential moment)**.** *Assume that $g : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{dim}$ fulfills $g \in \mathcal{L}(M, \chi, C_z)$. Define $\tau_2 := (\alpha M)^{-1}$. Suppose that Assumption 4.2.3 and 4.2.5 hold. Then it holds for arbitrary $q \geq 2$ that:*

$$\begin{aligned} \big\| |g(\tilde{Y}_t(u))|_2 \big\|_q &\leq |C_z|_2 E_{X,2} \cdot N_\alpha(qM)^M + |g(0)|_2, \\ \mathbb{E} \exp\Big(\frac{1}{2}\big(\frac{|g(\tilde{Y}_t(u))|_2}{|C_z|_2 E_{X,2} + |g(0)|_2}\big)^{\tau_2}\Big) &\leq C_E. \end{aligned}$$

In the special case $\alpha = 0$ it holds that $|g(\tilde{Y}_t(u))|_2 \leq |C_z|_2 E_{X,2} + |g(0)|_2$ a.s.

Proof of Lemma 4.5.3: Fix an index i. Then we have

$$\begin{aligned} \|g_i(\tilde{Y}_t(u)) - g_i(0)\|_q &\leq C_{z,i} \sum_{j=1}^{\infty} \chi_j \|\tilde{X}_{t-j+1}(u)\|_{qM} \cdot \big(1 + |\chi|_1^{M-1} \|\tilde{X}_t(u)\|_{qM}^{M-1}\big) \\ &\leq C_{z,i} |\chi|_1 C_X N_\alpha(qM) \cdot \big(1 + |\chi|_1^{M-1} C_X^{M-1} N_\alpha(qM)^{M-1}\big) \\ &\leq C_{z,i} N_\alpha(qM)^M \cdot E_{X,2}. \end{aligned}$$

For the second part, define $D_X := |C_z|_2 E_{X,2} + |g(0)|_2$ and $\lambda = (2D_X^{\tau_2})^{-1}$. It holds that $\mathbb{E} \exp\big(\lambda |g(\tilde{Y}_t(u))|_2^{\tau_2}\big) = \sum_{q=0}^{\infty} \frac{\lambda^q \||g(\tilde{Y}_t(u))|_2\|_{\tau_2 q}^{\tau_2 q}}{q!}$. If $\tau_2 q \geq 2$, we have

$$\||g(\tilde{Y}_t(u))|_2\|_{\tau_2 q}^{\tau_2 q} \leq D_X^{\tau_2 q} \cdot \Gamma(\alpha q \tau_2 M + 2) = D_X^{\tau_2 q} \Gamma(q+2).$$

This shows $\sum_{\tau_2 q \geq 2} \frac{\lambda^q \||g(\tilde{Y}_t(u))|_2\|_{\tau_2 q}^{\tau_2 q}}{q!} \leq \sum_{\tau_2 q \geq 0}^{\infty} (\lambda D_X^{\tau_2})^q \cdot \frac{\Gamma(q+2)}{\Gamma(q+1)} = \sum_{\tau_2 q \geq 2} \frac{q+1}{2^q} \leq 4$.
In the case $\tau_2 q < 2$, we have

$$\||g(\tilde{Y}_t(u))|_2\|_{\tau_2 q}^{\tau_2 q} \leq \||g(\tilde{Y}_t(u))|_2\|_2^{\tau_2 q} \leq D_X^{\tau_2 q} \Gamma(2\alpha M + 2)^{\tau_2 q/2} \leq D_X^{\tau_2 q} \Gamma(2\alpha M + 2).$$

This shows $\sum_{\tau_2 q < 2} \frac{\lambda^q \||g(\tilde{Y}_t(u))|_2\|_{\tau_2 q}^{\tau_2 q}}{q!} \leq \Gamma(2\alpha M + 2) \sum_{q=0}^{\infty} \frac{2^{-q}}{q!} = \exp(2^{-1})\Gamma(2\alpha M + 2)$. The result is obtained with $C_E := 4 + \exp(2^{-1})\Gamma(2\alpha M + 2)$. □

Lemma 4.5.4 (The empirical process approximation). *Assume that $g : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{dim}$ fulfills $g \in \mathcal{L}(M, \chi, C_z)$. Define $\tau = \tau(\alpha, M) := (\frac{1}{2} + \alpha M)^{-1}$. Suppose that Assumption 4.2.3 and 4.2.5 hold. Define*

$$E_{n,b}(g,u) := \frac{1}{K_{n,b}(u)} \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) \cdot \big\{g(\tilde{Y}_t(t/n)) - \mathbb{E} g(\tilde{Y}_t(t/n))\big\},$$

Then there exists $C_{E,1} > 0$ and $C_{E,2}(C_z) := E_{X,1}|C_z|_2 \sum_{k=0}^{\infty} \xi(\chi, k) > 0$ such that for all $\gamma > 0$:

$$\begin{aligned} \| |E_{n,b}(g,u)|_2 \|_q &\leq (q-1)^{1/2} C_{E,2} F_{n,b}(u)^{-1} \cdot N_\alpha(qM)^M, \\ \mathbb{P}(|E_{n,b}(g,u)|_2 > \gamma) &\leq C_{E,1} \exp\big(-(4e)^{-1}(C_{E,2}^{-1} \cdot F_{n,b}(u) \cdot \gamma)^\tau\big). \end{aligned}$$

Proof of Lemma 4.5.4: By the Hoelder inequality, we have for all $u \in [0,1]$, $\theta \in \Theta$ and each component $i = 1, ..., dim$:

$$\begin{aligned} &\big\| g_i(\tilde{Y}_t(u)) - g_i(\tilde{Y}_t^*(u)) \big\|_q \\ \leq\ & C_{z,i} \textstyle\sum_{j=1}^{t} \chi_j \|\tilde{X}_{t-j+1}(u) - \tilde{X}_{t-j+1}^*(u)\|_{qM} \cdot \big(1 + 2|\chi|_1^{M-1} C_X^{M-1} \cdot N_\alpha(qM)^{M-1}\big) \\ \leq\ & C_{z,i} \textstyle\sum_{j=1}^{t} \chi_j \delta(t-j+1) \cdot N_\alpha(qM)^M E_{X,1}. \end{aligned}$$

So we have shown that the dependence measure fulfills $\delta_q^{g(\tilde{Y}(u))}(k) \leq E_{X,1} C_{z,i} \cdot \xi(\chi, k) \cdot N_\alpha(qM)^M$ which is absolutely summable by Assumption 4.2.3. Note that for $q \geq 2$ and some random vector $v \in \mathbb{R}^d$, we have

$$\begin{aligned} \| |v|_2 \|_q &= \mathbb{E}\big[\big(\textstyle\sum_{j=1}^{d} |v_j|^2\big)^{q/2}\big]^{1/q} = \big\| \textstyle\sum_{j=1}^{d} |v_j|^2 \big\|_{q/2}^{1/2} \leq \big(\textstyle\sum_{j=1}^{d} \|v_j^2\|_{q/2}\big)^{1/2} \\ &\leq \big(\textstyle\sum_{j=1}^{d} \|v_j\|_q^2\big)^{1/2} = |\|v\|_q|_2. \end{aligned}$$

By Theorem 2.1 from Rio (2009) for $q > 2$ (and for $q = 2$ directly by calculating the variance of the following term), we have

$$\begin{aligned} \| |E_{n,b}(g,u)|_2 \|_q &\leq \Big| \Big\| \frac{1}{K_{n,b}(u)} \Big| \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) \big\{g(\tilde{Y}_t(t/n)) - \mathbb{E} g(\tilde{Y}_t(t/n))\big\} \Big\|_q \Big|_2 \\ &\leq \sum_{k=0}^{\infty} \frac{1}{K_{n,b}(u)} \Big| \Big\| \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) P_{t-k} g(\tilde{Y}_t(t/n)) \Big\|_q \Big|_2 \\ &\leq \sum_{k=0}^{\infty} \frac{1}{K_{n,b}(u)} (q-1)^{1/2} \Big| \Big(\sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big)^2 \|P_{t-k} g(\tilde{Y}_t(t/n))\|_q^2 \Big)^{1/2} \Big|_2 \\ &\leq (q-1)^{1/2} F_{n,b}(u)^{-1} \cdot |C_z|_2 E_{X,1} \sum_{k=0}^{\infty} \xi(\chi, k) \cdot N_\alpha(qM)^M. \end{aligned}$$

Define $D(u) := E_{X,1}|C_z|_2 F_{n,b}(u)^{-1} \cdot \sum_{k=0}^{\infty} \alpha(\chi, k)$. By Stirling's formula, we have for all $x \geq 1$:

$$\sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x} \leq \Gamma(x) \leq e^{1/12} \cdot \sqrt{2\pi} x^{x-\frac{1}{2}} e^{-x}.$$

By Markov's inequality, we have for $\varepsilon > 0$:

$$\mathbb{P}(|E_{n,b}(g,u)|_2 \geq \varepsilon) \leq e^{-\lambda\varepsilon^\tau}\mathbb{E}[e^{\lambda|E_{n,b}(g,u)|_2^\tau}] = e^{-\lambda\varepsilon^\tau}\sum_{q=0}^{\infty}\frac{\lambda^q\||E_{n,b}(g,u)|_2\|_{\tau q}^{\tau q}}{q!}.$$

In the case $\tau q \geq 2$, we have

$$\frac{\lambda^q\||E_{n,b}(g,u)|_2\|_{\tau q}^{\tau q}}{q!} \leq \frac{\lambda^q}{\Gamma(q+1)}(\tau q)^{\frac{\tau q}{2}}D(u)^{\tau q}\cdot\Gamma(\alpha M\tau q+2).$$

Note that $\alpha M\tau \leq 1$ and $\tau(\alpha M + \frac{1}{2}) = 1$, thus

$$\begin{aligned} q^{\frac{\tau q}{2}}\frac{\Gamma(\alpha M\tau q+2)}{\Gamma(q+1)} &\leq (q+2)^{\frac{\tau q}{2}}\cdot\frac{(\alpha M\tau q+2)^{\alpha M\tau q+\frac{3}{2}}e^{-(\alpha M\tau q+2)}e^{1/12}}{(q+1)^{q+\frac{1}{2}}e^{-(q+1)}} \\ &= e^{1/12}(q+2)\cdot\Big(\frac{q+2}{q+1}\Big)^{q+\frac{1}{2}}e^{-1}e^{q(1-\alpha M\tau)} \\ &\leq e^{1/12}(q+2)e^q. \end{aligned}$$

Define $\lambda := (4e)^{-1}D(u)^{-\tau}$. Note that $\tau \leq 2$, thus $\tau^{\tau/2} \leq 2$, this gives

$$\sum_{q\geq 2/\tau}\frac{\lambda^q\||E_{n,b}(g,u)|_2\|_{\tau q}^{\tau q}}{q!} \leq e^{1/12}\cdot\sum_{q\geq 2/\tau}(q+2)(\lambda\cdot 2eD(u)^\tau)^q \leq e^{1/12}\sum_{q\geq 2/\tau}\frac{q+2}{2^q} \leq 4e^{1/12}.$$

In the case $\tau q < 2$, we have

$$\begin{aligned} \frac{\lambda^q\||E_{n,h}(g,u)|_2\|_{\tau q}^{\tau q}}{q!} &\leq \frac{\lambda^q\||E_{n,b}(g,u)|_2\|_2^{\tau q}}{q!} \leq \frac{\lambda^q}{q!}D(u)^{\tau q}\cdot\Gamma(2\alpha M+2)^{\frac{\tau q}{2}} \\ &\leq \frac{(4e)^{-q}}{q!}\cdot\Gamma(2\alpha M+2), \end{aligned}$$

thus $\sum_{q<2/\tau}\frac{\lambda^q\||E_{n,b}(g,u)|_2\|_{\tau q}^{\tau q}}{q!} \leq \exp((4e)^{-1})\Gamma(2\alpha M+2)$. So the result is obtained with $C_{E,1} := 4e^{1/2} + \exp((4e)^{-1})\Gamma(2\alpha M+2)$ and $C_{E,2}(C_z)$ as given in the Lemma. □

Lemma 4.5.5 (The uniform empirical process approximation). *Assume that $g : \mathbb{R}^{\mathbb{N}} \times \Theta \to \mathbb{R}^{dim}$ fulfills $g \in \mathcal{L}(M,\chi,C_z,C_\theta)$. Suppose that Assumption 4.2.3 holds. Recall the definition of $C_{E,2}(C_z)$ from Lemma 4.5.4. Fix some $\gamma > 0$, and assume that*

$$b \geq b_*(C_z,\gamma) := \frac{\log(n)^{1+2\alpha M}}{n}\Big(\frac{|K|_\infty C_{E,2}(C_z)}{c_0\gamma}\cdot\big[4e\cdot(d+1)\big]^{\frac{1}{2}+\alpha M}\Big)^2. \qquad (4.5.4)$$

Then there exists a constant $C_{emp} = C_{emp}(\gamma) > 0$ not depending on b, n such that

$$\mathbb{P}\Big(\sup_{\theta\in\Theta}|E_{n,b}(g,u)|_2 > \gamma\Big) \leq C_{emp}\cdot n^{-1}.$$

Proof of Lemma 4.5.5: Choose $c_n := n^{-1}$. Let Θ_n be the smallest discretization of $\Theta \subset \mathbb{R}^d$ such that for each $\theta \in \Theta$ there exists $\theta' \in \Theta_n$ with $|\theta - \theta'|_2 \leq c_n$. Then $|\Theta_n| \leq (2 \cdot \mathrm{diam}(\Theta) + c_1^{-1})^d \cdot c_n^{-d} =: C_\Theta c_n^{-d}$, see Van de Geer (2000), Lemma 2.5. By Markov's inequality, we have

$$\Big\| \sup_{\theta \in \Theta} |E_{n,b}(g(\cdot,\theta),u)|_2 \Big\|_1 \leq |K|_\infty |C_\theta|_2 \cdot \Big(1 + |\chi|_1^M C_X^M\Big) \cdot c_n.$$

Thus

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\theta \in \Theta} |E_{n,b}(g(\cdot,\theta),u)|_2 > \gamma \Big) \\
\leq\ & \mathbb{P}\Big(\sup_{|\theta - \theta'|_2 \leq c_n} |E_{n,b}(g(\cdot,\theta),u) - E_{n,b}(g(\cdot,\theta'),u)|_2 > \gamma \Big) \\
& \qquad + |\Theta_n| \cdot \sup_{\theta \in \Theta} \mathbb{P}(|E_{n,b}(g(\cdot,\theta),u)|_2 > \gamma) \\
\leq\ & \frac{|K|_\infty}{\gamma} |C_\theta|_2 \cdot \big(1 + |\chi|_1^M C_X^M\big) \cdot c_n + C_\Theta C_{E,1} c_n^{-d} \cdot \exp\big(-(4e)^{-1} (C_{E,2}^{-1} \cdot F_{n,b}(u) \cdot \gamma)^\tau \big) \\
=\ & O(n^{-1}),
\end{aligned}$$

where the last equality is due to the fact that $F_{n,b}(u) \geq \frac{c_0}{|K|_\infty} \cdot (nb)^{1/2}$ and $b \geq b_*$. $\square$

4.5.2 Bias approximations

There are two possibilities where the bias approximation can take place. Usually it is more convenient to have a deterministic bias expansion, meaning that the expansion is done in the expectation. In this case a supremum over a parameter θ or the bandwidth b can be evaluated easily. However, there is a point in our derivations where we also need a stochastic bias expansion, meaning that the bias expansion is done when the underlying quantity is still random. This point arises naturally because we want to use a Bernstein inequality for martingale difference sequences to get a small penalization term which coincides with the penalization term of nonparametric regression if we look at this special case. The maximum likelihood expansion however gives a sum of stochastic terms of the form $\nabla \ell(Y_{t,n}^c, \theta_0(u))$ which only becomes a martingale difference sequence if we change $Y_{t,n}^c$ to $\tilde{Y}_t(u)$ which then forces us to use a stochastic bias expansion to discuss this term. In the following, we will use the abbreviation $l := l_\beta$ (recall the definition of l_β from $\theta_0 \in \Sigma(\beta, L)$).

Lemma 4.5.6 (The deterministic bias approximation)**.** *Assume that $g : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is l-times partially differentiable (with $l \geq 0$ a natural number) and $\partial_{i_1} ... \partial_{i_l} g \in \mathcal{L}(M - l, \chi, C_1(i_1) \cdot ... \cdot C_l(i_l))$ for each component of $\partial^l g$ where $C_1, ..., C_l$ are absolutely summable sequences. Furthermore assume that $|\partial_{i_k} ... \partial_{i_l} g(0)| \leq C_k(i_k) \cdot ... \cdot C_l(i_l)$ ($k = 1, ..., l+1$). Define*

$$B_{n,b}(g,u) := \frac{1}{K_{n,b}(u)} \sum_{t=1}^n K\Big(\frac{t/n - u}{b}\Big) \cdot \big\{ \mathbb{E} g(\tilde{Y}_t(t/n)) - \mathbb{E} g(u, \tilde{Y}_t(u)) \big\}.$$

Assume that Assumption 4.2.3 and 4.2.4 hold. Then there exist constants $C_B, C_{B,R} > 0$ such that for all $u \in [\frac{b}{2}, 1 - \frac{b}{2}]$:

$$|B_{n,b}(g,u)| \leq C_B b^\beta + C_{B,R} n^{-1}.$$

In the special case $l = 0$ the above holds for all $u \in [0,1]$.

Proof of Lemma 4.5.6: For this proof, define $\|Z\|_0 := 1$ and $\|Z\|_q := 0$ for $q < 0$. This is needed to include the case of constant functions Z in our proof technique. Define $\tilde{M} := M - l$. We only consider the case $l \geq 1$, the case $l = 0$ is easier. Use the abbreviation $f := \partial_{i_2} ... \partial_{i_l} g$ and $C_f(i_1) := C_1(i_1) \cdot ... \cdot C_l(i_l)$. Define $D(k) := \max\{\chi_k, C_f(k)\}$ (which is still absolutely summable since $D(k) \leq \chi(k) + C_f(k)$). We will now show that $f : (\mathbb{R}^\infty, |\cdot|_{D,1}) \to (\mathbb{R}, |\cdot|)$ is Frechet differentiable with derivative $f'(y)h := \sum_{j=1}^\infty \partial_j f(y) \cdot h_j$. Now choose $h \in \mathbb{R}^\infty$ with $|h|_{D,1} < \varepsilon$. Let $e_j \in \mathbb{R}^\infty$ be a sequence of zeros where only at the j-th position is a 1. By the mean value theorem in $\mathbb{R}$, there exists $s \in [0,1]$ such that

$$\begin{aligned}
|f(y+h) - f(y) - f'(y)h| &\leq \sum_{j=1}^\infty \big|f\big(y + \sum_{k=1}^j h_k e_k\big) - f\big(y + \sum_{k=1}^{j-1} h_k e_k\big) - \partial_j f(y) h_j\big| \\
&= \sum_{j=1}^\infty \big|\partial_j f\big(y + \sum_{k=1}^{j-1} h_k e_k + s h_j\big) - \partial_j f(y)\big| \cdot |h_j| \\
&\leq \sum_{j=1}^\infty C_f(j) |h|_{\chi,1} \cdot \big(1 + 2^{\tilde{M}} |y|_{\chi,1}^{\tilde{M}-1} + |h|_{\chi,1}^{\tilde{M}-1}\big) \cdot |h_j| \\
&\leq \varepsilon \cdot \big(1 + 2^{\tilde{M}} |y|_{\chi,1}^{\tilde{M}-1} + |h|_{\chi,1}^{\tilde{M}-1}\big).
\end{aligned}$$

This shows Frechet differentiability of f. This shows that $s \mapsto f(y + s \cdot (y' - y))$ is differentiable with derivative $\sum_{j=1}^\infty \partial_j f(y + s \cdot (y' - y)) \cdot (y'_j - y_j)$. By the fundamental theorem of analysis,

$$\begin{aligned}
|f(y') - f(y)| &\leq \int_0^1 \sum_{j=1}^\infty |\partial_j f(y + s \cdot (y' - y)) - \partial_j f(0)| \cdot |y'_j - y_j| \, \mathrm{d}s + |y' - y|_{C_f,1} \\
&\leq \Big\{\big(|y|_{\chi,1} + |y'|_{\chi,1}\big) \cdot \big(1 + 2^{\tilde{M}} |y|_{\chi,1}^{\tilde{M}-1} + 2^{\tilde{M}} |y'|_{\chi,1}^{\tilde{M}-1}\big) + 1\Big\} |y' - y|_{C_f,1} \\
&\leq \tilde{C}_1(\tilde{M}) C_2(i_2) \cdot ... \cdot C_l(i_l) |y' - y|_{C_1,1} \big(1 + |y|_{\chi,1}^{\tilde{M}} + |y'|_{\chi,1}^{\tilde{M}}\big)
\end{aligned}$$

with some constant $\tilde{C}_1(\tilde{M})$ dependent on $\tilde{M}$. This shows that

$$\partial_{i_2} ... \partial_{i_l} g \in \mathcal{L}(\tilde{M} + 1, \chi + C_1, \tilde{C}_1(\tilde{M}) C_2(i_2) \cdot ... \cdot C_l(i_l)).$$

Inductively we obtain Frechet differentiability of all partial derivatives $\partial_{i_{k+1}} ... \partial_{i_l} g$ ($k = 1, ..., l$) and

$$\partial_{i_{k+1}} ... \partial_{i_l} g \in \mathcal{L}\big(\tilde{M} + k, \chi^{(k)}, \tilde{C}^{(k)} \cdot C_{k+1}(i_{k+1}) \cdot ... \cdot C_l(i_l)\big).$$

where $\chi^{(k)} := \chi + \tilde{C}_1 + ... + \tilde{C}_k$ and $\tilde{C}^{(k)} := \tilde{C}_1(\tilde{M}) \cdot ... \cdot \tilde{C}_k(\tilde{M}+k-1)$.
Taking $q = \frac{\tilde{M}+l}{\tilde{M}+k}$, this shows that

$$\begin{aligned}
& \|\partial_{i_{k+1}}...\partial_{i_l} g(\tilde{Y}_t(u))\|_q \\
\leq\ & \|\partial_{i_{k+1}}...\partial_{i_l} g(\tilde{Y}_t(u)) - \partial_{i_{k+1}}...\partial_{i_l} g(0)\|_q + |\partial_{i_{k+1}}...\partial_{i_l} g(0)| \\
\leq\ & C_{k+1}(i_{k+1}) \cdot ... \cdot C_l(i_l) \Big(\tilde{C}^{(k)} \big\| |\tilde{Y}_t(u)|_{\chi^{(k)},1} \cdot \big(1 + |\tilde{Y}_t(u)|_{\chi^{(k)},1}^{\tilde{M}+k-1}\big) \big\|_q + 1 \Big) \\
\leq\ & C_{k+1}(i_{k+1}) \cdot ... \cdot C_l(i_l) \\
& \qquad \Big(\tilde{C}^{(k)} \big\| |\tilde{Y}_t(u)|_{\chi^{(k)},1} \big\|_{\tilde{M}+l} \cdot \Big(1 + \big\| |\tilde{Y}_t(u)|_{\chi^{(k)},1} \big\|_{\tilde{M}+l}^{\tilde{M}+k-1} \Big) + 1 \Big) \\
\leq\ & C_{k+1}(i_{k+1}) \cdot ... \cdot C_l(i_l) \cdot D^{(k)}
\end{aligned}$$

with $D^{(k)} := \tilde{C}^{(k)} |\chi^{(k)}|_1 \tilde{C}_X (1 + |\chi^{(k)}|_1 \tilde{C}_X^{\tilde{M}+k-1}) + 1$ and $\tilde{C}_X := C_X \vee 1$. Similarly, with $\tilde{D}^{(k)} := \tilde{C}^{(k)}(1 + |\chi^{(k)}|_1 \tilde{C}_X^{\tilde{M}+k-1})$, we have

$$\begin{aligned}
& \|\partial_{i_{k+1}}...\partial_{i_l} g(\tilde{Y}_t(u)) - \partial_{i_{k+1}}...\partial_{i_l} g(\tilde{Y}_t(u'))\|_q \\
\leq\ & C_{k+1}(i_{k+1}) \cdot ... \cdot C_l(i_l) \cdot \tilde{D}^{(k)} \cdot \sum_{j=1}^{\infty} \chi_j \|\tilde{X}_{t-j+1}(u) - \tilde{X}_{t-j+1}(u')\|_{\tilde{M}+l}. \qquad (4.5.5)
\end{aligned}$$

