THE UNIVERSITY OF CHICAGO

# ASYMPTOTIC THEORY FOR SIMULTANEOUS INFERENCE UNDER DEPENDENCE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
SAYAR KARMAKAR

CHICAGO, ILLINOIS
AUGUST 2018

To my parents Pannalal Karmakar, Debi Karmakar and my twin brother Sayak Karmakar for their unconditional love and support.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

My heartfelt thanks are due to my advisor, Professor Wei Biao Wu for introducing me to several interesting problems in time-series and for his continuous guidance, without which this thesis will not have seen the light of the day. He has been extremely encouraging and generous with his time and ideas −always ready to discuss and give feedbacks, often multiple times in a week. I am especially grateful for how accessible and helpful he has been throughout and feel blessed about the fact that he wants to help me even after I graduate. I would also like to thank Tracy and Jian for their time to be in my committee and revise this dissertation. My heartfelt gratitude is due to Tracy, Jonathan and my collaborator Maggie from NJIT for kindly agreeing to write my recommendation letters during my application. Also, I must thank Dr. Stefan Richter for his collaboration on what I consider a significant work in this thesis as I have learnt a lot in only 4 weeks of interaction.

I feel sad to be parted from the awesome journey that I had with Soudeep and Kushal for the last 10 years. I really enjoyed my time here with Somak, Rishideep, Subhajit, Vishwas, Ritabrata, Swarnali, Enakshi, Koushiki etc. I owe a lot to the vibrant and encouraging chats with some of my close friends from ISI such as Arka, Deepan, Rounak, Arkajyoti, Subhabrata, Moumanti, Avijit, Tamal, Rajarshi etc., although we probably lived in as many as 8 different time-zones across the globe. I would also like to thank my best friend Rudradev for staying beside me and being incredibly patient in hearing every bad news of my life in boring detail.

My family has been really supportive throughout my life and it is needless to say that my journey would surely have been impossible and incomplete without the sacrifices that my parents and my twin brother made for me. Lastly, I would like to thank Cynthia for leaving an indelible mark on my life. Her stupendous support and seamless encouragement for the last 6 years has not only helped this dissertation and motivated me to stay in academia, it has also taught me the miraculous strength of the wonderful institution called love.

# ABSTRACT

In this thesis we present some interesting new results for simultaneous inference under dependence. One of the key idea behind establishing such asymptotic theory is an invariance principle where the partial sums of a mean-zero process are approximated by a corresponding Gaussian analogue. The first work discusses such a Gaussian approximation result and extends achieving the popular KMT-type optimal bound to non-linear, non-stationary and weakly dependent vector-valued processes. In the second work, we use the invariance principle to construct simultaneous confidence intervals for time-varying coefficient models. Using Bahadur representation, it was possible to construct such confidence bands for models as complicated as ARMA-GARCH or generalized regression under a single framework. As a summary, this thesis is a systematic presentation of several interesting problems in simultaneous inference for vector-valued processes that can be explored with the help of a powerful invariance principle.

# CHAPTER 1

# INTRODUCTION

We observe departure from independence in almost all real datasets we intend to analyze. But many of the results in theoretical statistics are still suited for only independent error distribution. One of the way one can systematically extend the results from the independent case to a very general dependent error process is as follows:

- Linear process:
$$\epsilon_i = \sum_{k=0}^{\infty} a_k e_{i-k}.$$

- Non-linear process:
$$\epsilon_i = F(e_i, e_{i-1}, \ldots).$$

- Non-stationary process:
$$\epsilon_i = F_i(e_i, e_{i-1}, \ldots).$$

- Multivariate:
$$\epsilon_i = \mathbf{F}_i(e_i, e_{i-1}, \ldots).$$

Along with this direction of generalization, one also needs some measure to assess dependence. The functional dependence framework by Wu (2005, [76]) and its natural extensions to accommodate non-stationarity and other possible generalizations such as predictive dependence etc. has been extensively used to evaluate and measure the dependence. This framework has a very simple input-output interpretation and often is easier to verify and evaluate compared to the strong mixing conditions, the more popular way to formulate the dependent structure for weak dependent process. In this thesis, we will be using different functional dependence measure at different point and will define it in the corresponding context.

In Chapter 1, we start with the KMT-type Gaussian approximation result for non-stationary and multiple time-series. Komlós, Major and Tusnády (1975, 76, [38, 39]) established the very deep result for univariate independent process $e_i$ with partial sums $S_i'$,

$$\max_{1 \leq i \leq n} |S_i' - \sigma B(i)| = o_{\text{a.s.}}(\tau_n), \tag{1.1}$$

where $B(\cdot)$ is the standard Brownian motion and $S_n'$ is constructed on a richer space such that $(S_i)_{i \leq n} \overset{D}{=} (S_i')_{i \leq n}$, and the approximation rate $\tau_n = n^{1/p}$ is optimal. This has been extended to accommodate dependence in Berkes, Liu and Wu (2014, [5]) however, the extension to multiple time-series and non-stationary both come with its own challenge. Towards this direction, the best possible rate was roughly $n^{1/4}$. In chapter 1, we discuss how to extend this to $n^{1/p}$ rate for a $p > 4$. Based on the decay rate of functional dependence measure, we quantify the error bound of the Gaussian approximation based on the sample size $n$ and the moment condition. Under the assumption of $p$th finite moment, with $p > 2$, this can range from the worst $n^{1/2}$ to the optimal $n^{1/p}$ rate.

Chapter 2 focuses on simultaneous inference of time-varying models. A general class of time-varying regression models which cover general linear models as well as time series models is considered. We estimate the regression coefficients by using local linear M-estimation. For these estimators, Bahadur representations are obtained and are used to construct simultaneous confidence bands. For practical implementation, we propose a bootstrap based method to circumvent the slow logarithmic convergence of the theoretical simultaneous bands. Our results substantially generalize and unify the treatments for several time-varying regression and auto-regression models. The performance for ARCH and GARCH models is studied in simulations and a few real-life applications of our study are presented through analysis of some popular financial datasets.

# CHAPTER 2

# GAUSSIAN APPROXIMATION

## 2.1 Introduction

Functional central limit theorem (FCLT) or invariance principle plays an important role in statistics. Let $X_i, i \geq 1$, be independent and identically distributed (i.i.d.) random vectors in $\mathbb{R}^d$ with mean 0 and covariance matrix $\Sigma$, and $S_j = \sum_{i=1}^{j} X_i$. The FCLT asserts that

$$\{n^{-1/2} S_{\lfloor nu \rfloor},\, 0 \leq u \leq 1\} \Rightarrow \{\Sigma^{1/2} I\!B(u),\, 0 \leq u \leq 1\}, \tag{2.1}$$

where $\lfloor t \rfloor = \max\{i \in \mathbb{Z} : i \leq t\}$ and $I\!B$ is the standard Brownian motion in $\mathbb{R}^d$, namely it has independent increments and $I\!B(u+v) - I\!B(u) \sim N(0, vI_d)$, $u, v \geq 0$. In this paper we should substantially generalize (2.1) by developing a convergence rate of (2.1) for multiple time series which can be dependent and non-identically distributed.

The invariance principle was introduced by Erdös and Kac (1946, [25]). Doob (1949, [18]), Donsker (1952, [17]) and Prohorov (1956, [61]) furthered their ideas, which led to the theory of weak convergence of probability measures. There is an extensive literature concerning Gaussian approximation when the dimension $d = 1$. In this case optimal rates for independent random variables were obtained in [38] and [66], among others. When $d = 1$ and $X_i$ are i.i.d. with mean 0, variance $\sigma^2$ and have finite $p$-th moment, $p > 2$, Komlós, Major and Tusnády (1975, 76, [38, 39]) established the very deep result

$$\max_{1 \leq i \leq n} |S'_i - \sigma B(i)| = o_{\text{a.s.}}(\tau_n), \tag{2.2}$$

where $B(\cdot)$ is the standard Brownian motion and $S'_n$ is constructed on a richer space such that $(S_i)_{i \leq n} \overset{D}{=} (S'_i)_{i \leq n}$, and the approximation rate $\tau_n = n^{1/p}$ is optimal. Results of type (2.2) have many applications in statistics since one can use functionals involving Gaussian processes to approximate statistics of $(X_i)_{i=1}^n$ and thus exploit properties of Gaussian processes. Their result was generalized to independent random vectors by Einmahl (1987a, [20]; 1987b, [21]; 1989, [22]), Zaitsev (2001, [82]; 2002a, [83]; 2002b, [84]) and Götze and Zaitsev (2008, [31]), where optimal and nearly optimal results were obtained.

To generalize (2.2) to multiple time series, we shall consider the possibly non-stationary, $d$-dimensional, mean 0, vector-valued process

$$X_i = (X_{i1}, \ldots, X_{id})^{\mathsf{T}} = H_i(\mathcal{F}_i) = H_i(\epsilon_i, \epsilon_{i-1}, \ldots), \quad i \in \mathbb{Z}, \tag{2.3}$$

where $^{\mathsf{T}}$ denotes matrix transpose, $\mathcal{F}_i = (\epsilon_i, \epsilon_{i-1}, \ldots)$ and $\epsilon_i, i \in \mathbb{Z}$, are i.i.d. random variables. Here, $H_i(\cdot)$ is a measurable function so that $X_i$ is well-defined. We allow $H_i$ to be possibly non-linear in its argument $(\epsilon_i, \epsilon_{i-1}, \ldots)$ to capture a much larger class of processes. If $H_i(\cdot) \equiv H(\cdot)$ does not depend on $i$, (2.3) defines a stationary causal process. The latter framework is very general; see [73, 75, 60] among others. When $d = 1$, Wiener [74] considered representing stationary processes by functionals of i.i.d. random variables.

Lütkepohl [52] presented many applications of functional central limit theorems for multiple time series analysis. Wu and Zhao (2007, [80]) and Zhou and Wu (2010, [90]) applied Gaussian approximation results with sub-optimal approximation rates to trend estimation and functional regression models. For the class of weakly dependent processes (2.3), we shall show that there exists a probability space $(\Omega_c, A_c, P_c)$, on which we can define random vectors $X_i^c$ with the partial sum process $S_i^c = \sum_{t=1}^i X_t^c$ and a Gaussian process $G_i^c = \sum_{t=1}^i Y_t^c$ with $Y_t^c$ being mean 0 independent Gaussian vectors such that $(S_i^c)_{1 \le i \le n} \stackrel{D}{=} (S_i)_{1 \le i \le n}$ and

$$\max_{i \le n} |S_i^c - G_i^c| = o_P(\tau_n) \quad \text{in } (\Omega_c, A_c, P_c), \tag{2.4}$$

where the approximation bound $\tau_n$ is related with the dependence decaying rates. Our result is useful for asymptotic inference for multiple time series. As a primary contribution, we generalize and improve the existing results for Gaussian approximations in several directions. For some $p > 2$, we assume uniform integrability of $p$th moment and obtain an approximation bound $\tau_n$ in terms of $p$ and the decay rate of functional dependence measure. In particular, if the dependence decays fast enough, for $\tau_n$, we are able to achieve the optimal $o_P(n^{1/p})$ bound. In the current literature, optimal results were obtained for some special cases only. We start with a brief overview of them.

For stationary processes with $d = 1$, a sub-optimal rate was derived in Wu (2007, [77]) where the martingale approximation is applied. Berkes, Liu and Wu (2014, [5]) considered causal stationary process (2.3) above and obtained the $n^{1/p}$ bound for $p > 2$. It is considerably more challenging to deal with vector-valued processes. Eberlein (1986, [19]) obtained a Gaussian approximation result for dependent random vectors with approximation error $O(n^{1/2-\kappa})$, for some small $\kappa > 0$. The latter bound can be too crude for many statistical applications. The martingale approximation approach in [77] can not be applied to vector-valued processes since Strassen's embedding generally fails for vector-valued martingales [54]. For a stationary multiple time series with additional constraints, Liu and Lin (2009, [49]) obtained an important result on strong invariance principles for stationary processes with bounds of the order $n^{1/p}$ with $2 < p < 4$. Wu and Zhou (2011, [81]) obtained sub-optimal rates for a multiple non-stationary time series. A critical limitation of the result by [81, 49] was the restriction $2 < p < 4$. It is an open problem on whether the bound $n^{1/p}$ can be achieved when $p \geq 4$.

In this paper, we show that under proper decaying conditions on functional dependence measures for the process (2.3), we can indeed obtain the optimal bound $n^{1/p}$ for $p \geq 4$. Our condition is stated in the form of (2.7), which involves two parameters $\chi$ and $A$ to formulate the temporal dependence of the process. Generally speaking, larger values of $\chi$ and $A$ means the dependence decays faster. With proper conditions on $A$, we find optimal $\tau_n = \tau_n(\chi)$ for a general $\chi > 0$. In Corollary 2.1 in Berkes, Liu and Wu (2014, [5]) the authors discussed univariate and stationary processes. However, their focus was on larger values of $\chi$ that allows them to obtain $\tau_n = n^{1/p}$. In Theorem 2.2.1 we obtain a rate for any $\chi > 0$ and show that if $\chi$ increases from 0 to a certain number $\chi_0$, we obtain the optimal $\tau_n$ varying from the worst, $n^{1/2}$, to the optimal, $n^{1/p}$. This work is substantially useful for processes where dependence does not decay fast enough. For the borderline case $\chi = \chi_0$, we can have $o_P(n^{1/p})$ rate for $2 < p < 4$ and for $p \geq 4$ we have $o_P(n^{1/p} \log n)$ rate. However, if $\chi > \chi_0$ we can obtain the optimal $o_P(n^{1/p})$ bound for all $p > 2$.

Our sharp Gaussian approximation result is quite useful for simultaneous inference of curves where the unknown function is not even Lipschitz continuous. There is a huge literature of curve estimation assuming smooth or regular behavior of a function but not so much for functions that are not differentiable or not Lipschitz continuous. Our Gaussian approximation can play a key role in weakening the smoothness assumption and thus enlarging the scope of doing statistical inference. Some applications are mentioned in Karmakar and Wu (2017). Moreover, since the optimal $o_P(n^{1/p})$ bound for $2 < p < 4$ and stationary processes obtained in [49] has remained a popular choice over the past few years for a multivariate Gaussian approximation, we can apply our sharper invariance principle that generalize ([49])'s one in multiple directions and give optimal rates when $p \geq 4$.

The rest of the article is organized as follows. In Section 2.2, we introduce the functional dependence measure and present the main result. Theorem 2.2.1 is proved in Sections 2.4 and 2.6.1. The proof of Theorem 2.2.2 is given in Section 2.5. Applications to covariance processes and locally stationary processes are given in Section 2.3. In Section 2.4 we discuss the proof strategy briefly to give the readers a basic idea of our long and involved derivation. Some useful results are collected in Section 2.7.

We now introduce some notation. For a random vector $Y$, write $Y \in \mathcal{L}_p, p > 0$, if $\|Y\|_p := E(|Y|^p)^{1/p} < \infty$. If $Y \in \mathcal{L}_2$, $Var(Y)$ denotes the covariance matrix. For $\mathcal{L}_2$ norm write $\|\cdot\| = \|\cdot\|_2$. Throughout the text, we use $c_p$ for constants that depend only on $p$ and $c$ for universal constants. These might take different values in different lines unless otherwise specified. For two positive sequences $a_n$ and $b_n$, if $a_n/b_n \to 0$ (resp. $a_n/b_n \to \infty$), write $a_n \ll b_n$ (resp. $a_n \gg b_n$). Write $a_n \lesssim b_n$ if $a_n \leq cb_n$ for some $c < \infty$. The $d$-variate normal distribution with mean $\mu$ and covariance matrix $\Sigma$ is denoted by $N(\mu, \Sigma)$. Denote by $I_d$ the $d \times d$ identity matrix. For a matrix $A = (a_{ij})$, we define its Frobenius norm as $|A| = (\sum a_{ij}^2)^{1/2}$. For a positive semi-definite matrix $A$ with spectral decomposition $A = QDQ^{\mathsf{T}}$, where $Q$ is orthonormal and $D = (\lambda_1, \ldots, \lambda_d)$ with $\lambda_1 \geq \ldots \geq \lambda_d$, write the Grammian square root $A^{1/2} = QD^{1/2}Q^{\mathsf{T}}$, $\rho_*(A) = \lambda_d$ and $\rho^*(A) = \lambda_1$.

## 2.2 Main Results

We first introduce uniform functional dependence measure on the underlying process using the idea of coupling. Let $\epsilon_i', \epsilon_j, i, j \in \mathbb{Z}$, be i.i.d. random variables. Assume $X_i \in \mathcal{L}_p, p > 0$. For $j \geq 0$, $0 < r \leq p$, define the functional dependence measure

$$\delta_{j,r} = \sup_i \|X_i - X_{i,(i-j)}\|_r = \sup_i \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r, \tag{2.5}$$

where $\mathcal{F}_{i,(k)}$ is the coupled version of $\mathcal{F}_i$ with $\epsilon_k$ in $\mathcal{F}_i$ replaced by an i.i.d. copy $\epsilon_k'$,

$$\mathcal{F}_{i,(k)} = (\epsilon_i, \epsilon_{i-1}, \ldots, \epsilon_k', \epsilon_{k-1}, \ldots) \text{ and } X_{i,(i-j)} = H_i(\mathcal{F}_{i,(i-j)}).$$

Also, $\mathcal{F}_{i,(k)} = \mathcal{F}_i$ if $k > i$. Note that, $\|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r$ measures the dependence of $X_i$ on $\epsilon_{i-j}$. Since the physical mechanism function $H_i$ can possibly be different for a non-stationary process, we choose to define the functional dependence measure in an uniform manner. The quantity $\delta_{j,r}$ measures the uniform $j$-lag dependence in terms of the $r$th moment. Assume throughout the paper that

$$\Theta_{0,p} = \sum_{i=0}^{\infty} \delta_{i,p} < \infty. \tag{2.6}$$

This condition implies short range dependence in the sense that the cumulative dependence of $(X_j)_{j \geq k}$ on $\epsilon_k$ is finite. For presentational clarity, in this paper we assume there exists

$\chi > 0, A > 0$ such that the tail cumulative dependence measure

$$\Theta_{i,p} = \sum_{j=i}^{\infty} \delta_{j,p} = O\left(i^{-\chi}(\log i)^{-A}\right). \tag{2.7}$$

Larger $\chi$ or $A$ implies weaker dependence. Our Gaussian approximation rate $\tau_n$ (cf Theorems 2.2.1 and 2.2.2) depends on $\chi$ and $A$. Define functions $f_j(\cdot,\cdot)$ by

$$\begin{aligned} f_1 &= f_1(p,\chi) = p^2\chi^2 + p^2\chi,\ \ f_2 = 2p\chi^2 + 3p\chi - 2\chi, \\ f_3 &= p^3(1+\chi)^2 + 6f_1 + 4p\chi - 2,\ \ f_4 = 2p(2p\chi^2 + 3p\chi + p - 2), \\ f_5 &= p^2(p^2+4p-12)\chi^2 + 2p(p^3+p^2-4p-4)\chi + (p^2-p-2)^2. \end{aligned} \tag{2.8}$$

Assume that the process (2.3) satisfies the uniform integrability and the regularity condition on the covariance structure. The latter is frequently imposed in study of multiple time series.

(2.A) The series $(|X_i|^p)_{i\geq 1}$ is uniformly integrable: $\sup_{i\geq 1} E(|X_i|^p \mathbf{1}_{|X_i|\geq u}) \to 0$ as $u \to \infty$;

(2.B) (Lower bound on eigenvalues of covariance matrices of increment processes) There exists $\lambda_* > 0$ and $l_* \in \mathbb{N}$, such that for all $t \geq 1, l \geq l_*$,

$$\rho_*(Var(S_{t+l} - S_t)) \geq \lambda_* l.$$

**Theorem 2.2.1.** *Assume $E(X_i) = 0$, (2.A)-(2.B) and (2.7) holds with*

$$0 < \chi < \chi_0 = \frac{p^2 - 4 + (p-2)\sqrt{p^2+20p+4}}{8p}, \tag{2.9}$$

$$A > \frac{(2p+p^2)\chi + p^2 + 3p + 2 + f_5^{1/2}}{p(1+p+2\chi)}. \tag{2.10}$$

*Then (2.4) holds with the approximation bound $\tau_n = n^{1/r}$, where*

$$\frac{1}{r} = \frac{f_1 + p^2\chi + p^2 - 2p + f_2 - \chi\sqrt{(p-2)(f_3-3p)}}{f_4}. \tag{2.11}$$

**Theorem 2.2.2.** *Assume $E(X_i) = 0$, (2.A)-(2.B), (2.7). Recall (2.9) for $\chi_0$. (i) If $\chi > \chi_0$, and $A > 0$, we can achieve (2.4) with $\tau_n = n^{1/p}$ for all $p > 2$. For $\chi = \chi_0$, assume that $A$ satisfies (2.10). (ii) If $2 < p < 4$, we have $\tau_n = n^{1/p}$; (iii) if $p \geq 4$, we have $\tau_n = n^{1/p} \log n$.*

Theorems 2.2.1 and 2.2.2 concern the two cases $\chi < \chi_0$ and $\chi \geq \chi_0$, respectively, and they are proved in Sections 2.4 and 2.5. The proof of Theorem 2.2.2 requires a more refined treatment so that the optimal rate can be derived. For Theorem 2.2.1 and Theorem 2.2.2(i) and (iii), we apply Götze and Zaitsev (2008, [31]); see Proposition 2.7.1, while for Theorem 2.2.2(ii), Proposition 1 from Einmahl (1987, [20]) is applied. The expression of $r$ is complicated. Figure 2.1 plots the power $\max(1/r, 1/p)$. As $\chi \to 0$, $r \to 2$, and $r = p$ if $\chi = \chi_0$.

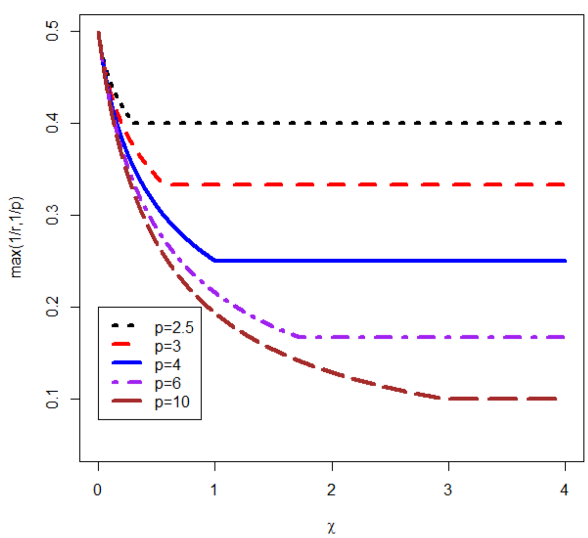



Figure 2.1: Optimal bound as a function of $\chi$

**Remark 2.2.3.** *The lower bound of $A$ for the case $\chi = \chi_0$ can be further simplified to*

$$A > \frac{p^2 + 8p + 4 + (p-2)\sqrt{p^2 + 20p + 4}}{6p}.$$

## 2.3 Applications

### 2.3.1 Covariance processes:

Assuming that $X_i$ is a vector linear process

$$X_i = \sum_{j=0}^{\infty} B_j \epsilon_{i-j}, \tag{2.12}$$

where $B_j$ are $d \times d$ coefficient matrix, and $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{id})^{\mathsf{T}}$, $\epsilon_{ir}$ are i.i.d. random variables with mean 0 and finite $q$th moment, $q > 4$. Let the $d(d+1)/2$ dimensional vector $W_i = (X_{ir}X_{is})_{1 \le r \le s \le d}$. Then $\bar{W}_n := \sum_{i=1}^{n} W_i/n$ gives sample covariances of $(X_i)_{i=1}^{n}$. Assume

$$\sum_{j=t}^{\infty} |B_j| = O(t^{-\chi}(\log t)^{-A}), \tag{2.13}$$

where $A$ satisfies (2.10). Write $p = q/2$. Note that, $\|X_i - X_{i,(i-j)}\|_q = O(|B_j|)$. Fix two co-ordinates $1 \le r \le s \le d$. Then,

$$\begin{aligned}
& \|X_{ir}X_{is} - X_{i,(i-j)r}X_{i,(i-j)s}\|_p \\
\le\ & \|X_{ir}X_{is} - X_{ir}X_{i,(i-j)s}\|_p + \|X_{ir}X_{i,(i-j)s} - X_{i,(i-j)r}X_{i,(i-j)s}\|_p \\
\le\ & \|X_{ir}\|_q \|X_{is} - X_{i,(i-j)s}\|_q + \|X_{ir} - X_{i,(i-j)r}\|_q \|X_{i,(i-j)s}\|_q \\
=\ & O(|B_j|),
\end{aligned}$$

where $B_{\cdot r}$ and $B_{\cdot s}$ and denote $r$th and $s$th row of the coefficient matrix $(B_j)_{j \ge 0}$. Thus the condition (2.13) translates to the condition (2.7) for the $W$ process. Condition (2.A) is trivially satisfied since $W$ is stationary and has finite $p$th moment.

Let $\Sigma = \sum_{k=-\infty}^{\infty} Cov(W_0, W_k)$ be the long-run covariance matrix of $(W_i)$. By Theorems

2.2.1 and 2.2.2, we have

$$\max_{i\le n}|i\bar{W}_i - iE(W_1) - \Sigma^{1/2}IB(i)| = o_P(\tau_n), \tag{2.14}$$

where $\tau_n$ takes the value $n^{1/r}$ (see (2.11)), and $n^{1/p}$ based on $\chi < \chi_0$ and $\chi > \chi_0$ respectively, $IB$ is a centered standard Brownian motion. Result (2.14) is helpful for change point inference for multiple time series based on covariances; see [3, 72] among others.