By Faa di Bruno's rule, we have for $k = 1, ..., l$:

$$\partial_u^k g(\tilde{Y}_t(u)) = \sum_{\pi \in \Pi_k} \sum_{i_1,...,i_{|\pi|}=1}^{\infty} \partial_{i_1} \cdots \partial_{i_{|\pi|}} g(\tilde{Y}_t(u)) \cdot \prod_{j=1}^{|\pi|} \partial_u^{|\pi_j|} \tilde{X}_{t-i_j+1}(u), \qquad (4.5.6)$$

where Π_k is the set of all partitions of $\{1, ..., k\}$ and $|\pi|$ denotes the number of elements of the partition and $|\pi_k|$ the number of elements in π_k. For convenience, let us define $C_{\partial X,max} := \max\{C_{\partial^k X} : k = 0, ..., l\}$, $C_{max}(i) := \max\{C_k(i) : k = 1, ..., l\}$ and $C_{s,max} := \sum_{i=1}^{\infty} C_{max}(i)$. By Hoelder's inequality,

$$\begin{aligned}
\|\partial_u^k g(\tilde{Y}_t(u))\|_1 \ \leq\ & \sum_{\pi \in \Pi_k} \sum_{i_1,...,i_{|\pi|}=1}^{\infty} \big\| \partial_{i_1} \cdots \partial_{i_{|\pi|}} g(\tilde{Y}_t(u)) \big\|_{\frac{\tilde{M}+l}{\tilde{M}+l-|\pi|}} \cdot \prod_{j=1}^{|\pi|} \big\| \partial_u^{|\pi_j|} \tilde{X}_{t-i_j+1}(u) \big\|_{\tilde{M}+l} \\
\leq\ & \sum_{\pi \in \Pi_k} \Big(D^{(l-|\pi|)} \sum_{i_1,...,i_{|\pi|}=1}^{\infty} C_{l-|\pi|+1}(i_1) \cdot ... \cdot C_l(i_{|\pi|}) \Big) \cdot \prod_{j=1}^{|\pi|} \big\| \partial_u^{|\pi_j|} \tilde{X}_0(u) \big\|_{\tilde{M}+l} \\
<\ & \infty.
\end{aligned}$$

Replacing $\tilde{Y}_t(t/n)$ by its differentiable modification, we have almost surely

$$g(\tilde{Y}_t(t/n)) = \sum_{k=0}^{l} \frac{\partial_u^k g(\tilde{Y}_t(u))}{k!} \big(\frac{t}{n} - u\big)^k + \int_u^{t/n} \frac{(s-u)^{l-1}}{(l-1)!} \big\{ \partial_u^l g(\tilde{Y}_t(s)) - \partial_u^l g(\tilde{Y}_t(u)) \big\} \, \mathrm{d}s. \qquad (4.5.7)$$

Thus

$$\begin{aligned}|B_{n,b}(g,u)| &\leq \frac{1}{c_0}\sum_{k=1}^{l}\frac{|\mathbb{E}\partial_u^k g(\tilde{Y}_t(u))|}{k!}\Big|\frac{1}{nb}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)\big(\frac{t}{n}-u\big)^k\Big| \\ &\quad +\frac{b^l}{c_0\cdot(l-1)!}\cdot|K|_\infty\sup_{|s-u|\leq b}\big\|\partial_u^l g(\tilde{Y}_t(s))-\partial_u^l g(\tilde{Y}_t(u))\big\|_1.\end{aligned}$$

Since K has order l and bounded variation B_K, we have for $u\in[\frac{b}{2},1-\frac{b}{2}]$ and $k=1,...,l$:

$$\Big|\frac{1}{nb}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)\big(\frac{t}{n}-u\big)^k-\frac{1}{b}\int_0^1 K\Big(\frac{y-u}{b}\Big)(y-u)^k\ \mathrm{d}y\Big|\leq\frac{B_K}{nb}\cdot b^k=O(n^{-1}),$$

since $\frac{1}{b}\int_0^1 K\Big(\frac{y-u}{b}\Big)(y-u)^k = b^k\int_{-u/b}^{(1-u)/b}K(z)z^k\ \mathrm{d}z = b^k\int_{-1/2}^{1/2}K(z)z^k\ \mathrm{d}z = 0$. By Assumption 4.2.4, we have $\|\partial_u^k\tilde{X}_0(u)-\partial_u^k\tilde{X}_0(u')\|_{\tilde{M}+l}\leq C_{\partial^{k+1}X}|u-u'|$ for all u,u' and $k=1,...,l-1$. Using (4.5.6) and (4.5.5), we obtain:

$$\begin{aligned}&\big\|\partial_u^l g(\tilde{Y}_t(s))-\partial_u^l g(\tilde{Y}_t(u))\big\|_1 \\ \leq& \sum_{\pi\in\Pi_l}\sum_{i_1,...,i_{|\pi|}=1}^{\infty}\Big\{\big\|\partial_{i_1}\cdots\partial_{i_{|\pi|}}g(\tilde{Y}_t(s))-\partial_{i_1}\cdots\partial_{i_{|\pi|}}g(\tilde{Y}_t(u))\big\|_{\frac{\tilde{M}+l}{\tilde{M}+l-|\pi|}} \\ &\qquad\times\prod_{j=1}^{|\pi|}\big\|\partial_u^{|\pi_j|}\tilde{X}_{t-i_j+1}(s)\big\|_{\tilde{M}+l} \\ &\quad+\big\|\partial_{i_1}\cdots\partial_{i_{|\pi|}}g(\tilde{Y}_t(u))\big\|_{\frac{\tilde{M}+l}{\tilde{M}+l-|\pi|}}\cdot\sum_{j=1}^{|\pi|}\prod_{1\leq j_2<j}\big\|\partial_u^{|\pi_{j_2}|}\tilde{X}_{t-i_{j_2}+1}(u)\big\|_{\tilde{M}+l} \\ &\qquad\times\big\|\partial_u^{|\pi_j|}\tilde{X}_{t-i_j+1}(s)-\partial_u^{|\pi_j|}\tilde{X}_{t-i_j+1}(u)\big\|_{\tilde{M}+l}\cdot\prod_{|\pi|\geq j_2>j}\big\|\partial_u^{|\pi_{j_2}|}\tilde{X}_{t-i_{j_2}+1}(s)\big\|_{\tilde{M}+l}\Big\} \\ \leq& \sum_{\pi\in\Pi_l}\Big(\tilde{D}^{(l-|\pi|)}|\chi^{(l-|\pi|)}|_1 C_{s,max}^{|\pi|}C_{\partial X,max}^{|\pi|}\cdot\|\tilde{X}_0(s)-\tilde{X}_0(u)\|_{\tilde{M}+l} \\ &\quad+D^{(l-|\pi|)}C_{s,max}^{|\pi|}\cdot|\pi|\cdot C_{\partial X,max}^{|\pi|-1}\cdot\sup_{k=1,...,l}\big\|\partial_u^k\tilde{X}_0(s)-\partial_u^k\tilde{X}_0(u)\big\|_{\tilde{M}+l}\Big) \qquad (4.5.8) \\ =& O(b^{\beta-l}).\end{aligned}$$

□

Lemma 4.5.7 (The non-deterministic bias expansion)**.** *Assume that the conditions of Lemma 4.5.6 hold and that* $\sum_{i=1}^{\infty}\chi_i\delta_{2M}^{\partial_u^k\tilde{X}(u)}(t-i+1)$ *and* $\sum_{i=1}^{\infty}C_{max}(i)\delta_{2M}^{\partial_u^k\tilde{X}(u)}(t-i+1)$ *are absolutely summable in* t *for all* $k=0,...,l$*, where* $C_{max}(i):=\max\{C_k(i):k=1,...,l\}$*. Define*

$$\tilde{B}_{n,b}(g,u):=\frac{1}{K_{n,b}(u)}\sum_{t=1}^{n}K\Big(\frac{t/n-u}{b}\Big)\cdot\big\{g(\tilde{Y}_t(t/n))-g(\tilde{Y}_t(u))\big\}.$$

Then there exist constants $C_B, C_{B,R} > 0$ such that for all $u \in [\frac{b}{2}, 1 - \frac{b}{2}]$:

$$\|\tilde{B}_{n,b}(g,u)\|_2 \leq C_{B,2} b^\beta + C_{B,R,2} n^{-1/2}.$$

In the special case $l = 0$ the above holds for all $u \in [0,1]$.

Proof of Lemma 4.5.7: First consider the special case $1 \geq \beta > 0$. Then we have

$$\|\tilde{B}_{n,b}(g,u)\|_2 \leq \frac{1}{K_{n,b}(u)} \sum_{t=1}^{n} \Big| K\Big(\frac{t/n - u}{b}\Big)\Big| \cdot \big\|g(\tilde{Y}_t(t/n)) - g(\tilde{Y}_t(u))\big\|_2,$$

and since $g \in \mathcal{L}(M,\chi,C)$, we have by Hoelder's inequality for $|\frac{t}{n} - u| \leq b$:

$$\begin{aligned} & \big\|g(\tilde{Y}_t(t/n)) - g(\tilde{Y}_t(u))\big\|_2 \\ \leq\ & C \cdot \sum_{j=1}^{\infty} \chi_j \cdot \|\tilde{X}_{t-j+1}(t/n) - \tilde{X}_{t-j+1}(u)\|_{2M} \\ & \quad \cdot \Big(1 + \Big(\sum_{j=1}^{\infty} \chi_j \|\tilde{X}_{t-j+1}(t/n)\|_{2M}\Big)^M + \Big(\sum_{j=1}^{\infty} \chi_j \|\tilde{X}_{t-j+1}(u)\|_{2M}\Big)^M\Big) \\ \leq\ & C|\chi|_1 \cdot C_{B,2} b^\beta \cdot \big(1 + 2|\chi|_1 C_X^M\big), \end{aligned}$$

which finally shows $\|\tilde{B}_{n,b}(g,u)\|_2 \leq b^\beta \cdot \frac{|K|_\infty}{c_0} C|\chi|_1 \cdot C_{B,2}\big(1 + 2|\chi|_1 C_X^M\big)$.
Now assume $\beta > 1$. It is already known from Lemma 4.5.6 that

$$|\mathbb{E}\tilde{B}_{n,b}(g,u)| \leq C_B b^\beta + C_{B,R} n^{-1}. \tag{4.5.9}$$

From (4.5.7) in the proof therein, we obtain

$$\begin{aligned} & \|\tilde{B}_{n,b}(g,u) - \mathbb{E}\tilde{B}_{n,b}(g,u)\|_2 \\ \leq\ & \frac{1}{K_{n,b}(u)} \sum_{k=1}^{l} \sum_{j=0}^{\infty} \Big\| \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) \frac{P_{t-j}\partial_u^k g(\tilde{Y}_t(u))}{k!} \big(\frac{t}{n} - u\big)^k \Big\|_2 \\ & + \frac{1}{K_{n,b}(u)} \sum_{j=0}^{\infty} \Big\| \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) \\ & \qquad \times P_{t-j} \int_u^{t/n} \frac{(s-u)^{l-1}}{(l-1)!} \big\{\partial_u^l g(\tilde{Y}_t(s)) - \partial_u^l g(\tilde{Y}_t(u))\big\}\ \mathrm{d}s \Big\|_2. \end{aligned}$$

Furthermore it holds that

$$\begin{aligned} & \frac{1}{K_{n,b}(u)} \Big\| \sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big) P_{t-j}\partial_u^k g(\tilde{Y}_t(u)) \big(\frac{t}{n} - u\big)^k \Big\|_2 \\ =\ & \Big(\sum_{t=1}^{n} K\Big(\frac{t/n - u}{b}\Big)^2 \|P_{t-j}\partial_u^k g(\tilde{Y}_t(u))\|_2^2 \Big|\frac{t}{n} - u\Big|^{2k}\Big)^{1/2} \leq b^k \cdot \delta_2^{\partial_u^k g(\tilde{Y}(u))}(j) \cdot F_{n,b}(u)^{-1}. \end{aligned}$$

Similarly,

$$\begin{aligned}
& \frac{1}{K_{n,b}(u)}\Big\| \sum_{t=1}^{n} K\Big(\frac{t/n-u}{b}\Big) P_{t-j} \int_u^{t/n} \frac{(s-u)^{l-1}}{(l-1)!}\{\partial_u^l g(\tilde{Y}_t(s)) - \partial_u^l g(\tilde{Y}_t(u))\} \,\mathrm{d}s\Big\|_2 \\
\le\ & 2\frac{b^l}{l!}\cdot \delta_2^{\partial_u^l g(\tilde{Y}_t(u))}(j)\cdot F_{n,b}(u)^{-1}
\end{aligned}$$

Since $F_{n,b}(u)^{-1}b^k \le \frac{|K|_\infty}{c_0}n^{-1/2}b^{k-1/2} \le \frac{|K|_\infty}{c_0}n^{-1/2}$, we conclude

$$\|\tilde{B}_{n,b}(g,u)-\mathbb{E}\tilde{B}_{n,b}(g,u)\|_2 \le n^{-1/2}\cdot\frac{|K|_\infty}{c_0}\Big(\sum_{k=1}^{l}\frac{1}{k!}\cdot\sum_{j=0}^{\infty}\delta_2^{\partial_u^k g(\tilde{Y}(u))}(j)+\frac{2}{l!}\cdot\sum_{j=0}^{\infty}\delta_2^{\partial_u^l g(\tilde{Y}(u))}(j)\Big),$$

which finally gives the result using (4.5.9) and $\|\tilde{B}_{n,b}(g,u)\|_2 \le \|\tilde{B}_{n,b}(g,u)-\mathbb{E}\tilde{B}_{n,b}(g,u)\|_2 + |\mathbb{E}\tilde{B}_{n,b}(g,u)|$. It remains to show that $\delta_2^{\partial_u^k g(\tilde{Y}(u))}(j)$ is absolutely summable for $k = 1,...,l$. Using similar techniques as in (4.5.8) in the proof of Lemma 4.5.6, we obtain for $k = 1,...,l$ and arbitrary $u \in [0,1]$:

$$\begin{aligned}
& \|\partial_u^k g(\tilde{Y}_t(u)) - \partial_u^k g(\tilde{Y}_t^*(u))\|_2 \\
\le\ & \sum_{\pi\in\Pi_k}\Big(\tilde{D}^{(l-|\pi|)}C_{s,max}^{|\pi|}C_{\partial X,max}^{|\pi|}\cdot\sum_{j=1}^{t}\chi_j\|\tilde{X}_{t-j+1}(u)-\tilde{X}_{t-j+1}^*(u)\|_{2(\tilde{M}+l)} \\
& \quad +D^{(l-|\pi|)}C_{s,max}^{|\pi|-1}\cdot|\pi|\cdot C_{\partial X,max}^{|\pi|-1} \\
& \qquad \times\sup_{k=1,...,l}\sum_{i=1}^{t}C_{max}(i)\cdot\big\|\partial_u^k\tilde{X}_{t-i+1}(u)-\partial_u^k\tilde{X}_{t-i+1}^*(u)\big\|_{2(\tilde{M}+l)}\Big) \\
\le\ & C\cdot\Big(\sum_{j=1}^{t}\chi_j\cdot\delta_{2M}^{\tilde{X}(u)}(t-j+1)+\sup_{k=1,...,l}\sum_{i=1}^{t}C_{max}(i)\cdot\delta_{2M}^{\partial_u^k\tilde{X}(u)}(t-i+1)\Big).
\end{aligned}$$

with some constant $C > 0$, which is absolutely summable in t by assumption. □

4.5.3 A weighted Euclidean norm representation of the Kullback-Leibler divergence

Here we show that the (misspecified) Kullback-Leibler divergence is approximately a weighted Euclidean norm.

Proposition 4.5.8. *Suppose that Assumption 4.2.3 and 4.2.2 hold. Let $\mathbb{P}_{\tilde{X}_t(u)|\tilde{Y}_{t-1}(u),\theta(\cdot)}$ be the conditional distribution of $\tilde{X}_t(u)$ given $\tilde{Y}_{t-1}(u)$ under the assumption that the true curve is $\theta(\cdot)$. Assume that $\ell(\tilde{X}_t(u),\tilde{Y}_t(u),u,\theta(u))$ is the negative logarithm of the corresponding density with respect to the Lebesgue measure. Then the Kullback-Leibler divergence of $\mathbb{P}_{\tilde{X}_t(u)|\tilde{Y}_{t-1}(u)|\theta_1(\cdot)}$ w.r.t. $\mathbb{P}_{\tilde{X}_t(u)|\tilde{Y}_{t-1}(u)|\theta_0(\cdot)}$ is given by*

$$\begin{aligned}
& KL(\mathbb{P}_{\tilde{X}_t(u)|\tilde{Y}_{t-1}(u),\theta_0(\cdot)},\mathbb{P}_{\tilde{X}_t(u)|\tilde{Y}_{t-1}(u),\theta_1(\cdot)}) \\
=\ & \mathbb{E}_{\theta_0}[\tilde{\ell}_t(u,\theta_1(u))]-\mathbb{E}_{\theta_0}[\tilde{\ell}_t(u,\theta_0(u))] = \frac{1}{2}|\theta_1(u)-\theta_0(u)|_{V(u)}^2 + O(|\theta_1(u)-\theta_0(u)|_2^3).
\end{aligned}$$

Proof of Proposition 4.5.8: A Taylor expansion of $L(u,\theta) = \mathbb{E}\tilde{\ell}_t(u,\theta)$ gives

$$L(u,\theta_1(u)) - L(u,\theta_0(u)) = \langle \nabla L(u,\theta_0(u)), \theta_1(u) - \theta_0(u)\rangle + \frac{1}{2}|\theta_1(u) - \theta_0(u)|^2_{\nabla^2 L(u,\theta_0(u))} + \frac{1}{2}|\theta_1(u) - \theta_0(u)|^2_{\nabla^2 L(u,\bar{\theta}(u)) - \nabla^2 L(u,\theta_0(u))}.$$

Since $\nabla\tilde{\ell}_t(u,\theta_0(u))$ is a martingale difference sequence by Assumption 4.2.2(iii), we have $\nabla L(u,\theta_0(u)) = 0$. Assumption 4.2.2(v) together with Lemma 4.5.15 gives $||\theta_1(u) - \theta_0(u)|^2_{\nabla^2 L(u,\bar{\theta}(u)) - \nabla^2 L(u,\theta_0(u))}| \le |\theta_1(u) - \theta_0(u)|_2^2 \cdot |\nabla^2 L(u,\bar{\theta}(u)) - \nabla^2 L(u,\theta_0(u))|_2 = O(|\theta_1(u) - \theta_0(u)|_2^3)$, leading to the result. □

4.5.4 Proofs of the examples

In this section we present two lemmata (somehow, they are more general examples) which help us to prove the statements in the examples.

Lemma 4.5.9 (Recursively defined time series)**.** *Suppose that Assumption 4.2.2(i) is fulfilled. Assume that*

$$X_{t,n} = \mu(Y_{t-1,n},\theta_0(t/n)) + \sigma(Y_{t-1,n},\theta_0(t/n))\varepsilon_t, \quad t = 1,...,n$$

where $Y_{t-1,n} = (X_{t-1,n},...,X_{t-p,n})$ *contains only finitely many past values in this example. Assume that* $\mu,\sigma : \mathbb{R}^p \times \Theta \to \mathbb{R}$ *satisfy*

$$\sup_\theta \sup_{y\neq y'} \frac{|\mu(y,\theta) - \mu(y',\theta)|}{|y-y'|_{\chi,1}} + \sup_\theta \sup_{y\neq y'} \frac{|\sigma(y,\theta) - \sigma(y',\theta)|}{|y-y'|_{\chi,1}} \|\varepsilon_0\|_q \le 1 \tag{4.5.10}$$

for all $q \ge 1$ *with some* $\chi \in \mathbb{R}^p_{\ge 0}$ *with* $|\chi|_1 < 1$*. Assume that* $\sigma(\cdot) \ge \sigma_0$ *with some constant* $\sigma_0 > 0$*. Assume that* $\nabla\sigma \neq 0$*, and*

(a) $\mathbb{E}\varepsilon_0 = 0$ *and* $\mathbb{E}\varepsilon_0^2 = 1$*. Either* $\sigma(\cdot,\theta) \equiv \sigma(\theta)$ *is constant and the Lebesgue density* $f_{|\varepsilon_0|}$ *of* $|\varepsilon_0|$ *fulfills* $f_{|\varepsilon_0|}(x) \le C_f \exp(-x^{1/\alpha})$ *for some* $\alpha, C_f > 0$*; or* $|\varepsilon_0| \le C_\varepsilon$ *a.s. (and set* $\alpha = 0$*).*

(b) $\sup_y \sup_{\theta\neq\theta'} \frac{|\mu(y,\theta)-\mu(y,\theta')|}{|\theta-\theta'|_2\cdot(1+|y|_1)} < \infty$ *and* $\sup_y \sup_{\theta\neq\theta'} \frac{|\sigma(y,\theta)-\sigma(y,\theta')|}{|\theta-\theta'|_2\cdot(1+|y|_1)} < \infty$*.*

(c) Omitting the arguments $\tilde{Y}_0(u)$*, there exists* $C_{L,\mu}(u) > 0$ *such that* $\mathbb{E}\big(\frac{\mu(\theta)-\mu(\theta_0(u))}{\sigma(\theta)}\big)^2 \ge \frac{1}{C_{L,\mu}(u)}|\theta - \theta_0(u)|_2^2$*.*

(d) There exists $C_{L,\sigma}(u) > 0$ *such that either* $\mathbb{E}\big(\frac{1}{\sigma^2(\theta)} - \frac{1}{\sigma^2(\theta_0(u))}\big)^2 \ge \frac{1}{C_{L,\sigma}(u)}|\theta - \theta_0(u)|_2^2$ *or* $\mathbb{E}\big(\sigma^2(\theta) - \sigma^2(\theta_0(u))\big)^2 \ge \frac{1}{C_{L,\sigma}(u)}|\theta - \theta_0(u)|_2^2$ *and* $\lim_{\delta\to\infty} \delta^2\mathbb{P}(\sup_\theta \sigma(\theta)^2 > \delta) = 0$*.*

(e) Define $v(u,\theta) := \sigma(\theta)^{-1}(\nabla\mu(\theta)', \nabla\sigma(\theta)')'$*. It holds that* $\inf_{\theta,u} \lambda_{min}(\mathbb{E}[v(u,\theta)\cdot v(u,\theta)']) > 0$*.*

Then Assumptions 4.2.2(i)-(iv), 4.2.3 and 4.2.6 are fulfilled for the Gaussian likelihood (4.1.7).
If $\sigma(\cdot) \equiv \sigma_0$ is constant and the conditions above still hold with appropriate changes (in (c) omit the second condition, in (d) drop the condition on Σ_I and use $v(u,\theta) := \sigma_0^{-1}\nabla\mu(\theta)$), and it holds that

(f) for $i = 1,2$, $\sup_\theta \sup_{y\neq y'} \frac{|\nabla^i \mu(y,\theta) - \nabla^i\mu(y',\theta)|}{|y-y'|_1} < \infty$ and $\sup_y \sup_{\theta\neq\theta'} \frac{|\nabla^i\mu(y,\theta)-\nabla^i\mu(y,\theta')|}{|\theta-\theta'|_2(1+|y|_1)} < \infty$,

then Assumptions 4.2.2, 4.2.3 and 4.2.6 are fulfilled for the Gaussian likelihood (4.1.7) with $M = 2$.

Proof of Lemma 4.5.9: Case $\nabla\sigma \neq 0$: (a) leads to either $\|\varepsilon_0\|_q^q \le C_\varepsilon^q$ or to

$$\begin{aligned}\|\varepsilon_0\|_q^q &\le C_f \int_0^\infty x^q \exp(-x^{1/\alpha})\,\mathrm{d}x = C_f\int_0^\infty u^{\alpha q+\alpha-1}e^{-u}\,\mathrm{d}u \\ &= C_f\alpha\Gamma(\alpha(q+1)) \le \Big(C_f\alpha(\alpha q+\alpha-1)^{\lceil\alpha-2\rceil\vee 0}\Big)\Gamma(\alpha q+2) \le C_\varepsilon^q N_\alpha(q)^q.\end{aligned}$$

with some $C_\varepsilon > 0$ depending on α, C_f. By (b), we have

$$\sup_{\theta\in\Theta}\|G_{\varepsilon_0}(0,\theta)\|_q \le \big(\sup_\theta|\mu(0,\theta)| + C_\varepsilon\sup_\theta|\sigma(0,\theta)|\big)N_\alpha(q).$$

Together with (4.5.10) and Proposition 3.1.3 in Chapter 3 (inspect the proof to get a specific representation of $\delta_q^{\tilde X(u)}(k)$) we obtain $\delta_q^{\tilde X(u)}(k) \le \delta(k)\cdot N_\alpha(q)$ with some $\delta(k) \le C_\delta\lambda^k$ where $C_\delta > 0, \lambda\in(0,1)$. Since $\theta_0 \in \Sigma(\beta, L)$, the conditions of Lemma 3.1.4 in Chapter 3 are fulfilled such that Assumption 4.2.3 holds.
In the following let $z = (x,y)$. In this case, we have that the inverse of $\varepsilon \mapsto G_\varepsilon(y,u,\theta)$ is given by $x \mapsto H(z,\theta,u) = \frac{x-\mu(y,\theta)}{\sigma(y,\theta)}$. Let us omit the arguments (y,θ,u) or $(\tilde Y_t(u),\theta,u)$ of $\mu(\cdot)$, $\sigma(\cdot)$ in the following. Then (4.1.7) with standard Gaussian density f takes the form $\ell(z,\theta) = \frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2 + \log\sigma$.
Since $\mu(\theta),\sigma(\theta)$ are $\mathcal{F}_{t-1}$-measurable and $\mathbb{E}\varepsilon_0 = 0$, $\mathbb{E}\varepsilon_0^2 = 1$ by (a), we have

$$L(u,\theta) - L(u,\theta_0(u)) = \mathbb{E}\Big(\frac{\mu(\theta)-\mu(\theta_0(u))}{\sigma(\theta)}\Big)^2 + \mathbb{E}\Big[\frac{\sigma^2(\theta_0(u))}{\sigma^2(\theta)} - \log\frac{\sigma^2(\theta_0(u))}{\sigma^2(\theta)} - 1\Big] \tag{4.5.11}$$

By a Taylor expansion of $x \mapsto x - \log(x) - 1$, we obtain that the second summand in (4.5.11) is lower bounded by

$$\begin{aligned}\frac{1}{4}\mathbb{E}\frac{(\sigma^2(\theta_0(u))-\sigma^2(\theta))^2}{(\sigma^2(\theta_0(u))-\sigma^2(\theta))^2+\sigma^4(\theta)} &= \frac{1}{4}\mathbb{E}\frac{\big(\frac{1}{\sigma^2(\theta)}-\frac{1}{\sigma^2(\theta_0(u))}\big)^2}{\big(\frac{1}{\sigma^2(\theta)}-\frac{1}{\sigma^2(\theta_0(u))}\big)^2+\frac{1}{\sigma^4(\theta)}} \\ &\ge \frac{1}{8\sigma_0^4}\cdot\mathbb{E}\Big(\frac{1}{\sigma^2(\theta)}-\frac{1}{\sigma^2(\theta_0(u))}\Big)^2.\end{aligned}$$

which then gives Assumption 4.2.2(ii) by conditions (c),(d). Alternatively, define $A_\delta := \{\sup_\theta\sigma(\theta)^2 \le \delta\}$. Then, for $\delta \ge \sigma_0$, the second summand of (4.5.11) is lower bounded by

$$\frac{1}{8\delta^2}\mathbb{E}\big(\sigma^2(\theta_0(u))-\sigma^2(\theta)\big)^2$$
$$+\frac{1}{4}\mathbb{E}\Big[\big(\sigma^2(\theta_0(u))-\sigma^2(\theta)\big)^2\cdot\Big(\frac{1}{\big(\sigma^2(\theta_0(u))-\sigma^2(\theta)\big)^2+\sigma^4(\theta)}-\frac{1}{2\delta^2}\Big)\mathbb{1}_{A_\delta^c}\Big].$$

The second summand is bounded by $\frac{1}{16\sigma_0^2}\mathbb{E}[(\sigma^2(\theta_0(u))-\sigma^2(\theta))^4]^{1/2}\cdot\mathbb{P}(A_\delta^c)^{1/2}$. If $\delta^2\mathbb{P}(A_\delta^c)^{1/2}=\delta^2\mathbb{P}(\sup_\theta\sigma(\theta)^2>\delta)\to 0$ for $\delta\to\infty$, we can find a $\delta=\delta_m$ large enough such that the second summand is bounded by the half of the first summand, so that the second summand of (4.5.11) is lower bounded by $\frac{1}{16\delta_m^2}\mathbb{E}(\sigma^2(\theta_0(u))-\sigma^2(\theta))^2$.
Omitting the arguments (z,u,θ), we have

$$\begin{aligned}\nabla\ell &= -\big(\frac{x-\mu}{\sigma}\big)\cdot\frac{\nabla\mu}{\sigma}+\frac{\nabla\sigma}{\sigma}\cdot\Big[1-\big(\frac{x-\mu}{\sigma}\big)^2\Big],\\ \nabla^2\ell &= \frac{\nabla\mu\nabla\mu'}{\sigma^2}+\big(\frac{x-\mu}{\sigma}\big)\cdot\Big[2\frac{\nabla\mu\nabla\sigma'+\nabla\sigma\nabla\mu'}{\sigma^2}-\frac{\nabla^2\mu}{\sigma}\Big]\\ &\quad+\frac{\nabla^2\sigma}{\sigma}\Big[1-\big(\frac{x-\mu}{\sigma}\big)^2\Big]+\frac{\nabla\sigma\nabla\sigma'}{\sigma^2}\Big[3\big(\frac{x-\mu}{\sigma}\big)^2-1\Big]. \qquad (4.5.12)\end{aligned}$$

Since $\frac{\tilde{X}_t(u)-\mu(\tilde{Y}_{t-1}(u),\theta_0(u))}{\sigma(\tilde{Y}_{t-1}(u),\theta_0(u))}=\varepsilon_t$ and $\mathbb{E}\varepsilon_t=0$, $\mathbb{E}\varepsilon_0^2=1$ by condition (a), we obtain the martingale difference sequence property of $\nabla\tilde{\ell}_t(u,\theta_0(u))$ and thus Assumption 4.2.2(iii). Define $\Sigma_V:=\left(\begin{smallmatrix}1&0\\0&2\end{smallmatrix}\right)$. Then we have

$$\begin{aligned}g_I(y,\theta) &= \sigma^{-2}\cdot\big(\mathbb{E}\varepsilon_0^2\cdot\nabla\mu\nabla\mu'+\mathbb{E}[\varepsilon_0^4-1]\cdot\nabla\sigma\nabla\sigma'+\mathbb{E}\varepsilon_0^3\cdot\big(\nabla\mu\nabla\sigma'+\nabla\sigma\nabla\mu'\big)\big),\\ &= \sigma^{-2}\mathrm{tr}\big\{(\Sigma_I\otimes I_d)\cdot\left(\begin{smallmatrix}\nabla\mu\\\nabla\sigma\end{smallmatrix}\right)\cdot\left(\begin{smallmatrix}\nabla\mu\\\nabla\sigma\end{smallmatrix}\right)'\big\},\\ g_V(y,\theta) &= \sigma^{-2}\cdot\big(\nabla\mu\nabla\mu'+2\nabla\sigma\nabla\sigma'\big)=\sigma^{-2}\mathrm{tr}\big\{(\Sigma_V\otimes I_d)\cdot\left(\begin{smallmatrix}\nabla\mu\\\nabla\sigma\end{smallmatrix}\right)\cdot\left(\begin{smallmatrix}\nabla\mu\\\nabla\sigma\end{smallmatrix}\right)'\big\}, \quad (4.5.13)\end{aligned}$$

where $\otimes$ denotes the Kronecker product. Condition (e) now implies Assumption 4.2.2(iv). If $\sigma(\cdot)\equiv\sigma_0$ is constant, Assumption 4.2.2(v) immediately follows from (f), $\delta(k)\le C_\delta\lambda^k$ and the representations (4.5.12) and (4.5.13). □

Proof of Example 4.2.9: In view of Example 4.5.9, note that $\sigma(y,\theta)=\langle\theta,m(y)\rangle\ge\rho_{min}m_0>0$ uniformly in y,θ. Furthermore, we have

$$\begin{aligned}|\sigma(y,\theta)^2-\sigma(y',\theta)^2| &\le \sum_{i=1}^d\theta_i|m_i(y)-m_i(y')|\\ &\le \sum_{i=1}^d\sqrt{\theta_i}|y-y|_{\chi_i,1}\cdot\big(\sqrt{\theta_i m_i(y)}+\sqrt{\theta_i m_i(y')}\big)\\ &\le \sum_{i=1}^d\sqrt{\theta_i}|y-y|_{\chi_i,1}\cdot\big(\sigma(y,\theta)+\sigma(y',\theta)\big),\end{aligned}$$

which shows that $\sup_\theta \sup_{y\neq y'} \frac{|\sigma(y,\theta)-\sigma(y',\theta)|}{|y-y'|_{\sum_{i=1}^d \sqrt{\theta_i}\chi_i,1}} \leq 1$, and

$$|\sigma(y,\theta)^2 - \sigma(y,\theta')^2| \leq \sum_{i=1}^d \frac{|\theta_i - \theta_i'|}{\sqrt{\theta_i}+\sqrt{\theta_i'}}\big(\sigma(y,\theta)+\sigma(y,\theta')\big)\cdot\sqrt{m_i(y)},$$

which shows that

$$\sup_\theta \sup_{y\neq y'} \frac{|\sigma(y,\theta)-\sigma(y',\theta)|}{|\theta-\theta'|_1(1+|y|_1)} \leq \frac{1}{2\sqrt{\rho_{min}}}|\theta-\theta'|_1 \max_{i=1,\dots,d} m_i(0)(1+\max_{i=1,\dots,d}|\chi_i|_1).$$

By Example 4.5.9 in connection with (4.5.1) it is known that $\|\tilde{X}_t(u)\|_q \leq C_X N_\alpha(q) = C_X$ for all $q \geq 1$. By Markov's inequality, we conclude that for each $\phi > 0$, we have $\mathbb{P}(|\tilde{X}_t(u)| > C_X(1+\phi)) \leq \frac{1}{(1+\phi)^q} \to 0$ $(q\to\infty)$ which shows that $|\tilde{X}_t(u)| \leq C_X$ a.s. for each $u \in [0,1]$. Thus it holds for all $i = 1,\dots,d$ that $\sqrt{m(\tilde{Y}_0(u))} \leq \sqrt{m(0)} + |\tilde{Y}_0(u)|_{\chi,1} \leq \sqrt{m(0)} + C_X|\chi|_1 =: D_X$ which shows that

$$\begin{aligned}\mathbb{E}\Big(\frac{1}{\sigma(\tilde{Y}_0(u),\theta)^2} - \frac{1}{\sigma(\tilde{Y}_0(u),\theta')^2}\Big)^2 &= \mathbb{E}\Big(\frac{\langle\theta-\theta', m(\tilde{Y}_0(u))\rangle}{\langle\theta, m(\tilde{Y}_0(u))\rangle\cdot\langle\theta', m(\tilde{Y}_0(u))\rangle}\Big)^2\\ &\geq \frac{1}{\Theta_{max}^2 D_X^4}|\theta-\theta'|_{\mathbb{E}[m(\tilde{Y}_0(u))m(\tilde{Y}_0(u))']}.\end{aligned}$$

In practise, C_X *may be obtained by solving the equation* $C_X = \sqrt{\langle m(C_X),\theta\rangle}C_\varepsilon$. *Note that* $|\langle m(y),\theta\rangle| \leq 2\sum_{i=1}^d \theta_i(\sqrt{m_i(y)} - \sqrt{m_i(0)}) + 2\sum_{i=1}^d \theta_i m_i(0) \leq 2\big(\sum_{i=1}^d \sqrt{\theta_i}|y|_{\chi_i,1}\big)^2 + 2|m(0)|_1\Theta_{max}$, *which shows that* $|\langle m(\tilde{Y}_0(u)),\theta\rangle| \leq 2(C_X\rho_{max}C_\varepsilon^{-1})^2 + 2|m(0)|_1\Theta_{max}$.

Since $m_1(\tilde{Y}_0(u)),\dots,m_d(\tilde{Y}_0(u))$ are linearly independent, we can now conclude that $\lambda_{min}(\mathbb{E}[m(\tilde{Y}_0(u))m(\tilde{Y}_0(u))']) > 0$ for all $u \in [0,1]$. Continuity properties of m and $u \mapsto \|\tilde{X}_0(u)\|_2$ (see Assumption 4.2.3) show that $\inf_{u\in[0,1]}\lambda_{min}(\mathbb{E}[m(\tilde{Y}_0(u))m(\tilde{Y}_0(u))']) > 0$. We have

$$\begin{aligned}\mathbb{E}\Big[\frac{\nabla\sigma(\tilde{Y}_0(u),\theta)\nabla\sigma(\tilde{Y}_0(u),\theta)'}{\sigma(\tilde{Y}_0(u),\theta)^2}\Big] &= \frac{1}{4}\mathbb{E}\Big[\frac{m(\tilde{Y}_0(u))m(\tilde{Y}_0(u))'}{\langle\theta, m(\tilde{Y}_0(u))\rangle^2}\Big]\\ &\succeq \frac{1}{4\Theta_{max}^2 D_X^4}\mathbb{E}[m(\tilde{Y}_0(u))m(\tilde{Y}_0(u))'].\end{aligned}$$

which is already known to have positive eigenvalues which are uniformly bounded away from 0. Finally, note that (omitting the arguments (x,y,θ) and y of ℓ,m, respectively):

$$\ell = \frac{x^2}{2\langle\theta,m\rangle} + \frac{1}{2}\log\langle\theta,m\rangle, \quad \nabla\ell = -\frac{x^2}{2}\cdot\frac{m}{\langle\theta,m\rangle^2} + \frac{1}{2}\frac{m}{\langle\theta,m\rangle}, \quad \nabla^2\ell = x^2\cdot\frac{mm'}{\langle\theta,m\rangle^3} - \frac{1}{2}\frac{mm'}{\langle\theta,m\rangle^2}.$$

Since $\frac{|m_i(y)|}{\langle\theta,m(y)\rangle} \leq \frac{1}{\rho_{min}}$ for all $i = 1,\dots,d$ and $g_I(y,\theta) = \frac{\mathbb{E}\varepsilon_0^4 - 1}{4}\frac{mm'}{\langle\theta,m\rangle^2}$, $g_V(y,\theta) = \frac{1}{2}\frac{mm'}{\langle\theta,m\rangle^2}$ it is easily seen that Assumption 4.2.2(v) is fulfilled with $M = 3$. □

In the following we discuss a model introduced in Dahlhaus and Polonik (2009). We provide conditions under which the theorems of this chapter are applicable.

Lemma 4.5.10 (Linear time series models)**.** *Suppose that Assumption 4.2.2(i) is fulfilled and there exists a superset $\tilde{\Theta} \supset \Theta$ which is convex. Assume that*

$$X_{t,n} = \sum_{k=0}^{\infty} a_{t,n}(k)\varepsilon_{t-k} \tag{4.5.14}$$

with some coefficients $a_{t,n}(k)$ and $a_\theta(k)$ satisfying

$$\sup_{t=1,\dots,n} |a_{t,n}(k) - a_{\theta_0(t/n)}(k)| \le C_B(k) n^{-\beta'}. \tag{4.5.15}$$

with some absolutely summable sequence $C_B(k)$. For $\theta \in \tilde{\Theta}$, $\lambda \in [-\pi,\pi]$, define $A_\theta(\lambda) := \sum_{k=0}^{\infty} a_\theta(k) e^{i\lambda k}$, the spectral density $f_\theta(\lambda) := \frac{1}{2\pi}|A_\theta(\lambda)|^2$ and real numbers $\gamma_\theta(k) := \frac{1}{2\pi}\int_{-\pi}^{\pi} A_\theta(\lambda)^{-1} e^{-i\lambda k}\, d\lambda$. Assume that

(a) $\mathbb{E}\varepsilon_0 = 0$, the Lebesgue density $f_{|\varepsilon_0|}$ of $|\varepsilon_0|$ fulfills $f_{|\varepsilon_0|}(x) \le C_f \exp(-x^{1/\alpha})$ for some $\alpha, C_f > 0$. If $\nabla\gamma_\theta(0) \neq 0$, assume that $\mathbb{E}\varepsilon_0^2 = 1$.

(b) $|A_\theta(\lambda)| \ge \delta_A > 0$ uniformly in $\theta \in \tilde{\Theta}, \lambda$. $A_\theta(\lambda)$ is three times continuously differentiable in $\theta \in \tilde{\Theta}$. There exist $\beta_A > 2, L_A > 0$ such that component-wise $\nabla^i A_\theta(\cdot) \in \Sigma(\beta_A, L_A)$ $(i = 0,1,2,3)$ uniformly in $\theta \in \tilde{\Theta}$.

(c) There exist a constant $C_{L,A} > 0$ such that for $\theta, \theta' \in \Theta$, $\int_{-\pi}^{\pi} \big|\frac{A_{\theta'}(\lambda)^{-1}}{\gamma_{\theta'}(0)} - \frac{A_\theta(\lambda)^{-1}}{\gamma_\theta(0)}\big|^2 \cdot f_\theta(\lambda)\, d\lambda + (\gamma_{\theta'}(0)^2 - \gamma_\theta(0)^2)^2 \ge \frac{1}{C_{L,A}}|\theta - \theta'|_2^2$.

(d) $\inf_{\theta\in\Theta} \lambda_{min}(\nabla\gamma_\theta(0)\nabla\gamma_\theta(0)') > 0$ and

$$\inf_{\theta\in\Theta} \lambda_{min}(\gamma_\theta(0)^2 \int_{-\pi}^{\pi} \nabla\big(\frac{A_\theta(\lambda)^{-1}}{\gamma_\theta(0)}\big)\nabla\big(\frac{A_\theta(\lambda)^{-1}}{\gamma_\theta(0)}\big)'\, d\lambda) > 0.$$

Then Assumptions 4.2.3 and 4.2.2 are fulfilled for the Gaussian likelihood (4.1.7). If additionally,

(e) $A_\theta(\lambda)$ is $l+1$-times continuously differentiable in $\theta \in \tilde{\Theta}$ and fulfills component-wise $\nabla^i A_\theta(\cdot) \in \Sigma(\beta_A, L_A)$ $(i = 0,\dots,l+1)$,

then Assumption 4.2.4 is fulfilled.