### *2.3.2 Nonlinear non-stationary time series:*

Consider the process

$$X_i = F(X_{i-1}, \epsilon_i, \theta(i/n)),\ 1 \le i \le n,$$

where $\epsilon_i$ are i.i.d. random variables, $F$ is a measurable function, $\theta : [0,1] \to \mathbb{R}$ is a parametric function such that $\max_{0\le u\le 1}\|F(x_0, \epsilon_i, \theta(u))\|_p < \infty$, and

$$\sup_{0\le u\le 1}\sup_{x\neq x'}\frac{\|F(x,\epsilon_i,\theta(u)) - F_i(x',\epsilon_i,\theta(u))\|_p}{|x-x'|} < 1. \tag{2.15}$$

Then the process $(X_i)$ satisfies the Geometric moment contraction: for some $0 < \beta < 1$,

$$\delta_{i,p} = O(\beta^i). \tag{2.16}$$

Thus (2.7) holds for any $\chi > 0$ and Theorem 2.2.2 is applicable with rate $\tau_n = n^{1/p}$. This facilitates inference for the unknown parametric function $\theta$. Time-varying ARCH and GARCH are prominent examples in this large class of models.

### 2.3.3 Non-Lipschitz trend function in presence of non-linear error:

Consider the trend-noise model with a non-Lipschitz trend and a non-linear non-stationary error process

$$\mathbf{X}_i = \mathbf{g}(i/n) + \mathbf{Y}_i. \tag{2.17}$$

One such example of $\mathbf{Y}_i$ can be a time-varying MGARCH process

$$\begin{aligned}
\mathbf{Y}_i &= \Sigma_i^{1/2} e_i \\
\Sigma_i &= A_i \Sigma_{i-1} A_i' + B_i \mathbf{Y}_i \mathbf{Y}_i^T B_i'
\end{aligned}$$

for some set of coefficient matrices $A_i$ and $B_i$. One can easily compute the functional dependence measure for such a process and thus can impose conditions on them so that conditions of Theorems 2.2.1 and 2.2.2 are true. For a general error process $\mathbf{Y}_i$, a Pristley-Chao type estimate for $\mathbf{g}$ reads

$$\hat{\mathbf{g}}(t) = \frac{1}{nb_n} \sum K\left(\frac{i/n - t}{b_n}\right) \mathbf{X}_i.$$

Assume $g$ is Hölder-$\alpha$ continuous for $\alpha < 1/2$. Towards constructing a simultaneous confidence band for $\mathbf{g}$, we bound $\hat{\mathbf{g}}(t) - \mathbf{g}(t)$ by

$$\begin{aligned}
\sqrt{nb_n}(\hat{\mathbf{g}}(t) - \mathbf{g}(t)) &= \frac{1}{\sqrt{nb_n}} \sum K\left(\frac{i/n - t}{b_n}\right)(\mathbf{g}(i/n) - \mathbf{g}(t)) \\
&+ \frac{1}{\sqrt{nb_n}} \{\sum K\left(\frac{i/n - t}{b_n}\right) - 1\}\mathbf{g}(t) \\
&+ \frac{1}{\sqrt{nb_n}} \sum K\left(\frac{i/n - t}{b_n}\right)(\mathbf{Y}_i - z_i) \\
&+ \frac{1}{\sqrt{nb_n}} \sum K\left(\frac{i/n - t}{b_n}\right) z_i
\end{aligned} \tag{2.18}$$

where $z_i$ is the approximating Gaussian process. We use Gaussian extreme value theory from Lindgren (1980, [46]) to obtain distributional theory for the last term. We show the other three terms are negligible. Using Hölder-$\alpha$ continuity of $\mathbf{g}$ and assuming bounded variation for $K$, one needs $n^{1/2}b_n^{1/2+\alpha} \to 0$. For the third term using the Gaussian approximation in 2.4 with $n^{1/r}$ rate, one has $n^{1/r}/\sqrt{nb_n} \to 0$. Clearly if $\alpha < 1/2$ then one need $n^{1/r}$ bound for $r > 4$ which we can achieve using Theorem 2.2.2 if the decay rate for the error process $\mathbf{Y}_i$ is fast enough.

## 2.4 Key ideas of the proof of Theorem 2.2.1

The proof of Theorem 2.2.1 is quite involved. Here we discuss a brief outline of the major components of the proof. In particular our discussion will emphasize the difficulties arise due to non-stationarity and vector-valued process and the techniques we use to circumvent those. Since these techniques allow us to solve this problem in such generality, we believe it might be of independent interest to the reader to at least have an overview of the major steps. The detailed proof is postponed to the appendix.

The first part of our proof consists of series of approximations to create almost independent blocks. The first of them, the truncation approximation will ensure the optimal $n^{1/p}$ bound. Secondly, we use the $m-$dependence approximation for a suitably chosen sequence $m_n$ in terms of the decay rate $\chi$. Lastly, the blocking approximation requires some sharp Rosenthal-type inequality that needs $\gamma$th moment of the block-sums in the numerator with $\gamma > p$. It is essential to use a power higher than $p$ to obtain a better rate.

To maintain clarity, we defer the exact choice of $\gamma$ and $m_n$ in terms of $\chi$ and $A$ to Subsection 2.4.4. Instead, in this subsection we come up with conditions (2.23), (2.26) and (2.27) to ensure $n^{1/r}$ rate and solve $\gamma, m_n$ and $r$ later to obtain the best possible choices for this sequences. Henceforth, we drop the suffix of $m_n$ for our convenience.

### *2.4.1 Outline of preparation step*

This step consists of truncation, m-dependence and blocking. All the three steps consist new techniques or approaches from what has been done in the literature for similarly named steps. These preparation steps are extremely important as they allow us to build a system of equation to solve for the approximation rate $\tau_n = n^{1/r}$ as a function of decay rate $\lambda$ in (2.7).

The truncation step is new compared to related literature as we introduced the operator of the type $T_b$ in (2.19). It differs from the treatment of Berkes, Liu and Wu (2014, [5]) because of the inclusion of the term $t_n$. This is necessary due to the non-stationarity. For the $m$-dependence approximation step, we use a generic sequence $m$ depending on how fast $\Theta_{i,p}$ decays in $\lambda$ according to rule (2.7). The blocking step needs $k$-dic decomposition where $k$ is possibly greater or equal to 3 to allow for non-stationarity.
Truncation approximation is necessary to allow higher moments manipulations. For $b > 0$ and $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, define

$$T_b(v) = (T_b(v_1), \ldots, T_b(v_d))^\top, \text{ where } T_b(w) = \min(\max(w, -b), b). \tag{2.19}$$

**Proposition 2.4.1.** *Assume Condition (2.A). It is possible to choose a sequence $t_n \to 0$ slow enough such that we have*

$$\max_{1 \le i \le n} |S_i - S_i^{\oplus}| = o_P(n^{1/p}), \textit{ where } S_l^{\oplus} = \sum_{i=1}^{l} [T_{t_n n^{1/p}}(X_i) - ET_{t_n n^{1/p}}(X_i)] \tag{2.20}$$

The $m$-dependence approximation is a very important tool that is extensively used in literature; see for example the Gaussian approximation in Liu and Lin (2009, [49]) and Berkes, Liu and Wu (2014, [5]). For a suitably chosen sequence $m$, we look at the conditional mean $E(X_i|\epsilon_i, \ldots \epsilon_{i-m})$. This gives a very simple yet effective way to handle the original process in terms of a collection of $\epsilon_i$'s. Define the partial sum process

$$\tilde{R}_{c,l} = \sum_{i=1+c}^{l+c} \tilde{X}_j, \text{ where } \tilde{X}_j = E(T_{t_n n^{1/p}}(X_j)|\epsilon_j, \ldots, \epsilon_{j-m}) - E(T_{t_n n^{1/p}}(X_j)). \tag{2.21}$$

Write $\tilde{R}_{0,i} = \tilde{S}_i$. From Lemma A1 in Liu and Lin (2009, [49]), we have

$$\| \max_{1 \le l \le n} |S_l^{\oplus} - \tilde{S}_l| \|_r \le c_r n^{1/2} \Theta_{1+m,r}. \tag{2.22}$$

The proofs in [49] are for stationary processes. Since our $\delta_{j,r}$ in (2.5) is defined in an uniform manner, the proof goes through for the non-stationary case as well. Assume

$$n^{1/2-1/r} \Theta_{m,r} \to 0. \tag{2.23}$$

By (2.22) and (2.23), we have $n^{1/r}$ convergence in the $m$-dependence approximation step

$$\max_{1 \le i \le n} |S_i^{\oplus} - \tilde{S}_i| = o_P(n^{1/r}). \tag{2.24}$$

Towards the blocking approximation, we approximate the partial sum process $\tilde{S}_i$ by sums of $A_j$ where, for $j \ge 0$,

$$A_{j+1} = \sum_{i=2jk_0m+1}^{(2k_0j+2k_0)m} \tilde{X}_i, \text{ where } k_0 = \lfloor \Theta_{0,2}^2/\lambda_* \rfloor + 2. \tag{2.25}$$

To this end, we will need the following two conditions, for some $\gamma > p$,

$$n^{1-\gamma/r} m^{\gamma/2-1} \to 0, \tag{2.26}$$

$$n^{1/p-1/\gamma} \sum_{j=m+1}^{\infty} \delta_{j,p}^{p/\gamma} \to 0. \tag{2.27}$$

We assume an almost polynomial rate for $m$: for some $0 < L < 1$,

$$m = \lfloor n^L t_n^k \rfloor, \quad 0 < k < (\gamma - p)/(\gamma/2 - 1). \tag{2.28}$$

**Proposition 2.4.2.** *Assume (2.26) and (2.27) for some $\gamma > p$. Moreover, assume (2.28) for the $m$-sequence and (2.7) for the decay rate of $\Theta_{i,p}$ with some $A > \gamma/p$. Then*

$$\max_{1 \le i \le n} |\tilde{S}_i - S_i^\diamond| = o_P(n^{1/r}), \text{ where } S_i^\diamond = \sum_{j=1}^{q_i} A_j, \ q_i = \lfloor i/(2k_0 m) \rfloor. \tag{2.29}$$

Summarizing (2.20), (2.24) and (2.30), we can work on $S_i^\diamond$ in view of

$$\max_{1 \le i \le n} |S_i - S_i^\diamond| = o_P(n^{1/r}). \tag{2.30}$$

In the next steps, we obtain a Gaussian approximation for $S_n^\diamond$; see backgrounds in Sections 2.4.2 and 2.4.3 and detailed argument in Section 2.6.1.

### *2.4.2 Outline of conditional Gaussian approximation:*

The blocks created in the first steps are not independent because two successive blocks share some $\epsilon_i$'s in their shared border. In this second stage, we look at the partial sum process conditioned on these borderline $\epsilon_i$'s, which implies conditional independence. Berkes, Liu and Wu (2014, [5]) did a similar treatment with triadic decomposition for stationary scalar processes and applied Sakhanenko's (2006, [66]) Gaussian approximation result on the conditioned process.

Since the result from Sakhanenko (2006, [66]) is only valid for $d = 1$, we need to use the Gaussian approximation result from Götze and Zaitsev (2008, [31]) (see Proposition 2.7.1) for $d \ge 2$. This comes with the cost of verifying a very technical sufficient condition on the covariance matrices of the independent vectors. Verification of such a condition is quite involved in our case since we are dealing with the conditional process. We opt for a $k$-dic decomposition instead of the triadic decomposition in [5]. This is necessary to accommodate the non-stationarity of the process. We need $k_0 > \Theta_{0,2}^2/\lambda_*$ (cf. (2.25)), where $\lambda_*$ is mentioned in Condition 2.B.

### 2.4.3 Outline of regrouping and unconditional Gaussian approximation:

In the last part of our proof, we obtain the Gaussian approximation for the unconditional process by applying Proposition 2.7.1 one more time. In the second part of our proof, we look at the conditional variance (cf $V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}) = Var(Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}))$ in (2.45) of Subsection 2.6.2) of the blocks. These conditional variances are 1-dependent. In order to apply Götze and Zaitsev (2008, [31])'s result, we rearrange the sums of these variances into sums of independent blocks (cf 2.47 in Subsection 2.6.2). Due to the non-stationarity, this regrouping is completely different and much more involved than Berkes, Liu and Wu (2014, [5]). In particular, the regrouping procedure leads to matrices that may not be positive definite hence cannot be used directly as possible covariance matrices of Gaussian processes. We overcome this obstacle by introducing a novel positive-definitization that does not affect the optimal rate.

### 2.4.4 Conclusion of the proof:

This subsection discusses the specific choice of the sequence $m, \gamma$ and the rate $\tau_n = n^{1/r}$ starting from (2.23), (2.26) and (2.27). Elementary calculations show that $r < p$ for $\chi < \chi_0$. Provided $1 - (\chi+1)p/\gamma < 0$, we have

$$\begin{aligned} \sum_{j=m+1}^{\infty} \delta_{j,p}^{p/\gamma} &\leq \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} \sum_{j=2^i}^{2^{i+1}-1} \delta_{j,p}^{p/\gamma} \leq \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} 2^{i(1-p/\gamma)} \Theta_{2^i,p}^{p/\gamma} \qquad (2.31) \\ &= \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} 2^{i(1-p/\gamma)} O(2^{-\chi ip/\gamma} i^{-Ap/\gamma}) = O(m^{1-p/\gamma-\chi p/\gamma} (\log m)^{-Ap/\gamma}). \end{aligned}$$

By (2.40) and (2.28) $\log m \asymp \log n$. Assume that,

$$\begin{aligned}
1/2 - 1/r - \chi L &= 0, \quad A > \gamma/p, && (2.32)\\
1 - \gamma/r + L(\gamma/2 - 1) &= 0, \quad 0 < k < (\gamma/2-1)^{-1}(\gamma - p) && (2.33)\\
1/p - 1/\gamma + (1 - (\chi+1)p/\gamma)L &= 0. && (2.34)
\end{aligned}$$

Then conditions (2.23), (2.26) and (2.27) hold. Solving equations in (2.32), (2.33) and (2.34), we obtain $r$ given in (2.11),

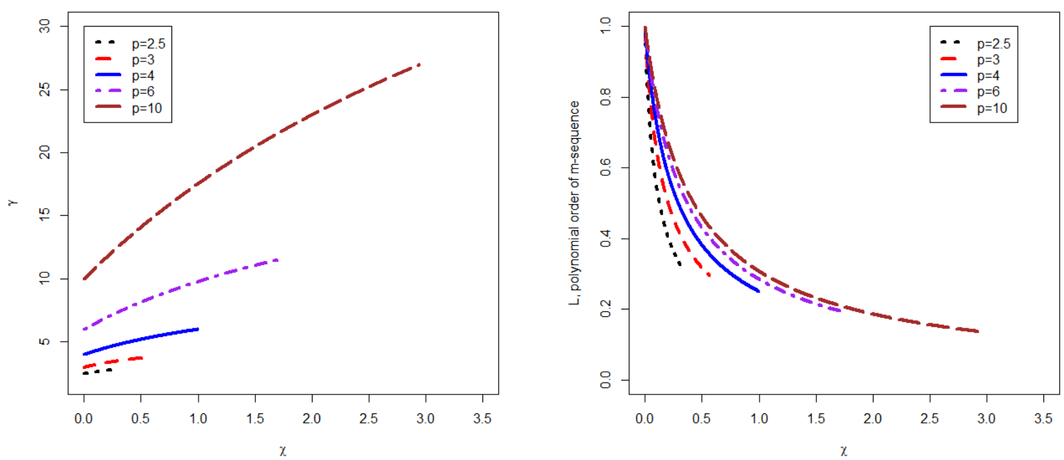
$$\begin{aligned}
\gamma &= \frac{(2p+p^2)\chi + p^2 + 3p + 2 + f_5^{1/2}}{2 + 2p + 4\chi},\\
L &= \frac{f_1 - f_2 + \chi\sqrt{(p-2)(f_3 - 3p)}}{\chi f_4},
\end{aligned}$$

with $f_1, \ldots, f_5$ given in (2.8). Moreover, we specifically choose $A > 2\gamma/p$ for a crucial step in the proof of our Gaussian approximation; see (2.63).

**Remark 2.4.3.** *Figure 2.2 depicts how $\gamma$ and $L$ change with $p$ and $\chi$ for $\chi < \chi_0$. Note that in Figure 2.2, $L$, the power of $n$ in the expression of $m$ is close to 1 if $\chi$ is small. This is intuitive since if dependence decays very slowly, to make blocks of size $m$ or a multiple of $m$ behave almost independently, one needs a larger $L$.*

## 2.5 Proof of Theorem 2.2.2

*Proof. Case 1 ($\chi > \chi_0$):-* Note that the optimal power $\gamma$ and the optimal bound $1/r$ increases and decreases with $\chi$ respectively (see also Figures 2.1 and 2.2). This is a motivation behind tweaking our proof for the verification of (2.70) to handle the $(\log n)$ term in choice of $l$ in (2.49). While using the Nagaev inequality to show (2.66), we use a power $\gamma' > \gamma$ while




(a) (b)

Figure 2.2: (a) $\gamma$ as a function of $\chi$, (b) $L$ as a function of $\chi$

keeping the choice of $l$ (cf 2.49) same as before. We form a set of new equations

$$\begin{aligned} 1/2 + 1/p - 2/r' + L'(1-(\chi+1)p/r') &= 0, \\ 1/p - 1/\gamma' + L' - L'(\chi+1)p/\gamma' &= 0, \\ 1 - \gamma'/r' + L'(\gamma'/2 - 1) &= 0. \end{aligned} \tag{2.35}$$

The intuition behind the first of these equations is to use a higher power than $p$ in the m-dependence approximation. However, we only defined moments up to $p$. So we use Lemma 2.7.3 to obtain a new equation corresponding to the $m$-dependence approximation using a power $r'$ that is little higher than $p$. The solution of (2.35) has the property

$$\gamma' < 2(1 + p + p\chi)/3. \tag{2.36}$$

for $\chi > \chi_0$. Also we observe $L' < L(\chi_0)$ (cf Figure 2.2) and hence $m^{1-\gamma'/2} \ll m'^{1-\gamma'/2}$ where $m'$ is taken as $n^{L'} t_n^k$ following (2.28). We apply the following version of Nagaev-type inequality from Liu, Xiao and Wu (2013, [51]) to obtain

$$\begin{aligned} P(|\tilde{S}_m| \geq \sqrt{lm}) \;\; &\lesssim \;\; \frac{m}{(lm)^{\gamma'/2}} \nu_R^{\gamma'+1} + \sum_{r=1}^{R} \exp\left(-c_{\gamma'} \frac{\lambda_r^2 l}{\tilde{\theta}_{r,2}^2}\right) + \frac{m^{\gamma'/2} \tilde{\Theta}_{m+1,\gamma'}^{\gamma'}}{(lm)^{\gamma'/2}} \\ &+ \;\; \frac{m \sup_i \|T_{t_n n^{1/p}}(X_i)\|_{\gamma'}^{\gamma'}}{(lm)^{\gamma'/2}} + \exp\left(-\frac{c_{\gamma'} l}{\sup_i \|T_{t_n n^{1/p}}(X_i)\|_2^2}\right), \end{aligned} \tag{2.37}$$

where $\nu_R = \sum_{r=1}^{R} \mu_r$, $\mu_r = (\tau_r^{\gamma'/2-1} \tilde{\theta}_{r,\gamma'}^{\gamma'})^{1/(\gamma'+1)}$, $\lambda_r = \mu_r/\nu_R$, $\tilde{\theta}_{r,t} = \sum_{i=1+\tau_{r-1}}^{\tau_r} \tilde{\delta}_{i,t}$ for some sequence $0 = \tau_0 < \tau_1 < \ldots < \tau_R = m$. For the choice $\tau_r = 2^{r-1}$ for $1 \leq r \leq R-1 = \lfloor \log_2 m \rfloor$, we obtain $\nu_R^{\gamma'+1} = O(n^{\gamma'/p-1} t_n^{\gamma'-p})$ using (2.36), (2.39) under the decay condition on $\Theta_{i,p}$ in (2.7). The third term and the exponential terms are straightforward to deal with. The fourth term is dealt similar to (2.75). Combining these in the view of our new set of equations in (2.35), we get $P(|\tilde{S}_m| \geq \sqrt{lm}) = o(m/n)$ which is sufficient to conclude the proof as proposed in (2.66).

The positive-definitization technique introduced in (2.54) is validated in Proposition 2.6.6. This step requires $\gamma > 4\chi$ for the case $\chi > \max(1/2, \chi_0)$. We observe that $\gamma' - 4\chi = 0$ has a root $\chi_1 > \chi_0$. This allows us to replace $\chi$ in the decay condition of $\Theta_{i,p}$ by $\min(\chi, \chi_1)$ and the proof goes through. The arguments for the rest of the proof of Theorem 2.2.1 remains valid.

*Case 2 ($\chi = \chi_0, 2 < p < 4$):-* We shall apply Proposition 1 from Einmahl (1987, [20]). He proved a Gaussian approximation result for independent but not necessarily identical vectors with diagonal covariance matrix. The two remarks following the proposition mention that the diagonal nature of every covariance matrix can be relaxed if these matrices have bounded eigenvalues. A careful check of his proof reveals that it can be further relaxed to the assumption of bounded eigenvalues of the covariance matrix of a normalized block sum only. This allows us to replace the $l$ (see (2.49)) to use the conclusion of Proposition 2.7.1 by $l'$ without the logarithm term ($\log n$) in the denominator and without the condition (2.71). Thus we obtain $o_P(n^{1/p})$ rate for all $2 < p < 4$.

*Case 3 ($\chi = \chi_0, p \geq 4$):-* In this case we do not have a similar optimal Gaussian approximation result for independent but not identically distributed random vectors. Instead we shall apply Proposition 2.7.1 again. The sufficient conditions in that result lead to an unavoidable $(\log n)$ term in choice of $l$ (see 2.49). This, in turn leads to $o_P(n^{1/p}\log n)$ rate. Note that, $\chi_0 > 1/2 - 1/p$ for all $p > 2$. From the proof for the case $0 < \chi < \chi_0$, consider (2.68), observe that if $\chi = \chi_0$, then

$$\frac{n}{m}P(|\tilde{S}_m| \geq \sqrt{lm}) = O((\log n)^p t_n^{k(p/\gamma - p/2)}),$$

which may diverge to $\infty$. To deal with this difficulty in this special case, we choose a different $m$ sequence. Our new set of conditions with $\tau_n = n^{1/p}(\log n)^\delta$ are

$$\begin{aligned}
n^{1/2-1/p}m^{-\chi}(\log n)^{-A-\delta} &\to 0,\\
n^{1/p-1/\gamma}m^{1-(\chi+1)p/\gamma}(\log n)^{-Ap/\gamma} &\to 0,\\
n^{1-\gamma/p}(\log n)^{-\gamma\delta}m^{\gamma/2-1} &\to 0,\\
(\log n)^{\gamma}m^{1-\gamma/2}n^{\gamma/p-1}t_n^{\gamma-p} &\to 0.
\end{aligned}$$

where the last one is obtained using $\gamma$th moment in (2.37). Let $m = \lfloor n^L(\log n)^{2\gamma/(\gamma-2)}t_n^k\rfloor$ with $0 < k < (\gamma/2-1)^{-1}(\gamma - p)$, we can achieve $\delta = 1$. We still have the same set of equations for $L, \gamma$ and $r$ as (2.32), (2.33) and (2.34). A careful check reveals that the rest of the proof goes through with this modified $m$ sequence. $\square$

## 2.6 Appendix A: Detailed Steps of the Proof of Theorem 2.2.1

The long detailed steps are in this section and the technical details are given in Appendix Section 2.6.1. All the lemmas are postponed to Appendix Section 2.7.

### 2.6.1 *Preparation:-Truncation, m-dependence and blocking approximations:*

## Truncation approximation:

*Proof.* of Proposition 2.4.1. We introduce a very slowly converging sequence $t_n \to 0$ based on the uniform integrability condition (2.A). For every $t > 0$, we have

$$\sup_i \frac{1}{t^p} E(|X_i|^p \mathbf{1}_{|X_i| > tn^{1/p}}) = 0 \text{ and } n \sup_i E \min(\frac{|X_i|^\gamma}{t^\gamma n^{\gamma/p}}, 1) \to 0 \text{ as } n \to \infty, \tag{2.38}$$

where $\gamma > p$. The second relation follows from Lemma 2.7.2. Clearly (2.38) implies that

$$\sup_i \frac{1}{t_n^p} E(|X_i|^p \mathbf{1}_{|X_i| > t_n n^{1/p}}) + n \sup_i E \min(\frac{|X_i|^\gamma}{t_n^\gamma n^{\gamma/p}}, 1) \to 0 \text{ as } n \to \infty, \tag{2.39}$$

holds for a sequence $t_n \to 0$ very slowly. Without loss of generality we can let

$$t_n \log\log n \to \infty \tag{2.40}$$

since otherwise we can replace $t_n$ by $\max(t_n, (\log\log n)^{-1/2})$ (say).

The truncation operator $T_b$ in (2.19) is Lipschitz continuous with Lipschitz constant 1. Let

$$R_{c,l} = \sum_{i=1+c}^{l+c} X_i^\oplus = \sum_{i=1+c}^{l+c} [T_{t_n n^{1/p}}(X_i) - ET_{t_n n^{1/p}}(X_i)]. \tag{2.41}$$

By (2.39), we have $P(\max_{i \le n} |S_i - \sum_{j=1}^i T_{t_n n^{1/p}}(X_j)| = 0) \to 1$ in view of

$$\sup_j P\left(|X_j| > t_n n^{1/p}\right) \le \sup_j \frac{1}{n t_n^p} E\left(|X_j|^p I\left(|X_j| > t_n n^{1/p}\right)\right) = o(1/n).$$

Also by (2.39), $\max_{j \le n} |E(X_j - T_{t_n n^{1/p}}(X_j))| = o(n^{1/p-1})$. Hence (2.20) follows. □

Blocking approximation:

We now define functional dependence measure for the truncated process $(T_{t_n n^{1/p}}(X_i))_{i\le n}$ as

$$\delta^{\oplus}_{j,l} = \sup_i \|T_{t_n n^{1/p}}(X_i) - T_{t_n n^{1/p}}(X_{i,(i-j)})\|_l, \text{ where } l \ge 2.$$

Similarly, define the functional dependence measure for the $m$-dependent process $(\tilde{X}_i)$ as

$$\tilde{\delta}_{j,l} = \sup_i \|\tilde{X}_i - \tilde{X}_{i,(i-j)}\|_l.$$

For these dependence measures, the following inequality holds for all $l \ge 2$:

$$\tilde{\delta}_{j,l} \le \delta^{\oplus}_{j,l} \le \delta_{j,l}. \tag{2.42}$$

We now proceed to proving Proposition 2.4.2, the blocking approximation result. As mentioned in the main text, we need to assume conditions (2.26) and (2.27) for this step. The almost-polynomial rate of $m$ as mentioned in (2.28) is also assumed.

Remark: We need another condition for the blocking approximation (see (2.73) in the proof of Lemma 2.7.4). However, we skip it here and choose $m$ and $\gamma$ such that conditions (2.23), (2.26) and (2.27) are met. These will automatically imply this fourth one in view of (2.7).

*Proof.* of Proposition 2.4.2: Let $\mathcal{S} = \{2ik_0m, 0 \le i \le q_n\}$, $\phi_n = (n^{1-\gamma/r}m^{\gamma/2-1})^{1/(2\gamma)}$. Then

$$\begin{aligned} P\left(\max_{1\le l\le n} |\tilde{R}_{0,l} - \sum_{j=1}^{\lfloor l/(2k_0m)\rfloor} A_j| \ge \phi_n n^{1/r}\right) &\le \frac{n}{2k_0m}\max_{c\in\mathcal{S}} P(\max_{1\le l\le 2k_0m} |\tilde{R}_{c,l}| \ge \phi_n n^{1/r}) \\ &\le n\max_{c\in\mathcal{S}} \frac{E(\max_{1\le l\le 2k_0m}|\tilde{R}_{c,l}|^{\gamma})}{2k_0m\phi_n^{\gamma}n^{\gamma/r}} = O(\phi_n^{\gamma}), \end{aligned}$$

from the assumption (2.26) and Lemma 2.7.4. Since $\phi_n \to 0$, (2.29) follows. $\square$

In this section we shall provide details of the arguments for steps mentioned in Sections 2.4.2 and 2.4.3. Section 2.6.2 presents the conditional Gaussian approximation, where we shall apply Proposition 2.7.1 stated in Section 2.7. Section 2.6.3 deals with unconditional Gaussian approximation and regrouping.

### *2.6.2 Conditional Gaussian approximation:*

The blocks $A_j$ created in (2.25) after the blocking approximation are weakly independent; except they share some dependence on the border. In this subsection, we look at the conditional process given the $\epsilon_i$ the blocks share in their borders. Demeaning the conditional process, we apply the Proposition 2.7.1 for the Gaussian approximation. For $1 \leq i \leq n$, let $\tilde{H}_i$ be a measurable function such that

$$\tilde{X}_i = \tilde{H}_i(\epsilon_i, \ldots, \epsilon_{i-m}). \tag{2.43}$$

Recall Proposition 2.4.2 for the definition of $q_i$. Let $q = q_n$. For $j = 1, \ldots, q$, define

$$\bar{a}_{2k_0 j} = \{a_{(2k_0 j-1)m+1}, \ldots, a_{2k_0 jm}\} \text{ and } a = \{\ldots, \bar{a}_0, \bar{a}_{2k_0}, \bar{a}_{4k_0}, \ldots\}.$$

Given $a$, define, for $2k_0 jm + 1 \leq i \leq (2k_0 j + 1)m$,

$$\tilde{X}_i(\bar{a}_{2k_0 j}) = \tilde{H}_i(\epsilon_i, \ldots, \epsilon_{2k_0 jm+1}, a_{2k_0 jm}, \ldots, a_{i-m})$$

and for $(2k_0 j + 2k_0 - 1)m + 1 \leq i \leq (2k_0 j + 2k_0)m$,

$$\tilde{X}_i(\bar{a}_{2k_0 j+2k_0}) = \tilde{H}_i(a_i, \ldots, a_{(2k_0 j+2k_0-1)m+1}, \epsilon_{(2k_0 j+2k_0-1)m}, \ldots, \epsilon_{i-m}).$$

Further, define the blocks as following,

$$
\begin{aligned}
F_{4j+1}(\bar{a}_{2k_0j}) &= \sum_{i=2k_0jm+1}^{(2k_0j+1)m} \tilde{X}_i(\bar{a}_{2k_0j}), \\
F_{4j+2} &= \sum_{i=(2k_0j+1)m+1}^{(2k_0j+k_0)m} \tilde{X}_i, \quad F_{4j+3} = \sum_{i=(2k_0j+k_0)m+1}^{(2k_0j+2k_0-1)m} \tilde{X}_i, \\
F_{4j+4}(\bar{a}_{2k_0j+2k_0}) &= \sum_{i=(2k_0j+2k_0-1)m+1}^{(2k_0j+2k_0)m} \tilde{X}_i(\bar{a}_{2k_0j+2k_0}).
\end{aligned} \tag{2.44}
$$

Similarly, for $j = 1, \ldots, q$, define

$$
\bar{\vartheta}_{2k_0j} = \{\epsilon_{(2k_0j-1)m+1}, \ldots, \epsilon_{2k_0jm}\} \text{ and } \vartheta = \{\ldots, \bar{\vartheta}_0, \bar{\vartheta}_{2k_0}, \bar{\vartheta}_{4k_0}, \ldots\}.
$$

Recall $A_j$ from (2.25). We have

$$
A_{j+1} = F_{4j+1}(\bar{\vartheta}_{2k_0j}) + F_{4j+2} + F_{4j+3} + F_{4j+4}(\bar{\vartheta}_{2k_0j+2k_0}).
$$

Define the mean functions

$$
\Lambda_{4j+1}(\bar{a}_{2k_0j}) = E^*(F_{4j+1}(\bar{a}_{2k_0j})) \text{ and } \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}) = E^*(F_{4j+4}(\bar{a}_{2k_0j+2k_0})),
$$

where $E^*$ refers to the conditional moment given $a$. In the sequel, with slight abuse of notation, we will simply use the usual $E$ to denote moments of random variables conditioned on $a$. Introduce the centered process

$$
\begin{aligned}
Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}) &= F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}) + F_{4j+2} \\
&\quad + F_{4j+3} + F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}).
\end{aligned} \tag{2.45}
$$

Following the definition of $S_n^{\diamond}$, we let

$$S_i(a) = \sum_{j=0}^{q_i-1} Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}).$$

The mean and variance function of $S_i(a)$ are respectively denoted by

$$\begin{aligned} M_i(a) &= \sum_{j=0}^{q_i-1} [\Lambda_{4j+1}(\bar{a}_{2k_0j}) + \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0})], \\ Q_i(a) &= \sum_{j=0}^{q_i-1} V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}), \end{aligned}$$

where $V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$ is the dispersion matrix of $Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$. Define

$$\begin{aligned} V_{j0}(\bar{a}_{2k_0j}) &= E(F_{4j-2}F_{4j-1}^{\top} + F_{4j-1}F_{4j-2}^{\top}) + Var(F_{4j-1} + F_{4j}(\bar{a}_{2k_0j}) - \Lambda_{4j}(\bar{a}_{2k_0j})) \\ &\quad + Var(F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}) + F_{4j+2}). \end{aligned} \tag{2.46}$$

Note that, the following identity holds for all $t$:

$$\sum_{j=0}^{t} V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}) = L(\bar{a}_0) + \sum_{j=1}^{t-1} V_{j0}(\bar{a}_{2k_0j}) + U_t(\bar{a}_{2k_0t+2k_0}), \tag{2.47}$$

where $L(\bar{a}_0) = Var(F_1(\bar{a}_0) + F_2)$ and

$$U_{t-1}(\bar{a}_{2k_0t}) = E(F_{4t-2}F_{4t-1}^{\top} + F_{4t-1}F_{4t-2}^{\top}) + Var(F_{4t-1} + F_{4t}(\bar{a}_{2k_0t}) - \Lambda_{4t}(\bar{a}_{2k_0t})). \tag{2.48}$$

Define

$$L_{\gamma}^{a} = \sum_{j=0}^{q-1} E(|Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})|^{\gamma}).$$

In the sequel, we suppress $Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$, $Y_j(\bar{\vartheta}_{2k_0j}, \bar{\vartheta}_{2k_0j+2k_0})$, $V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$, $V_{j0}(\bar{a}_{2k_0j})$, $V_j(\bar{\vartheta}_{2k_0j}, \bar{\vartheta}_{2k_0j+2k_0})$ and $V_{j0}(\bar{\vartheta}_{2k_0j})$ as just $Y_j^a$,$Y_j^{\vartheta}$, $V_j^a$, $V_{j0}^a$, $V_j^{\vartheta}$ and $V_{j0}^{\vartheta}$ respectively. We apply Proposition 2.7.1 to the independent mean 0 random vectors $Y_j^a$. We need to find a suitable sequence $\eta_k$ that allows us to get constants $C_1, C_2$ in (2.70) and $C_3$ in (2.71). There are roughly $q = n/(2k_0 m)$ many $Y_j^a$ random variables. Define

$$l = \lfloor q^{2/\gamma} / \log^2 q \rfloor. \tag{2.49}$$

To apply Proposition 2.7.1, we choose the sequence $\eta_k = kl$ and $s \asymp q/l$. This choice is justified by proving the following series of propositions.

**Proposition 2.6.1.** *Recall $\lambda_*$ and $A_j$ from (2.B) and (2.25) respectively. There exists a constant $\delta > 0$ such that*

$$2(\lambda_* + \delta)k_0 m \le \rho_*(Var(A_j)) \le \rho^*(Var(A_j)) \le \|A_j\|^2 \le 2k_0 m \Theta_{0,2}^2.$$

**Proposition 2.6.2.** *We can get positive constants $c_1$ and $c_2$ such that for all $j$,*

$$c_1 m \le \rho_*(Var(Y_j^\vartheta)) \le \rho^*(Var(Y_j^\vartheta)) \le E(|Y_j^\vartheta|^2) \le c_2 m. \tag{2.50}$$

**Proposition 2.6.3.** *For $l$ in (2.49), there exists constant $c_3$ such that,*

$$P\left(\max_{1 \le t \le q/l} |Var\left(\sum_{j=(t-1)l}^{tl-1} Y_j^a\right) - E\left(Var\left(\sum_{j=(t-1)l}^{tl-1} Y_j^a\right)\right)| \ge c_3 lm\right) \to 0.$$

**Proposition 2.6.4.** *We can get constants $c_4$ and $c_5$ such that*

$$P(c_4 q^{2/\gamma} m \le (L_\gamma^a)^{2/\gamma} \le c_5 q^{2/\gamma} m) \to 1.$$

**Proposition 2.6.5.** *Choose $\eta_k = kl$ with $l$ being defined in (2.49). Then we can get $C_1$ and $C_2$ such that (2.70) is satisfied. Moreover, with $l$ in (2.49), we can get $C_3$ such that (2.71) holds.*

Thus, we use Proposition 2.7.1 to construct $d$-variate mean 0 normal random vectors $N_j^a$ and random vectors $E_j^a$ such that

$$E_j^a \stackrel{D}{=} Y_j^a \text{ and } Var(N_j^a) = Var(Y_j^a), \quad 0 \le j \le q-1,$$

$$P_a\left(\max_{1\le i\le n} |\Pi_i^a - D_i^a| \ge c_0 z\right) \le C\frac{L_\gamma^a}{z^\gamma}, \text{ where } \Pi_i^a = \sum_{j=0}^{q_i-1} E_j^a,\ D_i^a = \sum_{j=0}^{q_i-1} N_j^a \tag{2.51}$$

and $C$ is a constant depending on $\gamma, c_1, \ldots, c_5$ and $C_3$. These constants are free of $a$. We can create a set $\mathcal{A}$ with $P(\mathcal{A}) \to 1$ so that $a \in \mathcal{A}$ implies the statements in Proposition 2.6.4 and Proposition 2.6.3 hold. Putting $z = n^{1/r}$ above in (2.51), by Lemma 2.7.4 and the restriction (2.33), we have, as $n \to \infty$,

$$E(L_\gamma^a n^{-\gamma/r}) \le \frac{q}{n^{\gamma/r}} c_\gamma \max_c E(|\tilde{R}_{c,2k_0m}|^\gamma) = O(n^{1-\gamma/r} m^{\gamma/2-1}) \to 0, \tag{2.52}$$

using

$$E(|Y_j(\bar{\vartheta}_{2k_0j}, \bar{\vartheta}_{2k_0j+2k_0})|^\gamma) \le c_\gamma \max_c E(|\tilde{R}_{c,2k_0m}|^\gamma) = O(m^{\gamma/2}).$$

Hence, conditioning on whether $a$ lies in $\mathcal{A}$ or not, from (2.52) we obtain,

$$\max_{i\le n} |\Pi_i^\vartheta - D_i^\vartheta| = o_P(n^{1/r}). \tag{2.53}$$

### *2.6.3 Unconditional Gaussian approximation and Regrouping:*

Here we shall work with the processes $\Pi_i^\vartheta, \mu_i^\vartheta$ and $D_i^\vartheta$. Note that, $V_{j0}(\bar{a}_{2k_0j})$ defined in (2.46) is a function of $\vartheta$ and might not be positive definite in an uniform fashion. For a constant

$0 < \delta_* < \lambda_*$, let

$$V_{j1}(\bar{a}_{2k_0j}) = \begin{cases} V_{j0}(\bar{a}_{2k_0j}) & \text{if } \rho_*(V_{j0}^a) \geq \delta_* m, \\ (\delta_* m) I_d & \text{otherwise}, \end{cases} \tag{2.54}$$

which is a positive-definitized version of $V_{j0}(\bar{a}_{2k_0j})$. The following proposition shows that partial sums of $V_{j0}(\bar{a}_{2k_0j})$ and $V_{j1}(\bar{a}_{2k_0j})$ are close to each other.

**Proposition 2.6.6.** *For some $\iota > 0$, we have*

$$\max_{i \leq n} E\left(\left|\sum_{j=1}^{max(1,q_i-1)} (V_{j0}(\bar{a}_{2k_0j}) - V_{j1}(\bar{a}_{2k_0j}))\right|\right) = o_P(n^{2/r-\iota}).$$

Henceforth in the sequel we will slightly abuse $max(1, q_i - 1) = max(1, \lfloor i/(2k_0m) \rfloor - 1)$ and simply use $q_i - 1 = \lfloor i/(2k_0m) \rfloor - 1$ for presentational clarity.

*Proof.* of Proposition 2.6.6. Recall (2.44) for the definition of $F_{4j+1}(.), F_{4j+2}$ etc. Define

$$F_{21} = \sum_{i=m+1}^{2m} \tilde{X}_i.$$

Define the projection operator $P_i$ by

$$P_i Y = E(Y|\mathcal{F}_i) - E(Y|\mathcal{F}_{i-1}), \quad Y \in \mathcal{L}_1.$$

For $1 \leq j \leq m$, $\|P_j F_{21}\| \leq \sum_{i=m+1-j}^{m} \delta_{i,2}$. Since $\|E(F_{21}^\top|\mathcal{F}_m)\|^2 = \sum_{j=1}^{m} \|P_j F_{21}\|^2$, we have

$$\begin{aligned} |E(F_1(\bar{a}_0)F_2^\top)| &= |E(F_1(\bar{a}_0)F_{21}^\top)| = |E(F_1(\bar{a}_0)E(F_{21}^\top|\mathcal{F}_m))| \\ &\leq \|F_1(\bar{a}_0)\|(\sum_{j=1}^{m}(\sum_{i=m+1-j}^{m} \delta_{i,2})^2)^{1/2}. \end{aligned} \tag{2.55}$$

Under the decay condition on $\Theta_{i,p}$ in (2.7), we have

$$E(|E(F_1(\bar{a}_0)F_{21}^\top)|^\gamma) = O(m^{\max(\gamma/2, \gamma-\chi\gamma)}).$$

We expand the last term of $V_{j0}(\bar{a}_{2k_0j})$ (see (2.46)). Also note that,

$$|E(F_{4j-2}F_{4j-1}^\top) + E(F_{4j-1}F_{4j-2}^\top)| \ll m \text{ and } \rho_*(Var(F_{4j+2})) \geq (k_0-1)\lambda_* m.$$

Then Proposition 2.6.6 follows from the fact that our solution of $\gamma$ from (2.32), (2.33), and (2.34) satisfy $\gamma > \max(2, 4\chi)$ for $\chi \leq \chi_0$ and

$$\begin{aligned} n \max_j P\left(\rho_*(V_{j0}^a) < \delta_* m\right) &\leq 2n \max_j P(|E(F_{4j+1}(\bar{a}_{2k_0j})F_{4j+2}^\top)| \geq -\theta m/2) \\ &= O(n)\frac{m^{\max(\gamma/2, \gamma-\chi\gamma)}}{m^\gamma} = o(n^{2/r-\iota}), \end{aligned}$$

for some $\iota > 0$ since we can choose $\delta_*$ such that $\theta = (k_0-1)\lambda_* - \delta_* > 0$. □

Recall (2.48) for the definition of $U_j$. By Lemma 2.7.4 and Jensen's inequality, we obtain $\max_j \|U_j(\bar{\vartheta}_{2k_0j+2k_0})\|_{\gamma/2} = O(m^{1/2})$. By (2.33), $\phi_n := q^{1/\gamma}m^{1/2}n^{-1/r} \to 0$. Then

$$\begin{aligned} P\left(\max_{0\leq j\leq q-1} |U_j(\bar{\vartheta}_{2k_0j+2k_0})| \geq \phi_n n^{2/r}\right) &\leq \sum_{j=0}^{q-1} P\left(|U_j(\bar{\vartheta}_{2k_0j+2k_0})| \geq \phi_n n^{2/r}\right) \\ &= O(\phi_n^{-\gamma/2} n^{1-\gamma/r} m^{\gamma/2-1}) = O(\phi_n^{\gamma/2}) \to 0. \end{aligned}$$

Similarly, $|L(\bar{\vartheta}_0)| = o_P(n^{2/r})$. Thus, by (2.47) and Proposition 2.6.6, since $Var(Y_j^a) = Var(N_j^a)$, one can construct i.i.d. $N(0, I_d)$ normal vectors $Z_l^a, l \in \mathbb{Z}$, such that

$$\max_{i\leq n} |D_i^\vartheta - \varsigma_i(\vartheta)| = o_P(n^{1/r}), \text{ where } \varsigma_i(a) = \sum_{j=1}^{q_i-1} V_{j1}^0(\bar{a}_{2k_0j})^{1/2} Z_j^a.$$

By (2.53), we have

$$\max_{i\le n}|\Pi_i^{\vartheta}-\varsigma_i(\vartheta)|=o_P(n^{1/r}).$$

Let $Z_l^*, l\in\mathbb{Z}$, independent of $(\epsilon_j)_{j\in\mathbb{Z}}$, be i.i.d. $N(0,I_d)$ and define

$$\Psi_i=\sum_{j=1}^{q_i-1}V_{j1}(\bar{\vartheta}_{2k_0j})^{1/2}Z_j^*.$$

From the distributional equality,

$$(\Pi_i^{\vartheta}+M_i(\vartheta))_{1\le i\le n}\overset{D}{=}(S_i^{\diamond})_{1\le i\le n}, \tag{2.56}$$

we need to prove Gaussian approximation for the process $\Psi_i+M_i(\vartheta)$. Define

$$B_j=V_{j1}(\bar{\vartheta}_{2k_0j})^{1/2}Z_j^*+\Lambda_{4j}(\bar{\vartheta}_{2k_0j})+\Lambda_{4j+1}(\bar{\vartheta}_{2k_0j}),$$

which are independent random vectors for $j=1,\dots,q$ and let

$$S_i^{\sharp}=\sum_{j=1}^{q_i-1}B_j \text{ and } W_i^{\sharp}=\Psi_i+M_i(\vartheta)-S_i^{\sharp}.$$

Note that,

$$\max_{i\le n}|W_i^{\sharp}|=\max_{i\le n}|\Lambda_{4q_i}(\vartheta_{2k_0q_i})+\Lambda_1(\vartheta_0)|=o_P(n^{1/r}). \tag{2.57}$$

Conditions (2.70) and (2.71) can be verified easily with this unconditional process $(S)_i^{\sharp}$ to use the Proposition 2.7.1. Thus, there exists $B_j^{new}$ and Gaussian random variable $B_j^{gau}$, such that $(B_j^{new})_{j\le q-1}\overset{D}{=}(B_j)_{j\le q-1}$ and corresponding $B_j^{gau}\sim N(0,Var(B_j))$, such that

$$\max_{i\le n}|\sum_{j=1}^{\lfloor i/2k_0m\rfloor-1}B_j^{new}-\sum_{j=1}^{\lfloor i/2k_0m\rfloor-1}B_j^{gau}| \;=\; o_P(n^{1/r}). \tag{2.58}$$

By (2.30), (2.56), (2.57) and (2.58), we can construct a process $S_i^c$ and $B_j^c$ such that $(S_i^c)_{i\le n} \overset{D}{=} (S_i)_{i\le n}$ and $(B_j^c)_{j\le q-1} \overset{D}{=} (B_j^{gau})_{j\le q-1}$ and

$$\max_{i\le n} |S_i^c - \sum_{j=1}^{\lfloor i/(2k_0 m)\rfloor -1} B_j^c| \;=\; o_P(n^{1/r}). \tag{2.59}$$

Relabel this final Gaussian process as

$$G_i^c = \sum_{j=1}^{\lfloor i/2k_0 m\rfloor -1} (Var(B_j))^{1/2} Y_j^c,$$

where $Y_j^c$ are i.i.d. $N(0, I_d)$. This concludes the proof of Theorem 2.2.1. □

*Proof.* of Proposition 2.6.1. Without loss of generality, we prove it for $j = 1$. Note that

$$2k_0 m\lambda_* \le \rho_*(Var(S_{2k_0 m})) \le \rho^*(Var(S_{2k_0 m})) \le \| \sum_{i=1}^{2k_0 m} X_i\|^2 \le 2k_0 m\Theta_{0,2}^2. \tag{2.60}$$

Recall $X_i^{\oplus}$ and $\tilde{X}_i$ from (2.41) and (2.21). The same upper bound works for $S_i^{\oplus}$ and $\tilde{S}_i$. Note that, $\|S_{2k_0 m}^{\oplus} - S_{2k_0 m}\| = o(m)$ and from [50], we have

$$\|A_1 - S_{2k_0 m}^{\oplus}\| = O(\sqrt{2k_0 m}\Theta_{m,2}) = o(\sqrt{2k_0 m}).$$

This concludes the proof using the Cauchy-Schwartz inequality. □

*Proof.* of Proposition 2.6.2. As $A_j$ is the block sum of the $m$-dependent processes with length $2k_0 m$, we have, using (2.60), for all $j$,

$$2k_0 m(\lambda_* + \delta) \le E(|A_j|^2) \le 2k_0 m\Theta_{0,2}^2,$$

for some small $\delta > 0$. We conclude the proof by using

$$|E(|Y_j^{\vartheta}|^2) - E(|A_{j+1}|^2)| = |\Lambda_{4j+1}(\bar{\vartheta}_{2k_0j})|^2 + |\Lambda_{4j+4}(\bar{\vartheta}_{2k_0j+2k_0})|^2 \leq 2m\Theta_{0,2}^2$$

and $k_0 > \Theta_{0,2}^2/\lambda_* + 1$. Using similar arguments, (2.50) follows. □

*Proof.* of Proposition 2.6.3. Note that, without loss of generality, we can assume $V_j^a$ to be independent for different $j$ since otherwise we can always break the probability statement in even and odd blocks and prove the statement separately. We use Corollary 1.6 and Corollary 1.7 from Nagaev (1979, [56]) respectively for the case $\gamma < 4$ and $\gamma \geq 4$ on $|V_j^a - E(V_j^a)|$ to deduce that it suffices to show the following

$$q \max_{1\leq t\leq q/l} \max_{t(l-1)+1\leq j\leq tl} P(|V_j^a - E(V_j^a)| \geq lm) \to 0. \tag{2.61}$$

We expand and write $V_j^a$ as follows:

$$\begin{aligned} V_j^a \;=\;& Var(F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j})) + Var(F_{4j+2} + F_{4j+3}) \qquad (2.62)\\ +\;& E((F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}))F_{4j+2}^{\mathsf{T}}) + E(F_{4j+2}(F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}))^{\mathsf{T}})\\ +\;& E(F_{4j+3}(F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}))^{\mathsf{T}})\\ +\;& E((F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}))F_{4j+3}^{\mathsf{T}})\\ +\;& Var(F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0})). \end{aligned}$$

Using derivation similar to (2.55), it suffices to show (2.61) for only the first and last term in (2.62). Moreover, we assume $d = 1$ and $j = 1$ to simplify notations. The proofs and the theorems used can be easily extended to vector-valued processes. Denote by $\tilde{S}_{m,\{j\}}$ for the sum $\tilde{S}_m$ with $\epsilon_j$ replaced by an i.i.d. copy $\epsilon_j'$. For the first term, by Burkholder's inequality,

$$\begin{aligned} E(|Var(F_1(\bar{a}_0)) - E(Var(F_1(\bar{a}_0)))|^{\gamma/2}) &= E(|E(\tilde{S}_m^2|a_{1-m},\ldots,a_0) - E(\tilde{S}_m^2)|^{\gamma/2}) \\ &= \|\sum_{j=-m}^{0} P_j\tilde{S}_m^2\|_{\gamma/2}^{\gamma/2} \le c_\gamma(\sum_{j=-m}^{0}\|P_j\tilde{S}_m^2\|_{\gamma/2}^2)^{\gamma/4} \end{aligned}$$