Proof of Lemma 4.5.10: Condition (a) implies that $\|\varepsilon_0\|_q \le C_\varepsilon N_\alpha(q)$ (see the proof of Example 4.5.9). The stationary process $\tilde{X}_t(u)$ satisfies (4.1.3) with $G_\varepsilon(y,u,\theta) = \frac{1}{\gamma_\theta(0)}\big(\varepsilon - \sum_{k=1}^{\infty}\gamma_\theta(k)y_k\big)$. By (4.5.15) we have for all $q \ge 1$:

$$\|X_{t,n} - \tilde{X}_t(t/n)\|_q \le \sum_{k=0}^{\infty} |a_{t,n}(k) - a_{\theta_0(t/n)}(k)|\|\varepsilon_{t-k}\|_q \le C_\varepsilon \sum_{k=0}^{\infty} C_B(k)\cdot N_\alpha(q).$$

It holds that $a_\theta(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A_\theta(\lambda)e^{-i\lambda k}\,\mathrm{d}\lambda$. By condition (b) and Katznelson (2004), chapter I, section 4, we have that $R(k) := \sup_\theta |\nabla a_\theta(k)|_2 = \sup_\theta \frac{1}{2\pi}\big|\int_{-\pi}^{\pi} \nabla A_\theta(\lambda)e^{-i\lambda k}\,\mathrm{d}\lambda\big|_2$ is absolutely summable in k and thus with the fundamental theorem of analysis:

$$\begin{aligned}
&\|\tilde{X}_t(u) - \tilde{X}_t(v)\|_q \\
\le\; & \frac{1}{2\pi}\sum_{k=0}^{\infty}\Big|\int_0^1 \big\langle \int_{-\pi}^{\pi} \nabla A_{\theta_0(v)+s(\theta_0(u)-\theta_0(v))}(\lambda)e^{-i\lambda k}\,\mathrm{d}\lambda, \theta_0(u)-\theta_0(v)\big\rangle\,\mathrm{d}s\Big| \cdot \|\varepsilon_{t-k}\|_q \\
\le\; & \sum_{k=0}^{\infty} R(k)\cdot|\theta_0(u)-\theta_0(v)|_2 \cdot C_\varepsilon N_\alpha(q). \qquad (4.5.16)
\end{aligned}$$

Furthermore we obtain that $\delta_q^{\tilde{X}(u)}(k) = |a_{\theta_0(u)}(k)|\cdot\|\varepsilon_0 - \varepsilon_0^*\|_q \le \delta(k)N_\alpha(q)$ with $\delta(k) := 2\sup_\theta |a_\theta(k)|\cdot C_\varepsilon$.
It was shown in **?**, Proposition 2.2 that (4.1.7) with the standard Gaussian density takes the form $\ell(z,\theta) = \frac{1}{2}H(z,\theta)^2 - \frac{1}{2}\log(2\pi\gamma_\theta(0)^2)$ with $H(z,\theta) = \sum_{k=0}^{\infty}\gamma_\theta(k)z_k$. Since it holds that $\tilde{X}_t(u) = \frac{1}{\gamma_\theta(0)}\cdot\big(\varepsilon_t - \sum_{k=1}^{\infty}\gamma_\theta(k)\tilde{X}_{t-k}(u)\big)$, we have

$$\begin{aligned}
&\nabla\ell(\tilde{Y}_t(u),\theta_0(u)) \\
=\; & \varepsilon_t\nabla H(\tilde{Y}_t(u),\theta_0(u)) - \frac{\nabla\gamma_\theta(0)}{\gamma_\theta(0)} \\
=\; & \varepsilon_t\cdot\sum_{k=1}^{\infty}\big\{\nabla\gamma_{\theta_0(u)}(k) - \frac{\nabla\gamma_{\theta_0(u)}(0)}{\gamma_{\theta_0(u)}(0)}\cdot\gamma_{\theta_0(u)}(k)\big\}\tilde{X}_{t-k}(u) + \frac{\nabla\gamma_{\theta_0(u)}(0)}{\gamma_{\theta_0(u)}(0)}\cdot\big(\varepsilon_t^2 - 1\big),
\end{aligned}$$

which shows that $\nabla\ell(\tilde{Y}_t(u),\theta_0(u))$ is a martingale difference sequence since $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_t^2 = 1$ or $\nabla\gamma_\theta(0) = 0$, thus Assumption 4.2.2(iii) holds.
In the situation of Example 4.5.9, we have $\mu(y,\theta) = -\frac{1}{\gamma_\theta(0)}\sum_{k=1}^{\infty}\gamma_\theta(k)y_k$ and $\sigma(y,\theta) = \frac{1}{\gamma_\theta(0)}$, thus (omitting the argument $\tilde{Y}_{t-1}(u)$):

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{\mu(\theta)-\mu(\theta_0(u))}{\sigma(\theta)}\Big)^2\Big] \\
=\; & \mathbb{E}\Big(\sum_{k=0}^{\infty}\Big(\frac{\gamma_\theta(k)}{\gamma_\theta(0)} - \frac{\gamma_{\theta_0(u)}(k)}{\gamma_{\theta_0(u)}(0)}\Big)\tilde{X}_{t-k}(u)\Big)^2 \\
=\; & \int_{-\pi}^{\pi}\Big|\frac{A_\theta(\lambda)^{-1}}{\gamma_\theta(0)} - \frac{A_{\theta_0(u)}(\lambda)^{-1}}{\gamma_{\theta_0(u)}(\lambda)}\Big|^2\cdot f_{\theta_0(u)}(\lambda)\,\mathrm{d}\lambda,
\end{aligned}$$

and, using the fact that $\partial_\varepsilon G_\varepsilon(y,\theta) = \frac{1}{\gamma_\theta(0)} \ge \delta_G > 0$, we have based on the results from Example 4.5.9:

$$\mathbb{E}\Big[\frac{(\sigma^2(\theta_0(u)) - \sigma^2(\theta))^2}{(\sigma^2(\theta_0(u)) - \sigma^2(\theta))^2 + \sigma^4(\theta)}\Big] \ge \frac{1}{8\delta_G^2}\big(\gamma_{\theta_0(u)}(0)^2 - \gamma_\theta(0)^2\big)^2.$$

This shows that Assumption 4.2.2(ii) is fulfilled under condition (c). Based on results of Example 4.5.9, we have with $d_\theta(\lambda) := \frac{A_\theta(\lambda)^{-1}}{\gamma_\theta(0)}$:

$$\begin{aligned}
\mathbb{E}[\sigma(\tilde{Y}_0(u),\theta)^{-2}\cdot\nabla\mu(\tilde{Y}_0(u),\theta)\nabla\mu(\tilde{Y}_0(u),\theta)'] &= \gamma_\theta(0)^2\int_{-\pi}^{\pi}\nabla d_\theta(\lambda)\nabla d_\theta(-\lambda)' f_{\theta_0(u)}(\lambda)\,\mathrm{d}\lambda,\\
\mathbb{E}[\sigma(\tilde{Y}_0(u),\theta)^{-2}\cdot\nabla\sigma(\tilde{Y}_0(u),\theta)\nabla\sigma(\tilde{Y}_0(u),\theta)'] &= \gamma_\theta(0)^{-2}\nabla\gamma_\theta(0)\nabla\gamma_\theta(0)'.
\end{aligned}$$

Together with the continuity of the occuring quantities and $f_{\theta_0(u)}(\lambda) \geq \frac{\delta_A^2}{2\pi}$, $\gamma_\theta(0)^{-2} \geq \delta_G^2$, condition (d) shows that Assumption 4.2.2(iv) holds. Lastly, note that Assumption 4.2.2(v) is fulfilled with $M = 2$, $\chi_k := \max_{i=0,1,2}\sup_{\theta\in\Theta}|\nabla^i\gamma_\theta(k)|$. Results of Katznelson (2004), chapter I, section 4 imply that there exist $C, \eta > 0$ such that for $i = 0, 1, 2$ and all $k \geq 0$ it holds that $\delta(k), \sup_{\theta\in\Theta}|\nabla^i\gamma_\theta(k)| \leq \frac{C}{k^{2+\eta}}$. This shows $\rho(t) \leq \sum_{j=t+1}^{\infty}\chi_j \leq \frac{C}{t^{1+\eta/2}}\sum_{j=1}^{\infty}\frac{1}{j^{1+\eta/2}}$ which is absolutely summable in t. Finally, we have $|j|\cdot|t-j+1| \geq \frac{t}{2}$ for $j = 1, ..., t$ which leads to $\xi(t) := \sum_{j=1}^{t}\chi_j\delta(t-j+1) \leq \frac{C^2 2^{1+\eta/2}}{t^{1+\eta/2}}\sum_{j=1}^{\infty}\frac{1}{j^{1+\eta/2}}$ which is absolutely summable in t.
Under condition (e) we have that $\theta \mapsto a_\theta(k)$ is $l+1$-times differentiable and thus for $k = 1, ..., l$: $\partial_u^k\tilde{X}_t(u) = \sum_{k=0}^{\infty}\partial_u^k a_{\theta_0(u)}(k)\varepsilon_{t-k}$. This leads to $\|\partial_u^k\tilde{X}_t(u)\|_{2M} \leq \sum_{k=0}^{\infty}|\partial_u^k a_{\theta_0(u)}(k)|\cdot\|\varepsilon_0\|_{2M}$ which is finite since $|\nabla^k a_\theta(k)|$ ($k = 1, ..., l$) is (component-wise) absolutely summable by condition (e). Furthermore, one can prove similar to (4.5.16) that (4.2.5) is fulfilled. Lastly, note that Assumption 4.2.4 is fulfilled with $\psi = \chi$. $\square$

Proof of Example 4.2.10: Denote by $w_1, ..., w_r$ the zeros of $p_\theta(w)$. Since $|w_i^{-1}| \leq \frac{1}{1+\rho}$ uniformly for all $\theta \in \Theta$, it is easy to see that $p_\theta(w) = \prod_{i=1}^{r}(1 - w_i^{-1}w)$ implies that there exists $C_a > 0$ such that for all $\theta \in \Theta$, $|a_i| \leq C_a$. C_a can be chosen large enough such that $|b_i| \leq C_a$ holds, too. We conclude that $\Theta \subset [-C_a, C_a]^{r+s} \times [\sigma_{min}, \sigma_{max}]$ is bounded. Since $p_\theta(w), q_\theta(w)$ are continuous in θ,w, there exists $\rho_3 > 0$ such that for $|w| \leq 1+\rho$,

$$p_\theta(w) \neq 0, q_\theta(w) \neq 0 \quad\Leftrightarrow\quad |p_\theta(w)| \geq \rho_3, |q_\theta(w)| \geq \rho_3.$$

This shows that Θ is closed and thus compact. Proposition 2.4 in Dahlhaus and Polonik (2009) shows that there exists a solution $X_{t,n}$ of the form (4.1.5) with $a_{t,n}(j) = \left(\prod_{l=0}^{j-1}A\left(\frac{t-l}{n}\right)\right)_{11}\cdot\sigma\left(\frac{t-j}{n}\right)$, where $A(u) := \begin{pmatrix} (-a_i(u))_{i=1,...,r-1} & -a_r(u)\\ \mathrm{Id}_{r-1} & 0\end{pmatrix}$ and $|a_{t,n}(j)| \leq C(\rho')^j$ with some constants $C > 0$, $1 > \rho' > 0$. We have

$$A_\theta(\lambda) = \sigma\cdot\frac{q_\theta(e^{i\lambda})}{p_\theta(e^{i\lambda})}, \quad \gamma_\theta(0) = \frac{1}{2\pi\sigma}\int_{-\pi}^{\pi}\frac{p_\theta(e^{i\lambda})}{q_\theta(e^{i\lambda})}\,\mathrm{d}\lambda = \frac{1}{2\pi i\sigma}\int_{\{e^{i\lambda}:\lambda\in[-\pi,\pi]\}}\frac{p_\theta(z)}{q_\theta(z)}\frac{1}{z}\,\mathrm{d}z = \sigma^{-1},$$

by the residue theorem. Therefore it is obvious that $A_\theta(\lambda)$ is infinitely differentiable on $(\theta, \lambda) \in \Theta \times [-\pi, \pi]$ with bounded derivatives. By definition of Θ we have that $p_\theta(e^{i\lambda}), q_\theta(e^{i\lambda}) \neq 0$ for all θ, λ which by continuity implies that $\delta_{max} \geq |p_\theta(e^{i\lambda})|, |q_\theta(e^{i\lambda})| \geq \delta_{min}$ with some $\delta_{max} > \delta_{min} > 0$ uniformly in θ, λ. We conclude that $\inf_{\theta,\lambda}|A_\theta(\lambda)| \geq \sigma_{min}\frac{\delta_{min}}{\delta_{max}} =: \delta_A$. Note that here,

$$(\gamma_{\theta'}(0)^2 - \gamma_\theta(0)^2)^2 = ((\sigma')^{-2} - \sigma^{-2})^2 \geq \frac{4}{\sigma_{max}^6}\cdot(\sigma' - \sigma)^2,$$

and

$$\int_{-\pi}^{\pi} \Big|\frac{A_{\theta'}(\lambda)^{-1}}{\gamma_{\theta'}(0)} - \frac{A_{\theta}(\lambda)^{-1}}{\gamma_{\theta}(0)}\Big|^2 \cdot f_\theta(\lambda) \, \mathrm{d}\lambda = \int_{-\pi}^{\pi} \Big|\frac{p_\theta(e^{i\lambda})}{q_\theta(e^{i\lambda})} - \frac{p_{\theta'}(e^{i\lambda})}{q_{\theta'}(e^{i\lambda})}\Big|^2 \cdot f_\theta(\lambda) \, \mathrm{d}\lambda$$
$$\geq \frac{\delta_A^2}{2\pi\delta_{max}^2} \int_{-\pi}^{\pi} \big|p_\theta(e^{i\lambda})q_{\theta'}(e^{i\lambda}) - p_{\theta'}(e^{i\lambda})q_\theta(e^{i\lambda})\big|^2 \, \mathrm{d}\lambda. \qquad (4.5.17)$$

We have (defining $a_j := a_j' := 0$ for $j \notin \{0, ..., r\}$, $b_j := b_j' := 0$ for $j \notin \{0, ..., s\}$):

$$p_\theta(e^{i\lambda})q_{\theta'}(e^{i\lambda}) - p_{\theta'}(e^{i\lambda})q_\theta(e^{i\lambda}) = \sum_{j,k\geq 0} (a_j b_k' - a_j' b_k)e^{i\lambda(j+k)} = \sum_{d\geq 0} \tilde{f}(d)e^{i\lambda d},$$

where $\tilde{f}(d) := \sum_{j\geq 0}(a_j b_{d-j}' - a_j' b_{d-j}) = \sum_{k\geq 0} a_{d-k}(b_k' - b_k) + \sum_{j\geq 0} b_{d-j}(a_j - a_j')$. This shows that (4.5.17) is lower bounded by

$$\frac{\delta_A^2}{\delta_{max}^2} \sum_{d\geq 0} \tilde{f}(d)^2 = \frac{\delta_A^2}{\delta_{max}^2} \begin{pmatrix} \underline{a} - \underline{a}' \\ \underline{b}' - \underline{b} \end{pmatrix}' P(\theta) \begin{pmatrix} \underline{a} - \underline{a}' \\ \underline{b}' - \underline{b} \end{pmatrix},$$

where $\underline{a} := (a_1, ..., a_r)', \underline{b} := (b_1, ..., b_s)'$ ($\underline{a}', \underline{b}'$ are similarly defined), and $P(\theta) := \begin{pmatrix} A(p_\theta, p_\theta) & A(p_\theta, q_\theta) \\ A(q_\theta, p_\theta) & A(q_\theta, q_\theta) \end{pmatrix}$, where $(A(\psi_1, \psi_2))_{jk} := \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_1(e^{i\lambda})\overline{\psi_2(e^{i\lambda})} e^{i\lambda(j-k)} \, \mathrm{d}\lambda$. Note that for some vector $\tilde{x} = (u, v) \in \mathbb{R}^{r+s}$, we have $\tilde{x}'P(\theta)\tilde{x} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \big|p_\theta(e^{i\lambda}) \sum_{j=1}^r u_j e^{i\lambda j} + q_\theta(e^{i\lambda}) \sum_{k=1}^s v_k e^{i\lambda k}\big|^2 \, \mathrm{d}\lambda \geq 0$ with equality if and only if

$$p_\theta(e^{i\lambda}) \sum_{j=1}^r u_j e^{i\lambda j} = q_\theta(e^{i\lambda}) \sum_{k=1}^s (-v_k)e^{i\lambda k}.$$

Since p_θ, q_θ have no common zeros for $\theta \in \Theta$, this can not be achieved. Thus, $P(\theta)$ is positive definite for all $\theta \in \Theta$ and thus has only positive eigenvalues. Since $P(\theta)$ is continuous in θ, the minimal eigenvalue $\lambda_{min}(P(\theta)) = \inf_{|x|_2=1} x'P(\theta)x$ is continuous in θ, too. By compactness of Θ, we conclude that $\inf_{\theta\in\Theta} \lambda_{min}(P(\theta)) \geq \lambda_0$ with some $\lambda_0 > 0$, leading to the lower bound $\frac{\delta_A^2}{\delta_{max}^2}\lambda_0\big(|\underline{a}' - \underline{a}|_2^2 + |\underline{b}' - \underline{b}|_2^2\big)$ for (4.5.17).

To show the conditions of Assumption 4.2.6, define $B(\theta) := \begin{pmatrix} (-b_i)_{i=1,...,s-1} & -b_s \\ \mathrm{Id}_{s-1} & 0 \end{pmatrix}$ for all $\theta = (a_1, ..., a_r, b_1, ..., b_s, \sigma) \in \Theta$. It is easily seen that

$$H(x, y, (\theta^{(k)})_{k\geq 0}) = \frac{1}{\sigma^{(0)}} \cdot \Big(x + \sum_{k=1}^{\infty} y_k \cdot \sum_{j=(k-p)\vee 0}^{k} \Big(\prod_{l=0}^{j-1} B(\theta^{(l)})\Big)_{11} a_{k-j}^{(0)}\Big),$$

and $\ell := \frac{1}{2}H^2 + \log \sigma^{(0)}$, where the empty product is defined as 1. We now use a similar argumentation as in the proof of Proposition 2.4 in Dahlhaus and Polonik (2009). Since $\det(\lambda \cdot \mathrm{Id}_s - B(\theta)) = \lambda^p \sum_{j=0}^s b_j \lambda^{-j})$, it follows that $\lambda_{\max}^{|\cdot|}(B(\theta)) \leq \frac{1}{1+\rho}$ for all $\theta \in \Theta$, where $\lambda_{max}^{|\cdot|}(B) := \max\{|\lambda| : \lambda \text{ eigenvalue of } B\}$. By Householder (1964), page 46, there exists a positive definite matrix $M(\theta)$ with $\|B\|_{M(\theta)} \leq \lambda_{max}^{|\cdot|}(B) + \varepsilon$ for every $\varepsilon > 0$ and

$\theta \in \Theta$. Here, $\lvert A \rvert_M := \sup\{\lvert A\tilde{x}\rvert_M : \lvert \tilde{x}\rvert_M = 1\}$ and $\lvert \tilde{x}\rvert_M := |M^{-1}\tilde{x}|_1$. Since $B = B(\theta)$ is continuous in $\theta \in \Theta$, we can find a finite partition $\Theta = \Theta_1 \cup ... \cup \Theta_{\tilde{m}}$ and matrices $M_k := M(\theta_k)$ with $\theta_i \in \Theta_i$ such that $\lvert B(\theta)\rvert_{M_i} \leq \tilde{\rho} := (1 + \frac{\varrho}{2})^{-1} < 1$. There exists a constant c_0 such that $|B|_1 \leq c_0 \lvert B\rvert_{M_k}$ for all k since M_k is positive definite. It is now easy to see that for each $j > 0$ we have

$$\begin{aligned}
\Big(\prod_{l=0}^{j-1} B(\theta^{(l)})\Big)_{11} &\leq \Big|\prod_{l=0}^{j-1} B(\theta^{(l)})\Big|_1 \leq \prod_{k=1}^{\tilde{m}}\Big| \prod_{\theta^{(l)}\in\Theta_k, 0\leq l\leq j-1} B(\theta^{(l)})\Big|_1 \\
&\leq c_0^{\tilde{m}} \prod_{k=1}^{\tilde{m}}\Big\lvert \prod_{\theta^{(l)}\in\Theta_k, 0\leq l\leq j-1} B(\theta^{(l)})\Big\rvert_{M_k} \leq c_0^{\tilde{m}} \prod_{k=1}^{\tilde{m}} \tilde{\rho}^{\#\{0\leq l\leq j-1:\theta^{(l)}\in\Theta_k\}} \leq c_0^{\tilde{m}} \cdot \tilde{\rho}^j.
\end{aligned}$$

Let $\nabla_{\theta^{(m)}}$ denote the derivative with respect to $\theta^{(m)}$. Obviously, ℓ is differentiable with

$$\nabla_{\theta^{(m)}}\ell = H\nabla_{\theta^{(m)}}H - \mathbb{1}_{\{m=0\}} \cdot (0, ..., 0, \frac{1}{\sigma^{(0)}})'. \tag{4.5.18}$$

Furthermore, $\partial_{\sigma^{(m)}}H = -\frac{1}{\sigma^{(0)}}H\mathbb{1}_{\{m=0\}}$, $\partial_{a_i^{(m)}}H = \mathbb{1}_{\{m=0\}}\frac{1}{\sigma^{(0)}}\sum_{k=i}^{\infty} y_k\cdot\Big(\prod_{l=0}^{k-i-1} B(\theta^{(l)})\Big)_{11}$ and

$$\begin{aligned}
\partial_{b_i^{(m)}}H &= \frac{1}{\sigma^{(0)}}\sum_{k=m+1}^{\infty} y_k \cdot \sum_{j=(k-p)\vee(m+1)}^{k} \Big(\Big(\prod_{l=0}^{m-1} B(\theta^{(l)})\Big)\cdot \partial_{b_i^{(m)}}B(\theta^{(m)}) \\
&\qquad \times\Big(\prod_{l=m+1}^{j-1} B(\theta^{(l)})\Big)\Big)_{11} a_{k-j}^{(0)},
\end{aligned}$$

where $\partial_i^{(m)}B(\theta^{(m)}) = \begin{pmatrix} (-\mathbb{1}_{\{\nu=i\}})_{\nu=1,...,s} \\ 0 \end{pmatrix}$. It is easy to see that $\nabla_{\theta^{(0)}}H(z,\theta)$ with $z = (x,y)$ lies in $\mathcal{L}(1, (\tilde{\rho}^j)_{j\geq 1}, C_z, C_\theta)$ with some constant vectors C_z, C_θ. For $m > 0$, we have

$$\begin{aligned}
&\big|\partial_{b_i^{(m)}}H(z, (\theta_k)_{k\geq 0}) - \partial_{b_i^{(m)}}H(z', (\theta_k)_{k\geq 0})\big| \\
\leq\ & \frac{1}{\sigma^{(0)}}\sum_{k=m+1}^{\infty} |y_k - y_k'| \cdot \sum_{j=(k-p)\vee(m+1)}^{k} c_0^{2\tilde{m}}\tilde{\rho}^{j-1}|a_{k-j}^{(0)}| \\
\leq\ & \frac{C_a c_0^{2\tilde{m}} p}{\sigma_{min}} \sum_{k=m+1}^{\infty} |y_k - y_k'| \cdot \tilde{\rho}^{k-p-1}.
\end{aligned}$$

Similarly, for $m_2 > 0$,

$$\begin{aligned}
&\big|\partial_{b_i^{(m)}}H(z, (\theta_k)_{k\geq 0}) - \partial_{b_i^{(m)}}H(z, (\theta_k + (\tilde{\theta}_k - \theta_k)\mathbb{1}_{\{k=m_2\}})_{k\geq 0})\big| \\
\leq\ & \frac{C_a c_0^{2\tilde{m}} p}{\sigma_{min}} \sum_{\nu=1}^{s} |b_\nu^{(m_2)} - \tilde{b}_\nu^{(m_2)}| \cdot \sum_{k=m+1}^{\infty} |y_k|\tilde{\rho}^{k-p-2}.
\end{aligned}$$

More calculations of the same kind together with (4.5.18) imply that Assumption 4.2.6 is fulfilled. □

4.5.5 Proof of the lower bound

Let us define a class of functions $\mathcal{L}$ which is similar to $\mathcal{L}$ but does not ask for Lipschitz continuity in θ.