For $-m \le j \le 0$, $\|P_j\tilde{S}_m^2\|_{\gamma/2} \le \|\tilde{S}_m^2 - \tilde{S}_{m,\{j\}}^2\|_{\gamma/2} \le \|\tilde{S}_m - \tilde{S}_{m,\{j\}}\|_\gamma\|\tilde{S}_m + \tilde{S}_{m,\{j\}}\|_\gamma$. Note that $\|\tilde{S}_m\|_\gamma = O(m^{1/2})$ and $\|\tilde{S}_m - \tilde{S}_{m,\{j\}}\|_\gamma \le \sum_{r=1}^m \tilde{\delta}_{r-j,\gamma}$. By Lemma 2.7.3, $\tilde{\delta}_{k,\gamma} \le 2n^{1/p-1/\gamma}t_n^{1-p/\gamma}\delta_{k,p}^{p/\gamma}$. Then since $3 > 2(\chi+1)p/\gamma$ for $\chi \le \chi_0$, we have

$$\begin{aligned} \sum_{j=-m}^{0}\|P_j\tilde{S}_m^2\|_{\gamma/2}^2 &= O(m)\sum_{j=-m}^{0}\sum_{r=1}^{m}(\tilde{\delta}_{r-j,\gamma})^2 \\ &= O(m)n^{2/p-2/\gamma}t_n^{2-2p/\gamma}\sum_{j=0}^{m}(\sum_{r=1}^{m}\delta_{r+j,p}^{p/\gamma})^2 \\ &= O(m)n^{2/p-2/\gamma}t_n^{2-2p/\gamma}m^{3-2(\chi+1)p/\gamma}(\log m)^{-2Ap/\gamma}, \end{aligned} \tag{2.63}$$

by (2.7) and the Hölder inequality. Then, since $A > 2\gamma/p$ and $\log m \asymp \log q \asymp \log n$,

$$\begin{aligned} & qE(|Var(F_1(\bar{a}_0)) - E(Var(F_1(\bar{a}_0)))|^{\gamma/2}) \\ \lesssim\ & qm^{\gamma-(\chi+1)p/2}n^{\gamma/2p-1/2}t_n^{\gamma/2-p/2}(\log n)^{-Ap/2} = o((lm)^{\gamma/2}), \end{aligned} \tag{2.64}$$

using (2.40), (2.34) and the choice of $l$ in (2.49). For the last term in (2.62), we view $E(F_4(\bar{a}_{2k_0})^2)$ as

$$E(F_4(\bar{a}_{2k_0})^2) = E((\tilde{S}_{2k_0m} - \tilde{S}_{(2k_0-1)m})^2|a_{(2k_0-1)m+1},\ldots a_{2k_0m})$$

and show that it is close to $(\tilde{S}_{2k_0 m} - \tilde{S}_{(2k_0-1)m})^2$. Let $\mathcal{F}_j^m = (\epsilon_j, \ldots, \epsilon_m)$. Note that,

$$\begin{aligned}
\|\tilde{S}_m^2 - E(\tilde{S}_m^2 | a_m, \ldots, a_1)\|_{\gamma/2}^{\gamma/2} &\lesssim \Big(\sum_{j=-m-1}^{0} \|E(\tilde{S}_m^2|\mathcal{F}_j^m) - E(\tilde{S}_m^2|\mathcal{F}_{j+1}^m)\|_{\gamma/2}^2\Big)^{\gamma/4} \\
&\leq c m^{\gamma-(\chi+1)p/2} n^{\gamma/2p-1/2} t_n^{\gamma/2-p/2} (\log m)^{-Ap/2} \\
&= o(q^{-1}(lm)^{\gamma/2}),
\end{aligned} \tag{2.65}$$

similar to the derivation in (2.63). By (2.64) and (2.65), it suffices to show that

$$\frac{n}{m} P(|\tilde{S}_m| \geq \sqrt{lm}) \to 0. \tag{2.66}$$

Using the Nagaev-type inequality from Wu and Wu (2016, [79]) we obtain

$$P(|\tilde{S}_m| \geq \sqrt{lm}) \leq C_1 \frac{m^{\max\{1, p(1/2-\chi)\}}}{(lm)^{p/2}} + C_2 \exp(-C_3 l), \tag{2.67}$$

where $C_1, C_2$ and $C_3$ depend on $\chi$ and $p$. The second term in (2.67) is $o(m/n)$ since $e^{-l} \to 0$ very fast. For the first term in (2.67), if $\chi < 1/2 - 1/p$, then

$$\frac{n}{m} \frac{m^{p(1/2-\chi)}}{(lm)^{p/2}} = (\log n)^p n^{1-p/\gamma+L(p/\gamma-p\chi-1)} t_n^{k(p/\gamma-p\chi-1)} = o(1),$$

as from (2.34) we have $1 - p/\gamma + L(p/\gamma - p\chi - 1) = L(p/\gamma - 1)(\chi p + p + 1) < 0$. If $1/2 - 1/p \leq \chi < \chi_0$ and consequently $r < p$, then we have, for the first term in (2.67),

$$\frac{n}{m} \frac{m}{(lm)^{p/2}} = (\log n)^p n^{p(1/p-1/\gamma+L(1/\gamma-1/2))} t_n^{k(p/\gamma-p/2)} = o(1), \tag{2.68}$$

using (2.40), $r < p$ and the fact that $r$ satisfy $1/r - 1/\gamma + L(1/\gamma - 1/2) = 0$. □

*Proof.* of Proposition 2.6.4. By Lemma 2.7.4, $E(L_\gamma^a) \asymp q m^{\gamma/2}$. Then it suffices to prove

$$P(|L_\gamma^a - E(L_\gamma^a)| \geq c q m^{\gamma/2} / \log q) \to 0, \tag{2.69}$$

holds for some constant $c > 0$. Note that $E(|Y_j^a|^\gamma)$ are even indices $j$ (also for odd indices $j$). Thus we can prove the statement separately by breaking $L_\gamma^a$ in sum of even and odd $E(|Y_j^a|^\gamma)$. Without loss of generality, we assume all $E(|Y_j^a|^\gamma)$ are independent and proceed. Define $J_j = (2k_0m)^{-\gamma/2}E(|\tilde{S}_{2k_0mj} - \tilde{S}_{2k_0m(j-1)}|^\gamma|\bar{a}_{2k_0(j-1)}, \bar{a}_{2k_0j})$ and $\theta = l^{\gamma/2} = q/(\log q)^\gamma$. Recall the truncation operator $T$ from (2.19). Noting $E(J_j) = O(1)$ from Lemma 2.7.4, we have

$$P(|\sum_{j=1}^{q} T_\theta(J_j) - E(T_\theta(J_j))| \geq \phi) \leq \frac{q}{\phi^2}\max_j E(T_\theta(J_j)^2) = O(\theta q/\phi^2) = o(1),$$

where $\phi = q/\log q$, and

$$\max_j P(J_j \geq \theta) \leq \max_j P(E(|\tilde{S}_{2k_0mj} - \tilde{S}_{2k_0m(j-1)}|^2|\bar{a}_{2k_0(j-1)}, \bar{a}_{2k_0j}) \geq 2k_0lm) = o(q^{-1}),$$

from (2.64), (2.65) and (2.66). Thus $P(|\sum_{j=1}^q J_j - \sum_{j=1}^q E(J_j)| \geq \phi) \to 0$ which is a restatement of (2.69). □

*Proof.* of Proposition 2.6.5. We showed in Proposition 2.6.4 that

$$P(cqm^{\gamma/2} \leq L_\gamma \leq Cqm^{\gamma/2}) \to 1,$$

for some constants $c$ and $C$. Let $l$ be as given in (2.49). Let $S = \{0, l, 2l, \cdots\}$. Proposition 2.6.2 and Proposition 2.6.3 show that, for some constants $c$ and $C$,

$$P(clk_0m \leq \min_{i\in S}\rho_*\left(Var\left(\sum_{j=i}^{i+l-1} Y_j^a\right)\right) \leq \max_{i\in S}\rho^*\left(Var\left(\sum_{j=i}^{i+l-1} Y_j^a\right)\right) \leq Clk_0m) \to 1.$$

We choose $\eta_k = kl$ and $s \asymp q/l$. Starting with the conditional block sum process $Y_j^a$ for $0 \leq j \leq q-1$, this choice of $\eta_k$ satisfies (2.70) for a given $a$ with probability going to 1. The other condition, (2.71) can be easily verified for such a choice of $\eta$-sequence using ideas similar to the proof of Proposition 2.6.4. We skip the details of that derivation. □

## 2.7 Appendix C: Some Useful Results

Proposition 2.7.1 concerns Gaussian approximation for independent vectors. There are several types of Gaussian approximations in literature for independent vectors. We find the following result by Götze and Zaitsev (2008, [31]) particularly useful since it provides an explicit and good approximation bound for the partial sums. This has been used several times in our proof.

**Proposition 2.7.1.** *Let $\xi_1, \ldots, \xi_n$ be independent $\mathbb{R}^d$-valued mean 0 random vectors. Assume that there exist $s \in \mathbb{N}$ and a strictly increasing sequence of non-negative integers $\eta_0 = 0 < \eta_1 < \ldots < \eta_s = n$ satisfying the following conditions. Let*

$$\zeta_k = \xi_{\eta_{k-1}+1} + \ldots + \xi_{\eta_k}, \quad Var(\zeta_k) = B_k, \quad k = 1, \ldots, s$$

*and $L_\gamma = \sum_{j=1}^n E(|\xi_j|^\gamma)$, $\gamma \geq 2$, and assume that, for all $k = 1, \ldots, s$,*

$$C_1 w^2 \leq \rho_*(B_k) \leq \rho^*(B_k) \leq C_2 w^2, \tag{2.70}$$

*where $w = (L_\gamma)^{1/\gamma} / \log^* s$, with some positive constants $C_1$ and $C_2$. Suppose the quantities*

$$\lambda_{k,\gamma} = \sum_{j=\eta_{k-1}+1}^{\eta_k} E\|\xi_j\|^\gamma, \quad k = 1, \ldots s,$$

*satisfy, for some $0 < \epsilon < 1$ and constant $C_3$,*

$$C_3 d^{\gamma/2} s^\epsilon (\log^* s)^{\gamma+3} \max_{1 \leq k \leq s} \lambda_{k,\gamma} \leq L_\gamma. \tag{2.71}$$

*Then one can construct on a probability space independent random vectors $X_1, \ldots, X_n$ and a corresponding set of independent Gaussian vectors $Y_1, \ldots, Y_n$ so that $(X_j)_{j=1}^n \overset{D}{=} (\xi_j)_{j=1}^n$,*

*$E(Y_j) = 0$, $Var(Y_j) = Var(X_j), 1 \le j \le n$, and for any $z > 0$,*

$$P\left(\max_{t\le n} |\sum_{i=1}^{t} X_i - \sum_{i=1}^{t} Y_i| \ge z\right) \le C_* L_\gamma z^{-\gamma}.$$

*where $C_*$ is a constant that depends on $d, \gamma, C_1, C_2$ and $C_3$.*

**Lemma 2.7.2.** *Let $p < \gamma$. Assume (2.A). Then $\sup_i E \min\{|X_i|^\gamma n^{-\gamma/p}, 1\} = o(n^{-1})$.*

*Proof.* Choose $k_n = \lfloor 2(\log n)/((p+\gamma)\log 2)\rfloor$. Then $n = o(2^{\gamma k_n})$ and $2^{pk_n} = o(n)$. Let $Z = |X_i| n^{-1/p}$. The lemma follows from

$$\begin{aligned} E(\min\{Z^\gamma, 1\}) &\le P(Z \ge 1) + \sum_{k=0}^{k_n} 2^{-k\gamma} P(2^{-1-k} \le Z < 2^{-k}) + 2^{-\gamma(k_n+1)} \\ &\le E(Z^p \mathbf{1}_{Z\ge 1}) + \sum_{k=0}^{k_n} 2^{p(k+1)-k\gamma} E(Z^p \mathbf{1}_{Z \ge 2^{-1-k}}) + 2^{-\gamma(k_n+1)} = o(n^{-1}), \end{aligned}$$

in view of the uniform integrability condition (2.A) and $n^{1/2}/2^{k_n} \to \infty$. □

**Lemma 2.7.3.** *The functional dependence measures defined on the truncated process $(X_i^\oplus)$ and the m-dependent process $(\tilde{X}_i)$, satisfy $\tilde{\delta}_{j,\gamma} \le \delta^\oplus_{j,\gamma} \le 2n^{1/p-1/\gamma} t_n^{1-p/\gamma} \delta_{j,p}^{p/\gamma}$.*

*Proof.* Since the truncation operator $T$ is Lipschitz continuous,

$$\begin{aligned} (\delta^\oplus_{j,\gamma})^\gamma &= \sup_i E(|T_{t_n n^{1/p}}(X_i) - T_{t_n n^{1/p}}(X_{i,(i-j)})|^\gamma) \\ &= n^{\gamma/p} t_n^\gamma \sup_i E\left(\left|\min\left(2, \left|\frac{X_i - X_{i,(i-j)}}{t_n n^{1/p}}\right|\right)\right|^\gamma\right) \le 2^\gamma n^{\gamma/p-1} t_n^{\gamma-p} \delta^p_{j,p}. \end{aligned}$$

The first inequality $\tilde{\delta}_{j,\gamma} \le \delta^\oplus_{j,\gamma}$ follows from (2.42). □

**Lemma 2.7.4.** *<u>Rosenthal Type Moment Bound</u> Recall (2.39) and (2.40) for $t_n$. Assume (2.23), (2.26), (2.27) along with (2.10) on A related to the restriction on $\Theta_{i,p}$ as mentioned in (2.7). Moreover, assume $m = \lfloor n^L t_n^k \rfloor$ with $k$ satisfying $k < (\gamma/2-1)^{-1}(\gamma - p)$. Then,* *we have,*

$$\max_t E(\max_{1\le l\le m} |\tilde{R}_{t,l}|^\gamma) = O(m^{\gamma/2}). \tag{2.72}$$

*Proof.* Since the functional dependence measure is defined in an uniform manner, we can ignore the $\max_t$ in (2.72) and use the Rosenthal-type inequality for stationary processes in Liu, Xiao and Wu (2013, [51]). By [51], there is a constant $c$, depending only on $\gamma$, such that

$$\begin{aligned} \| \max_{1\le l\le m} |\tilde{R}_{t,l}| \|_\gamma &\le cm^{1/2}[\sum_{j=1}^m \tilde{\delta}_{j,2} + \sum_{j=1+m}^\infty \tilde{\delta}_{j,\gamma} + \sup_i \|T_{t_n n^{1/p}}(X_i)\|] \\ &\quad + cm^{1/\gamma}[\sum_{j=1}^m j^{1/2-1/\gamma} \tilde{\delta}_{j,\gamma} + \sup_i \|T_{t_n n^{1/p}}(X_i)\|_\gamma] \\ &\le c(I + II + III + IV), \end{aligned}$$

where

$$\begin{aligned} I &= m^{1/2} \sum_{j=1}^m \tilde{\delta}_{j,2} + m^{1/2} \|X_1\|_2, \\ II &= m^{1/2} \sum_{j=m+1}^\infty \tilde{\delta}_{j,\gamma}, \quad III = m^{1/\gamma} \sum_{j=1}^\infty j^{1/2-1/\gamma} \tilde{\delta}_{j,\gamma}, \\ IV &= m^{1/\gamma} \sup_i \|T_{t_n n^{1/p}}(X_i)\|_\gamma. \end{aligned}$$

For the first term $I$, since $\sum_{j=1}^\infty \delta_{j,2} + \sup_i \|X_i\|_2 \le 2\Theta_{0,2}$ and $\tilde{\delta}_{j,2} \le \delta_{j,2}$, we have $I = O(m^{1/2})$. Starting with $II$, we apply Lemma 2.7.3 to obtain

$$II = m^{1/2} \sum_{j=m+1}^\infty \tilde{\delta}_{j,\gamma} \lesssim m^{1/2} n^{1/p-1/\gamma} t_n^{1-p/\gamma} \sum_{j=m+1}^\infty \delta_{j,p}^{p/\gamma}.$$

The rest follows from the derivation in (2.31) and (2.34). For the third term, we have

$$\begin{aligned} III &\lesssim m^{1/\gamma} n^{1/p-1/\gamma} t_n^{1-p/\gamma} \sum_{j=1}^{m} j^{1/2-1/\gamma} \delta_{j,p}^{p/\gamma} \\ &\leq m^{1/\gamma} n^{1/p-1/\gamma} t_n^{1-p/\gamma} \sum_{l=1}^{\lfloor \log_2 m \rfloor + 1} \sum_{j=2^{l-1}}^{2^l-1} j^{1/2-1/\gamma} \delta_{j,p}^{p/\gamma} \\ &\leq m^{1/\gamma} n^{1/p-1/\gamma} t_n^{1-p/\gamma} \sum_{l=1}^{\lfloor \log_2 m \rfloor + 1} 2^{l(3/2-1/\gamma-p/\gamma)} O(2^{-l\chi p/\gamma} l^{-Ap/\gamma}). \end{aligned} \tag{2.73}$$

Recall the definition of $\chi_0$ from (2.9). If $\chi \leq \chi_0$, then our solution for $\gamma$ satisfies

$$3/2 - 1/\gamma - (\chi + 1)p/\gamma \geq 0,$$

with equality holding only for $\chi = \chi_0$. Hence, if $\chi < \chi_0$, we have

$$m^{-1/2} III = m^{1-(\chi+1)p/\gamma} n^{1/p-1/\gamma} t_n^{1-p/\gamma} (\log n)^{-Ap/\gamma} O(1) = o(1),$$

from (2.34), (2.28) and (2.40). If $\chi = \chi_0$, since $A > \gamma/p$ from (2.10) [The lower bound for $A$ there is just $2\gamma/p$ as mentioned in (2.32)], we have

$$m^{-1/2} III = m^{1/\gamma-1/2} n^{1/p-1/\gamma} t_n^{1-p/\gamma} O(1) = o(1), \tag{2.74}$$

since (2.33) is true. Also for the case of $\chi > \chi_0$ in the proof of Theorem 2.2.2, the way we define our three conditions in (2.35) the new solution also satisfy $\gamma' = 2(1 + p + p\chi)/3$ and thus (2.74) holds. For the fourth term $IV$, in the light of (2.33), we use (2.39) to derive

$$\begin{aligned} m^{-\gamma/2} IV^{\gamma} = m^{1-\gamma/2} \sup_i \|T_{t_n n^{1/p}}(X_i)\|^{\gamma} &\leq m^{1-\gamma/2} t_n^{\gamma} n^{\gamma/p} \sup_i E\left(\min\{\frac{|X_i|^{\gamma}}{t_n^{\gamma} n^{\gamma/p}}, 1\}\right) \\ &= m^{1-\gamma/2} t_n^{\gamma} n^{\gamma/p-1} o(1) = o(1), \end{aligned}$$

$\square$

# CHAPTER 3

# SIMULTANEOUS INFERENCE ON TIME-VARYING MODELS

## 3.1 Introduction

Time-varying dynamical systems have been studied extensively in the literature of statistics, economics and related fields. For stochastic processes observed over a long time horizon, stationarity is often an over-simplified assumption that ignores systematic deviations of parameters from constancy. For example, in the context of financial datasets, empirical evidence shows that external factors such as war, terrorist attacks, economic crisis, some political event etc. introduce such parameter inconstancy. As Bai [4] points out, 'failure to take into account parameter changes, given their presence, may lead to incorrect policy implications and predictions'. Thus functional estimation of unknown parameter curves using time-varying models has become an important research topic recently. In this paper, we propose a general setting for simultaneous inference of local linear M-estimators in semi-parametric time-varying models. Our formulation is general enough to allow unifying time-varying models from the usual linear regression, generalized regression and several auto-regression type models together. Before discussing our new contributions in this paper, we provide a brief overview of some previous works in these areas.

In the regression context, time-varying models are discussed over the past two decades to describe non-constant relationships between the response and the predictors; see, for instance, Fan and Zhang [26], Fan and Zhang [27], Hoover et al. [33], Huang et al. [34], Lin and Ying [44], Ramsay and Silverman [62], Zhang et al. [88] among others. Consider the following two regression models

$$\text{Model I: } y_i = x_i^{\mathsf{T}}\theta_i + e_i, \quad \text{Model II: } y_i = x_i^{\mathsf{T}}\theta_0 + e_i, \qquad i = 1, \ldots, n,$$

where $x_i \in \mathbb{R}^d$ $(i = 1, \ldots, n)$ are the covariates, $^{\mathsf{T}}$ is the transpose, $\theta_0$ and $\theta_i = \theta(i/n)$ are the regression coefficients. Here, $\theta_0$ is a constant parameter and $\theta : [0,1] \to \mathbb{R}^d$ is a smooth function. Estimation of $\theta(\cdot)$ has been considered by Hoover et al. [33], Cai [9]) and Zhou and Wu [91] among others. Hypothesis testing is widely used to choose between model I and model II, see, for instance, Zhang and Wu [86], Zhang and Wu [87], Chow [11], Brown et al. [8], Nabeya and Tanaka [55], Leybourne and McCabe [42], Nyblom [57], Ploberger et al. [59], Andrews [2] and Lin and Teräsvirta [43]. Zhou and Wu [91] discussed obtaining simultaneous confidence bands (SCB) in model I, i.e. with additive errors. However their treatment is heavily based on the closed-form solution and it does not extend to processes defined by a more general recursion. Our framework allows us to perform inference on a much larger class of regression settings. Moreover, it can also accommodate generalized linear models as shown in Section 3.5. Little has been known for time-varying models in this direction previously.

The results from time-varying linear regression can be naturally extended to time-varying AR, MA or ARMA processes. However, such an extension is not obvious for conditional heteroscedastic (CH) models. These are difficult to estimate but also often more useful in analyzing and predicting financial datasets. Since Engle [23] introduced the classical ARCH model and Bollerslev [7] extended it to a more general GARCH model, these have remained primary tools for analyzing and forecasting certain trends for stock market datasets. As the market is vulnerable to frequent changes, non-uniformity across time is a natural phenomenon. The necessity of extending these classical models to a set-up where the parameters can change across time has been pointed out in several references; for example Stărică and Granger [68], Engle and Rangel [24] and Fryzlewicz et al. [29]. Towards time-varying parameter models in the CH setting, numerous works discussed the CUSUM-type procedure, for instance, Kim et al. [36] for testing change in parameters of GARCH(1,1). Kulperger et al. [41] studied the high moment partial sum process based on residuals and applied it to residual CUSUM tests in GARCH models. Interested readers can find some more changepoint detection results in the context of CH models in James Chu [35], Chen and Gupta [10], Lin et al. [45], Kokoszka et al. [37] or Andreou and Ghysels [1].

Historically in the analysis of financial datasets, the common practice to account for the time-varying nature of the parameter curves was to transfer a stationary tool/method in some ad hoc way. For example, in Mikosch and Stărică [53], the authors analyzed S&P500 data from 1953-1990 and suggested that time-varying parameters are more suitable due to such a long time-horizon. They re-estimated the parameters for every block of 100 sample points and to account for the abrupt fluctuation of the coefficients, they generated re-estimates of parameters for samples of size $100, 200, \ldots$. This treatment suffers from different degree of reliability of the estimators at different parts of the time horizon. There are examples outside the analysis of economic datasets, where similar approach of splitting the time-horizon has been adapted to fit CH type models. For example, in Giacometti et al. [30], the authors analyzed Italian mortality rates from 1960-2003 using an AR(1)-ARCH(1) model and observed abrupt behavior of yearwise coefficients. Our framework can simultaneously capture these models and provide significant improvements over such heuristic treatments.

A time-varying framework and a pointwise curve estimation using M-estimators for locally stationary ARCH models was provided by Dahlhaus and Subba Rao [15]. Since then, while several pointwise approaches were discussed in the tvARMA and tvARCH case (cf. Dahlhaus and Polonik [14], Dahlhaus and Subba Rao [15], Fryzlewicz et al. [29]), pointwise theoretical results for estimation in tvGARCH processes were discussed in Rohan and Ramanathan [65] and Rohan [64] for GARCH(1,1) and GARCH($p$,$q$) models.

The goals of this paper are twofold. We provide a unifying framework that binds linear regression models, generalized regression models and many popularly used auto-regressive models including CH type processes. Moreover, we use Bahadur representations, a Gaussian approximation theorem from Zhou and Wu [89] and extreme value Gaussian theory to obtain SCBs for contrasts of the parameter curves. These intervals provide a generalization from testing parameter constancy to testing any particular parametric form such as linear, quadratic, exponential etc. A very general recursion model (cf. (3.1)) is considered and asymptotic results for a local linear M-estimator are provided. To deal with bias expansions, we use the theory about derivative processes which was recently formalized in Dahlhaus et al. [13].

The rest of the article is organized as follows. In Section 3.2, we introduce our framework, the functional dependence measure, the assumptions and the M-estimators of the parameter curves. Section 3.3 consists of the results about the Bahadur representation and the SCBs of the related contrasts. Section 3.4 is dedicated to practical issues which arise when using the SCBs like estimation of the dispersion matrix of the estimator, bandwidth selection and a wild Bootstrap procedure to overcome the slow logarithmic convergence from the theoretical SCB. We discuss some examples to show the general applicability of our framework in Section 3.5. Some summarized simulation studies and real data applications can be found in Section 3.6. We defer all the proofs to the end.

## 3.2 Model assumptions and estimators

### *3.2.1 The model*

For some known family of real-valued (possibly stochastic) functions $F_i$, we consider the model with time-varying parameter curve

$$Y_i = F_i(X_i, \theta(i/n)), \quad i = 1, \ldots, n, \tag{3.1}$$

where $n$ is the number of observations, $X_i = (X_{ij})_{j\in\mathbb{N}}$ and $Y_i$ represent a possibly infinite-dimensional covariate process and the real-valued response process respectively. Here $\theta : [0,1] \to \Theta \subset \mathbb{R}^{d_\Theta}$ is a time varying parameter curve. To cover important time series models, we assume that not $X_i$ itself but some truncated version $X_i^c = (X_{ij}^c)_{j\in\mathbb{N}}$ is observed. We assume that both $(Y_i)$ and $(X_i)$ are locally stationary processes in the following sense: Let $\zeta_i$, $i \in \mathbb{Z}$ be independent and identically distributed random variables and $\mathcal{F}_i := (\ldots, \zeta_{i-1}, \zeta_i)$.

We assume the following form for $Y_i$ and $X_i$

$$X_i = G_i(\mathcal{F}_i), \qquad Y_i = H_i(\mathcal{F}_i), \quad i = 1, \ldots, n, \tag{3.2}$$

where $G_i(\cdot) = (G_{ij}(\cdot))_{j \in \mathbb{N}}$ and $H_i(\cdot)$ are measurable functions.

It is worth noting that we do not necessarily need the representation (3.1) as it is only needed in an optional condition (3.13). Some more general formulations may still fit in the setting of this paper. There are some important special cases of (3.1):

(a) Time-varying time series models: Assume that, $(\varepsilon_i)_{i \in \mathbb{Z}}$ are i.i.d., choose $\zeta_i = \varepsilon_i$ and $X_i = (Y_{i-1}, Y_{i-2}, \ldots)$. Then (3.1) translates to

$$Y_i = F((Y_{i-1}, Y_{i-2}, \ldots), \theta(i/n), \varepsilon_i),$$

which for instance covers tvARMA, tvARCH, tvGARCH processes. In this context, one usually has $X_i^c = (Y_{i-1}, \ldots, Y_1, 0, 0, \ldots)$ since only $Y_1, \ldots, Y_n$ are observed.