Definition 4.5.11 (The class $\mathcal{L}(M,\chi,C_z)$). *We say that a function $g:\mathbb{R}^{\mathbb{N}}\to\mathbb{R}^p$ is in the class $\mathcal{L}(M,\chi,C_z)$ if $\tilde{g}:\mathbb{R}^{\mathbb{N}}\times\Theta\to\mathbb{R}^p, \tilde{g}(z,\theta):=g(z)$ is in $\mathcal{L}(M,\chi,C_z,1)$.*

Proof of Theorem 4.1.3: Fix some $u\in(0,1)$. Note that

$$\inf_{\hat{\theta}\in\sigma(X_{1,n},\dots,X_{n,n})}\sup_{\theta_0\in\Sigma(\beta,L)}\mathbb{E}_{\theta_0}|\hat{\theta}(u)-\theta_0(u)|^2\geq\inf_{\hat{\theta}\in\sigma(X_{t,n}:t\in\mathbb{Z},t\leq n)}\sup_{\theta_0\in\Sigma(\beta,L)}\mathbb{E}_{\theta_0}|\hat{\theta}(u)-\theta_0(u)|^2,$$

because the second infimum is taken over a larger set of estimators. We now use the general reduction scheme from Tsybakov (2009), section 2.2. Let $\mathbb{P}_{\tilde{Y}_0(0)|\theta_0}:\mathbb{R}^{\mathbb{N}}\to\mathbb{R}$ denote the stationary distribution of $\tilde{Y}_0(0)$ which is only dependent on $\theta_0(0)$. Note that the log density of $X_{t,n}$ given the infinite past $Y_{t-1,n}$ is given by

$$\log f_{X_{t,n}|Y_{t-1,n},\theta_0}(X_{t,n},Y_{t-1,n})=-\ell\big(Y_{t,n},\theta_0\big(\frac{t-j}{n}\vee 0\big)_{j\geq 0}\big).$$

Let $\mathbb{P}_{Y_{n,n}|\theta_0}$ denote the distribution of $Y_{n,n}$. Then we have $\mathbb{P}_{Y_{n,n}|\theta_0}\ll\mathbb{P}_{\tilde{Y}_0(0)|\theta_0}$ with density

$$\frac{d\mathbb{P}_{Y_{n,n}|\theta_0}}{d\mathbb{P}_{\tilde{Y}_0(0)|\theta_0}}(Y_{n,n}) \;=\; \prod_{t=1}^{n}\exp\big\{-\ell\big(Y_{t,n},\theta_0\big(\frac{t-j}{n}\vee 0\big)_{j\geq 0}\big)\big\}.$$

Define $b=b_n=\gamma n^{-1/(2\beta+1)}$ where γ is specified below. Choose n large enough such that $u\in[\frac{b}{2},1-\frac{b}{2}]$. Define $\theta_1:[0,1]\to\Theta,\theta_1(v)=\theta_0(v)+\mathbb{1}Lb^{\beta}K\big(\frac{v-u}{b}\big)$, where $\mathbb{1}=(1,...,1)'\in\mathbb{R}^d$ and $K:\mathbb{R}\to\mathbb{R}$ is infinitely often continuously differentiable (i.e. in $C^{\infty}(\mathbb{R})$) with compact support $[-\frac{1}{2},\frac{1}{2}]$, for instance $K(v)=a\exp(-1/(1-4v^2))\cdot\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(v)$ with $a>0$ small enough. The function θ_1 is well-defined for n large enough since K is uniformly bounded by $|K|_{\infty}$.
Define $\psi_n:=n^{\frac{-\beta}{2\beta+1}}$. Note that

$$|\theta_0(u)-\theta_1(u)|_2=\sqrt{d}Lb^{\beta}\cdot K(0)=2A\psi_n$$

if we define $A:=\sqrt{d}\frac{L}{2}\gamma^{\beta}K(0)$. By construction, we have $\theta_1(0)=\theta_0(0)+Lb^{\beta}K\left(\frac{-u}{b}\right)=\theta_0(0)$. This shows $\mathbb{P}_{\tilde{Y}_0(0)|\theta_0}=\mathbb{P}_{\tilde{Y}_0(0)|\theta_1}$. The Kullback-Leibler divergence between the models with true curves θ_0 or θ_1, respectively, reads (let $e_k:=(\delta_{jk})_{j\geq 0}$ where $\delta_{jk}=1$ if and only if $j = k$)

$$
\begin{aligned}
& K(\mathbb{P}_{Y_{n,n}|\theta_0}, \mathbb{P}_{Y_{n,n}|\theta_1}) \\
= \;& \mathbb{E}_{\theta_0}\Big[\log \frac{d\mathbb{P}_{Y_{n,n}|\theta_0}}{d\mathbb{P}_{\tilde{Y}_0(0)|\theta_0}}(Y_{n,n}) - \log \frac{d\mathbb{P}_{Y_{n,n}|\theta_1}}{d\mathbb{P}_{\tilde{Y}_0(0)|\theta_1}}(Y_{n,n})\Big] \\
= \;& \sum_{t=1}^{n} \mathbb{E}_{\theta_0}\Big[\ell\big(Y_{t,n}, \theta_1\big(\frac{t-j}{n} \vee 0\big)_{j\geq 0}\big) - \ell\big(Y_{t,n}, \theta_0\big(\frac{t-j}{n} \vee 0\big)_{j\geq 0}\big)\Big] \\
= \;& \sum_{t=1}^{n}\sum_{k=0}^{t-1}\Big\{\mathbb{E}_{\theta_0}\Big[\big\langle \nabla_{\theta_k}\ell\big(Y_{t,n}, \theta_0\big(\frac{t-j}{n} \vee 0\big)_{j\geq 0}\big), \theta_1\big(\frac{t-k}{n}\big) - \theta_0\big(\frac{t-k}{n}\big)\big\rangle\Big] \\
& + \mathbb{E}_{\theta_0}\Big[\int_0^1 \big\langle \nabla_{\theta_k}\ell\big(Y_{t,n}, \theta_0\big(\frac{t-j}{n} \vee 0\big)_{j\geq 0} + \sum_{j=0}^{k-1}\big\{\theta_1\big(\frac{t-j}{n}\big) - \theta_0\big(\frac{t-j}{n}\big)\big\} \\
& \qquad + s\cdot e_k\big\{\theta_1\big(\frac{t-k}{n}\big) - \theta_0\big(\frac{t-k}{n}\big)\big\}\big) - \nabla_{\theta_k}\ell\big(Y_{t,n}, \theta_0\big(\frac{t-j}{n} \vee 0\big)_{j\geq 0}\big), \\
& \qquad \theta_1\big(\frac{t-k}{n}\big) - \theta_0\big(\frac{t-k}{n}\big)\big\rangle\Big]\Big\}
\end{aligned} \tag{4.5.19}
$$

For the first expectation in (4.5.19), note that component-wise

$$
\begin{aligned}
& \big|\nabla_{\theta_k}\ell(Y_{t,n}, \theta_0\big(\frac{t-j}{n} \vee 0\big)_{j\geq 0}) - \nabla_{\theta_k}\ell(Y_{t,n}, \theta_0\big(\frac{t}{n}\big)_{j\geq 0})\big| \\
\leq \;& C_{\nabla\ell,1}(k)\sum_{j=0}^{\infty} C_{\nabla\ell,2}(j)\cdot\big|\theta_0\big(\frac{t-j}{n} \vee 0\big) - \theta_0\big(\frac{t}{n}\big)\big|_2 \cdot \big(1 + |Y_{t,n}|_{\chi,1}^M\big) \\
\leq \;& C_{\nabla\ell,1}(k)\big(1 + |Y_{t,n}|_{\chi,1}^M\big)\cdot |L_{\theta_0}|_2 \cdot n^{-\beta'}\sum_{j=0}^{\infty} jC_{\nabla\ell,2}(j).
\end{aligned}
$$

Note furthermore that

$$
\begin{aligned}
& \||Y_{t,n}|_{\chi,1}\|_M \leq \sum_{j=1}^{\infty}\chi_j\|X_{t-j+1,n}\|_M \\
\leq \;& \sum_{j=1}^{\infty}\chi_j\|X_{t-j+1,n} - \tilde{X}_{t-j+1}\big(\frac{t-j+1}{n}\big)\|_M + \sum_{j=1}^{\infty}\chi_j\|\tilde{X}_{t-j+1}\big(\frac{t-j+1}{n}\big)\|_M \\
\leq \;& \sum_{j=1}^{\infty}\chi_j\cdot\big(C_{B,1} + C_X N_\alpha(M)\big) =: D_X.
\end{aligned}
$$

This shows that the first term in (4.5.19) can be replaced by

$$
\sum_{t=1}^{n}\sum_{k=0}^{t-1}\mathbb{E}_{\theta_0}\Big[\big\langle\nabla_{\theta_k}\ell\big(Y_{t,n}, \theta_0\big(\frac{t}{n}\big)_{j\geq 0}\big), \theta_1\big(\frac{t-k}{n}\big) - \theta_0\big(\frac{t-k}{n}\big)\big\rangle\Big] \tag{4.5.20}
$$

with a replacement error bounded by

$$n^{-\beta'}b^\beta nb \cdot L|L_{\theta_0}|_2 \cdot \sum_{j=0}^\infty jC_{\nabla\ell,2}(j) \cdot \sum_{k=0}^\infty C_{\nabla\ell,1}(k) \cdot |K|_\infty \big(1 + D_X^M\big) = o(1),$$

since $n^{-\beta'}b^\beta nb = o(1)$. Note that

$$\big|\nabla_{\theta_k}\ell(Y_{t,n}, \theta_0(\frac{t}{n})_{j\geq 0})\big|_1 \leq d \cdot C_{\nabla\ell,1}(k)\big(1 + |Y_{t,n} - y_0|_{\chi,1}^M\big) + C_{\nabla\ell,1}(k).$$

Therefore we can replace (4.5.20) by

$$\begin{aligned} & \sum_{t=1}^n \mathbb{E}_{\theta_0}\Big[\sum_{k=0}^\infty \big\langle \nabla_{\theta_k}\ell\big(Y_{t,n}, \theta_0(\frac{t}{n})_{j\geq 0}\big), \theta_1(\frac{t}{n}) - \theta_0(\frac{t}{n})\big\rangle\Big] \\ = & \sum_{t=1}^n \mathbb{E}_{\theta_0}\Big[\big\langle \nabla\ell\big(Y_{t,n}, \theta_0(\frac{t}{n})\big), \theta_1(\frac{t}{n}) - \theta_0(\frac{t}{n})\big\rangle\Big]. \end{aligned} \tag{4.5.21}$$

with replacement error bounded by

$$\begin{aligned} & Lb^\beta \cdot \sum_{t=1}^n \sum_{k=0}^\infty \Big|K\Big(\frac{(t-k)/n - u}{b}\Big) - K\Big(\frac{t/n - u}{b}\Big)\Big| \cdot C_{\nabla\ell,1}(k) \\ & \qquad \times\Big(1 + d\big(1 + 2^M(D_X^M + |y_0|_{\chi,1}^M)\big)\Big) \\ \leq \ & Lb^\beta \cdot B_K \sum_{k=0}^\infty k \cdot C_{\nabla\ell,1}(k) \cdot \Big(1 + d\big(1 + 2^M(D_X^M + |y_0|_{\chi,1}^M)\big)\Big) = O(b^\beta), \end{aligned}$$

where B_K is the variation of K. Finally, since $\nabla\ell\big(\tilde{Y}_t(\frac{t}{n}), \theta_0(\frac{t}{n})\big)$ is a martingale difference sequence, we can replace (4.5.21) by

$$\sum_{t=1}^n \mathbb{E}_{\theta_0}\Big[\big\langle \nabla\ell\big(\tilde{Y}_t(\frac{t}{n}), \theta_0(\frac{t}{n})\big), \theta_1(\frac{t}{n}) - \theta_0(\frac{t}{n})\big\rangle\Big] = 0$$

with replacement error

$$\begin{aligned} Lb^\beta \sum_{t=1}^n \big\|\nabla\ell\big(Y_{t,n}, \theta_0(\frac{t}{n})\big) - \nabla\ell\big(\tilde{Y}_t(\frac{t}{n}), \theta_0(\frac{t}{n})\big)\big\|_1 \cdot \Big|K\Big(\frac{t/n - u}{b}\Big)\Big| \ &\leq \ Lb^\beta|K|_\infty \cdot nb \cdot Cn^{-\beta'} \\ &= \ o(1). \end{aligned}$$

with some constant $C > 0$ (the proof is similar to Lemma 4.5.1). The second expectation in (4.5.19) is bounded by

$$
\begin{aligned}
& d\cdot\sum_{t=1}^{n}\sum_{k=0}^{t-1} C_{\nabla\ell,1}(k)\cdot\big(1+\||Y_{t,n}|_{\chi,1}\|_M^M\big)\\
&\qquad\qquad \times\sum_{j=0}^{\infty} C_{\nabla\ell,2}(j)\cdot\big|\theta_1\big(\frac{t-j}{n}\big)-\theta_0\big(\frac{t-j}{n}\big)\big|_2\cdot\big|\theta_1\big(\frac{t-k}{n}\big)-\theta_0\big(\frac{t-k}{n}\big)\big|_2\\
\leq\ & dL^2b^{2\beta}\cdot\big(1+D_X^M\big)\sum_{t=1}^{n}\sum_{k=0}^{t-1} C_{\nabla\ell,1}(k)\big|K\big(\frac{(t-k)/n-u}{b}\big)\big|\\
&\qquad\qquad\qquad\qquad \times\sum_{j=0}^{\infty} C_{\nabla\ell,2}(j)\cdot\big|K\big(\frac{(t-j)/n-u}{b}\big)\big|\\
\leq\ & dL^2nb^{2\beta+1}\cdot\big(1+D_X^M\big)|K|_\infty^2\cdot\sum_{k=0}^{\infty} C_{\nabla\ell,1}(k)\cdot\sum_{j=0}^{\infty} C_{\nabla\ell,2}(j).
\end{aligned}
$$

So for arbitrarily chosen $\alpha > 0$, we can achieve $K(\mathbb{P}_{Y_{n,n}|\theta_0},\mathbb{P}_{Y_{n,n}|\theta_1}) \leq \alpha$ if we choose $\gamma := \big(\frac{\alpha/2}{dL^2(1+D_X^M)|K|_\infty^2\sum_{k=0}^{\infty}C_{\nabla\ell,1}(k)\sum_{j=0}^{\infty}C_{\nabla\ell,2}(j)}\big)^{\frac{1}{2\beta+1}}$ and choose n large enough.

□

4.5.6 Maximum likelihood basic inequality

Since an explicit representation of the maximum likelihood estimator $\hat{\theta}_b(u)$ is usually not available and the well-known representation $\hat{\theta}_b(u)-\theta_0(u) = -\nabla^2 L_{n,h}(u,\bar{\theta}_n(u))^{-1}\cdot\nabla L_{n,b}(u,\theta_0(u))$ is only available if $\hat{\theta}_b(u)$ is a point in the interior of the parameter space Θ, we first have to develop a 'basic inequality' for the difference $\hat{\theta}_b(u)-\theta_0(u)$ which is only based on the behavior of the likelihood $L_{n,b}(u,\theta)$ itself. To obtain rates, Assumption 4.2.2 is crucial.

Lemma 4.5.12 (Basic inequality and maximum likelihood rate)**.** *Suppose that Assumption 4.2.2, 4.2.5 and 4.2.3 hold. Fix some $\gamma > 0$. Assume that $b \geq b_*(M,\chi,C_\ell,\frac{\gamma}{8C_L(u)})$, where b_* is defined in Lemma 4.5.5. Then there exists a set $\hat{A}_{n,b}$ and a constant $C_{mle} = C_{mle}(\gamma) > 0$ such that $\mathbb{P}(\hat{A}_{n,b}) \leq C_{mle}\cdot\big(n^{-2\beta'}+n^{-1}\big)$ and*

$$\big\{|\hat{\theta}_b(u)-\theta_0(u)|_2^2 > \gamma\big\} \subset \hat{A}_{n,b}\cup\big\{\sup_{\theta\in\Theta}|B_{n,b}(\ell(\cdot,\theta),u)| > \frac{\gamma}{8C_L(u)}\big\} \tag{4.5.22}$$

Proof of Lemma 4.5.12: We follow the approach of Van de Geer (2000), page 248, before Lemma 12.1. Because $\hat{\theta}_b(u)$ is a minimizer of $\theta \mapsto L_{n,b}(u,\theta)$, we have

$$\begin{aligned} 0 &\geq L_{n,b}(u,\hat{\theta}_b(u)) - L_{n,b}(u,\theta_0(u)) \\ &= \big\{S_{n,b}(\ell(\cdot,\hat{\theta}_b(u)),u) - S_{n,b}(\ell(\cdot,\theta_0(u)),u)\big\} \\ &\quad +\big\{C_{n,b}(\ell(\cdot,\hat{\theta}_b(u)),u) - C_{n,b}(\ell(\cdot,\theta_0(u)),u)\big\} \\ &\quad +\big\{E_{n,b}(\ell(\cdot,\hat{\theta}_b(u)),u) - E_{n,b}(\ell(\cdot,\theta_0(u)),u)\big\} \\ &\quad +\big\{B_{n,b}(\ell(\cdot,\hat{\theta}_b(u)),u) - B_{n,b}(\ell(\cdot,\theta_0(u)),u)\big\} \\ &\quad +\big\{L(u,\hat{\theta}_b(u)) - L(u,\theta_0(u))\big\}. \end{aligned}$$

Note that by Assumption 4.2.5, we have $F_{n,b}(u) \geq \frac{c_0}{|K|_\infty}(nb)^{1/2}$. By Assumption 4.2.2, we have $L(u,\hat{\theta}_b(u)) - L(u,\theta_0(u)) \geq \frac{1}{C_L(u)}|\hat{\theta}_b(u)-\theta_0(u)|_2^2$. This shows (4.5.22) with

$$\begin{aligned} \hat{A}_{n,b} &= \big\{\sup_{\theta\in\Theta}|S_{n,b}(\ell(\cdot,\theta),u)| > \frac{\gamma}{8C_L(u)}\big\} \cup \big\{\sup_{\theta\in\Theta}|C_{n,b}(\ell(\cdot,\theta),u)| > \frac{\gamma}{8C_L(u)}\big\} \\ &\quad \cup\big\{\sup_{\theta\in\Theta}|E_{n,b}(\ell(\cdot,\theta),u)| > \frac{\gamma}{8C_L(u)}\big\}. \end{aligned}$$

An application of Markov's inequality to the first two sets and Lemma 4.5.1, 4.5.2 and 4.5.5 gives the result. □

4.5.7 Proof of the upper bound

Proof of Theorem 4.1.4: Let $\gamma_I(\Xi), \gamma_V(\Xi), \gamma_\theta(\Xi)$ be the variables from Lemma 4.5.13. Define $\gamma_{mle}(\Xi,u) := \frac{\delta_\Theta}{2} \wedge \frac{\lambda_{min}(V(u))}{4|C_{\nabla^2\ell,\theta}|_2(1+|\chi|_1^M C_X^M N_\alpha(M)^M)} \wedge \gamma_\theta$. Define $\Omega_{1,b}(u) := \{|\hat{\theta}_b(u) - \theta_0(u)|_2 \leq \gamma_{mle}(\Xi,u)\}$, $\Omega_{2,b}(u) := \{\sup_{\theta\in\Theta}|\nabla^2 L_{n,b}(u,\theta) - \nabla^2 L(u,\theta)|_2 \leq \frac{\lambda_{min}(V(u))}{4}\}$ and recall the definition of $\Omega_{pen,\Xi,b}(u)$ from Lemma 4.5.13. Define $A_\Xi(b) := \bigcap_{b'\in B_n, b'\leq b}\big(\Omega_{1,b'}\cap \Omega_{2,b'}\cap\Omega_{pen,\Xi,b'}\big)$. We now show that in both cases $\Xi = \mathrm{Id}$ and $\Xi = V(u)$ an inequality of the form

$$\begin{aligned} |\hat{\theta}_{\hat{b}_\Xi} - \theta_0|_2^2 &\leq C_1(\Xi)\max_{b'\in B_n, b'\leq b}\big\{|\Xi^{1/2}V^{-1}\nabla L_{n,b'}(\theta_0)|_2^2 - 32P_n(b')\big\}_+ \\ &\quad + C_2(\Xi)C_P\max_{b'\in B_n, b'\geq b}\hat{P}_n(b') + C_3(\Xi)\mathbb{1}_{A(b)^c} \end{aligned} \tag{4.5.23}$$

with some (numerical) constants $C_i(\Xi)$, $i=1,2,3$ holds for all $b\in B_n$. In the following we assume that $A_\Xi(b)$ holds and that $b'\in B_n, b'\leq b$. To keep the notation simple, the arguments Ξ and $u\in[0,1]$ are omitted in the following case distinction.
Case $\Xi = \mathrm{Id}$: Note that the penalization term $P_n(b)$ is not monotone. For $b\in B_n$ it holds that

$$|\hat{\theta}_{\hat{b}} - \theta_0|_2^2 \leq 3\Big(|\hat{\theta}_{\hat{b}} - \hat{\theta}_{b\vee\hat{b}}|_2^2 + |\hat{\theta}_{b\vee\hat{b}} - \hat{\theta}_b|_2^2 + |\hat{\theta}_b - \theta_0|_2^2\Big).$$

Now we have

$$\begin{aligned} |\hat{\theta}_{\hat{b}} - \hat{\theta}_{b\vee\hat{b}}|_2^2 &\leq \big\{|\hat{\theta}_{\hat{b}} - \hat{\theta}_{b\vee\hat{b}}|_2^2 - \widehat{\mathrm{pen}}_n(\hat{b})\big\}_+ + \widehat{\mathrm{pen}}_n(\hat{b}) \\ &\leq \max_{b'\in B_n}\big\{|\hat{\theta}_{b'} - \hat{\theta}_{b\vee b'}|_2^2 - \widehat{\mathrm{pen}}_n(b')\big\}_+ + \widehat{\mathrm{pen}}_n(\hat{b}) \leq Y(b) + \widehat{\mathrm{pen}}_n(\hat{b}). \end{aligned}$$

Using similar arguments we can prove that $|\hat{\theta}_{b\vee\hat{b}} - \hat{\theta}_b|_2^2 \leq Y(\hat{b}) + \widehat{\text{pen}}_n(b)$, thus (note that $\hat{b}$ minimizes $b \mapsto Y(b) + \widehat{\text{pen}}_n(b)$):

$$|\hat{\theta}_{\hat{b}} - \theta_0|_2^2 \leq 6\Big(Y(b) + \widehat{\text{pen}}_n(b)\Big) + 3|\hat{\theta}_b - \theta_0|_2^2 \tag{4.5.24}$$

Now, we upper bound $Y(b)$. We have:

$$\begin{aligned}
Y(b) &\leq \max_{b'\in B_n, b'\leq b}\big\{|\hat{\theta}_b - \theta_0|_2^2 - \max_{b''\in B_n, b''\geq b'} C_P\hat{P}_n(b'')\big\}_+ \\
&\qquad\qquad + \max_{b'\in B_n, b'\leq b}\big\{|\hat{\theta}_{b'} - \theta_0|_2^2 - C_P\hat{P}_n(b')\big\}_+ \\
&\leq \big\{|\hat{\theta}_b - \theta_0|_2^2 - C_P\hat{P}_n(b)\big\}_+ + \max_{b'\in B_n, b'\leq b}\big\{|\hat{\theta}_{b'} - \theta_0|_2^2 - C_P\hat{P}_n(b')\big\}_+ \\
&\leq 2\max_{b'\in B_n, b'\leq b}\big\{|\hat{\theta}_{b'} - \theta_0|_2^2 - C_P\hat{P}_n(b')\big\}_+.
\end{aligned}$$

In the same way, we have

$$|\hat{\theta}_b - \theta_0|_2^2 \leq \big\{|\hat{\theta}_b - \theta_0|_2^2 - C_P\hat{P}_n(b)\big\}_+ + C_P\hat{P}_n(b).$$

We finally conclude that for all $b \in B_n$:

$$|\hat{\theta}_{\hat{b}} - \theta_0|_2^2 \leq 15\cdot\big(\max_{b'\in B_n, b'\leq b}\big\{|\hat{\theta}_{b'} - \theta_0|_2^2 - C_P\hat{P}_n(b')\big\}_+ + \max_{b'\in B_n, b'\geq b} C_P\hat{P}_n(b')\big). \tag{4.5.25}$$

We now discuss the first summand in (4.5.25). By Lemma 4.5.15, we have

$$|\nabla^2 L(\theta) - \nabla^2 L(\theta')|_2 \leq |C_{\nabla^2\ell,\theta}|_2\cdot|\theta - \theta'|_2\cdot\big(1 + |\chi|_1^M C_X^M\big).$$

Since $A(b') \subset \Omega_{1,b'}(u)$ and thus $|\hat{\theta}_{b'}(u) - \theta_0(u)|_2 \leq \frac{\delta_\Theta}{2}$, $\hat{\theta}_{b'}(u)$ is in a ball around $\theta_0(u)$ which is completely contained in Θ. Thus, with some $\bar{\theta}(u) \in \Theta$ with $|\bar{\theta}(u) - \theta_0(u)|_2 \leq |\hat{\theta}_{b'}(u) - \theta_0(u)|_2$:

$$-\nabla L_{n,b'}(u, \theta_0(u)) = \nabla^2 L_{n,b'}(u, \bar{\theta}(u))\cdot(\hat{\theta}_{b'}(u) - \theta_0(u)). \tag{4.5.26}$$

Recall that $V(u) = \nabla^2 L(u, \theta_0(u))$. Because $A(b') \subset \Omega_{1,b'}(u)\cap\Omega_{2,b'}(u)$ it holds that

$$\begin{aligned}
&|\nabla^2 L_{n,b'}(u, \bar{\theta}(u)) - V(u)|_2 \\
\leq\ & |\nabla^2 L_{n,b'}(u, \bar{\theta}(u)) - \nabla^2 L(u, \bar{\theta}(u))|_2 + |\nabla^2 L(u, \bar{\theta}(u)) - V(u)|_2 \\
\leq\ & \frac{\lambda_{min}(V(u))}{2}.
\end{aligned}$$

Lemma 4.5.17 implies that $\nabla^2 L_{n,b'}(u, \bar{\theta}(u))$ is invertible and, by (4.5.26),

$$|\hat{\theta}_{b'}(u) - \theta_0(u)|_2 \leq 2|V(u)^{-1}\nabla L_{n,b'}(u, \theta_0(u))|_2.$$

Define $\tilde{L}_{n,b}(u,\theta) := \frac{1}{K_{n,b}(u)}\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)\tilde{\ell}_t(u,\theta)$. We conclude that

$$\begin{aligned}
&\max_{b'\in B_n, b'\leq b}\big\{|\hat{\theta}_{b'} - \theta_0|_2^2 - C_P\hat{P}_n(b')\big\}_+ \\
\leq\ & \max_{b'\in B_n, b'\leq b}\big\{4|V^{-1}\cdot\nabla L_{n,b'}(\theta_0)|_2^2 - \frac{C_P}{2}P_n(b')\big\}_+\mathbb{1}_{A(b)} + \Theta_{max}^2\mathbb{1}_{A(b)^c}.
\end{aligned}$$

By (4.5.25), we have shown that for all $b \in B_n$:

$$\begin{aligned}|\hat{\theta}_{\hat{b}} - \theta_0|_2^2 &\leq 15 \max_{b' \in B_n, b' \leq b} \big\{4|V^{-1} \cdot \nabla L_{n,b'}(\theta_0)|_2^2 - \frac{C_P}{2} P_n(b')\big\}_+ \mathbb{1}_{A(b)} \\ &\quad +15 \max_{b' \in B_n, b' \geq b} C_P \hat{P}_n(b') + \Theta_{max}^2 \mathbb{1}_{A(b)^c},\end{aligned}$$

so (4.5.23) is fulfilled with $C_1(\Xi) = 60$, $C_2(\Xi) = 15 C_P(\Xi)$, where $C_P(\Xi) = 2^8$.
Case $\Xi = V(u)$: Note that $\tilde{V}_{n,b} \succeq 0$ for all $b \in B_n$. Similar to (4.5.24) in the proof of Theorem 4.1.4 and using $\tilde{V}_{n,b} \vee \tilde{V}_{n,b'} \succeq \tilde{V}_{n,b}, \tilde{V}_{n,b'}$, we obtain for all $b \in B_n$:

$$|\hat{\theta}_{\hat{b}} - \theta_0|_{\tilde{V}_{n,b}}^2 \leq 6\Big(Y(b) + \widehat{\text{pen}}_n(b)\Big) + 3|\hat{\theta}_b - \theta_0|_{\tilde{V}_{n,b}}^2. \tag{4.5.27}$$

Step 1: Upper bound $Y(b)$ and $|\hat{\theta}_b - \theta_0|_{\tilde{V}_{n,b}}^2$ in (4.5.27): Let $A(b)$ be fulfilled. It holds that

$$Y(b) \leq \max_{b' \in B_n, b' \leq b} \big\{|\hat{\theta}_b - \hat{\theta}_{b'}|_{\tilde{V}_{n,b}}^2 - \frac{1}{2}\widehat{\text{pen}}_n(b')\big\}_+ + \max_{b' \in B_n, b' \leq b} \big\{|\hat{\theta}_b - \hat{\theta}_{b'}|_{\tilde{V}_{n,b'}}^2 - \frac{1}{2}\widehat{\text{pen}}_n(b')\big\}_+. \tag{4.5.28}$$

Since $A(b) \subset \Omega_{pen,\Xi,b'}$, it holds that $|\hat{\theta}_b - \hat{\theta}_{b'}|_{\tilde{V}_{n,b}}^2 \leq \frac{3}{2}|\hat{\theta}_b - \hat{\theta}_{b'}|_V^2$ (see also Step 3). We therefore have

$$\begin{aligned}Y(b) &\leq 2 \max_{b' \in B_n, b' \leq b} \big\{\frac{3}{2}|\hat{\theta}_b - \hat{\theta}_{b'}|_V^2 - \frac{1}{2}\widehat{\text{pen}}_n(b')\big\}_+ \\ &\leq 4 \max_{b' \in B_n, b' \leq b} \big\{\frac{3}{2}|\hat{\theta}_{b'} - \theta_0|_V^2 - \frac{C_P}{2}\hat{P}_n(b')\big\}_+.\end{aligned}$$

Together with $3|\hat{\theta}_b - \theta_0|_{\tilde{V}_{n,b}}^2 \leq 3 \cdot \big\{\frac{3}{2}|\hat{\theta}_b - \theta_0|_V^2 - \frac{1}{2}\hat{P}_n(b)\big\}_+ + \frac{3}{2}\hat{P}_n(b)$ and (4.5.27) we conclude

$$|\hat{\theta}_{\hat{b}} - \theta_0|_{\tilde{V}_{n,b}}^2 \leq 27 \max_{b' \in B_n, b' \leq b} \big\{\frac{3}{2}|\hat{\theta}_{b'} - \theta_0|_V^2 - \frac{C_P}{2}\hat{P}_n(b')\big\}_+ + \frac{27}{2} \sup_{b' \in B_n, b' \geq b} C_P \hat{P}_n(b').$$

Step 2: Estimation of $|\hat{\theta}_{b'} - \theta_0|_V^2$ for some $b' \in B_n, b' \leq b$: Let $A(b)$ be fulfilled. Since $A(b) \subset \Omega_{1,b'}$ we have $\hat{\theta}_{b'} - \theta_0 = -\nabla^2 L_{n,b'}(\bar{\theta})^{-1}\nabla L_{n,b'}(\theta_0)$, where $\bar{\theta} \in \Theta$ is an intermediate value with $|\bar{\theta} - \theta_0|_2 \leq |\hat{\theta}_{b'} - \theta_0|_2$. We conclude

$$\begin{aligned}|\hat{\theta}_{b'} - \theta_0|_V^2 &= |V^{1/2}\nabla^2 L_{n,b'}(\bar{\theta})^{-1}\nabla L_{n,b'}(\theta_0)|_2^2 \\ &\leq |V^{1/2}\nabla^2 L_{n,b'}(\bar{\theta})^{-1/2}|_{spec}^2 \cdot |\nabla^2 L_{n,b'}(\bar{\theta})^{-1/2}\nabla L_{n,b'}(\theta_0)|_2^2.\end{aligned}$$

Since $A(b) \subset \Omega_{1,b'} \cap \Omega_{2,b'}$ it holds that $|\nabla^2 L_{n,b'}(\bar{\theta}) - V|_2 \leq \frac{\lambda_{min}(V)}{2}$. Lemma 4.5.17(ii) yields $|\nabla^2 L_{n,b'}(\bar{\theta})^{1/2} - V^{1/2}|_2 \leq \frac{1}{\sqrt{2}}\frac{\lambda_{min}(V^{1/2})}{2}$ and thus by Lemma 4.5.17(i), we have $|\nabla^2 L_{n,b'}(\bar{\theta})^{-1/2}\nabla L_{n,b'}(\theta_0)|_2 \leq (1 + \frac{1}{\sqrt{2}})|V^{-1/2}\nabla L_{n,b'}(\theta_0)|_2$ and $|V^{1/2}\nabla^2 L_{n,b'}(\bar{\theta})^{-1/2}|_{spec} \leq 1 + \frac{1}{\sqrt{2}}$ and thus $|\hat{\theta}_{b'} - \theta_0|_V^2 \leq \big(1 + \frac{1}{\sqrt{2}}\big)^4 \cdot |V^{-1/2}\nabla L_{n,b'}(\theta_0)|_2^2$.
Step 3: Get a bound for $|\hat{\theta}_{\hat{b}} - \theta_0|_V^2$: For each $b \in B_n$, we have

$$\begin{aligned}|\hat{\theta}_{\hat{b}} - \theta_0|_V^2 &= |\hat{\theta}_{\hat{b}} - \theta_0|_V^2 \mathbb{1}_{A(b)} + |\hat{\theta}_{\hat{b}} - \theta_0|_V^2 \mathbb{1}_{A(b)^c} \\ &\leq |\hat{\theta}_{\hat{b}} - \theta_0|_{V-\tilde{V}_{n,b}}^2 \mathbb{1}_{A(b)} + |\hat{\theta}_{\hat{b}} - \theta_0|_{\tilde{V}_{n,b}}^2 \mathbb{1}_{A(b)} + |V|_{spec}\Theta_{max}^2 \mathbb{1}_{A(b)^c}.\end{aligned}$$

Since $A(b) \subset \Omega_{pen,\Xi,b}$, we have $V \succeq \lambda_{min}(V) I_{d\times d}$ and thus $|\hat{\theta}_{\hat{b}} - \theta_0|^2_{V-\tilde{V}_{n,b}} \le |\hat{\theta}_{\hat{b}} - \theta_0|_2^2 \cdot |V - \tilde{V}_{n,b}|_2 \le |\hat{\theta}_{\hat{b}} - \theta_0|_2^2 \frac{\lambda_{min}(V)}{2} \le \frac{1}{2}|\hat{\theta}_{\hat{b}} - \theta_0|_V^2$. With the results of Step 1, Step 2, and the fact that $A(b) \subset \{|\hat{P}_n(b') - P_n(b')| \le \frac{P_n(b')}{2}\}$ this finally leads to

$$\begin{aligned}
|\hat{\theta}_{\hat{b}} - \theta_0|_V^2 &\le 2|\hat{\theta}_{\hat{b}} - \theta_0|^2_{\tilde{V}_{n,b}} \mathbb{1}_{A(b)} + 2|V|_{spec}\Theta_{max}^2 \mathbb{1}_{A(b)^c} \\
&\le 54 \max_{b' \in B_n, b' \le b} \Big\{\frac{3}{2}\big(1 + \frac{1}{\sqrt{2}}\big)^2 |V^{-1/2}\nabla L_{n,b'}(\theta_0)|_2^2 - \frac{C_P}{4} P_n(b')\Big\}_+ \\
&\qquad +27 \sup_{b' \in B_n, b' \ge b} C_P \hat{P}_n(b') + 2|V|_{spec}\Theta_{max}^2 \mathbb{1}_{A(b)^c}
\end{aligned}$$

for all $b \in B_n$. Thus, (4.5.23) is fulfilled with $C_1(\Xi) = 54 \cdot \frac{3}{2}(1 + \frac{1}{\sqrt{2}})^2$ and $C_2(\Xi) = 27C_P(\Xi)$, where $C_P(\Xi) = 32 \cdot 4 \cdot \frac{3}{2}(1 + \frac{1}{\sqrt{2}})^2 \le 498$.
Bounds for (4.5.23): In the following we use a generic constant C in front of terms which will be shown to be of (negligible) order $\log(n)(n^{-1} + n^{-2\beta'})$. For the definitions of b_*, $b_*^{(2)}$ and $b_*^{(3)}$, see Lemma 4.5.5 and 4.5.14. Assume that

$$\begin{aligned}
b &\ge b_*(C_{\nabla^2\ell,z}, \frac{\lambda_{min}(V)}{16}) \vee b_*(C_{\ell,z}, \frac{\gamma_{mle}(\Xi)^2}{8C_L}) \vee b_*(C_{I,z}, \frac{\gamma_I(\Xi)}{8}) \vee b_*(C_{V,z}, \frac{\gamma_V(\Xi)}{8}) \\
&\qquad \vee \max_{j=1,\dots,d} b_*^{(2)}(e_j'|\Xi^{1/2}V^{-1}|C_{I,z}|V^{-1}\Xi^{1/2}|e_j, (\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}) \vee b_*^{(3)}(\Xi) \\
&=: c_b(\Xi) \cdot \frac{\log(n)^{1+2\alpha M}}{n}, \qquad (4.5.29)
\end{aligned}$$

which gives a definition of $c_b(\Xi)$. Define $\tilde{L}_{n,b}(u,\theta) := \frac{1}{K_{n,b}(u)} \sum_{t=1}^n K(\frac{t/n-u}{b}) \tilde{\ell}_t(u,\theta)$. It holds that $\nabla L_{n,b'}(\theta_0) = S_{n,b'}(\nabla \ell(\cdot,\theta_0)) + C_{n,b'}(\nabla\ell(\cdot,\theta_0)) + \tilde{B}_{n,b'}(\nabla\ell(\cdot,\theta_0)) + \nabla\tilde{L}_{n,b'}(\theta_0)$. Note that B_n has at most $\frac{\log(n)}{\log(a)}$ elements. By Lemma 4.5.1, 4.5.2 and 4.5.14, we have with $(w_1+w_2)^2 \le 2(w_1^2 + w_2^2)$ and $(w_1+w_2+w_3)^2 \le 3(w_1^2+w_2^2+w_3^2)$:

$$\begin{aligned}
&\mathbb{E} \max_{b' \in B_n, b' \le b} \big\{|\Xi^{1/2}V^{-1}\nabla L_{n,b'}(\theta_0)|_2^2 - 32P_n(b')\big\}_+ \\
\le& \sum_{b' \in B_n, b' \le b} \Big(2\mathbb{E}\big\{|\Xi^{1/2}V^{-1}\nabla\tilde{L}_{n,b'}(\theta_0)|_2^2 - 16P_n(b')\big\}_+ + 6\mathbb{E}|S_{n,b'}(\nabla\ell(\cdot,\theta_0))|_2^2 \\
&\qquad +6\mathbb{E}|C_{n,b'}(\nabla\ell(\cdot,\theta_0))|_2^2 + 6\mathbb{E}|\tilde{B}_{n,b'}(\nabla\ell(\cdot,\theta_0))|_2^2\Big) \\
\le& C\log(n)(n^{-1} + n^{-2\beta'}) + 6\sum_{b' \in B_n, b' \le b} \mathbb{E}|\tilde{B}_{n,b'}(\nabla\ell(\cdot,\theta_0))|_2^2.
\end{aligned}$$

Furthermore, the results of Lemma 4.5.13 (for $\Omega_{pen,\Xi,b}$), 4.5.12 (for $\Omega_{1,b}$) and 4.5.1, 4.5.2 and 4.5.5 (for $\Omega_{2,b}$) imply that

$$\mathbb{E}C_3(\Xi)\mathbb{1}_{A(b)^c} \le C\log(n)(n^{-1} + n^{-2\beta'}) + W_{n,1,\Xi}(b)$$

with

$$
\begin{aligned}
& W_{n,1,\Xi}(b) \\
:= \;& C_3(\Xi) \sup_{b' \in B_n, b' \leq b} \Big\{ \mathbb{1}\Big(\sup_{\theta \in \Theta} |B_{n,b'}(\nabla^2 \ell(\cdot,\theta), u)|_2 > \frac{\lambda_{min}(V)}{16} \Big) \\
& \quad + \mathbb{1}\Big(\sup_{\theta\in\Theta} |B_{n,b'}(\ell(\cdot,\theta),u)| > \frac{\gamma_{mle}(\Gamma)^2}{8C_L} \Big) + \mathbb{1}\Big(\sup_{\theta\in\Theta} |B_{n,b'}(g_V(\cdot,\theta),u)|_2 > \frac{\gamma_V(\Gamma)}{8} \Big) \\
& \quad + \mathbb{1}\Big(\sup_{\theta\in\Theta} |B_{n,b'}(g_I(\cdot,\theta),u)|_2 > \frac{\gamma_I(\Gamma)}{8} \Big) \Big\}.
\end{aligned}
$$

We now discuss the second term in (4.5.23). First, we have

$$
\sup_{b' \in B_n, b' \geq b} \hat{P}_{n,\Xi}(b') \leq \frac{3}{2} \sup_{b' \in B_n, b' \geq b} P_{n,\Xi}(b') + \sup_{b' \in B_n, b' \geq b} \hat{P}_{n,\Xi}(b') \cdot \mathbb{1}_{\Omega^c_{pen,\Xi,b'}}. \tag{4.5.30}
$$

It holds that $\hat{V}_{n,b} \vee V_0 \succeq V_0$ and $I_m \succeq \hat{I}_{n,b} \wedge I_m$, thus Lemma 4.5.16(i) implies that $V_0^{-1} - (\hat{V}_{n,b} \vee V_0)^{-1}$ is positive semidefinite, thus $\mathrm{tr}(\tilde{V}_{n,b}^{-1} \tilde{I}_{n,b} \tilde{V}_{n,b}^{-1}) \leq \mathrm{tr}(V_0^{-1} I_m V_0^{-1})$ or $\mathrm{tr}(\tilde{I}_{n,b}\tilde{V}_{n,b}^{-1}) \leq \mathrm{tr}(I_m V_0^{-1})$, respectively. Application of Lemma 4.5.12 and Lemma 4.5.13 to $\Omega^c_{pen,\Xi,b'} \subset (\Omega^c_{pen,\Xi,b'} \cap \Omega_{1,b'}) \cup \Omega^c_{1,b'}$ leads to

$$
\begin{aligned}
& \mathbb{E} \sup_{b' \in B_n, b' \geq b} \hat{P}_{n,\Xi}(b') \cdot \mathbb{1}_{\Omega^c_{pen,\Xi,b'}} \\
\leq \;& \frac{|K|_\infty^2}{c_0^2} \big\{ \mathrm{tr}(V_0^{-1} I_m V_0^{-1}) \vee \mathrm{tr}(I_m V_0^{-1}) \big\} \cdot \mathbb{E} \sup_{b' \in B_n, b' \geq b} \frac{|\log(b')|}{nb'} \cdot \mathbb{1}_{\Omega^c_{pen,\Xi,b'}} \\
\leq \;& C \frac{|\log(b)|}{nb} \sum_{b' \in B_n, b' \geq b} \big(n^{-1} + n^{-2\beta'} \big) + \frac{1}{C_2(\Xi)} W_{n,2,\Xi}(b),
\end{aligned}
$$

where

$$
\begin{aligned}
& W_{n,2,\Xi}(b) \\
:= \;& C_2(\Xi) \sup_{b' \in B_n, b' \geq b} \frac{|\log(b')|}{nb'} \cdot \Big[\mathbb{1}\Big(\sup_{\theta\in\Theta} |B_{n,b'}(g_V(\cdot,\theta),u)|_2 > \frac{\gamma_V(\Xi)}{8} \Big) \\
& \quad + \mathbb{1}\Big(\sup_{\theta\in\Theta} |B_{n,b'}(g_I(\cdot,\theta),u)|_2 > \frac{\gamma_I(\Xi)}{8} \Big) + \mathbb{1}\Big(\sup_{\theta\in\Theta} |B_{n,b'}(\ell(\cdot,\theta),u)| > \frac{\gamma_{mle}(\Xi)^2}{8C_L} \Big) \Big].
\end{aligned}
$$

Recall that B_n has at most $\frac{\log(n)}{\log(a)}$ elements. $b \in B_n$ implies $b \geq \underline{b}_n = c_b \frac{\log(n)^{1+2\alpha M}}{n}$. This finally gives

$$
\mathbb{E} \sup_{b' \in B_n, b' \geq b} \hat{P}_n(b') \cdot \mathbb{1}_{\Omega^c_{pen,\Xi,b'}} \leq C \log(n)(n^{-1} + n^{-2\beta'}) + \frac{1}{C_2(\Xi)} W_{n,2,\Xi}(b).
$$

Now define $W_{n,\Xi}(b) := W_{n,1,\Xi}(b) + W_{n,2,\Xi}(b)$, which gives the desired representation (4.1.15).