(b) The generalized linear model: By using $F_i(x, \theta) = g_i(x^\top \theta)$, where $g_i : \mathbb{R} \to \mathbb{R}$ serves as a (probably stochastic) link function, (3.1) has the form

$$Y_i = g_i(X_i^\top \theta(i/n)).$$

An important example is logistic regression which is assumed to be time-varying in the following sense:

$$Y_i \sim Bin(m, \pi_i), \quad \log\big(\frac{\pi_i}{1 - \pi_i}\big) = X_i^\top \theta(i/n),$$

where $Y_i$ could possibly be lagged values of $X_i$ as well. Such autoregressive logistic models are commonly used in conjunction with longitudinal data from several scientific research fields. For medical research and biology, see de Vries et al. [16], Kowsar et al. [40] etc; for climatology, see Guanche et al. [32]; for risk management analysis see

Taylor and Yu [71] etc.

In either case, our goal is to estimate $\theta(\cdot)$ from the observations $Z_i^c = (Y_i, X_i^c)$, $i = 1, \ldots, n$.

### 3.2.2 The estimator

In this paper, we focus on local M-estimation: Let $K(\cdot) \in \mathcal{K}$, where $\mathcal{K}$ is the family of non-negative symmetric kernels with support $[-1, 1]$ which are continuously differentiable on $[-1, 1]$ such that $\int_{-1}^{1} |K'(u)|^2 du > 0$. Let $\ell(z, \theta)$ be an objective function. A usual choice is the negative log conditional (Gaussian) likelihood of the model which leads to a minimum distance estimator. Define the local linear likelihood function

$$L_{n,b_n}^c(t, \theta, \theta') := (nb_n)^{-1} \sum_{i=1}^{n} K_{b_n}(t - i/n) \ell(Z_i^c, \theta + \theta'(i/n - t)), \tag{3.3}$$

where $K_{b_n}(\cdot) := K(\cdot/b_n)$. Let $\Theta' := [-R, R]^k$ with some $R > 0$. A local linear estimator of $\theta(t)$, $\theta'(t)$ is given by

$$(\hat{\theta}_{b_n}(t), \widehat{\theta'}_{b_n}(t)) = \arg \min_{(\theta,\theta') \in \Theta \times \Theta'} L_{n,b_n}^c(t, \theta, \theta'), \qquad t \in [0, 1]. \tag{3.4}$$

In Examples 3.5.1 and 3.5.2, we discuss applications and choices of $\ell$ for general recursively defined locally stationary time series models and tvGARCH processes. In Example 3.5.3, we consider a time-varying logistic regression model with a Binomial likelihood function $\ell$.

### 3.2.3 The functional dependence measure

To state the structure of dependence we use throughout the paper, we introduce a functional dependence measure on the underlying process using the idea of coupling as done in Wu [75]. Assume that a stationary process $Z_i$ has mean 0, $Z_i \in \mathcal{L}_q, q > 0$ and it admits the causal representation

$$Z_i = J(\zeta_i, \zeta_{i-1}, \dots). \tag{3.5}$$

Suppose that $(\zeta_i^*)_{i\in\mathbb{Z}}$ is an independent copy of $(\zeta_i)_{i\in\mathbb{Z}}$. For some random variable $Z$, let $\|Z\|_q := (\mathbb{E}|Z|^q)^{1/q}$ denote the $\mathcal{L}_q$-norm of $Z$. For $j \geq 0$, define the functional dependence measure

$$\delta_q^Z(i) = \|Z_i - Z_i^*\|_q, \tag{3.6}$$

where $\mathcal{F}_i^*$ is a coupled version of $\mathcal{F}_i$ with $\zeta_0$ in $\mathcal{F}_i$ replaced by $\zeta_0^*$,

$$\mathcal{F}^* = (\zeta_i, \zeta_{i-1}, \cdots, \zeta_1, \zeta_0^*, \zeta_{-1}, \zeta_{-2}, \cdots), \tag{3.7}$$

and $Z_i^* = J(\mathcal{F}_i^*)$. Note that $\delta_q^Z(i)$ measures the dependence of $Z_i$ on $\zeta_0$ in terms of the $q$th moment. The tail cumulative dependence measure $\Delta_q^Z(j)$ for $j \geq 0$ is defined as

$$\Delta_q^Z(j) = \sum_{i=j}^{\infty} \delta_q^Z(i). \tag{3.8}$$

### *3.2.4 The class $\mathcal{H}(M_y, M_x, \chi, \bar{C})$*

To prove uniform convergence of $L_{n,b_n}^c$ and its derivatives w.r.t. $\theta$, we require $\ell$ to be Lipschitz continuous in direction of $\theta$ and to grow at most polynomially in direction of $z = (y, x)$, where the degree is measured by integers $M_y, M_x \geq 1$. We will therefore ask $\ell$ and its derivatives to be in the class $\mathcal{H}(M_y, M_x, \chi, \bar{C})$ which is defined as follows: Let $\chi = (\chi_i)_{i=1,2,\dots}$ be a sequence of nonnegative real numbers with $|\chi|_1 := \sum_{i=1}^{\infty} \chi_i < \infty$, and $\bar{C} > 0$ be some constant. Define $|x|_{\chi,1} := \sum_{i=1}^{\infty} \chi_i |x_i|$. Put $\hat{\chi} = (1, \chi)$, and for nonnegative integers $d_x, d_y$, define the 'polynomial rest'

$$R_{d_y,d_x}(z) := \sum_{\substack{k=0 \\ k+l\le \max\{d_x,d_y\}}}^{d_y} \sum_{l=0}^{d_x} |y|^k |x|_{\chi,1}^l.$$

A function $g : \mathbb{R}\times\mathbb{R}^{\mathbb{N}}\times\Theta \to \mathbb{R}$ is in $\mathcal{H}(M_y, M_x, \chi, \bar{C})$ if $\sup_{\theta\in\Theta}|g(0,\theta)| \le \bar{C}$,

$$\sup_z \sup_{\theta\neq\theta'} \frac{|g(z,\theta)-g(z,\theta')|}{|\theta-\theta'|_1 R_{M_y,M_x}(z)} \le \bar{C}$$

and

$$\sup_\theta \sup_{z\neq z'} \frac{|g(z,\theta)-g(z',\theta)|}{|z-z'|_{\hat{\chi},1}\cdot\{R_{M_y-1,M_x-1}(z)+R_{M_y-1,M_x-1}(z')\}} \le \bar{C}.$$

If $g$ is vector- or matrix-valued, $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ means that every component of $g$ is in $\mathcal{H}(M_y, M_x, \chi, \bar{C})$. In Section 3.5, we will see that a large class of log Gaussian likelihoods and the usual logistic regression likelihood belongs to $\mathcal{H}(M_y, M_x, \chi, \bar{C})$. In case of time series it often holds that $M = M_x = M_y$, which allows to use a simplified version $R_{M_y,M_x}(z) = 1 + |z|_{\hat{\chi},1}^M$.

### *3.2.5 Assumptions*

In this paper, we prove Bahadur representations and construct simultaneous confidence bands for $\hat{\theta}_{b_n}(\cdot)$ and $\widehat{\theta'_{b_n}}(\cdot)$. Clearly, more smoothness assumptions on $\theta(\cdot)$ and $\ell$ are needed to prove results for the latter one which is postponed to Assumption 3.2.2.

In the following, we will assume the existence of measurable functions $H, G$ such that $\tilde{Y}_i(t) = H(t,\mathcal{F}_i) \in \mathbb{R}$ and $\tilde{X}_i(t) = G(t,\mathcal{F}_i) \in \mathbb{R}^{\mathbb{N}}$ are well-defined for all $t \in [0,1]$. These processes will serve as stationary approximations of $Y_i$, $X_i$ if $|i/n - t| \ll 1$. For brevity, define $\tilde{Z}_i(t) := (\tilde{Y}_i(t), \tilde{X}_i(t)^{\mathsf{T}})^{\mathsf{T}}$ and $Z_i := (Y_i, X_i^{\mathsf{T}})^{\mathsf{T}}$. The constant $r \ge 2$ in the following assumption is connected to the number of moments that are assumed for $Z_i$ (cf. (A5) and

(A7)), while $\gamma > 1$ is a measure of decay of the dependence which is present in the model.

**Assumption 3.2.1.** *Suppose that for some $r \geq 2$ and some $\gamma > 1$,*

*(A1) (Smoothness in $\theta$-direction) $\ell$ is twice continuously differentiable w.r.t. $\theta$. It holds that $\ell, \nabla_\theta \ell, \nabla_\theta^2 \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ for some $M_y, M_x \geq 1$, $\bar{C} > 0$ and $\chi = (\chi_i)_{i=1,2,\ldots}$ with $\chi_i = O(i^{-(1+\gamma)})$.*

*(A2) (Assumptions on unknown parameter curve) $\Theta$ is compact and for all $t \in [0,1]$, $\theta(t)$ lies in the interior of $\Theta$. Each component of $\theta(\cdot)$ is in $C^3[0,1]$.*

*(A3) (Correct model specification) For all $t \in [0,1]$, the function $\theta \mapsto L(t,\theta) := \mathbb{E}\ell(\tilde{Z}_0(t), \theta)$ is uniquely minimized by $\theta(t)$.*

*(A4) The eigenvalues of the matrices*

$$\begin{aligned} V(t) &= \mathbb{E}\nabla_\theta^2 \ell(\tilde{Z}_0(t), \theta(t)), && (3.9)\\ I(t) &= \mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) \cdot \nabla_\theta \ell(\tilde{Z}_0(t), \theta(t))^{\mathsf{T}}], && (3.10)\\ \Lambda(t) &= \sum_{j\in\mathbb{Z}} \mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) \cdot \nabla_\theta \ell(\tilde{Z}_j(t), \theta(t))^{\mathsf{T}}], && (3.11) \end{aligned}$$

*are bounded from below by some $\lambda_0 > 0$, uniformly in $t$.*

*(A5) (Stationary approximation) Let $M = \max\{M_x, M_y\}$. There exist $C_A, C_B, D > 0$ such that for all $n \in \mathbb{N}$, $i = 1, \ldots, n$, $t, t' \in [0,1]$, $j \in \mathbb{N}$:*

$$\max\{\|Y_i\|_{rM},\ \|\tilde{Y}_0(t)\|_{rM},\ \|X_{ij}\|_{rM},\ \|\tilde{X}_{0j}(t)\|_{rM}\} \leq D,$$

*and*

$$\|X_{ij} - \tilde{X}_{ij}(i/n)\|_{rM} \leq C_A n^{-1}, \qquad \|\tilde{X}_{0j}(t) - \tilde{X}_{0j}(t')\|_{rM} \leq C_B |t-t'|,$$

*and* *<u>either</u>*

$$\|Y_i - \tilde{Y}_i(i/n)\|_{rM} \leq C_A n^{-1}, \qquad \|\tilde{Y}_0(t) - \tilde{Y}_0(t')\|_{rM} \leq C_B|t-t'| \tag{3.12}$$

*<u>or</u>* *(with $\chi$ from (A1))*

$$\sup_{x \neq x'} \frac{\|F_i(x,\theta) - F_i(x',\theta)\|_{M_y}}{|x-x'|_{\chi,1}} < \infty. \tag{3.13}$$

*(A6) (Negligibility of the truncation) For all $i,j$: $|X_{ij}^c| \leq |X_{ij}|$. For $1 \leq j \leq i$, $X_{ij} = X_{ij}^c$.*

*(A7) (Weak dependence) It holds that $\sup_{t\in[0,1]} \delta_{rM}^{\tilde{X}(t)}(k) = O(k^{-(1+\gamma)})$ and <u>either</u> (3.13) <u>or</u> $\sup_{t\in[0,1]} \delta_{rM}^{\tilde{Y}(t)}(k) = O(k^{-(1+\gamma)})$ holds.*

Note that (A2), (A3) and (A4) are typical assumptions in M-estimation theory to guarantee convergence of the estimator towards the correct parameter and to use Taylor expansions and bias expansions. The condition on $L$ in (A3) directly implies $0 = \nabla_\theta L(t,\theta(t)) = \mathbb{E}\nabla_\theta \ell(\tilde{Z}_0(t),\theta(t))$ under the imposed smoothness conditions, which will be used in the proofs. In many special cases in time series analysis (cf. Example 3.5.1), it may even occur that $\nabla_\theta \ell(\tilde{Z}_0(t),\theta(t))$ is a martingale difference sequence or at least an uncorrelated sequence. In these cases, $\Lambda(t) = I(t)$ such that the verification of (A4) is simplified.

Asking the objective function $\ell$ to be twice continuously differentiable w.r.t. $\theta$ as done in (A1) is a typical condition and is needed to use Taylor expansions. We additionally ask $\ell$ and its derivatives w.r.t. $\theta$ to be in $\mathcal{H}(M_y, M_x, \chi, \tilde{C})$. This is exploited in two ways: It allows quantification of the order of dependence of $\ell(Y_i, X_i, \theta)$ based on the dependence of $X_i, Y_i$, and it allows to deal with local stationarity by replacing $X_i, Y_i$ by its stationary counterparts. In this context, we especially need a decay condition on the coefficients $x_i$ which appear in $\ell$. This decay is quantified by the sequence $\chi = (\chi_i)_{i\in\mathbb{N}}$. We use this rate to show that the observed truncated values $X_i^c$ are negligible compared to $X_i$ and that the overall dependence of $\ell(Y_i, X_i, \theta)$ has the same order as the original sequences $Y_i, X_i$ (cf.

(A7)). Lastly, condition (A1) implicitly implies continuity of the matrices appearing in (A4) such that it is enough to show pointwise positive definiteness.

To eliminate bias terms, we state (A5) which asks for smoothness of the processes $X_i, Y_i$ in time direction and the existence of a stationary approximation. Here we consider two different cases. The case (3.13) is dedicated to general linear models which may have discretely distributed observations $Y_i$ and thus would not fulfill a condition like (3.12) for $rM \geq 2$. To prove central limits theorems and to use strong Gaussian approximations, we need a weak dependence assumption which is given in (A7). Let us emphasize the fact that all conditions besides (A5) are formulated for the stationary approximation $\tilde{Z}_i(t) = (\tilde{Y}_i(t), \tilde{X}_i(t))$ which in general allows easier verification and the possibility to use earlier results obtained for stationary settings.

To prove a typical second-order bias decomposition for $\widehat{\theta}'_{b_n}(t)$, we need that the stationary approximations $\tilde{Z}_i(t)$ are differentiable w.r.t. $t$. The concept of derivative processes in the context of locally stationary processes was originally introduced in Dahlhaus [12] and Dahlhaus and Subba Rao [15] and formalized in Dahlhaus et al. [13] especially for processes with Markov structure.

**Assumption 3.2.2** (Differentiability assumptions)**.** *Suppose that there exist $M_y', M_x' \geq 2$ such that $M' := \max\{M_x', M_y'\}$ fufills $M' \leq rM$ and*

*(B1)* $\theta(\cdot) \in C^4[0,1]$.

*(B2)* $\nabla_\theta^2 \ell(z,\theta)$ *is continuously differentiable. It holds that* $\nabla_\theta^3 \ell \in \mathcal{H}(M_y', M_x', \chi, \bar{C})$, *and for all* $l \in \mathbb{N}_0$, $\partial_{z_l} \nabla_\theta^2 \ell \in \mathcal{H}(M_y' - 1, M_x' - 1, \chi', \bar{C}\hat{\chi}_l)$ *with some absolutely summable sequence* $\chi' = (\chi_i')_{i=1,2,\dots}$.

*(B3)* $t \mapsto \tilde{Z}_0(t)$ *is continuously differentiable and* $\sup_{t\in[0,1]} \sup_{j\in\mathbb{N}_0} \|\tilde{Z}_{0j}(t)\|_{M'} \leq D$,

$$\sup_{j\in\mathbb{N}_0} \sup_{t\neq t'} \frac{\|\partial_t \tilde{Z}_{0j}(t) - \partial_t \tilde{Z}_{0j}(t')\|_{M'}}{|t-t'|} \leq C_B.$$

Note that the condition $\partial_{x_l}\nabla_\theta^2 \ell \in \mathcal{H}(M_y', M_x', \chi', \bar{C}\chi_l)$ asks $\nabla_\theta^2 \ell$ to be dependent on $x_l$ with a factor of at most $\chi_l$ which is a stronger condition than the corresponding condition on $\nabla_\theta^2 \ell$ in (A1).

## 3.3 Main results

### *3.3.1 Consistency and asymptotic normality*

For $l \geq 0$, define

$$\mu_{K,l} := \int K(x)x^l dx, \qquad \sigma_{K,l}^2 := \int K(x)^2 x^l dx.$$

Under weaker assumptions than those needed for the proof of SCBs, we can obtain pointwise consistency and asymptotic normality of the estimators $\hat{\theta}_{b_n}$ and $\widehat{\theta}'_{b_n}$. For matrices $A, B$, let $A \otimes B$ denote the Kronecker product and

$$A^{\otimes k} = A \otimes \ldots \otimes A \tag{3.14}$$

denote the $k$-fold Kronecker product.

**Theorem 3.3.1.** *Fix $t \in (0,1)$. Let Assumption 3.2.1 hold with $r = 2$. Assume that $nb_n \to \infty$, $b_n \to 0$.*

*(i) (Consistency) It holds that $\hat{\theta}_{b_n}(t) - \theta(t) = o_{\mathbb{P}}(1)$.*

*If additionally $nb_n^3 \to \infty$, it holds that $\widehat{\theta}'_{b_n}(t) - \theta'(t) = o_{\mathbb{P}}(1)$.*

*Assume that $\sup_{j\in\mathbb{N}_0} \sup_{t\in[0,1]} \|Z_{0j}(t)\|_{(2+a)M} < \infty$ for some $a > 0$.*

*(ii) If $nb_n^7 \to 0$, then*

$$\sqrt{nb_n}(\hat{\theta}_{b_n}(t) - \theta(t) - b_n^2 \frac{\mu_{K,2}}{2}\theta''(t)) \xrightarrow{d} N(0, \sigma_{K,0}^2 \cdot V(t)^{-1} I(t) V(t)^{-1}). \tag{3.15}$$

*(iii) If additionally, Assumption 3.2.2 is fulfilled and* $nb_n^9 \to 0$, *then*

$$\begin{pmatrix} \sqrt{nb_n}(\hat{\theta}_{b_n}(t) - \theta(t) - b_n^2 \frac{\mu_{K,2}}{2}\theta''(t)) \\ \sqrt{nb_n^3}(\widehat{\theta}'_{b_n}(t) - \theta'(t) - b_n^2 \frac{\mu_{K,4}}{2\mu_{K,2}} bias(t)) \end{pmatrix}$$

$$\xrightarrow{d} \quad N\Big(0, \sigma_{K,0}^2 \begin{pmatrix} 1 & 0 \\ 0 & \mu_{K,2}^{-2} \end{pmatrix} \otimes \{V(t)^{-1}I(t)V(t)^{-1}\}\Big), \tag{3.16}$$

*where* $bias(t) = \frac{1}{3}\theta^{(3)}(t) + V(t)^{-1}\mathbb{E}[\partial_t \nabla_\theta^2 \ell(\tilde{Z}_0(t), \theta(t))]\theta''(t)$.

**Remark 3.3.2.** *The condition* $\sup_{j \in \mathbb{N}_0} \|\tilde{Z}_{0j}(t)\|_{(2+a)M} < \infty$ *is needed to prove a Lindeberg-type condition. As pointed out in the proof of Theorem 2.9 in Dahlhaus et al. [13], it can be dropped if instead* $\sup_{j\in\mathbb{N}_0} \|\sup_{t\in[0,1]} |\tilde{Z}_{0j}(t)|\|_{2M} < \infty$ *is assumed.*

**Remark 3.3.3** (About local constant estimation)**.** *If instead of (3.3) and (3.4), a local constant estimation via*

$$L_{n,b_n,const}^c(t,\theta) := (nb_n)^{-1} \sum_{i=1}^n K_{b_n}(i/n - t)\ell(Z_i^c, \theta)$$

*and* $\hat{\theta}_{b_n,const}(t) = \arg\min_{\theta\in\Theta} L_{n,b_n}^c(t,\theta)$ *is used, one needs more smoothness assumptions on the underlying process to obtain a similar result as in (3.15). If for instance twice differentiability of* $t \mapsto \tilde{Z}_0(t)$ *is assumed, one obtains*

$$\sqrt{nb_n}(\hat{\theta}_{b_n}(t) - \theta(t) - b_n^2 \frac{\mu_{K,2}}{2} V(t)^{-1}\mathbb{E}[\partial_t^2 \nabla_\theta \ell(\tilde{Z}_0(t), \theta)\big|_{\theta=\theta(t)}]) \\ \xrightarrow{d} N(0, \sigma_{K,0}^2 \cdot V(t)^{-1}I(t)V(t)^{-1}).$$

*Note that the bias term changes significantly.*

### 3.3.2 Bahadur representation

In the following, we obtain a Bahadur representation of $\hat{\theta}_{b_n}$ and $\widehat{\theta}'_{b_n}$ which will be used to construct simultaneous confidence bands. In general, Bahadur representations are important for the asymptotic analysis of estimators by approximating them by linear forms. Due to the general setup, the result may be of independent interest. The first part of Theorem 3.3.4(i) shows that $\hat{\theta}_{b_n}(t) - \theta(t)$ can be approximated by the expression $V(t)^{-1}\nabla_\theta L^c_{n,b_n}(t, \theta(t), \theta'(t))$ as expected due to a standard Taylor argument. The second part of Theorem 3.3.4(i) deals with approximating this term by a weighted sum of $t$-free terms, namely

$$(nb_n)^{-1}\sum_{i=1}^{n} K_{b_n}(i/n - t)h_i, \qquad h_i := \nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n)),$$

which is necessary to apply some earlier results from Zhou and Wu [91]. Similar results are obtained for $\widehat{\theta}'_{b_n}$ in Theorem 3.3.4(ii). Let $\mathcal{T}_n := [b_n, 1 - b_n]$. For some vector or matrix $x$, let $|x| := |x|_2$ denote its Euclidean or Frobenius norm, respectively.

**Theorem 3.3.4** (Bahadur representation of $\hat{\theta}_{b_n}$, $\widehat{\theta}'_{b_n}$)**.** *Let $\beta_n = (nb_n)^{-1/2}b_n^{-1/2}\log(n)^{1/2}$ and put $\tau_n^{(j)} = (\beta_n + b_n)((nb_n)^{-1/2}\log(n) + b_n^{1+j})$ for $j = 1, 2$. Let Assumption 3.2.1 be fulfilled.*

*(i) It holds that*

$$\sup_{t\in\mathcal{T}_n}\Big|V(t)\cdot\big\{\hat{\theta}_{b_n}(t) - \theta(t)\big\} - \nabla_\theta L^c_{n,b_n}(t, \theta(t), \theta'(t))\Big| = O_{\mathbb{P}}(\tau_n^{(1)}), \tag{3.17}$$

$$\sup_{t\in\mathcal{T}_n}\Big|\nabla_\theta L^c_{n,b_n}(t, \theta(t), \theta'(t)) - b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) - (nb_n)^{-1}\sum_{i=1}^{n} K_{b_n}(i/n - t)h_i\Big| = O_{\mathbb{P}}(\beta_n b_n^2 + b_n^3 + (nb_n)^{-1}). \tag{3.18}$$

*(ii) If additionally Assumption 3.2.2 is fulfilled, then*

$$\sup_{t\in\mathcal{T}_n}\left|\mu_{K,2}V(t)\cdot b_n\{\widehat{\theta}'_{b_n}(t)-\theta'(t)\}-b_n^{-1}\nabla_{\theta'}L^c_{n,b_n}(t,\theta(t),\theta'(t))\right| = O_{\mathbb{P}}(\tau_n^{(2)}), \tag{3.19}$$

$$\sup_{t\in\mathcal{T}_n}\Big|b_n^{-1}\nabla_{\theta'}L^c_{n,b_n}(t,\theta(t),\theta'(t))-b_n^3\frac{\mu_{K,4}}{2}V(t)bias(t) \tag{3.20}$$

$$-(nb_n)^{-1}\sum_{i=1}^n K_{b_n}(i/n-t)\frac{(i/n-t)}{b_n}h_i\Big| = O_{\mathbb{P}}(\beta_n b_n^2+b_n^4+(nb_n)^{-1}).$$

### *3.3.3 Simultaneous confidence bands*

Based on the Bahadur result, we use results from Wu and Zhou [81] to obtain a Gaussian analogue of

$$\frac{1}{nb_n}\sum_{i=1}^n K_{b_n}(t-i/n)\nabla_\theta \ell(\tilde{Z}_i(i/n),\theta(i/n)) = \frac{1}{nb_n}\sum_{i=1}^n K_{b_n}(t-i/n)\tilde{h}_i(i/n).$$

For the following results, let us assume that there exists some measurable function $\tilde{H}(\cdot,\cdot)$ such that for each $t\in[0,1]$, $\tilde{h}_i(t)=\tilde{H}(t,\mathcal{F}_i)\in\mathbb{R}^s$ is well-defined. Let $h_i=\tilde{h}_i(i/n)$. Put $S_h(i) := \sum_{j=1}^i h_j$. For a positive semidefinite matrix $A$ with eigendecomposition $A = QDQ^\mathsf{T}$, where $Q$ is orthonormal and $D$ is a diagonal matrix, define $A^{1/2}=QD^{1/2}Q^\mathsf{T}$, where $D^{1/2}$ is the elementwise root of $D$.

**Theorem 3.3.5** (Theorem 1 and Corollary 2 from Wu and Zhou [81])**.** *Assume that for each component $j=1,\ldots,s$:*

*(a)* $\sup_{t\in[0,1]}\|\tilde{h}_0(t)_j\|_4<\infty$,

*(b)* $\sup_{t\neq t'\in[0,1]}\|\tilde{h}_0(t)_j-\tilde{h}_0(t')_j\|_2/|t-t'|<\infty$,

*(c)* $\sup_{t\in[0,1]}\delta_4^{\tilde{h}(t)_j}(k)=O(k^{-(\gamma+1)})$ *with some* $\gamma\geq 1$.