Discussion of $W_{n,\Xi}(b)$: It is known from Lemma 4.5.6 that it holds (component-wise) that $\sup_{\theta\in\Theta}|B_{n,b'}(g(\cdot,\theta),u)| \leq C_B\cdot(b')^{\beta} + C_{B,R}n^{-1}$ for $g\in\{\ell,\nabla^2\ell,g_I,g_V\}$. With

$$\sup_{b'\in B_n, b'\leq b}\mathbb{1}\Big(\sup_{\theta\in\Theta}|B_{n,b'}(g(\cdot,\theta),u)|_2>\gamma\Big)\leq \mathbb{1}\big(b^{\beta}>\frac{\gamma}{2C_B}\big)+\frac{2}{\gamma}C_{B,R}n^{-1},$$

we can bound $W_{n,1,\Xi}\leq C\big\{\mathbb{1}\big(b^{\beta}>c_1\big)+n^{-1}\big\}$ with some constant $c_1>0$. Using Markov's inequality, we obtain

$$\mathbb{1}\Big(\sup_{\theta\in\Theta}|B_{n,b'}(g(\cdot,\theta),u)|_2>\gamma\Big)\leq \big(\frac{\gamma}{2C_B}\big)^{-\frac{1}{\beta'}}\cdot b'+\frac{2}{\gamma}C_{B,R}n^{-1}.$$

and thus can show that $W_{n,2,\Xi}(b)\leq C(\sup_{b'\in B_n,b'\geq b}\frac{|\log(b')|}{n}+n^{-1})\leq C\log(n)n^{-1}$ (where C here is dependent on β if $\beta<1$). The inequality $\mathbb{E}|\tilde{B}_{n,b'}(\Xi^{1/2}V^{-1}\nabla\ell(\cdot,\theta_0(u)),u)|_2^2\lesssim(b')^{2\beta}+n^{-1}$ follows directly from Lemma 4.5.7. $\square$

Lemma 4.5.13 (The penalization term approximation)**.** *To keep the notation simple, we will omit the argument $u\in[0,1]$ in the following. Assume that Assumption 4.2.2, 4.2.5, 4.2.3 are fulfilled. Define*

$$\Omega_{pen,\Xi,b}(u):=\big\{|\hat{P}_{n,\Xi}(b)-P_{n,\Xi}(b)|\leq\frac{1}{2}P_{n,\Xi}(b)\big\}\cap\{|\tilde{V}_{n,b}-V|_2\leq\frac{\lambda_{min}(V)}{2}\},$$

where $P_{n,\Xi},\hat{P}_{n,\Xi}$ are from (4.1.10). Then there exists some set $\tilde{A}_{n,\Xi,b}$ and constants $\gamma_\theta(\Xi)$, $\gamma_I(\Xi)$, $\gamma_V(\Xi)$, $C_{pen}>0$ (which may differ for MSE / KL) such that for all

$$b\geq b_*(C_{I,z},\gamma_I(\Xi)/8)\vee b_*(C_{V,z},\gamma_V(\Xi)/8).$$

it holds that $\mathbb{P}(\tilde{A}_{n,\Xi,b})\leq C_{pen}(n^{-1}+n^{-2\beta'})$ and

$$\begin{aligned}&\Omega_{pen,\Xi,b}(u)^c\cap\{|\hat{\theta}_b-\theta_0|_2\leq\gamma_\theta(\Xi)\}\\ \subset\;&\tilde{A}_{n,\Xi,b}\cup\big\{\sup_{\theta\in\Theta}|B_{n,b}(g_V(\cdot,\theta),u)|_2>\frac{\gamma_V(\Xi)}{8}\big\}\cup\big\{\sup_{\theta\in\Theta}|B_{n,b}(g_I(\cdot,\theta),u)|_2>\frac{\gamma_I(\Xi)}{8}\big\}.\end{aligned}$$

Proof of Lemma 4.5.13: Define the set $R_{n,b}:=\{|\tilde{V}_{n,b}-V|_2\leq\frac{\lambda_{min}(V)}{2}\}$. It holds that

$$\Omega^c_{pen,\Xi,b}\subset\big\{\big|\mathrm{tr}\big(\tilde{\Xi}_{n,b}\tilde{V}_{n,b}^{-1}\tilde{I}_{n,b}\tilde{V}_{n,b}^{-1}\big)-\mathrm{tr}\big(\Xi V^{-1}IV^{-1}\big)\big|>\frac{1}{2}\mathrm{tr}\big(\Xi V^{-1}IV^{-1}\big)\big\}\cup R^c_{n,b}.$$

Consider the case $\Xi=\mathrm{Id}$: Assume that for all $b'\in B_n, b'\geq b$, we have $|\tilde{V}_{n,b}-V|_2\leq\frac{\lambda_{min}(V)}{2}$. With Lemma 4.5.17 and the rules $|\mathrm{tr}(AB)|\leq|A|_2|B|_2$ and $|AB|_2\leq|A|_{spec}|B|_2$, we conclude

$$\begin{aligned}&\big|\mathrm{tr}\big(\tilde{V}_{n,b}^{-1}\tilde{I}_{n,b}\tilde{V}_{n,b}^{-1}\big)-\mathrm{tr}\big(V^{-1}IV^{-1}\big)\big|\\ \leq\;&|\tilde{V}_{n,b}^{-1}-V^{-1}|_2|I|_2|\tilde{V}_{n,b}^{-1}|_{spec}+|\tilde{V}_{n,b}^{-1}|_2|\tilde{I}_{n,b}-I|_2|\tilde{V}_{n,b}^{-1}|_{spec}+|V^{-1}|_{spec}|I|_2|\tilde{V}_{n,b}^{-1}-V^{-1}|_2\\ \leq\;&\big(2|V^{-1}|_2|V^{-1}|_{spec}|\tilde{V}_{n,b}-V|_2\big)\cdot|I|_2\cdot(2|V^{-1}|_{spec})\\ &+(2|V^{-1}|_2)\cdot|\tilde{I}_{n,b}-I|_2\cdot(2|V^{-1}|_{spec})+|V^{-1}|_{spec}|I|_2\cdot(2|V^{-1}|_2|V^{-1}|_{spec}|\tilde{V}_{n,b}-V|_2)\\ \leq\;&4|V^{-1}|_2|V^{-1}|_{spec}\cdot|\tilde{I}_{n,b}-I|_2+6|I|_2|V^{-1}|_2|V^{-1}|^2_{spec}\cdot|\tilde{V}_{n,b}-V|_2.\end{aligned}$$

Now consider $\Xi = V(u)$: The proof for $\Omega^c_{pen,\Xi,b} \subset \big\{\big|\mathrm{tr}\big(\tilde{I}_{n,b}\tilde{V}^{-1}_{n,b}\big) - \mathrm{tr}\big(IV^{-1}\big)\big| > \frac{1}{2}\mathrm{tr}\big(IV^{-1}\big)\big\}$ is easier and therefore omitted. Here, we have

$$\begin{aligned} |\mathrm{tr}(\tilde{I}_{n,b}\tilde{V}^{-1}_{n,b}) - \mathrm{tr}(IV^{-1})| &\leq |\tilde{I}_{n,b} - I|_2|\tilde{V}^{-1}_{n,b}|_2 + |\tilde{V}^{-1}_{n,b} - V^{-1}|_2|I|_2 \\ &\leq 2|V^{-1}|_2|\tilde{I}_{n,b} - I|_2 + 2|V^{-1}|_2|V^{-1}|_{spec}|I|_2|\tilde{V}_{n,b} - V|_2. \end{aligned}$$

We conclude in both cases that

$$\begin{aligned} &\big|\mathrm{tr}\big(\tilde{\Xi}_{n,b}\tilde{V}^{-1}_{n,b}\tilde{I}_{n,b}\tilde{V}^{-1}_{n,b}\big) - \mathrm{tr}\big(\Xi V^{-1}IV^{-1}\big)\big| \\ \leq\ & 6\big(|V^{-1}|_2|\Xi V^{-1}|_{spec} \cdot |\tilde{I}_{n,b} - I|_2 + |I|_2|V^{-1}|_2|\Xi V^{-1}|_{spec}|\tilde{V}_{n,b} - V|_2\big). \end{aligned}$$

Define $\gamma_I(\Xi) := \frac{\mathrm{tr}(V^{-1}IV^{-1})}{24|V^{-1}|_2|\Xi V^{-1}|_{spec}}$, $\gamma_V(\Xi) := \frac{\lambda_{min}(V)}{2} \wedge \frac{\mathrm{tr}(V^{-1}IV^{-1})}{24|V^{-1}|_2|\Xi V^{-1}|^2_{spec}|I|_2}$. By Assumption 4.2.2, $g_V \in \mathcal{L}(M,\chi,C_{V,z},C_{V,\theta})$, $g_I \in \mathcal{L}(M,\chi,C_{I,z},C_{I,\theta})$. Define $A := \{|\hat{\theta}_b - \theta_0| \leq \gamma_\theta(\Xi)\}$, where $\gamma_\theta(\Xi) := \frac{(\gamma_I(\Xi)/|C_{I,\theta}|_2)\wedge(\gamma_V(\Xi)/|C_{V,\theta}|_2)}{2(1+|\chi|_1^M C_X^M N_\alpha(M)^M)}$. By Lemma 4.5.15 it holds on A that $|G_V(\hat{\theta}_b) - G_V(\theta_0)|_2 \leq \frac{\gamma_V}{2}$ and similarly for G_I. By Lemma 4.5.16(ii) and $(V \vee V_0)\wedge V_m = V$, $(I \vee I_0) \wedge I_m = I$ (see Assumption 4.2.2) we have $|\tilde{V}_{n,b} - V|_2 \leq |\hat{V}_{n,b} - V|_2$, $|\tilde{I}_{n,b} - I|_2 \leq |\hat{I}_{n,b} - I|_2$. We conclude

$$\begin{aligned} \Omega^c_{pen,\Xi,b} \cap A &\subset \big(\big\{|\tilde{I}_{n,b} - I|_2 > \gamma_I(\Xi)\big\} \cup \big\{|\tilde{V}_{n,b} - V|_2 > \gamma_V(\Xi)\big\} \cup R^c_{n,b}\big) \cap A \\ &\subset \big\{|\hat{I}_{n,b} - G_I(\hat{\theta}_b)|_2 > \frac{\gamma_I(\Xi)}{2}\big\} \cup \big\{|\hat{V}_{n,b} - G_V(\hat{\theta}_b)|_2 > \frac{\gamma_V(\Xi)}{2}\big\}. \end{aligned}$$

which gives the result in view of Lemmas 4.5.1, 4.5.2 and 4.5.5 and $b \geq b_*(C_{I,z},\gamma_I(\Xi)/8) \vee b_*(C_{V,z},\gamma_V(\Xi)/8)$. □

4.5.8 A crucial inequality

In the following we will use a Bernstein inequality for martingale difference sequences to obtain a inequality which guarantees that the penalization term has a sufficient rate.

Lemma 4.5.14 (A crucial inequality)**.** *Fix some $u \in [0,1]$. Define the stationary likelihood $\tilde{L}_{n,b}(u,\theta) := \frac{1}{K_{n,b}(u)}\sum_{t=1}^n K(\frac{t/n-u}{b})\tilde{\ell}_t(u,\theta)$. The argument u will be suppressed in the following. Suppose that Assumptions 4.2.3, 4.2.2 and 4.2.5 hold. Define $D_P := 16$. Assume that for all $b \in B_n$ it holds that*

$$\begin{aligned} b &\geq b^{(2)}_* := \max_{j=1,\dots,d} b^{(2)}_*(e_j'|\Xi^{1/2}V^{-1}|C_{I,z}|V^{-1}\Xi^{1/2}|e_j, (\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}), \\ b &\geq b^{(3)}_* := b^{(3)}_*(\Xi) = \frac{\log(n)^{1+2\alpha M}}{n} \\ &\quad\times\Big(\frac{4^{2\alpha M}\cdot 8|K|^2_\infty}{c_0^2} \cdot \max_{j=1,\dots,d}\frac{(e_j'|\Xi^{1/2}V^{-1}|C_{\nabla\ell,z})\cdot E_{X,2} + |e_j'\Xi^{1/2}V^{-1}\nabla\ell(0,u,\theta_0(u))|}{(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}}\Big), \end{aligned}$$

where $b^{(2)}_*(C_z,\gamma) := \frac{\log(n)^{1+2\alpha M}}{n}\big(\frac{|K|^2_\infty C_{E,2}(C_z)}{c_0^2\gamma}\cdot(8e)^{\frac{1}{2}+\alpha M}\big)^2$. Then there exists some constant $C_{be} > 0$ independent of b, n such that for all $n \geq 3$:

$$\sum_{b\in B_n}\mathbb{E}\Big(|\Xi^{1/2}V^{-1}\cdot\nabla\tilde{L}_{n,b}(\theta_0)|_2^2 - D_P\cdot P_{n,\Xi}(b)\Big)_+ \leq \frac{C_{be}\log(n)}{n}.$$

Proof of Lemma 4.5.14: Recall the definitions $\tau = (\frac{1}{2} + \alpha M)^{-1}$ and $\tau_2 = (\alpha M)^{-1}$. Define $\tilde{\ell}_t(u,\theta) := \ell(\tilde{Y}_t(u), u, \theta)$. Since $\tilde{\ell}_t(u, \theta_0(u))$ is stationary and a martingale difference w.r.t. $\mathcal{F}_t$ by Assumption 4.2.2, we have

$$\mathbb{E}|\Xi^{1/2} V^{-1} \nabla \tilde{L}_{n,b}(\theta_0)|_2^2 = \mathbb{E}\nabla \tilde{L}_{n,b}(\theta_0) V^{-1} \Xi V^{-1} \nabla \tilde{L}_{n,b}(\theta_0)' = F_{n,b}(u)^{-2} \cdot \mathrm{tr}(\Xi V^{-1} I V^{-1}).$$

which shows that $P_{n,\Xi}(b) = |\log(b)| \cdot \sum_{j=1}^d \mathbb{E}\big(\Xi^{1/2} V^{-1} \nabla \tilde{L}_{n,b}(\theta_0)\big)_j^2$. Define $P_{n,\Xi,j}(b) := |\log(b)| \cdot \mathbb{E}\big(\Xi^{1/2} V^{-1} \nabla \tilde{L}_{n,b}(\theta_0)\big)_j^2$.

Step 1: Conditional variance truncation. Define $Z_{t,j} := e_j' \Xi^{1/2} V^{-1} \nabla \tilde{\ell}_t(\theta_0)$ and

$$\hat{R}_{n,b,j}^2 := \sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)^2 \cdot \mathbb{E}\big[\big(\Xi^{1/2} V^{-1} \nabla \tilde{\ell}_t(\theta_0)\big)_j^2 \big| \mathcal{F}_{t-1}\big] = \sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2 \cdot \mathbb{E}[Z_{t,j}^2 | \mathcal{F}_{t-1}],$$

$R_{n,b,j}^2 := \mathbb{E}\hat{R}_{n,b,j}^2 = \sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2 (\Xi^{1/2} V^{-1} I V^{-1} \Xi^{1/2})_{jj}$, $\hat{R}_{n,b}^2 := \sum_{j=1}^d \hat{R}_{n,b,j}^2$ and $R_{n,b}^2 := \mathbb{E}\hat{R}_{n,b}^2$. We have

$$\begin{aligned}
& \mathbb{E}\big\{|\Xi^{1/2} V^{-1} \cdot \nabla \tilde{L}_{n,b}(\theta_0)|_2^2 - D_P P_{n,\Xi}(b)\big\}_+ \\
\leq\; & \sum_{j=1}^d \mathbb{E}\big\{\big(\Xi^{1/2} V^{-1} \cdot \nabla \tilde{L}_{n,b}(\theta_0)\big)_j^2 - D_P P_{n,\Xi,j}(b)\big\}_+ \\
\leq\; & \sum_{j=1}^d \mathbb{E}\big\{\big(\Xi^{1/2} V^{-1} \cdot \nabla \tilde{L}_{n,b}(\theta_0)\big)_j^2 - D_P P_{n,\Xi,j}(b)\big\}_+ \mathbb{1}_{\{\hat{R}_{n,b,j}^2 \leq 2R_{n,b,j}^2\}} \\
& \quad + \sum_{j=1}^d \mathbb{E}\big\{\big(\Xi^{1/2} V^{-1} \cdot \nabla \tilde{L}_{n,b}(\theta_0)\big)_j^2 - D_P P_{n,\Xi,j}(b)\big\}_+ \mathbb{1}_{\{\hat{R}_{n,b,j}^2 > 2R_{n,b,j}^2\}} \\
=:\; & \sum_{j=1}^d (W_{1,j} + W_{2,j}).
\end{aligned}$$

Let e_j denote the j-th unit vector of $\mathbb{R}^d$. By the Cauchy-Schwarz inequality and Lemma 4.5.3, we have $\|Z_{t,j}\|_4 \leq C_j$ with some $C_j > 0$ and thus

$$W_{2,j} \leq \big(\frac{|K|_\infty}{c_0} C_j\big)^2 \cdot \mathbb{P}(|\hat{R}_{n,b,j}^2 - R_{n,b,j}^2|^2 > R_{n,b,j}^2)^{1/2}.$$

Define the function $g_j(y) := e_j' \Xi^{1/2} V(u)^{-1} g_I(y, u, \theta_0(u)) V(u)^{-1} \Xi^{1/2} e_j$. We have that $\mathbb{E}[Z_{t,j}^2 | \mathcal{F}_{t-1}] = g_j(\tilde{Y}_{t-1}(u))$.
By Assumption 4.2.2, we have $g_j \in \mathcal{L}(M, \chi, e_j' |\Xi^{1/2} V^{-1}| C_{I,z} |V^{-1} \Xi^{1/2}| e_j)$. Note that

$$\begin{aligned}
c_0 nb \;&\leq\; \textstyle\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big) \leq \big(\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2\big)^{1/2} \cdot \big(\sum_{t=1}^n \mathbb{1}_{\{|t/n-u| \leq b/2\}}\big)^{1/2} \\
&\leq\; \big(\textstyle\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2\big)^{1/2} \cdot (nb)^{1/2},
\end{aligned}$$

which implies $c_0^2 \leq \sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2$. We can apply Lemma 4.5.4 (with K^2 instead of K), where the assumption $b \geq b_*^{(2)}$ leads to

$$\begin{aligned}
&\mathbb{P}(|\hat{R}_{n,b,j}^2 - R_{n,b,j}^2| > R_{n,b,j}^2)\\
= \ &\mathbb{P}\Big(\Big|\Big(\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2\Big)^{-1}\cdot\sum_{t=1}^n K\big(\frac{t/n-u}{b}\big)^2\cdot\big\{\mathbb{E}[g_j(\tilde{Y}_t(u))|\mathcal{F}_{t-1}] - \mathbb{E}g_j(\tilde{Y}_t(u))\big\}\Big|\\
&\qquad > (\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}\Big) \leq C_{E,1}(g_j)n^{-2}.
\end{aligned}$$

This shows

$$\sum_{b\in B_n}\sum_{j=1}^d n\cdot W_{2,j} \leq \big(\frac{|K|_\infty}{c_0}\big)^2\cdot\sum_{j=1}^d C_j^2 C_{E,1}(g_j)^{1/2}\cdot\log(a)^{-1}\cdot\log(n).$$

Step 2: High value truncation. Define $h_j(z) := e_j'\Xi^{1/2}V(u)^{-1}\nabla\ell(z,u,\theta_0(u))$. By Assumption 4.2.2, we have $h_j \in \mathcal{L}(M,\chi,e_j|\Xi^{1/2}V(u)^{-1}|C_{\nabla\ell,z})$. Set $M_n = M_{n,j} := D_{X,j}\cdot(4\log(n))^{1/\tau_2}$, where $D_{X,j} := (e_j'|\Xi^{1/2}V^{-1}|C_{\nabla\ell,z})\cdot E_{X,2} + |e_j'\Xi^{1/2}V^{-1}\nabla\ell(0,u,\theta_0(u))|$. Now, we use the fact that $\nabla\tilde{\ell}_t(\theta_0)$ and thus $Z_{t,j}$ is a martingale difference sequence to decompose

$$\big(V^{-1}\nabla\tilde{L}_{n,b}(\theta_0)\big)_j =: V_{n,b,j}^{(\leq M_n)} + V_{n,b,j}^{(>M_n)},$$

where $Z_{t,j}^{(\leq M_n)} := Z_{t,j}\mathbb{1}_{\{|Z_{t,j}|\leq M_n\}} - \mathbb{E}[Z_{t,j}\mathbb{1}_{\{|Z_{t,j}|\leq M_n\}}|\mathcal{F}_{t-1}]$, similarly $Z_{t,j}^{(>M_n)}$ and

$$V_{n,b,j}^{(\leq M_n)} = \frac{1}{K_{n,b}(u)}\sum_{t=1}^n K\left(\frac{t/n-u}{b}\right)\cdot Z_{t,j}^{(\leq M_n)},$$

and $V_{n,b,j}^{(>M_n)}$ similarly. According to this, write $W_{1,j} \leq 2(W_{1,j}^{(\leq M_n)} + W_{1,j}^{(>M_n)})$. Note that $\mathbb{E}\big(V_{n,b,j}^{(>M_n)}\big)^2 \leq \big(\frac{|K|_\infty}{c_0}\big)^2\cdot\|Z_{t,j}^{(>M_n)}\|_2^2$. By the projection property of the conditional expectation, the Cauchy-Schwarz inequality and Lemma 4.5.3 we have

$$\|Z_{t,j}^{(>M_n)}\|_2^2 \ \leq \ \|Z_{t,j}\mathbb{1}_{\{|Z_{t,j}|>M_n\}}\|_2^2 \leq \|Z_{t,j}\|_4^2\cdot\mathbb{P}(|Z_{t,j}|>M_n)^{1/2} \leq C_j^2\cdot C_E(h_j)^{1/2}\cdot n^{-1}.$$

This shows

$$\sum_{b\in B_n}\sum_{j=1}^d n\cdot W_{1,j}^{(>M_n)} \leq \sum_{b\in B_n}\sum_{j=1}^d n\cdot\mathbb{E}\big(V_{n,b,j}^{(>M_n)}\big)^2 \leq \big(\frac{|K|_\infty}{c_0}\big)^2\sum_{j=1}^d C_j^2 C_E(h_j)^{1/2}\cdot\log(a)^{-1}\cdot\log(n).$$

Step 3: Application of a Bernstein inequality for martingale differences. Note that

$$\begin{aligned}
n\cdot W_{1,j}^{(\leq M_n)} &= n\cdot\mathbb{E}\Big((V_{n,b}^{(\leq M_n)})_j^2 - D_P P_{n,\Xi,j}(b)\Big)_+\mathbb{1}_{\{\hat{R}_{n,b,j}^2\leq 2R_{n,b,j}^2\}}\\
&= \int_0^\infty \mathbb{P}\left(n\big((V_{n,b,j}^{(\leq M_n)})^2 - D_P P_{n,\Xi,j}(b)\big) \geq t, \hat{R}_{n,b,j}^2 \leq 2R_{n,b,j}^2\right)\,\mathrm{d}t\\
&\leq \int_0^\infty \mathbb{P}\Big(\big|V_{n,b,j}^{(\leq M_n)}\big| \geq \sqrt{D_P P_{n,\Xi,j}(b) + \frac{t}{n}}, \hat{R}_{n,b,j}^2 \leq 2R_{n,b,j}^2\Big)\,\mathrm{d}t.
\end{aligned}$$

We now use a Bernstein-type inequality for martingales from Van de Geer (2000), Lemma 8.9.: Note that $K\big(\frac{t/n-u}{b}\big)Z_{t,j}^{(\le M_n)}$ is a martingale difference sequence w.r.t. $\mathcal{F}_t$. By the projection property of the conditional expectation, we have $\mathbb{E}[|Z_{t,j}^{(\le M_n)}|^2|\mathcal{F}_{t-1}] \le \mathbb{E}[Z_{t,j}^2|\mathcal{F}_{t-1}]$. This shows that

$$\sum_{t=1}^n K\left(\frac{t/n-u}{b}\right)^2 \mathbb{E}\big[|Z_{t,j}^{(\le M_n)}|^2\big|\mathcal{F}_{t-1}\big] \le \hat{R}_{n,b,j}^2.$$

Define $K_{max} := 2|K|_\infty M_n$. We conclude that

$$\sum_{t=1}^n \mathbb{E}\Big[\Big|K\Big(\frac{t/n-u}{b}\Big)Z_{t,j}^{(\le M_n)}\Big|^m\Big|\mathcal{F}_{t-1}\Big] \le K_{max}^{m-2}\cdot \hat{R}_{n,b,j}^2.$$