*Then on a richer probability space, there are i.i.d.* $V_1,V_2,\ldots\sim N(0,I_{s\times s})$ *and a process*

$S_h^0(i) = \sum_{j=1}^{i} \Sigma_h(j/n)V_j$ *such that* $(S_h(i))_{i=1,\ldots,n} \overset{d}{=} (S_h^0(i))_{i=1,\ldots,n}$ *and*

$$\max_{i=1,\ldots,n} |S_h(i) - S_h^0(i)| = O_{\mathbb{P}}(\pi_n).$$

*where* $\pi_n = n^{(1+\gamma)/(1+4\gamma)} \log(n)^{(5\gamma)/(1+4\gamma)}$ *and*

$$\Sigma_h(t) = \big(\sum_{j\in\mathbb{Z}} \mathbb{E}[\tilde{h}_0(t)\tilde{h}_j(t)^{\mathsf{T}}]\big)^{1/2}.$$

Based on this theorem, we are able to prove the following asymptotic statement for simultaneous confidence bands for $\theta(\cdot)$:

**Theorem 3.3.6** (Simultaneous confidence bands for $\theta(\cdot)$ and $\theta'(\cdot)$)**.** *Let* $C$ *be a fixed* $k \times s$ *matrix with rank* $s \leq k$*. Define* $\hat{\theta}_{b_n,C}(t) := C^{\mathsf{T}}\hat{\theta}_{b_n}(t)$, $\hat{\theta}'_{b_n,C}(t) := C^{\mathsf{T}}\widehat{\theta}'_{b_n}(t)$ *and* $\theta_C(t) := C^{\mathsf{T}}\theta(t)$, $A_C(t) := V(t)^{-1}C$, $\Sigma_C(t) := A_C^{\mathsf{T}}(t)\Lambda(t)A_C(t)$.

*Let Assumption 3.2.1 be fulfilled. Assume that* $\log(n)\big(b_n n^{(2\gamma-1)/(1+4\gamma)}\big)^{-1} \to 0$.

*(i) If* $nb_n^7\log(n) \to 0$*, then*

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\sqrt{nb_n}}{\sigma_{K,0}} \sup_{t\in\mathcal{T}_n} \Big|\Sigma_C^{-1}(t)\Big\{\hat{\theta}_{b_n,C}(t) - \theta_C(t) - b_n^2\frac{\mu_{K,2}}{2}C^{\mathsf{T}}\theta''_C(t)\Big\}\Big| - B_K(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)), \quad (3.21)$$

*(ii) If additionally, Assumption 3.2.2 is fulfilled and* $nb_n^9\log(n) \to 0$*, then with* $\hat{K}(x) = K(x)x$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\sqrt{nb_n^3}\mu_{K,2}}{\sigma_{K,2}} \sup_{t\in\mathcal{T}_n} \Big|\Sigma_C^{-1}(t)\Big\{\hat{\theta}'_{b_n,C}(t) - \theta'_C(t) - b_n^2\frac{\mu_{K,4}}{2\mu_{K,2}}C^{\mathsf{T}}bias(t)\Big\}\Big| - B_{\hat{K}}(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)), \quad (3.22)$$

*where in both cases* $\mathcal{T}_n = [b_n, 1-b_n]$, $m^* = 1/b_n$ *and*

$$B_K(m^*) = \sqrt{2\log(m^*)} + \frac{\log(C_K) + (s/2-1/2)\log(\log(m^*)) - \log(2)}{\sqrt{2\log(m^*)}}, \tag{3.23}$$

*with*

$$C_K = \frac{\left\{ \int_{-1}^{1} |K'(u)|^2 du/\sigma_{K,0}^2 \pi \right\}^{1/2}}{\Gamma(s/2)}.$$

**Remark 3.3.7.** *The conditions on* $b_n$ *are fulfilled for bandwidths* $b_n = n^{-\alpha}$, *where* $\alpha \in (0,1)$ *satisfies:*

*(i)* $1/7 < \alpha < (2\gamma-1)/(1+4\gamma)$ *in case (i),*

*(ii)* $1/9 < \alpha < (2\gamma-1)/(1+4\gamma)$ *in case (ii),*

*i.e.* $\gamma > 1$ *satisfies that bandwidths* $b_n = cn^{-1/5}$ *are covered.*

Note that for practical use of the SCB in (3.21) and (3.22), one needs to estimate the bias term, choose a proper bandwidth $b_n$ and estimate $\Sigma_C(t)$. Furthermore, the theoretical SCB only has slow logarithmic convergence, thus one requires huge $n$ to achieve the desired coverage probability. To tackle these type of problems, we discuss practical issues in the next Section 3.4.

## 3.4 Implementational issues

In this section, we discuss some issues which arise by implementing the procedure from Theorem 3.3.6. We focus on estimation of $\hat{\theta}_{b_n}$ and optimization of the corresponding SCBs; the results for $\widehat{\theta'}_{b_n}$ can be obtained accordingly.

### *3.4.1 Bias correction*

There are several possible ways to eliminate the bias term in (3.21). A natural way is to estimate $\theta''(t)$ by using a local quadratic estimation routine with some bandwidth $b_n' \geq b_n$.

However the estimation of $\theta''(t)$ may be unstable due to the convergence condition $nb_n^5 \to \infty$ which may be hard to realize together with $nb_n^7 \log(n) \to 0$ from Theorem 3.3.6 in practice. Here instead we propose a bias correction via the following jack-knife method: We define

$$\tilde{\theta}_{b_n}(t) := 2\hat{\theta}_{b_n/\sqrt{2}}(t) - \hat{\theta}_{b_n}(t). \tag{3.24}$$

Since the Bahadur representation from Theorem 3.3.4(i) holds both for $\hat{\theta}_{b_n/\sqrt{2}}$ and $\hat{\theta}_{b_n}(t)$, we obtain

$$\sup_{t \in \mathcal{T}_n} \left| V(t) \cdot \{\tilde{\theta}_{b_n}(t) - \theta(t)\} - (nb_n)^{-1} \sum_{i=1}^{n} \tilde{K}_{b_n}(i/n - t) h_i \right| = O_{\mathbb{P}}(\tau_n^{(1)} + \beta_n b_n^2 + b_n^3 + (nb_n)^{-1}),$$

where $\tilde{K}(x) := 2\sqrt{2}K(\sqrt{2}x) - K(x)$. Note that the bias term of order $b_n^2$ is eliminated by construction. This shows that Theorem 3.3.6(i) still holds true for $\tilde{\theta}_{b_n}(\cdot)$ with kernel $K$ replaced by the fourth-order kernel $\tilde{K}$ and with no bias term of order $b_n^2$.

### *3.4.2 Estimation of the covariance matrix* $\Sigma_C(t)$

In this subsection, we discuss the estimation of $\Sigma_C(t)$ (namely, $V(t)$ and $\Lambda(t)$) since this term is generally unknown but arises in the SCB in Theorem 3.3.6. In Examples 3.5.1, 3.5.2 and 3.5.3 it can be seen that in many time series and independent regression models where the objective function $\ell$ is given by a (conditional) maximum likelihood approach, it holds that $\Lambda(t) = I(t)$ due to the fact that the $\nabla_\theta \ell(\tilde{Z}_i(t), \theta(t))$, $i \in \mathbb{Z}$ are uncorrelated. In the case that the objective function $\ell$ coincides with the true log conditional likelihood, one has even $V(t) = I(t)$. As it can be seen in Examples 3.5.1 and 3.5.2, even in the misspecified case it may often hold that $V(t) = c_0 \cdot I(t)$ with some constant $c_0 > 0$ only dependent on properties of the i.i.d. innovations $\zeta_0$ which can be calculated by further assumptions on $\zeta_0$.

Therefore, it may often hold that $\Sigma_C(t) = C^\mathsf{T} V(t)^{-1} \Lambda(t) V(t)^{-1} C$ obeys one of the two equalities

$$\Sigma_C(t) = C^\mathsf{T} V(t)^{-1} I(t) V(t)^{-1} C, \qquad \text{or} \tag{3.25}$$

$$\Sigma_C(t) = C^\mathsf{T} I(t)^{-1} C / c_0 \quad \text{with some known constant } c_0. \tag{3.26}$$

We therefore focus on estimation of $V(t)$ and $I(t)$. We propose the estimators

$$\hat{V}_{b_n}(t) := (nb_n)^{-1} \sum_{i=1}^{n} K_{b_n}(i/n - t) \nabla_\theta^2 \ell(Z_i^c, \hat{\theta}_{b_n}(t) + (i/n - t)\widehat{\theta}'_{b_n}(t)), \tag{3.27}$$

$$\hat{I}_{b_n}(t) := (nb_n)^{-1} \sum_{i=1}^{n} K_{b_n}(i/n - t) \nabla_\theta \ell(Z_i^c, \hat{\theta}_{b_n}(t) + (i/n - t)\widehat{\theta}'_{b_n}(t)) \times \nabla_\theta \ell(Z_i^c, \hat{\theta}_{b_n}(t) + (i/n - t)\widehat{\theta}'_{b_n}(t))^\mathsf{T}. \tag{3.28}$$

The convergence of these estimators is given in the next Proposition and easily follows from Lemma 3.7.4 and Lemma 3.7.5 in the appendix. Note that the following Proposition also holds if $\widehat{\theta}'_{b_n}$ in (3.27) and (3.28) is replaced by 0.

**Proposition 3.4.1.** *Let Assumption 3.2.1 hold with some $r \geq 2$. Let $nb_n^2 \log(n)^{-2d_\Theta} \to \infty$ and $b_n \to 0$. Then*

*(i)* $\sup_{t \in \mathcal{T}_n} |\hat{V}_{b_n}(t) - V(t)| = O_\mathbb{P}((\log n)^{-1})$.

*(ii) If $r \geq 4$, then* $\sup_{t \in \mathcal{T}_n} |\hat{I}_{b_n}(t) - I(t)| = O_\mathbb{P}((\log n)^{-1})$.

This shows uniform consistency of $\hat{V}_{b_n}(\cdot)$, $\hat{I}_{b_n}(\cdot)$ if $nb_n^2 \log(n)^{-2d_\Theta} \to \infty$ and $b_n \to 0$. Note that in (ii), we need more moments to discuss $\nabla_\theta \ell \cdot \nabla_\theta \ell^\mathsf{T} \in \mathcal{H}(2M_y, 2M_x, \chi, \bar{\bar{C}})$ $(\bar{\bar{C}} > 0)$. In either case (3.25) or (3.26), we define $\hat{\Sigma}_C(t)$ by replacing $V(t), I(t)$ by the corresponding estimators $\hat{V}_{b_n}(t)$, $\hat{I}_{b_n}(t)$.

If no relations are known between $V(t)$ and $\Lambda(t)$, one has to use a more general approach to estimate $\Lambda(t)$. We do not want to focus on this situation since the applications we have in mind (cf. Section 3.5) are kept by (3.25) or (3.26). Therefore, we only adopt a method from

Zhou and Wu [91] to estimate $\Lambda(t)$. Define $\tilde{D}_i := \nabla_\theta \ell(Z_i^c, \hat{\theta}_{b_n}(i/n))$, $\tilde{Q}_i := \sum_{j=-m}^{m} \tilde{D}_{i+j}$ and $\tilde{\Phi}_i := \tilde{Q}_i \tilde{Q}_i^\top/(2m+1)$. Let $\tau_n$ be some bandwidth, and put $\gamma_n := \tau_n + (m+1)/n$. For $t \in \mathcal{I}_n := [\gamma_n, 1-\gamma_n] \subset (0,1)$, define

$$\tilde{\Lambda}(t) := \frac{\sum_{i=1}^n K_{\tau_n}(i/n - t)\tilde{\Phi}_i}{\sum_{i=1}^n K_{\tau_n}(i/n-t)}.$$

Note that $\tilde{\Lambda}(t)$ is always positive semi-definite. We have the following convergence result.

**Theorem 3.4.2.** *Suppose that Assumption 3.2.1 holds with $r = 4$. Assume that $\omega_n = o(1)$, where $\omega_n = n^{1/4}\sqrt{m}\log(n)\{(nb_n)^{-1/2}\log(n) + b_n^2\}$. Then with $\rho = 1$,*

$$\sup_{t \in \mathcal{I}_n} |\tilde{\Lambda}(t) - \Lambda(t)| = O_{\mathbb{P}}\Big(\omega_n + \sqrt{\frac{m}{n\tau_n^2}} + m^{-1} + \tau_n^\rho\Big).$$

*If additionally Assumption 3.2.2(B1),(B3) is fulfilled with $M' = 2M$ and $\nabla_\theta \ell$ is continuously differentiable with $\partial_{z_j}\nabla_\theta \ell \in \mathcal{H}(M_y - 1, M_x - 1, \chi', \hat{\chi}_j \bar{C})$ for all $j \in \mathbb{N}_0$, then one can choose $\rho = 2$.*

Let us shortly discuss the choices of $\tau_n$, $b_n$ and $m$ in the above setting. For two positive sequences $(r_n)$, $(s_n)$ we write $r_n \asymp s_n$ if $r_n/s_n$ and $s_n/r_n$ are bounded for all $n$ large enough. If one chooses $m \asymp n^{q_1}$, $b_n \asymp n^{-q_2}$ and $\tau_n = n^{-q_3}$ with some $q_1, q_2, q_3 > 0$, we obtain from Theorem 3.4.2 that $\sup_{t\in\mathcal{I}_n} |\tilde{\Lambda}(t) - \Lambda(t)| = O_{\mathbb{P}}(n^{-\nu}) = O_{\mathbb{P}}((\log n)^{-1})$ with some $\nu > 0$ if $q_1/2 + 1/4 < \min\{2q_2, 1/2 - q_2/2\}$ and $q_1 < 1 - 2q_3$. In the special case $q_2 = 1/5$, this reduces to the condition $q_1 < \min\{3/10, 1-2q_3\}$.

### *3.4.3 Bandwidth selection*

Based on the asymptotic result (3.15) in Theorem 3.3.1 under Assumption 3.2.1, the MSE global optimal bandwidth choice reads

$$\hat{b}_n = n^{-1/5} \cdot \Big(\frac{\sigma_{K,0}^2 \int_0^1 \mathrm{tr}(V(t)^{-1}I(t)V(t)^{-1})dt}{\mu_{K,2}^2 \int_0^1 |\theta''(t)|^2 dt}\Big)^{1/5}. \qquad (3.29)$$

We therefore adapt a model-based cross validation method from Richter and Dahlhaus [63], which was shown to work even if the underlying parameter curve is only Hölder continuous and $\nabla_\theta \ell(\tilde{Z}_i(t), \theta(t))$ is uncorrelated. Here, we reformulate this selection procedure for the local linear setting. For $j = 1, \ldots, n$, define the leave-one-out local linear likelihood

$$L^c_{n,b_n,-j}(t, \theta, \theta') := (nb_n)^{-1} \sum_{i=1, i\neq j}^{n} K_{b_n}(i/n - t)\ell(Z_i^c, \theta + (i/n - t)\theta') \tag{3.30}$$

and the corresponding leave-one-out estimator

$$(\hat{\theta}_{b_n,-j}(t), \hat{\theta}'_{b_n,-j}(t)) = \arg \min_{\theta\in\Theta, \theta'\in\Theta'} L^c_{n,b_n,-j}(t, \theta, \theta').$$

The bandwidth $\hat{b}_n^{CV}$ is chosen via minimizing

$$CV(b) := n^{-1} \sum_{i=1}^{n} \ell(Z_i^c, \hat{\theta}_{b_n,-i}(i/n))w(i/n), \tag{3.31}$$

where $w(\cdot)$ is some weight function to exclude boundary effects. A possible choice is $w(\cdot) := \mathbb{1}_{[\gamma_0, 1-\gamma_0]}$ with some fixed $\gamma_0 > 0$. Note that it is important to use the modified local linear approach due to the different bias terms (cf. Remark 3.3.3). In Richter and Dahlhaus [63], it was shown that the local constant version of this procedure selects asymptotically optimal bandwidths and works even if a model misspecification is present, i.e. if the function $\ell$ leads to estimators $\hat{\theta}_{b_n}$ which are not consistent. This motivates that a similar behavior should hold for the local constant version.

### *3.4.4 Bootstrap method*

The SCB for $\theta_C(t)$ obtained in Theorem 3.3.6 provides a slow logarithmic rate of convergence to the Gumbel distribution. Thus, for even moderately large values of sample size $n$, it is practically infeasible to use such a theoretical SCB as the coverage will possibly be lower than the specified nominal level. We circumvent this convergence issue in this subsection by proposing a wild bootstrap algorithm. Recall the jackknife-based bias corrected estimator of $\tilde{\theta}_{b_n}$ from (3.24). Let $\tilde{\theta}_C(t) = C^T \tilde{\theta}_{b_n}(t)$. We have the following proposition as the key idea behind the bootstrap method.

**Proposition 3.4.3.** *Suppose that Assumption 3.2.1 holds with $r = 4$. Furthermore, assume that $b_n = O(n^{-\kappa})$ with $1/7 < \kappa < (2\gamma - 1)/(1 + 4\gamma)$. Then on a richer probability space, there are i.i.d. $V_1, V_2, \ldots, \sim N(0, Id_s)$ such that*

$$\sup_{t \in \mathcal{T}_n} |\tilde{\theta}_C(t) - \theta_C(t) - W(t)| = O_{\mathbb{P}}\left(\frac{n^{-\nu}}{\sqrt{nb_n}\log(n)^{1/2}}\right), \tag{3.32}$$

*where $\nu = \min\{(2\gamma - 1)/(2 + 8\gamma) - \kappa/2, 7\kappa/2 - 1/2, \kappa/2\} > 0$ and*

$$W(t) = \Sigma_C(t)\mu_{b_n}(t) \quad \textit{with} \quad \mu_{b_n}(t) = \frac{1}{nb_n}\sum_{i=1}^{n} V_i K_{b_n}(i/n - t).$$

The proof of Proposition 3.4.3 is immediate from the approximation rates (3.93), (3.94), (3.95) and (3.97) which, ignoring the $\log(n)$ terms, are of the form $c_n \cdot (nb_n)^{-1/2}\log(n)^{-1/2}$ with

$$c_n \in \{\big(b_n n^{(2\gamma-1)/(1+4\gamma)}\big)^{-1/2}, b_n^{1/2}, b_n, (nb_n^7)^{1/2}, (nb_n^2)^{-1/2}\}.$$

One can interpret (3.32) in the sense that $\Sigma_C(t)\mu_{b_n}(t)$ approximates the stochastic variation in $\tilde{\theta}_C(t) - \theta_C(t)$ uniformly over $t \in \mathcal{T}_n$ and thus it can be used as margin of errors to construct confidence bands, provided one can consistently estimate $\Sigma_C(t)$. Motivated by this interpretation, one can create a large number of i.i.d. copies of $\mu_{b_n}(t)$ as

$$\mu_{b_n}^{boot}(t) = \frac{1}{nb_n}\sum_{i=1}^{n} V_i^* K_{b_n}(i/n - t), \tag{3.33}$$

where $V_1^*, V_2^*, \ldots,$ are i.i.d. $N(0, I_{s\times s})$-distributed random variables. Next, we compute the quantiles of $\mu_{b_n}(t)$ by generating a large number of copies $\mu_{b_n}^{boot}(t)$ and determining the corresponding empirical quantile. Then one can use Proposition 3.4.3 to construct the confidence band for $\theta_C(t)$. For convenience of the readers, we provide a summarized algorithm of the above discussion.

**Algorithm for constructing SCBs of $\theta_C(t)$**

- Compute the appropriate bandwidth $b_n$ based on the cross validation method in Subsection 3.4.3 and compute $\tilde{\theta}_C(t)$ based on the jackknife-based estimator from 3.4.1.

- For $r = 1, \ldots, N$ with some large $N$, generate $n$ i.i.d. $N(0, I_{s\times s})$ random variables $V_1^*, \ldots, V_n^*$ and compute $q_r = \sup_{t\in[0,1]} |\mu_{b_n}^{boot}(t)|$, where $\mu_{b_n}^{boot}(t)$ is computed according to (3.33).

- Repeat the above step for a large number of times and compute $u_{1-\alpha} = q_{\lfloor(1-\alpha)N\rfloor}$, the empirical $(1-\alpha)$th quantile of $\sup_{t\in[0,1]} |\mu_{b_n}(t)|$.

- Calculate $\hat{\Sigma}_C(t) = \{C^T \hat{V}(t)^{-1}\hat{\Lambda}(t)\hat{V}(t)^{-1}C\}^{1/2}$ with the estimators proposed in Subsection 3.4.2. As mentioned there, $V(t)^{-1}\Lambda(t)V(t)^{-1}$ can often be simplified.

- The SCB for $\theta_C(t)$ is $\tilde{\theta}_{C,b_n}(t) + \hat{\Sigma}_C(t)u_{1-\alpha}\mathcal{B}_s$, where $\mathcal{B}_s = \{x \in \mathbb{R}^s : |x| \le 1\}$ is the unit ball in $\mathbb{R}^s$.

## 3.5 Examples

We now apply our theory to a large class of recursively defined time series models, GARCH processes and, as an important special case of general linear models, logistic regression models. Due to the general formulations, we do not focus on obtaining optimal moment conditions or minimal restrictions on parameter spaces here.

**Example 3.5.1** (Time-varying recursively defined time series models). *Assume that* $X_i = (Y_{i-1}, \ldots, Y_{i-p}, 0, \ldots)^\mathsf{T}$, $X_i^c = (Y_{i-1}, \ldots, Y_{1\vee(i-p)}, 0, \ldots)^\mathsf{T}$ *and consider*

$$Y_i = \mu(X_i, \theta(i/n)) + \sigma(X_i, \theta(i/n))\zeta_i, \tag{3.34}$$

*where* $\theta = (\alpha_1, \ldots, \alpha_k, \beta_0, \ldots, \beta_l)^\mathsf{T}$ *and*

$$\mu(x,\theta) := \sum_{i=1}^{k} \alpha_i m_i(x), \qquad \sigma(x,\theta) := \big(\sum_{i=0}^{l} \beta_i \nu_i(x)\big)^{1/2},$$

*with some functions* $m_i : \mathbb{R}^p \to \mathbb{R}$, $\nu_i : \mathbb{R}^p \to \mathbb{R}_{\geq 0}$. *Assume that*

1. $\zeta_i$ *are i.i.d. with* $\mathbb{E}\zeta_i = 0$, $\mathbb{E}\zeta_i^2 = 1$ *and* $\mathbb{E}\zeta_i^{4M} < \infty$ *(*$M$ *is defined below).*

2. *For all* $t \in [0,1]$*, the sets*

$$\{m_1(\tilde{X}_0(t)), \ldots, m_k(\tilde{X}_0(t))\}, \qquad \{\nu_0(\tilde{X}_0(t)), \ldots, \nu_l(\tilde{X}_0(t))\}$$

   *are (separately) linearly independent in* $\mathcal{L}_2$.

3. *There exist* $(\kappa_{ij}) \in \mathbb{R}_{\geq 0}^{k \times p}$, $(\rho_{ij}) \in \mathbb{R}_{\geq 0}^{(l+1) \times p}$ *such that for all* $i$*:*

$$\sup_{x \neq x'} \frac{|m_i(x) - m_i(x')|}{|x - x'|_{\kappa_{i\cdot},1}} \leq 1, \qquad \sup_{x \neq x'} \frac{|\sqrt{\nu_i(x)} - \sqrt{\nu_i(x')}|}{|x - x'|_{\rho_{i\cdot},1}} \leq 1. \tag{3.35}$$

   *Let* $\nu_{min} > 0$ *be some constant such that for all* $x \in \mathbb{R}$, $\nu_0(x) \geq \nu_{min}$. *With some* $\beta_{min} > 0$, *choose* $\tilde{\Theta} \subset \mathbb{R}^k \times \mathbb{R}^{l+1}_{\geq \beta_{min}}$ *such that*

$$\sum_{j=1}^{p} \Big( \sup_{\theta \in \tilde{\Theta}} \sum_{i=1}^{k} |\alpha_i| \kappa_{ij} + \|\zeta_0\|_{4M} \cdot \sup_{\theta \in \tilde{\Theta}} \sum_{i=0}^{l} \sqrt{\beta_i} \rho_{ij} \Big) < 1. \tag{3.36}$$

4. *Assumption 3.2.1 (A2) is valid with some* $\Theta \subset \tilde{\Theta}$.

*Then Assumption 3.2.1 is fulfilled for* $\ell$ *chosen to be proportional to the negative log Gaussian conditional likelihood,*

$$\ell(y,x,\theta) = \frac{1}{2}\Big[\Big(\frac{y - \mu(x,\theta)}{\sigma(x,\theta)}\Big)^2 + \log \sigma(x,\theta)^2\Big],$$

*with $M = 3$ and $\Lambda(t) = I(t)$. In the special case $\sigma(x,\theta)^2 \equiv \beta_0$, one can choose $M = 2$. If (i) $\mathbb{E}\zeta_0^3 = 0$, or (ii) $\mu(x,\theta) \equiv 0$ or (iii) $\sigma(x,\theta) \equiv \beta_0$ and $\mathbb{E}m(\tilde{X}_0(t)) = 0$, then*

$$I(t) = \left(\begin{smallmatrix} I_k & 0 \\ 0 & (\mathbb{E}\zeta_0^4 - 1)I_{l+1}/2 \end{smallmatrix}\right) \cdot V(t),$$

*where $I_d$ denotes the d-dimensional identity matrix.*

*If additionally, Assumption 3.2.2 (B1) is fulfilled and $m_i$, $\nu_i$ are differentiable such that for all $j = 1,\dots,p$ and all $i$,*

$$\sup_{x \neq x'} \frac{|\partial_{x_j} m_i(x) - \partial_{x_j} m_i(x')|}{|x - x'|_1} < \infty, \qquad \sup_{x \neq x'} \frac{|\partial_{x_j} \nu_i(x) - \partial_{x_j} \nu_i(x')|}{|x - x'|_1} < \infty,$$

*then Assumption 3.2.2 is fulfilled for $\ell$.*

Note that in many special cases as the tvAR or the tvARCH processes, the restrictive conditions on the parameter space (3.36) can be relaxed by using matrix arguments, see also the proof techniques in Example 3.5.2.