A careful inspection of the proof of Van de Geer (2000), Lemma 8.9 shows that the following modification holds for arbitrary $a \ge 0$:

$$\begin{aligned}
&\mathbb{P}\Big(K_{n,b}(u)\cdot V_{n,b,j}^{(\le M_n)} \ge a, \hat{R}_{n,b,j}^2 \le 2R_{n,b,j}^2\Big)\\
\le\ &\mathbb{P}\Big(\sum_{t=1}^n K\Big(\frac{t/n-u}{b}\Big)Z_{t,j}^{(\le M_n)} \ge a, \hat{R}_{n,b,j}^2 \le 2R_{n,b,j}^2\Big) \le \exp\Big[-\frac{a^2}{2(aK_{max}+2R_{n,b,j}^2)}\Big].
\end{aligned}$$

Because $-V_{n,b,j}^{(\le M_n)}$ is a martingale, too, we can extend the bound to $|V_{n,b,j}^{(\le M_n)}|$ by introducing a factor 2 on the right hand side. With $a(t) := \big(D_P|\log(b)|\frac{R_{n,b,j}^2}{K_{n,b}(u)} + \frac{K_{n,b}(u)}{n}t\big)^{1/2}$, $\gamma := 4K_{max}/\sqrt{K_{n,b}(u)}, \beta := 8\frac{R_{n,b,j}^2(u)}{K_{n,b}(u)}$ we conclude

$$\begin{aligned}
&\mathbb{P}\big(K_{n,b}(u)^{1/2}\big|(V_{n,b}^{(\le M_n)})_j\big| \ge a, R_{n,b,j}^2 \le 2R_{n,b,j}^2\big)\\
\le\ & 2\exp\Big[-\frac{a^2}{2\big(\frac{aK_{max}}{\sqrt{K_{n,b}(u)}} + 2\frac{R_{n,b,j}^2}{K_{n,b}(u)}\big)}\Big] = \exp\Big[-\frac{a(t)^2}{2(a(t)\cdot\gamma/4+\beta/4)}\Big]\\
\le\ & \exp\Big[-\Big(\frac{a(t)}{\gamma}\wedge\frac{a^2(t)}{\beta}\Big)\Big] \le \begin{cases}\exp(-a^2(t)/\beta), & a(t)\le \beta/\gamma\\ \exp(-a(t)/\gamma), & a(t)\ge\beta/\gamma.\end{cases}
\end{aligned}$$

Note that $a^{-1}(s) = \frac{n}{K_{n,b}(u)}\big(s^2 - D_P|\log(b)|\frac{R_{n,b,j}^2}{K_{n,b}(u)}\big)$. We conclude

$$n\cdot W_{1j}^{(\le M_n)} \le 2\int_{\{a(t)\le\beta/\gamma\}}\exp(-a(t)^2/\beta)\ \mathrm{d}t + 2\int_{\{a(t)\ge\beta/\gamma\}}\exp(-a(t)/\gamma)\ \mathrm{d}t. \quad (4.5.31)$$

The first term on the right hand side of (4.5.31) is bounded by

$$\begin{aligned}
&2\exp\Big(-\frac{D_P|\log(b)|}{8}\Big)\int_0^\infty \exp\Big(-\frac{1}{8}\frac{F_{n,b}(u)^2}{n}\frac{1}{(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}}t\Big)\ \mathrm{d}t\\
=\ & \exp\Big(-\frac{D_P|\log(b)|}{8}\Big)\cdot 16\frac{n}{F_{n,b}(u)^2}(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}\\
\le\ & \exp\Big(-\frac{D_P|\log(b)|}{8}\Big)\cdot 16(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}\Big(\frac{|K|_\infty}{c_0}\Big)^2\cdot\frac{1}{b}\\
\le\ & b\cdot 16(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}\Big(\frac{|K|_\infty}{c_0}\Big)^2.
\end{aligned}$$

In the second term of the right hand side of (4.5.31) we use the substitution $u := \frac{a(t)}{\gamma}$ to get

$$\begin{aligned} 4\gamma^2 \frac{n}{K_{n,b}(u)} \int_{\{u \ge \beta/\gamma^2\}} u \cdot e^{-u} \, \mathrm{d}u &= 4\gamma^2 \frac{n}{K_{n,b}(u)} \cdot \Big[- (u+1)e^{-u} \Big]_{\beta/\gamma^2}^{\infty} \\ &= 4\gamma^2 \frac{n}{K_{n,b}(u)} \Big(\frac{\beta}{\gamma^2} + 1 \Big) \cdot e^{-\beta/\gamma^2}. \end{aligned}$$

Because $b \ge b_*^{(3)}$, we have $\beta/\gamma^2 = \frac{R_{n,b,j}^2}{2K_{max}^2} = \frac{(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}}{8|K|_\infty^2} \frac{\sum_{t=1}^n K((t/n-u)/b)^2}{M_n^2} \ge \frac{(\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj} c_0^2}{8|K|_\infty^2} \frac{nb}{M_n^2} \ge \log(n)$. Note that $\frac{4\gamma^2}{K_{n,b}(u)} \le \frac{4^4|K|_\infty^2}{c_0^2} \cdot \frac{M_n^2}{(nb)^2}$. Since $x \mapsto (x+1)e^{-x}$ is non-decreasing, we conclude for $n \ge 3$:

$$\begin{aligned} & 2\int_{\{a(t) \ge \beta/\alpha\}} \exp(-a(t)/\alpha) \, \mathrm{d}t \le \frac{1}{nb} \cdot \frac{4^4|K|_\infty^2}{c_0^2} \frac{M_n^2}{nb} (\log(n) + 1) \\ \le\ & \frac{1}{nb} \cdot 32 (\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj} \cdot \frac{\log(n)+1}{\log(n)} \le \frac{64}{nb} (\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj}. \end{aligned}$$

Finally,

$$\begin{aligned} \sum_{b \in B_n} \sum_{j=1}^d n \cdot W_{1j}^{(\le M_n)} &\le 4^2 \sum_{j=1}^d (\Xi^{1/2}V^{-1}IV^{-1}\Xi^{1/2})_{jj} \Big[\big(\frac{|K|_\infty}{c_0}\big)^2 \cdot \sum_{b \in B_n} b + \frac{4}{n} \sum_{b \in B_n} b^{-1} \Big] \\ &\le 4^2 \mathrm{tr}(\Xi V^{-1} I V^{-1}) \cdot \frac{a}{a-1} \Big[\big(\frac{|K|_\infty}{c_0}\big)^2 + 4 \Big]. \end{aligned}$$

□

4.5.9 Elementary results

The proof of the following lemma is elementary and therefore omitted.

Lemma 4.5.15 (Standard approximation)**.** *Assume that $g : \mathbb{R}^{\mathbb{N}} \times \Theta \to \mathbb{R}^p$ is such that each component $g_i \in \tilde{\mathcal{L}}(M, \chi, C_i)$ where $C_i = (C_{i,z}, C_{i,\theta})$. Define $G(u,\theta) := \mathbb{E}[g(\tilde{Y}_t(u), \theta)]$. Then*

$$|G(u,\theta) - G(u,\theta')|_2 \le |C_{\cdot,\theta}|_2 \cdot |\theta - \theta'|_2 \cdot (1 + |\chi|_1^M C_X^M N_\alpha(M)^M).$$

Lemma 4.5.16. *Recall that we write $A \succeq B$ or $A \succ B$ if $A - B$ is positive semidefinite or positive definite, respectively. Let V, V', V_0 be symmetric $d \times d$-matrices.*

(i) Assume that $V \succeq V_0 \succ 0$. Then $V \succ 0$ and $V_0^{-1} \succeq V^{-1}$.

Let us define $V \vee V_0$ as follows: Since $V - V_0$ is symmetric, there exists a spectral decomposition $V - V_0 = S\Lambda S^{-1}$ where Λ is a diagonal matrix containing the eigenvalues. Define $\tilde{\Lambda} := \max\{\Lambda, 0\}$ where the maximum is taken component-wise and $V \vee V_0 := S\tilde{\Lambda}S^{-1} + V_0$. Furthermore define $V \wedge V_0 := -((-V) \vee (-V_0))$.

(ii) It holds that $V \vee V_0 \succeq V_0$ and furthermore, $|V \vee V_0 - V' \vee V_0|_2 \leq |V - V'|_2$.

Proof of Lemma 4.5.16: (i) Since $V - V_0$ is positive semidefinite, we have

$$\lambda_{min}(V) = \inf_{|x|_2=1} x'Vx \geq \inf_{|x|_2=1} x'(V - V_0)x + \inf_{|x|_2=1} x'V_0x \geq \lambda_{min}(V_0)$$

which shows that V is positive definite. The positive semidefiniteness of $V - V_0$ implies that $I - V^{-1/2}V_0V^{-1/2}$ is positive semidefinite. Since $V^{-1/2}V_0V^{-1/2}$ is matrix similar to $V_0^{1/2}V^{-1}V_0^{1/2}$, also $I - V_0^{1/2}V^{-1}V_0^{1/2}$ is positive semidefinite and thus $V_0^{-1} - V^{-1}$ is positive semidefinite.
(ii) We have $V \vee V_0 - V_0 = S\tilde{\Lambda}S^{-1}$ which is obviously positive semidefinite, thus $V \vee V_0 \succeq V_0$. Furthermore, with the spectral decompositions $V - V_0 = S\Lambda S^{-1}$, $V' - V_0 = S'\Lambda'(S')^{-1}$ we have

$$|V \vee V_0 - V' \vee V_0|_2 \leq |S\tilde{\Lambda}S^{-1} - S'\tilde{\Lambda}'(S')^{-1}|_2.$$

By Theorem 1.1. in Wihler (2009) applied to $f(x) = \max\{x, 0\}$ this is bounded by $|(V - V_0) - (V' - V_0)|_2 = |V - V'|_2$. $\square$

Lemma 4.5.17. *Let A be a positive definite symmetric $d \times d$-matrix, B a symmetric $d \times d$-matrix with $|A - B|_2 \leq x \cdot \frac{\lambda_{min}(A)}{2}$ for some $0 < x \leq 1$. Then, we have $\lambda_{min}(B) \geq \frac{\lambda_{min}(A)}{2}$ and*

(i) For all $v \in \mathbb{R}^d$: $|A^{-1} - B^{-1}|_2 \leq \frac{2x|A^{-1}|_2}{\lambda_{min}(A)} \cdot |A - B|_2$ and $|B^{-1}v|_2 \leq (1 + x)|A^{-1}v|_2$, and $|AB^{-1}|_{spec} \leq 1 + x$.

(ii) $|A^{1/2} - B^{1/2}|_2 \leq \frac{x}{2}\sqrt{\frac{\lambda_{min}(A)}{2}}$.

Proof of Lemma 4.5.17: (i) We have $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$, thus with the rules $|\mathrm{tr}(CD)| \leq |C|_2|D|_2$ and $|CD|_2 \leq |C|_{spec}|D|_2$:

$$|A^{-1} - B^{-1}|_2 \leq |A^{-1}|_2|B^{-1}|_{spec}|A - B|_2. \tag{4.5.32}$$

and

$$|B^{-1}v|_2 \leq |A^{-1}v|_2 + |B^{-1}(A - B)A^{-1}v|_2 \leq \big(1 + |B^{-1}|_{spec}|B - A|_2\big) \cdot |A^{-1}v|_2. \tag{4.5.33}$$

Basic properties of the Rayleigh quotient and the fact $|\lambda_{max}(C)| = |C|_{spec} \leq |C|_2$ for symmetric C give

$$\begin{aligned} \lambda_{min}(B) &= \inf_{|v|_2=1} v'Bv \geq \inf_{|v|_2=1} v'(B - A)v + \inf_{|v|_2=1} x'Ax \geq -|\lambda_{max}(B - A)| + \lambda_{min}(A) \\ &\geq -|B - A|_2 + \lambda_{min}(A) \geq \frac{\lambda_{min}(A)}{2}. \end{aligned}$$

This shows $|B^{-1}|_{spec} = \lambda_{max}(B^{-1}) = \frac{1}{\lambda_{min}(B)} \leq \frac{2}{\lambda_{min}(A)}$. Plugging this into (4.5.32) and (4.5.33) prove the first two inequalities. For the third inequality note that $B^{-1} = A^{-1} + A^{-1}(A - B)B^{-1}$ and thus $|AB^{-1}|_{spec} \leq 1 + |A - B|_2|B^{-1}|_{spec} \leq 1 + x$.
(ii) Applying Theorem 1.1 of Wihler (2009) with $f : [\frac{\lambda_{min}(A)}{2}, \infty) \to \mathbb{R}, f(x) = \sqrt{x}$ yields $|A^{1/2} - B^{1/2}|_2 \leq \frac{1}{2}\sqrt{\frac{2}{\lambda_{min}(A)}}|A - B|_2 \leq \frac{x}{2}\sqrt{\frac{\lambda_{min}(A)}{2}}$. $\square$

Chapter 5

Conclusion

In this thesis, we have dealt with data adaptive bandwidth selectors for maximum likelihood estimators in locally stationary processes. Before this work, the theoretical behavior of such quantities was nearly untouched in the literature. Furthermore, no general proposals for selection routines in this context were available. In Chapters 2 and 4 we invented two selectors for large classes of locally stationary processes and proved their consistency: A global bandwidth selector inspired by cross validation and a local bandwidth selector motivated by a contrast minimization approach. Our thesis can also be seen as a contribution to bandwidth selection theory in nonparametric statistics since the popular i.i.d. regression model is a special case of the processes where we can apply our selectors. Due to the general formulation, our results also give a hint how to define bandwidth selectors in multivariate locally stationary time series models or more sophisticated situations. The simulation results of both methods show that they behave stable if the model is correctly specified which suggests their use in practice. An application of the cross validation bandwidth selector may be a good starting point in applications since this estimator does not need further choices of tuning constants.

An essential difficulty in the proofs was the discussion of the bias terms for recursively defined locally stationary processes. To solve this problem, we developed a general approximation theory for such processes in Chapter 3. Based on ideas in Dahlhaus and Subba Rao (2006) and Dahlhaus (2011), we introduced so called stationary approximations and derivative processes. Besides the more general formulation, our main contribution here was the invention of a theory of existence and uniqueness even when no explicit representation of the process is available (as it was the case in earlier publications). Derivative processes allowed us to expand locally stationary processes into stationary processes, making them a powerful tool for proofs. Using these expansions, we proved some laws of large numbers and central limit theorems with bias expansions under minimal moment conditions. We used these results to obtain an easily applicable asymptotic theory for maximum likelihood estimators.

Our results offer several possibilities for further research. Regarding bandwidth selection it may be useful to generalize the theoretical results to multivariate time series. The bandwidth selectors defined in this thesis may depend on unknown properties like the variance or the fourth moment of the underlying i.i.d. error sequence - it seems to be a very challenging problem to find ad-hoc estimators of these quantities which achieve parametric rates. Regarding Chapter 3, let us note that we set up a bunch of conditions for the existence of derivative processes and definitions of interesting functionals which may be generalized. One of the most interesting problems is the question if differentiability of the recursion function is necessary to guarantee the existence of a derivative process. In case of a positive answer it would be possible to apply the theory of derivative processes to an even larger class of processes, for instance the tvTAR process. The idea of derivative processes and stationary approximations is not restricted to discrete recursively defined time series models. Stochastic differential equations may be a field where a similar theory could be invented.

Bibliography

[Arkoun (2010)] O. Arkoun: Sequential adaptive estimators in nonparametric autoregressive models, arXiv:1004.5199

[Arkoun and Pergamenchtchikov (2014)] O. Arkoun, S. Pergamenchtchikov: Nonparametric estimation for an autoregressive model, arXiv:0806.3012

[Brockwell and Davis (1987)] P. J. Brockwell, R. A. Davis (1987): Time Series: Theory and Methods, Springer New York Inc.

[Chiu (1991)] S-T. Chiu (1991): Bandwidth Selection for kernel density estimation, The Annals of Statistics, Vol. 19, No. 4, 1883-1905

[Dahlhaus and Giraitis (1998)] R. Dahlhaus and L. Giraitis (1998): On the optimal segment length for parameter estimates for locally stationary time series, Journal of Time Series Analysis Vol. 19, No. 6, 629-655

[Dahlhaus (1996)] R. Dahlhaus (1996): On the Kullback-Leibler information divergence of locally stationary processes, Stochastic Processes and their Applications 62, 139-168

[Dahlhaus (1997)] R. Dahlhaus (1997): Fitting time series models to nonstationary processes, The Annals of Statistics, Vol. 25, No. 1, 1-37

[Dahlhaus (2000)] R. Dahlhaus (2000): A likelihood approximation for locally stationary processes, Annals of Statistics, Vol. 28, No. 6, 1762-1794

[Dahlhaus and Polonik (2006)] R. Dahlhaus, W. Polonik (2006): Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes, Annals of Statistics, Vol. 34, No. 6, 2790-2824

[Dahlhaus and Polonik (2009)] R. Dahlhaus, W. Polonik (2009): Empirical spectral processes for locally stationary time series, Bernoulli 15(1), 1-39

[Dahlhaus (2009)] R. Dahlhaus (2009): Local inference for locally stationary time series based on the empirical spectral measure, Journal of Econometrics 151, 101-112

[Dahlhaus (2011)] R. Dahlhaus (2011): Locally Stationary Processes, Handbook of Statistics, http://arxiv.org/abs/1109.4174

[Dahlhaus and Subba Rao (2006)] R. Dahlhaus and S. Subba Rao (2006): Statistical inference for time-varying ARCH processes, The Annals of Statistics Vol. 34, No. 3, 1075-1114

[Duflo (1997)] M. Duflo (1997): Random Iterative Models, Springer Berlin Heidelberg New York.

[Fryzclewicz, Sapatinas and Subba Rao (2008)] P. Fryzclewicz, T. Sapatinas, S. Subba Rao (2008): Normalized least-squares estimation in time-varying ARCH models, Annals of Statistics, Vol. 36, No. 2, 742-786.

[Giraud, Roueff and Sanchez-Perez (2015)] C. Giraud, F. Roueff, A. Sanchez-Perez (2015): Aggregation of predictors for nonstationary sub-linear processes and online adaptive forecasting of time varying autoregressive processes, The Annals of Statistics, 43, No. 6, 2412-2450

[Goldenshluger and Lepski (2011)] A. Goldenshluger and O. Lepski: Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality, The Annals of Statistics, 39:1608-1632

[Hall, Marron and Park (1992)] P. Hall, J. S. Marron, B. U. Park (1992): Smoothed cross-validation, Probability Theory and Related Fields 92, 1-20

[Härdle and Marron (1985)] W. Härdle, J. S. Marron (1985): Optimal bandwidth selection in nonparametric regression function estimation, Annals of Statistics, Vol. 13, No. 4, 1465-1481

[Härdle, Hall and Marron (1988)] W. Härdle, P. Hall, J. S. Marron (1988): How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?, Journal of the American Statistical Association, Vol. 83, No. 401, pp. 86 - 95

[Householder (1964)] A. S. Householder (1964): The Theory of Matrices in Numerical Analysis, New York: Blaisdell.

[Jones (1978)] D. A. Jones (1978): Nonlinear Autoregressive Processes, Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, Vol. 360, No. 1700, 71-95

[Karmakar and Wu (2016)] S. Karmakar, W. B. Wu (2016): An Invariance Principle for Multivariate Processes, arXiv.

[Katznelson (2004)] Y. Katznelson (2004): An Introduction to harmonic analysis, Cambridge Mathematical Library, Third Edition

[Kreiss and Paparoditis (2015)] J.-P. Kreiss, E. Paparoditis (2015): Bootstrapping locally stationary processes, J. R. Statist. Soc. B 77, Part 1, pp. 267-290.

[Lepski, Mammen and Spokoiny (1997)] O. V. Lepski, E. Mammen, V. G. Spokoiny (1997): Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors, The Annals of Statistics, Vol. 25, No. 3, 929-947

[Lepski and Spokoiny (1997)] O. V. Lepski, V. G. Spokoiny (1997): Optimal pointwise adaptive methods in nonparametric estimation, Vol. 25, No. 6, 2512-2546

[Liu, Xiao and Wu (2013)] W. Liu, H. Xiao, W. B. Wu (2013): Probability and moment inequalities under dependence, Statistica Sinica 23, 1257-1272

[Marron (1986)] J. S. Marron (1986): Random Approximations to Some Measures of Accuracy in Nonparametric Curve Estimation, Journal of multivariate Analysis 20, 91-113

[Martin und Flandrin (1985)] W. Martin, P. Flandrin (1985): Wigner-Ville spectral analysis of nonstationary processes, IEEE Trans. Acoust. Speech Signal Process. 33, 1461-1470

[Priestley (1965)] M. B. Priestley (1965): Evolutionary spectra and non-stationary processes, Journal of the Royal Statistical Society Series B 27

[Priestley (1981)] M. B. Priestly (1981): Spectral Analysis and Time Series, Academic Press, London, Vol. 2

[Rice (1984)] J. Rice (1984): Bandwidth choice for nonparametric regression, The Annals of Statistics, Vol.12, No. 4, 1215-1230

[Rio (2009)] E. Rio (2009): Moment Inequalities for Sums of Dependent Random Variables under Projective Conditions, Journal of Theoretical Probability 22: 146-163

[Shao and Wu (2007)] X. Shao, W. B. Wu (2007): Asymptotic spectral theory for nonlinear time series, The Annals of Statistics, Vol. 35, No. 4, 1773-1801

[Subba Rao (2006)] S. Subba Rao (2006): On some nonstationary, nonlinear random processes and their stationary approximations, Advances in Applied Probability, Vol. 38, 1155-1172

[Subba Rao (2010)] S. Subba Rao (2010): A note on quadratic forms of nonstationary stochastic processes, technical report

[Tsybakov (2009)] A. Tsybakov (2009): Introduction to Nonparametric Estimation, Springer Science+Business Media, LLC

[Van de Geer (2000)] S. van de Geer (2000): Empirical processes in M-estimation, Cambridge University Press

[Van der Vaart (2009)] A. W. van der Vaart (1998): Asymptotic Statistics, Cambridge University Press

[Vogt (2012)] M. Vogt (2012): Nonparametric regression for locally stationary time series, The Annals of Statistics, Vol. 40, No. 5, 2601-2633

[Whittle (1960)] P. Whittle (1960): Bounds for the moments of linear and quadratic forms in independent variables, Theory of Probability and its applications, Volume V, Number 3

[Wihler (2009)] T. Wihler: On the Hoelder continuity of matrix functions for normal matrices, Journal of inequalities in pure and applied mathematics, Vol. 10 (2009), Issue 4

[Woodroofe (1970)] M. Woodroofe (1970): On choosing a delta-sequence, The Annals of Mathematical Statistics, Vol. 41, No. 5, 1665-1671

[Wu and Shao (2004)] W. B. Wu, X. Shao (2004): Limit theorems for iterated random functions, Journal of Applied Probability 41, 425-436

[Wu (2005)] W. B. Wu (2005): Nonlinear system theory: Another look at dependence, PNAS, vol. 102, no.40, 14150-14154

[Wu (2011)] W. B. Wu (2011): Asymptotic theory for stationary processes, Statistics and its Interface, Volume 0 (2011) 1-20

[Zhou and Wu (2009)] Z. Zhou and W. B. Wu (2009): Local linear quantile estimation for nonstationary time series, The Annals of Statistics Vol. 37, No. 5B, 2696-2729