In the following, we consider the tvGARCH model. This model was for instance studied in the stationary case in Francq and Zakoïan [28]. More recently, pointwise asymptotic results were obtained in Rohan and Ramanathan [65]. For a matrix $A$, we define $\|A\|_q := (\|A_{ij}\|_q)_{ij}$ as a component-wise application of $\|\cdot\|_q$. Recall the Kronecker product from (3.14).

**Example 3.5.2** (tvGARCH models)**.** *For $i = 1,\dots,n$, consider the recursion*

$$\begin{aligned} Y_i &= \sigma_i^2 \zeta_i^2, \\ \sigma_i^2 &= \alpha_0(i/n) + \sum_{j=1}^m \alpha_j(i/n) Y_{i-j} + \sum_{j=1}^l \beta_j(i/n)\sigma_{i-j}^2, \end{aligned}$$

*where $\theta = (\alpha_0,\dots,\alpha_m,\beta_1,\dots,\beta_l) : [0,1] \to \Theta \subset \mathbb{R}^{m+l+1}$. Let $f(\theta) = (\alpha_1,\dots,\alpha_m,\beta_1,\dots,\beta_l)^\mathsf{T}$ and let $e_j = (0,\dots,0,1,0,\dots,0)^\mathsf{T}$ be the unit column vector with $j$th element being 1, $1 \le j \le l+m$. Define $M_i(\theta) = (f(\theta)\zeta_i^2, e_1,\dots,e_{m-1}, f(\theta), e_{m+1},\dots,e_{m+l-1})^\mathsf{T}$. Let*

*$\alpha_{min} > 0$, and*

$$\tilde{\Theta} = \{\theta \in \mathbb{R}_{\geq 0}^{m+l+1} : \alpha_0 \geq \alpha_{min}, \lambda_{max}(\mathbb{E}[M_0(\theta)^{\otimes 16}]) < 1, \lambda_{max}(\|M_0(\theta)\|_{16}) < 1\}.$$

*Suppose that*

*(i) Assumption 3.2.1(A2) is fulfilled with $\Theta \subset \tilde{\Theta}$ and each component of $\theta(\cdot)$ is in $C^4[0,1]$,*

*(ii) $\zeta_i$ are i.i.d. with $\mathbb{E}\zeta_i = 0$, $\mathbb{E}\zeta_i^2 = 1$ and $\mathbb{E}\zeta_i^{32} < \infty$.*

*Then Assumptions 3.2.1 and 3.2.2 are fulfilled with the choices $X_i = (Y_{i-1}, Y_{i-2}, \ldots)$ and $X_i^c = (Y_{i-1}, Y_{i-2}, \ldots, Y_1, 0, 0, \ldots)$ for the conditional quasi likelihood*

$$\ell(y, x, \theta) = \frac{1}{2}\Big[\frac{y}{\sigma(x,\theta)^2} + \log(\sigma(x,\theta)^2)\Big],$$

*where $\sigma(x,\theta)^2$ is recursively defined via $\sigma(x,\theta)^2 = \alpha_0 + \sum_{j=1}^m \alpha_j x_j + \sum_{j=1}^l \beta_j \sigma(x_{j\to}, \theta)^2$ and $x_{j\to} := (x_{j+1}, x_{j+2}, \ldots)$. It holds that $\Lambda(t) = I(t) = ((\mathbb{E}\zeta_0^4 - 1)/2)V(t)$.*

We conjecture that the moment conditions can be weakened to $\mathbb{E}\zeta_0^{12} < \infty$, if instead the parameter space is more restricted (see the proof techniques used in Example 3.5.1).

Let $q \in \mathbb{N}$. In the important GARCH(1,1) case, Ling [47] proved that $\lambda_{max}(\mathbb{E}[M_i(\theta)^{\otimes q}]) < 1$ is equivalent to the condition

$$\sum_{j=0}^{q} \binom{q}{j} \|\zeta_0\|_{2j}^{2j} \alpha_1^j \beta_1^{q-j} < 1,$$

given in Theorem 2 in Bollerslev [7]. It is easy to see that $\lambda_{max}(\|M_0(\theta)\|_q) = \beta_1 + \alpha_1 \|\zeta_0\|_{2q}^2$. Due to the binomial theorem, we have $\lambda_{max}(\|M_0(\theta)\|_q) < 1$ if and only if

$$\sum_{j=0}^{q} \binom{q}{j} \|\zeta_0\|_{2q}^{2j} \alpha_1^j \beta_1^{q-j} < 1,$$

which indicates that $\lambda_{max}(\|M_0(\theta)\|_q) < 1$ is a stronger condition than $\lambda_{max}(\mathbb{E}[M_i(\theta)^{\otimes q}]) < 1$ and therefore

$$\tilde{\Theta} = \{\theta \in \mathbb{R}^3_{\geq 0} : \alpha_0 \geq \alpha_{min}, \beta_1 + \alpha_1 \|\zeta_0\|_{32}^2 < 1\}.$$

We conjecture that similar implications hold for general tvGARCH$(m, l)$-models.

Lastly, let us consider a locally stationary logistic regression model which could be used to check if effects of certain covariates change over time. We only consider one population of size $m$ for simplicity.

**Example 3.5.3** (Logistic regression)**.** *Fix $m \in \mathbb{N}$. Assume that $\zeta_i = (\zeta_{i,0}, \zeta_{i,1}, \ldots, \zeta_{i,m})^{\mathsf{T}}$, where $\zeta_{i,j}$, $i \in \mathbb{Z}$, $j = 0, \ldots, m$ are i.i.d. uniformly distributed on $[0,1]$. Let $X_i \in \mathbb{R}^p$ be a vector of covariates, $X_i = G(i/n, \mathcal{G}_i)$ with $\mathcal{G}_i = (\ldots, \zeta_{i-1,0}, \zeta_{i,0})$. For $i = 1, \ldots, n$,*

$$Y_i = \sum_{j=1}^{m} \mathbb{1}_{\{\zeta_{i,j} \leq \pi(X_i^{\mathsf{T}}\theta(i/n))\}}, \qquad i.e. \qquad Y_i | X_i \sim Bin\big(m, \pi(X_i^{\mathsf{T}}\theta(i/n))\big),$$

*where $\pi(w)$ is given by $logit(w) = w$ and $\theta : [0,1] \to \Theta \subset \mathbb{R}^p$ is the parameter curve which we want to estimate.*

*We use the typical maximum likelihood approach based on*

$$\ell(y, x, \theta) = m \cdot \log\big(1 + \exp\big(x^{\mathsf{T}}\theta\big)\big) - y \cdot \big(x^{\mathsf{T}}\theta\big).$$

*Assume that:*

1. *Assumption 3.2.1(A2) is fulfilled with some compact $[-D, D]^{p+1} \subset \Theta \subset \mathbb{R}^{p+1}$, $D > 0$,*
2. *$\tilde{X}_i(t) = G(t, \mathcal{G}_i)$ fulfills $\sup_{t\in[0,1]} \|\tilde{X}_0(t)\|_8 < \infty$ and $\sup_{t\in[0,1]} \delta_8^{\tilde{X}(t)}(k) = O(k^{-(1+\gamma)})$ with some $\gamma > 1$.*
3. *For all $t, t' \in [0,1]$ it holds that, with some constant $C_B > 0$,*

$$\|\tilde{X}_0(t) - \tilde{X}_0(t')\|_8 \leq C_B |t - t'|.$$

*4. For each $t \in [0,1]$, $\mathbb{E}[\tilde{X}_0(t)\tilde{X}_0(t)^{\mathsf{T}}]$ is positive definite.*

*Then Assumption 3.2.1 is fulfilled and $\Lambda(t) = I(t) = V(t)$.*

Note that it is not possible to fulfill Assumption 3.2.2 in our setting of Example 3.5.3 since the condition of the existence of an a.s. derivative of $t \mapsto \tilde{X}_0(t)$ is too strong. It was discussed in Dahlhaus et al. [13], that differentiability in $L^1$ should be enough to show the bias expansions for which Assumption 3.2.2 is needed, i.e. we conjecture that the results for $\widehat{\theta}'_{b_n}$ of this paper also hold true for this example.

## 3.6 Simulation results and Applications

This section consists of some summarized simulations and some real data applications related to our theoretical results. Because of the generality of our theoretical framework, it is impossible to report simulation performance even for the most prominent examples in these different classes. Therefore we restrict ourselves to conditional heteroscedasticity (CH) models for simulations and real data applications. For the time-varying simultaneous band, to the best of our knowledge, there is no or little simulation results reported. For the tvAR, tvMA, tvARMA and tvRegressions we obtained satisfactory results but they are omitted here to keep this discussion concise.

### *3.6.1 Simulations*

In this section, we study the finite sample coverage probabilities of our SCBs for theoretical coverage $\alpha = 0.9$ and $\alpha = 0.95$ in the following tvARCH(1) and tvGARCH(1,1) models:

(a) $X_i = \sqrt{\alpha_0(i/n) + \alpha_1(i/n)X_{i-1}^2}\zeta_i$, where $\alpha_0(t) = 0.5 + 0.2\sin(2\pi t)$, $\alpha_1(t) = 0.4 + 0.1\sin(\pi t)$,

(b) $X_i = \sigma_i\zeta_i$, $\sigma_i^2 = \alpha_0(i/n) + \alpha_1(i/n)X_{i-1}^2 + \beta_1(i/n)\sigma_{i-1}^2$, where $\alpha_0(t) = 1.0 + 0.2\sin(2\pi t)$, $\alpha_1(t) = 0.45 + 0.1\sin(\pi t)$ and $\beta_1(t) = 0.1 + 0.1\sin(\pi t)$, where $\zeta_i$ is i.i.d. standard normal distributed. For estimation, we choose $K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{[-1,1]}(x)$ to be the Epanechnikov kernel, $n = 2000$ for (a) and $n = 5000$ for (b) and $b_n$ ranging from 0.175 to 0.375 in steps of 0.025 (the optimal bandwidths (3.29) are given by $\hat{b}_n^{(a)} \approx 0.27$ for model (a) and by $\hat{b}_n^{(b)} \approx 0.32$ for model (b)). For each situation, $N = 2000$ replications are performed and it is checked if the obtained SCB contains the true curves in $\mathcal{T}_n = [b_n, 1 - b_n]$. In both models we have $\Lambda(t) = I(t) = V(t)$ and therefore estimate $\Sigma_C(t) = C^\mathsf{T} I(t)^{-1} C$ via replacing $I(t)$ by $\hat{I}_{b_n}(t)$ from (3.28). We obtained the results given in Tables 3.1 and 3.2. It can be seen that the empirical coverage probabilities are reasonably close to the nominal level for bandwidths close to the optimal ones and they do not differ too much for other bandwidths as well.

Table 3.1: Coverage probabilities of the SCB in (a) for $n = 2000$

| | $\alpha = 90\%$ | | | $\alpha = 95\%$ | | |
|---|---|---|---|---|---|---|
| $b_n$ | $\alpha_0$ | $\alpha_1$ | $(\alpha_0, \alpha_1)^\mathsf{T}$ | $\alpha_0$ | $\alpha_1$ | $(\alpha_0, \alpha_1)^\mathsf{T}$ |
| 0.175 | 0.883 | 0.823 | 0.829 | 0.933 | 0.890 | 0.893 |
| 0.200 | 0.882 | 0.871 | 0.889 | 0.932 | 0.914 | 0.939 |
| 0.225 | 0.881 | 0.891 | 0.883 | 0.928 | 0.930 | 0.930 |
| 0.250 | 0.887 | 0.883 | 0.901 | 0.936 | 0.922 | 0.950 |
| 0.275 | 0.875 | 0.893 | 0.893 | 0.931 | 0.935 | 0.939 |
| 0.300 | 0.900 | 0.911 | 0.906 | 0.947 | 0.948 | 0.947 |
| 0.325 | 0.879 | 0.930 | 0.918 | 0.926 | 0.959 | 0.946 |
| 0.350 | 0.886 | 0.921 | 0.913 | 0.935 | 0.946 | 0.946 |

Table 3.2: Coverage probabilities of the SCB in (b) for $n = 5000$

| | $\alpha = 90\%$ | | | | $\alpha = 95\%$ | | | |
|---|---|---|---|---|---|---|---|---|
| $b_n$ | $\alpha_0$ | $\alpha_1$ | $\beta_1$ | $(\alpha_0, \alpha_1, \beta_1)^\mathsf{T}$ | $\alpha_0$ | $\alpha_1$ | $\beta_1$ | $(\alpha_0, \alpha_1, \beta_1)^\mathsf{T}$ |
| 0.200 | 0.840 | 0.895 | 0.813 | 0.785 | 0.905 | 0.938 | 0.895 | 0.853 |
| 0.225 | 0.863 | 0.906 | 0.804 | 0.801 | 0.913 | 0.940 | 0.885 | 0.873 |
| 0.250 | 0.875 | 0.898 | 0.831 | 0.836 | 0.926 | 0.945 | 0.892 | 0.891 |
| 0.275 | 0.877 | 0.931 | 0.848 | 0.855 | 0.924 | 0.962 | 0.893 | 0.898 |
| 0.300 | 0.897 | 0.931 | 0.873 | 0.877 | 0.941 | 0.960 | 0.915 | 0.912 |
| 0.325 | 0.917 | 0.930 | 0.901 | 0.897 | 0.949 | 0.961 | 0.934 | 0.940 |
| 0.350 | 0.912 | 0.942 | 0.910 | 0.906 | 0.948 | 0.968 | 0.946 | 0.940 |
| 0.375 | 0.913 | 0.946 | 0.922 | 0.912 | 0.946 | 0.970 | 0.947 | 0.947 |

### *3.6.2 Applications*

In this section, we consider a few real-data applications of our procedure. As mentioned in Section 3.1, there are abundant results in the literature about time-varying regression but the results for time-varying autoregressive conditional heteroscedastic models are scarce. Thus it is important to evaluate the performance of our constructed SCBs for these type of models in both theoretical and real data scenarios. Among the popular heteroscedastic models, usually GARCH type models are most difficult to estimate due to the recursion of the variance term.

We consider two examples from the class of conditional heteroscedastic models with two types of financial datasets: one foreign exchange and one stock market daily pricing data. As Fryzlewicz et al. [29] proposed, ARCH models have good forecasting ability for currency exchange type data whereas for data coming from the stock market, GARCH models are preferred. Typically, these daily closing price data show unit root behavior and thus instead of using the daily price data, we model the log-return data. The log-return is defined as follows and is close to the relative return

$$Y_i = \log P_i - \log P_{i-1} = \log\left(1 + \frac{P_i - P_{i-1}}{P_{i-1}}\right) \approx \frac{P_i - P_{i-1}}{P_{i-1}},$$

where $P_i$ is the closing price on the $i^{th}$ day. Because of the apparent time-varying nature of volatility these log-return data typically show, conditional heteroscedastic models are used for analysis and forecasting.

## Real data application I: USD/GBP rates

For the first application, we consider a tvARCH($p$) model with $p = 1, 2$. It has the following form

$$Y_i^2 = \sigma_i^2 \zeta_i^2, \qquad \sigma_i^2 = \alpha_0(i/n) + \alpha_1(i/n) Y_{i-1}^2 + \ldots + \alpha_p(i/n) Y_{i-p}^2.$$

Many different exchange rates from 1990-1999 for USD with other currencies were analyzed in [29] using tvARCH($p$) models with $p = 0, 1, 2$. The authors suggested choosing $p = 1$ for USD-GBP exchange rates. To fit tvARCH(1) and tvARCH(2) models, we collect the data from *www.federalreserve.gov/releases/h10/Hist/default1999.htm*. This is a sample of size 2514 and we use cross-validated bandwidth 0.15 and 0.16 for the two models. We only show the results for the fit with tvARCH(1) here. We observed that the estimates for the parameter curves $\alpha_0(\cdot)$ and $\alpha_1(\cdot)$ for tvARCH(2) model are very similar to that from the tvARCH(1) fit and thus it indicates against including the extra $\alpha_2(\cdot)$ parameter in our model. We also provide the plots for the log-returns and ACF plot of squared sample that shows evidence of conditional heteroscedasticity.

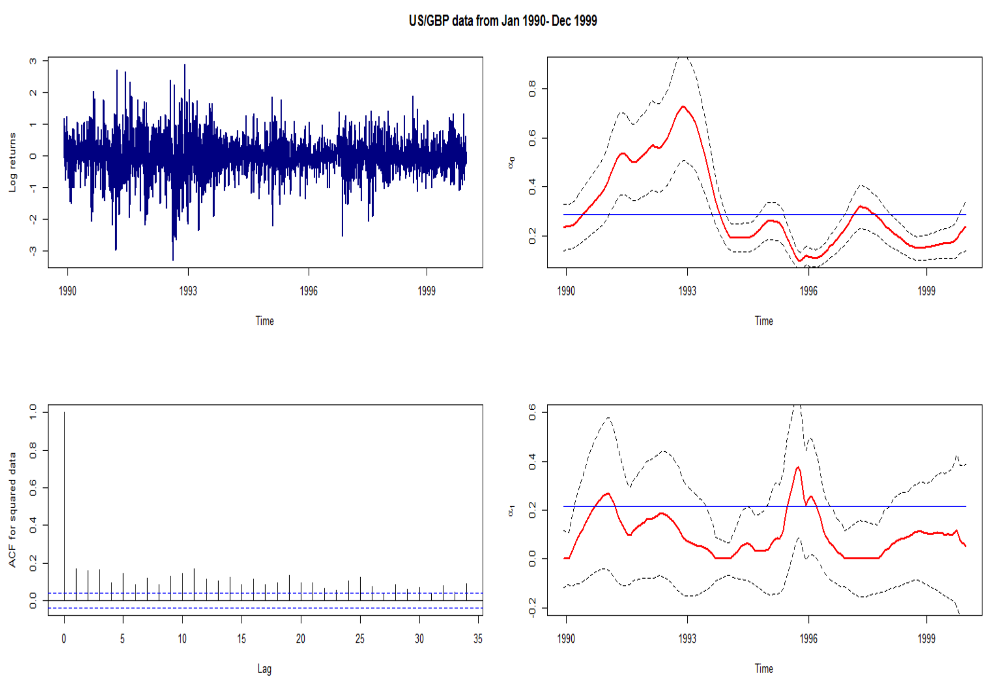



Figure 3.1: Analysis of USD/GBP data from Jan 1990 to Dec 1999. Top left: Log-returns. Bottom left: ACF plot. Right panel: Estimates of the parameters $\alpha_0(\cdot)$, $\alpha_1(\cdot)$, respectively (red) with 95% SCBs (dashed) and estimates of the parameters assuming constancy (blue).

Based on Figure 3.1 time-constancy for both the parameter curves is rejected at 5% level of significance. For $\alpha_1(\cdot)$, the estimate generally stays below the stationary fit. Also, one can see from the plot of actual log-returns that there are large shocks from 1990 to 1993 compared to those seen in 1993-1999. This can be explained through the high (low) values shown for the estimated curve $\alpha_0(\cdot)$ for the time-period 1990-1993 (1993-1999).

## Real data application II: Merval index data

In the empirical analysis of log-return for stock market data, however, as Palm [58] and others suggest, lower order GARCH have been often found to account sufficiently for the conditional heteroscedasticity. Moreover, GARCH(1,1) and in a very few cases GARCH(1,2) and GARCH(2,1) models are used and higher order GARCH models are typically not necessary. Another advantage of using GARCH(1,1) over ARCH($p$) models is that one need not worry about choosing a proper lag $p$ as GARCH(1,1) can be thought as an ARCH model with $p = \infty$. In this subsection, we implement a time-varying version of GARCH(1,1) and obtain the bootstrapped SCB. A tvGARCH(1,1) model has the following form:

$$Y_i^2 = \sigma_i^2 \zeta_i^2, \qquad \sigma_i^2 = \alpha_0(i/n) + \alpha_1(i/n) Y_{i-1}^2 + \beta_1(i/n) \sigma_{i-1}^2.$$

As our second example, we choose to analyze the log returns of Merval index data from Argentina for the time period January 2010 to October 2017. In Tagliafichi Ricardo [70], the author considered daily returns for the period 1990-2000 and mentioned how time-varying nature can be present in the parameters of the GARCH(1,1) model he fits. In particular, he chose to split this time horizon in 3 parts and computed the estimates separately to compare with the overall estimates. This index was remodelled in 2000 and has increased about 1000% in each five years. We considered daily log returns from January 2010 to October 2017 in this analysis. Our cross-validated bandwidth is 0.29 for this data of size 1919. As one can see from Figure 3.2, the time series show significant lags in its ACF plot after squaring; indicating conditional heteroscedasticity.

One can see that the estimates for $\alpha_1(\cdot)$ and $\beta_1(\cdot)$ generally over-estimates and under-estimates their corresponding time-constant fits respectively. Moreover, the bands for $\alpha_1(\cdot)$ show somewhat bell-curved shape which is an important find in terms of economic impli-

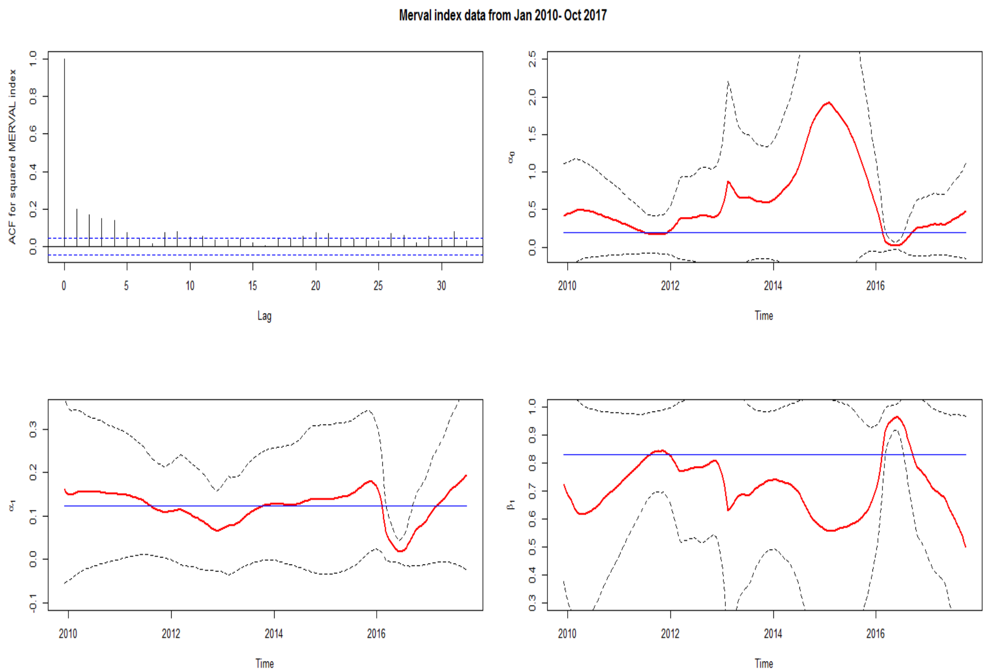



Figure 3.2: Analysis of MERVAL index data from Jan 2010 to Oct 2017. Top left: ACF plot. Top right, bottom left, bottom right: Estimates of the parameters $\alpha_0(\cdot)$, $\alpha_1(\cdot)$ and $\beta_1(\cdot)$, respectively (red) with SCBs (dashed) and estimates of the parameters assuming constancy (blue).

cations. For all the three parameters, however, since it is possible to find a horizontal line passing through the corridor created by the bands, the hypothesis of time-constancy cannot be rejected at 5% level of significance. But specific patterns such as those seen in the simultaneous bands for $\alpha_1(\cdot)$ and $\beta_1(\cdot)$ cannot be implied from just a time-constant fit.

## 3.7 Proofs

For $\eta = (\eta_1, \eta_2) \in \Theta \times (\Theta' \cdot b_n) =: E_n$, define

$$L^{\circ,c}_{n,b_n}(t,\eta) := (nb_n)^{-1} \sum_{i=1}^{n} K_{b_n}(i/n - t)\ell(Z_i^c, \eta_1 + \eta_2(i/n - t)b_n^{-1})$$

and $\hat{L}^{\circ}_{n,b_n}$, $L^{\circ}_{n,b_n}$ similarly as $L^{\circ,c}_{n,b_n}$ but with $Z_i^c$ replaced by $\tilde{Z}_i(i/n)$ or $Z_i$, respectively. We define $\eta_{b_n}(t) = (\theta(t)^{\mathsf{T}}, b_n\theta'(t)^{\mathsf{T}})^{\mathsf{T}}$ as the value which should be estimated by $\hat{\eta}_{b_n}(t) = (\hat{\theta}_{b_n}(t)^{\mathsf{T}}, b_n\widehat{\theta}'_{b_n}(t)^{\mathsf{T}})^{\mathsf{T}}$, the minimizer of $L^{\circ}_{n,b_n}(t,\eta)$. In the proof of Theorem 3.3.1, it is shown that $L^{\circ}_{n,b_n}(t,\eta)$ converges to $L^{\circ}(t,\eta) := \int_{-1}^{1} K(x)L(t, \eta_1 + \eta_2 x)dx$. If $\chi \in \mathbb{R}^{\mathbb{N}}$, recall that $\hat{\chi} = (1, \chi) \in \mathbb{R}^{\mathbb{N}_0}$.

### 3.7.1 Proofs of Section 3.3

*Proof of Theorem 3.3.1.* The proof is similar to the proof of Theorems 5.2 and 5.4 in [13].

(i) Fix $t \in (0,1)$. By Lemma 3.7.4(ii) applied to $\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we have

$$\sup_{\eta \in E_n} |\hat{L}^{\circ}_{n,b_n}(t,\eta) - \mathbb{E}\hat{L}^{\circ}_{n,b_n}(t,\eta)| = o_{\mathbb{P}}(1).$$

Applying Lemma 3.7.5 to $\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we obtain

$$\sup_{\eta \in E_n} |\mathbb{E}\hat{L}^{\circ}_{n,b_n}(t,\eta) - L^{\circ}(t,\eta)| = O(b_n + (nb_n)^{-1}) = o(1),$$

where $L^{\circ}(t,\eta) = \int_{-1}^{1} K(x)L(t, \eta_1 + \eta_2 x)dx$. By Lemma 3.7.4(i), we have

$$\Big\| \sup_{\eta \in E_n} |L^{\circ,c}_{n,b_n}(t,\eta) - \hat{L}^{\circ}_{n,b_n}(t,\eta)| \Big\|_1 = O((nb_n)^{-1}),$$

and thus

$$\sup_{\eta \in E_n} |L^{\circ,c}_{n,b_n}(t,\eta) - L^{\circ}(t,\eta)| = o_{\mathbb{P}}(1).$$

By Lemma 3.7.1, $\eta \mapsto L^{\circ}(t,\eta)$ is Lipschitz continuous. Since $\theta(t)$ is the unique minimizer of $\theta \mapsto L(t,\theta)$, we conclude that $(\eta_1,\eta_2) = (\theta(t),0)$ is the unique minimizer of $\eta \mapsto L^{\circ}(t,\eta)$. Since $\hat{\eta}_{b_n}(t) = (\theta_{b_n}(t)^{\mathsf{T}}, b_n\theta'_{b_n}(t)^{\mathsf{T}})^{\mathsf{T}}$ is a minimizer of $L^{\circ,c}_{n,b_n}(t,\eta)$, standard arguments yield

$$\hat{\eta}_{b_n}(t) = (\hat{\theta}_{b_n}(t)^{\mathsf{T}}, b_n\widehat{\theta'}_{b_n}(t)^{\mathsf{T}})^{\mathsf{T}} = (\theta(t)^{\mathsf{T}},0)^{\mathsf{T}} + o_{\mathbb{P}}(1). \tag{3.37}$$

We now show that $\widehat{\theta'}_{b_n}(t) - \theta'(t) = o_{\mathbb{P}}(1)$ if $nb_n^3 \to \infty$. The following argumentation is also a preparation for the proof of (ii),(iii). By (3.37), we have that $\hat{\eta}_{b_n}(t)$ is in the interior of $\Theta \times (\Theta' \cdot b_n)$ with probability tending to 1 (since it converges to $(\theta(t)^{\mathsf{T}},0)$ in probability), thus $\nabla_\eta L^{\circ,c}_{n,b_n}(t,\hat{\eta}_{b_n}(t)) = 0$ with probability tending to 1. By a Taylor expansion we obtain

$$\hat{\eta}_{b_n}(t) - \eta_{b_n}(t) = -\big[\nabla^2_\eta L^{\circ,c}_{n,b_n}(t,\bar{\eta}(t))\big]^{-1} \cdot \nabla_\eta L^{\circ,c}_{n,b_n}(t,\eta_{b_n}(t)), \tag{3.38}$$

with some $\bar{\eta}(t) \in \Theta \times (\Theta' \cdot b_n)$ satisfying $|\bar{\eta}(t) - \eta_{b_n}(t)| \le |\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)|$. Let $V(t,\theta) := \mathbb{E}\nabla^2_\theta \ell(\tilde{Z}_0(t),\theta)$. Since $g = \nabla^2_\theta \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we can use similar arguments as in (i) to obtain

$$\sup_{\eta \in E_n} |\nabla^2_\eta L^{\circ,c}_{n,b_n}(t,\eta) - V^{\circ}(t,\eta)| = o_{\mathbb{P}}(1), \tag{3.39}$$

where

$$V^{\circ}(t,\eta) = \int_{-1}^{1} K(x) \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \otimes V(t,\eta_1 + \eta_2 x)dx. \tag{3.40}$$

Let

$$V^{\circ}(t) := \begin{pmatrix} 1 & 0 \\ 0 & \mu_{K,2} \end{pmatrix} \otimes V(t). \tag{3.41}$$

From (i), we have $|\bar{\eta}(t) - \eta_{b_n}(t)| \le |\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)| = o_{\mathbb{P}}(1)$, i.e. $\bar{\eta}_1(t) = \theta(t) + o_{\mathbb{P}}(1)$ and $\bar{\eta}_2(t) = b_n\theta'(t) + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$. By continuity of $\theta \mapsto V(t,\theta)$ and (3.39), we conclude that

$$\nabla^2_\theta L^{\circ,c}_{n,b_n}(t,\bar{\eta}(t)) = V^{\circ}(t,\bar{\eta}(t)) + o_{\mathbb{P}}(1) = V^{\circ}(t) + o_{\mathbb{P}}(1). \tag{3.42}$$

By Lemma 3.7.4(i), we have

$$\|\nabla_\eta L^{\circ,c}_{n,b_n}(t,\eta_{b_n}(t)) - \nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t))\|_1 = O((nb_n)^{-1}). \tag{3.43}$$

With (3.38), (3.42) and (3.43) we obtain

$$\begin{aligned}
&\begin{pmatrix} \sqrt{nb_n}(\hat{\theta}_{b_n}(t)-\theta(t)) \\ \sqrt{nb_n^3}(\widehat{\theta}'_{b_n}(t)-\theta'(t)) \end{pmatrix} = \sqrt{nb_n}(\hat{\eta}_{b_n}(t)-\eta_{b_n}(t)) \\
= \quad & -V^\circ(t)^{-1}\sqrt{nb_n}\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) + o_{\mathbb{P}}(1) \\
= \quad & -V^\circ(t)^{-1}\sqrt{nb_n}\big\{\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t))\big\} \\
& \qquad -V^\circ(t)^{-1}\begin{pmatrix} \sqrt{nb_n}\mathbb{E}\nabla_{\eta_1}\hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) \\ \sqrt{nb_n^3}b_n^{-1}\mathbb{E}\nabla_{\eta_2}\hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) \end{pmatrix} + o_{\mathbb{P}}(1).
\end{aligned} \tag{3.44}$$

By (3.44), it is enough to show the two stochastic convergences

$$b_n^{-1}\big\{\nabla_{\eta_2} L^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t))\big\} \;=\; o_{\mathbb{P}}(1), \tag{3.45}$$

$$b_n^{-1}\mathbb{E}\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) \;=\; o_{\mathbb{P}}(1). \tag{3.46}$$

Using (3.73) from the proof of Lemma 3.7.4(ii) (applied to each component of $\nabla\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ with $\hat{K}(x) = K(x)x$), we obtain

$$\big\|\nabla_{\eta_2} L^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t))\big\|_2 = O((nb_n)^{-1/2}),$$

which shows (3.45) due to $nb_n^3 \to \infty$. Using the intermediate result (3.83) in the proof of 3.7.5, we have

$$\mathbb{E}\nabla_\eta \hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) = O(b_n^3 + n^{-1} + (nb_n)^{-1}),$$

we obtain (3.46) due to $nb_n^3 \to \infty$, which completes the proof of (i).

(ii),(iii) Our aim is to show asymptotic normality of the term in the second to last line of (3.44). Define $U_{i,n}(t) := (K_{b_n}(i/n-t), K_{b_n}(i/n-t)(i/n-t)b_n^{-1})^{\mathsf{T}}$. Following the proof idea of Theorem 3(ii) in [75], let $m \geq 1$ and define

$$S_{n,b_n,m}(t) := \sum_{l=0}^{m-1} (nb_n)^{-1/2} \sum_{i=1}^{n} U_{i,n}(t) \otimes P_{i-l} \nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(t) + \theta'(t)(i/n-t)).$$

Recall $\eta_{b_n}(t) = (\theta(t)^{\mathsf{T}}, b_n\theta'(t)^{\mathsf{T}})^{\mathsf{T}}$. Write shortly LIM for $\limsup_{n\to\infty} \limsup_{m\to\infty}$. Then we have for each component $j = 1, \ldots, 2d_\Theta$, that

$$\begin{aligned}
& \text{LIM } \|S_{n,b_n,m}(t)_j - (nb_n)^{1/2}\big\{\nabla_{\eta_j}\hat{L}^\circ_{n,b_n}(t, \eta_{b_n}(t)) - \mathbb{E}\nabla_{\eta_j}\hat{L}^\circ_{n,b_n}(t, \eta_{b_n}(t))\big\}\|_2 \\
\leq\ & \text{LIM } (nb_n)^{-1/2} \sum_{l=m}^{\infty} \Big\| \sum_{i=1}^{n} (U_{i,n}(t) \otimes P_{i-l}\nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(t) + \theta'(t)(i/n-t)))_j \Big\|_2 \\
=\ & \text{LIM } (nb_n)^{-1/2} \Big( \sum_{i=1}^{n} \big\|(U_{i,n}(t) \otimes P_{i-l}\nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(t) + \theta'(t)(i/n-t)))_j\big\|_2^2 \Big)^{1/2} \\
\leq\ & |K|_\infty \text{LIM} \sum_{l=m}^{\infty} \sup_{t\in[0,1]} \sup_{i,\theta} \delta_2^{\nabla_{\theta_i}\ell(\tilde{Z}(t),\theta)}(l) = 0, \qquad (3.47)
\end{aligned}$$

by Lemma 3.7.3(i). Define $M_i(t) := (nb_n)^{-1/2} \sum_{l=0}^{m-1} U_{i,n}(t) \otimes P_i \nabla_\theta \ell(\tilde{Z}_{i+l}((i+l)/n), \theta(t) + \theta'(t)(i/n-t))$ and $\tilde{S}_{n,b_n,m}(t) := \sum_{i=1}^{n} M_i(t)$. Since $m$ is finite and $\nabla_\theta \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, a straightforward calculation shows that for each component $j = 1, \ldots, 2d_\Theta$,

$$\|S_{n,b_n,m}(t)_j - \tilde{S}_{n,b_n,m}(t)_j\|_2 = O((nb_n)^{-1/2}). \qquad (3.48)$$

Let $a = (a_1^{\mathsf{T}}, a_2^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{d_\Theta} \times \mathbb{R}^{d_\Theta}$. We want to apply a central limit theorem for martingale differences to $a^{\mathsf{T}}\tilde{S}_{n,b_n,m}(t)$. Put

$$\Sigma_m := \sum_{l_1,l_2=0}^{m-1} \mathbb{E}\big[P_0\nabla_\theta \ell(\tilde{Z}_{l_1}(t), \theta(t)) P_0 \nabla_\theta \ell(\tilde{Z}_{l_2}(t), \theta(t))^{\mathsf{T}}\big] = \mathrm{Cov}\Big(\sum_{l=0}^{m-1} P_0 \nabla_\theta \ell(\tilde{Z}_l(t), \theta(t))\Big).$$

Since $\nabla_\theta \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we obtain by Lemma 3.7.1:

$$\begin{aligned}
& \sup_{|i/n-t|\le b_n} \|P_i \nabla_\theta \ell(\tilde{Z}_{i+l_1}((i+l_1)/n), \theta(t) + \theta'(t)(i/n-t)) - P_i \nabla_\theta \ell(\tilde{Z}_{i+l_1}(t), \theta(t))\|_2 \\
\le & \sup_{|i/n-t|\le b_n} \|\nabla_\theta \ell(\tilde{Z}_0((i+l_1)/n), \theta(t) + \theta'(t)(i/n-t)) - \nabla_\theta \ell(\tilde{Z}_0(t), \theta(t))\|_2 \\
= & O(b_n + n^{-1})
\end{aligned}$$

and

$$\sup_i \|P_i \nabla_\theta \ell(\tilde{Z}_{i+l_2}((i+l_2)/n), \theta(t) + \theta'(t)(i/n-t))\|_2 \le \sup_\theta \sup_{t\in[0,1]} \|\nabla_\theta \ell(\tilde{Z}_0(t), \theta)\|_2 < \infty.$$

We therefore have by Hölder's and Markov's inequality that

$$\begin{aligned}
& \sum_{i=1}^n M_i(t) M_i(t)^\mathsf{T} \\
= & \sum_{l_1,l_2=0}^{m-1} (nb_n)^{-1} \sum_{i=1}^n K_{b_n}(i/n-t)^2 \begin{pmatrix} 1 & (i/n-t)b_n^{-1} \\ (i/n-t)b_n^{-1} & (i/n-t)^2 b_n^{-2} \end{pmatrix} \\
& \qquad \otimes \{P_i \nabla_\theta \ell(\tilde{Z}_{i+l_1}(t), \theta(t)) \cdot P_i \nabla_\theta \ell(\tilde{Z}_{i+l_2}(t), \theta(t))^\mathsf{T}\} + O_\mathbb{P}(b_n + n^{-1}) \\
= & \begin{pmatrix} \sigma_{K,0}^2 & 0 \\ 0 & \sigma_{K,2}^2 \end{pmatrix} \otimes \Sigma_m + o_\mathbb{P}(1).
\end{aligned}$$

The last step is due to Lemma A.2 in [15]. It remains to show a Lindeberg-type condition for $M_i(t)$. Put $\tilde{M}_{ij,l} := P_i \nabla_{\theta_j} \ell(\tilde{Z}_{i+l}((i+l)/n), \theta(t) + \theta'(t)(i/n-t))$. There exists some constant $C > 0$ such that for $j = 1, \ldots, d_\Theta$ and $\iota > 0$,

$$\begin{aligned}
& \sum_{i=1}^n \mathbb{E}\big[M_i(t)_j^2 \mathbb{1}_{\{|M_i(t)_j| > \iota\}}\big] \\
\le & C(nb_n)^{-1} \sum_{l=0}^{m-1} \sum_{i=1}^n K_{b_n}(i/n-t)^2 \mathbb{E}\big[\tilde{M}_{ij,l}^2 \mathbb{1}_{\{|K|_\infty |\tilde{M}_{ij,l}| > \iota(nb_n)^{1/2}\}}\big]. \qquad (3.49)
\end{aligned}$$

Using Hölder's inequality we have

$$\mathbb{E}\big[\tilde{M}_{ij,l}^2 \mathbb{1}_{\{|K|_\infty |\tilde{M}_{ij,l}| > \iota(nb_n)^{1/2}\}}\big] \le \mathbb{E}[|\tilde{M}_{ij,l}|^{2+a}]^{2/(2+a)} \mathbb{P}(|K|_\infty |\tilde{M}_{ij,l}| > \iota(nb_n)^{1/2})^{a/(2+a)},$$

which tends to zero using Markov's inequality, Lemma 3.7.1(i) applied to $\nabla_\theta \ell$ and the assumption $\sup_{j\in\mathbb{N}_0}\sup_{t\in[0,1]}\|\tilde{Z}_0(t)_j\|_{(2+a)M} < \infty$. This shows that (3.49) is tending to 0. The proof for $j = d_\Theta + 1, \dots, 2d_\Theta$ is similar. From Theorem 18.2 in Billingsley [6] and the Cramer-Wold device we obtain

$$\tilde{S}_{n,b_n,m}(t) \xrightarrow{d} N\Big(0, \begin{pmatrix} \sigma_{K,0}^2 & 0 \\ 0 & \sigma_{K,2}^2 \end{pmatrix} \otimes \Sigma_m\Big). \tag{3.50}$$

Using (3.47), (3.48), (3.50) and $\Sigma_m \to \Lambda(t)$ $(m \to \infty)$, we obtain

$$(nb_n)^{1/2}\{\nabla_\eta \hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t))\} \xrightarrow{d} N\Big(0, \begin{pmatrix} \sigma_{K,0}^2 & 0 \\ 0 & \sigma_{K,2}^2 \end{pmatrix} \otimes \Lambda(t)\Big). \tag{3.51}$$

Using (3.51), the expansion (3.44) and Lemma 3.7.6, we obtain the result provided that $nb_n^7 \to 0$ for (ii) and $nb_n^9 \to 0$ for (iii). □

*Proof of Theorem 3.3.4.* (i),(ii) By Lemma 3.7.4(i),(iii) and Lemma 3.7.5 applied to $g = \ell$, we have that

$$\sup_{t\in\mathcal{T}_n}\sup_{\eta\in E_n} |L_{n,b_n}^{\circ,c}(t,\eta) - L^\circ(t,\eta)| = O_{\mathbb{P}}(\beta_n + (nb_n)^{-1}) + O(b_n),$$

i.e. $L_{n,b_n}^{\circ,c}(t,\eta)$ converges to $L^\circ(t,\eta)$ uniformly in $t, \eta$ if $b_n = o(1)$ and $\beta_n = o(1)$. It was already seen in the proof of Theorem 3.3.1 that $L^\circ(t,\eta)$ is continuous w.r.t. $\eta$ and uniquely minimized by $\eta = (\theta(t)^\mathsf{T}, 0)^\mathsf{T}$. Standard arguments give

$$\sup_{t\in\mathcal{T}_n} |\hat{\eta}_{b_n}(t) - \eta_n(t)| = o_{\mathbb{P}}(1). \tag{3.52}$$

Thus for $n$ large enough, $\hat{\eta}_{b_n}(t)$ is in the interior of $E_n$ uniformly in $t$. By a Taylor expansion, we obtain for each $t \in \mathcal{T}_n$:

$$\hat{\eta}_{b_n}(t) - \eta_{b_n}(t) = -\big[V^\circ(t) + R_{n,b_n}(t)\big]^{-1} \cdot \nabla_\eta L^{\circ,c}_{n,b_n}(t, \eta_{b_n}(t)), \tag{3.53}$$

where $R_{n,b_n}(t) = \nabla^2_\theta L^{\circ,c}_{n,b_n}(t, \bar{\eta}(t)) - V^\circ(t)$ with some $\bar{\eta}(t) \in E_n$ satisfying $|\bar{\eta}(t) - \eta_{b_n}(t)|_1 \leq |\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)|_1$ and $V^\circ(t)$ is defined in (3.41).

By Lemma 3.7.4(i),(iii) and Lemma 3.7.5 applied to $g = \nabla^2_\theta \ell$ and $\hat{K}(x) = K(x)$ or $\hat{K}(x) = K(x)x^2$, respectively, we have

$$\sup_{t \in \mathcal{T}_n} \sup_{\theta \in \Theta} |\nabla^2_\eta L^{\circ,c}_{n,b_n}(t,\eta) - V^\circ(t,\eta)| = O_{\mathbb{P}}(\beta_n + (nb_n)^{-1}) + O(b_n), \tag{3.54}$$

where $V^\circ(t,\eta)$ is defined in (3.40).

Note that $\mathbb{E}\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) = 0$ by Assumption 3.2.1(A3) in connection with 3.2.1(A1). By Lemma 3.7.3, we have $\sup_{\theta,t} \delta_4^{\nabla_{\theta_j}\ell(\tilde{Z}(t),\theta)}(i) = O(i^{-(1+\gamma)})$ for each $j = 1, \dots, d_\Theta$. Using Lemma 3.7.1, we see that the assumptions of Lemma 3.7.9 are fulfilled and thus:

$$\sup_{t \in \mathcal{T}_n} \big|(nb_n)^{-1} \sum_{i=1}^n K_{b_n}(i/n - t) \nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n))\big| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n)). \tag{3.55}$$

With Lemma 3.7.7, we obtain

$$\sup_{t \in \mathcal{T}_n} \big|\nabla_\eta \hat{L}^\circ_{n,b_n}(t, \eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^\circ_{n,b_n}(t, \eta_{b_n}(t))\big| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n) + \beta_n b_n^2).$$

Since $\mathbb{E}\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) = 0$, we obtain with Lemma 3.7.6(i),(ii) and Lemma 3.7.4(i):

$$\sup_{t \in \mathcal{T}_n} |\nabla_{\eta_j} L^{\circ,c}_{n,b_n}(t, \eta_{b_n}(t))| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n) + (nb_n)^{-1} + \beta_n b_n^2 + b_n^{1+j}), \tag{3.56}$$

where $j = 1, 2$. Since $\theta \mapsto V(t,\theta) = \mathbb{E}\nabla^2_\theta \ell(\tilde{Z}_0(t), \theta)$ is Lipschitz continuous due to $\nabla^2 \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ and Lemma 3.7.1, the same holds for $\eta \mapsto V^\circ(t,\eta)$. We conclude that with some constant $C > 0$,

$$\sup_{t\in\mathcal{T}_n} |R_{n,b_n}(t)| \leq \sup_{t\in\mathcal{T}_n} \sup_{\eta\in E_n} |\nabla^2_\eta L^{\circ,c}_{n,b_n}(t,\eta) - V^\circ(t,\eta)| + C \sup_{t\in\mathcal{T}_n} |\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)|. \tag{3.57}$$

Inserting (3.56), (3.57) and (3.52) into (3.53), we obtain

$$\sup_{t\in\mathcal{T}_n} |\hat{\eta}_{b_n,j}(t) - \eta_{b_n,j}(t)| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n) + (nb_n)^{-1} + \beta_n b_n^2 + b_n^{1+j}), \tag{3.58}$$

where $j = 1, 2$. Inserting (3.58), (3.54) into (3.57), we get $\sup_{t\in\mathcal{T}_n} |R_{n,b_n}(t)| = O_{\mathbb{P}}(\beta_n + b_n + (nb_n)^{-1})$. Together with

$$\begin{aligned}
& \big|V^\circ(t)\big(\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)\big) - \nabla L^{\circ,c}_{n,b_n}(t, \eta_{b_n}(t))\big| \\
\leq\ & \big|\big[I_{2k\times 2k} + V^\circ(t)^{-1}R_{n,b_n}(t)\big]^{-1} - I^{-1}_{k\times k}\big| \cdot |\nabla_\eta L^{\circ,c}_{n,b_n}(t, \eta_n(t))| \\
\leq\ & \big|\big[I_{2k\times 2k} + V^\circ(t)^{-1}R_{n,b_n}(t)\big]^{-1}\big| \cdot \big|V^\circ(t)^{-1}R_{n,b_n}(t)\big| \cdot |\nabla_\eta L^{\circ,c}_{n,b_n}(t, \eta_{b_n}(t))|,
\end{aligned}$$

and (3.56) we have (3.17) and (3.19). The second result (3.18) and (3.20) follows from Lemma 3.7.4(i), Lemma 3.7.7 and Lemma 3.7.6. □

**Lemma 3.7.1.** *Let $q > 0$. Let $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ and $M := \max\{M_x, M_y\}$. Let $\hat{Y}, \hat{Y}'$ be random variables and $\hat{X} = (\hat{X}_j)_{j\in\mathbb{N}}$, $\hat{X}' = (\hat{X}'_j)_{j\in\mathbb{N}}$ be sequences of random variables. Assume that there exists some $D > 0$ such that uniformly in $j \in \mathbb{N}$,*

$$\|\hat{Y}\|_{qM}, \qquad \|\hat{Y}'_j\|_{qM}, \qquad \|\hat{X}_j\|_{qM}, \qquad \|\hat{X}'_j\|_{qM} \leq D.$$

*Let $\hat{Z} = (\hat{Y}, \hat{X})$, $\hat{Z}' = (\hat{Y}', \hat{X}')$. Then there exists some constant $C > 0$ only dependent on $M_x$, $M_y$, $D$, $\chi$ and $\tilde{D}$ (only in (ii)) such that*

*(i)*

$$\|\sup_{\theta\in\Theta}|g(\hat{Z},\theta)-g(\hat{Z}',\theta)|\|_q \leq \bar{C}\cdot C\sum_{j=0}^{\infty}\hat{\chi}_j\|\hat{Z}_j-\hat{Z}_j'\|_{qM}, \tag{3.59}$$

$$\|\sup_{\theta\neq\theta'}\frac{|g(\hat{Z},\theta)-g(\hat{Z},\theta')|}{|\theta-\theta'|_1}\|_q \leq \bar{C}\cdot C, \tag{3.60}$$

$$\|\sup_{\theta\in\Theta}|g(\hat{Z},\theta)|\|_q \leq \bar{C}\cdot C. \tag{3.61}$$

*(ii) If additionally,* $\mathbb{E}[|\hat{Y}-\hat{Y}'|^{qM_y}|\sigma(\hat{X},\hat{X}')]\leq\tilde{D}|\hat{X}-\hat{X}'|_{\chi,1}^{qM_y}$ *with some constant* $\tilde{D}>0$*, then*

$$\|\sup_{\theta\in\Theta}|g(\hat{Z},\theta)-g(\hat{Z}',\theta)|\|_q\leq\bar{C}\cdot C\sum_{j=1}^{\infty}\chi_j\|\hat{X}_j-\hat{X}_j'\|_{qM}. \tag{3.62}$$

*Proof of Lemma 3.7.1.* During the proofs, we consider $M_y, M_x\geq 2$ and thus $M\geq 2$. In the case $M_y=1$ or $M_x=1$, the proofs are easier since some terms do not show up.

(i) Note that $R_{M_y-1,M_x-1}$ is a polynomial in $|x|_{\chi,1}$, $|y|$ with (joint) degree at most $M-1$. Since

$$R_{M_y-1,M_x-1}(\hat{Z})=\sum_{k+l\leq M-1,0\leq k\leq M_y-1,0\leq l\leq M_x-1}|\hat{Y}|^k|\hat{X}|_{\chi,1}^l,$$

we have by Hölder's inequality,

$$\begin{aligned}
&\|R_{M_y-1,M_x-1}(\hat{Z})\|_{qM/(M-1)}\\
\leq\ &\sum_{k+l\leq M-1,0\leq k\leq M_y-1,0\leq l\leq M_x-1}\Big(\sum_{i=1}^{\infty}\chi_i\|\hat{X}_i\|_{qM}\Big)^l\|\hat{Y}\|_{qM}^k\\
\leq\ &\sum_{0\leq k+l\leq M-1}(|\chi|_1 D)^l D^k\leq(1+D(|\chi|_1+1))^{M-1}.
\end{aligned}$$

Therefore:

$$
\begin{aligned}
& \| \sup_{\theta\in\Theta} |g(\hat{Z},\theta) - g(\hat{Z}',\theta)| \|_q \\
\leq\ & \bar{C}\big\| |\hat{Y} - \hat{Y}'| (R_{M_y-1,M_x-1}(\hat{Y},\hat{X}) + R_{M_y-1,M_x-1}(\hat{Y}',\hat{X}))\big\|_q \\
& \quad + \big\| |\hat{X} - \hat{X}'|_{\chi,1} \cdot (R_{M_y-1,M_x-1}(\hat{Y},\hat{X}) + R_{M_y-1,M_x-1}(\hat{Y},\hat{X}'))\big\|_q \qquad (3.63) \\
\leq\ & \bar{C}\|\hat{Y} - \hat{Y}'\|_{qM} \cdot \big( \|R_{M_y-1,M_x-1}(\hat{Y},\hat{X})\|_{qM/(M-1)} \\
& \qquad\qquad + \|R_{M_y-1,M_x-1}(\hat{Y}',\hat{X})\|_{qM/(M-1)} \big) \\
& \quad + \bar{C}\| |\hat{X} - \hat{X}'|_{\chi,1}\|_{qM} \big( \|R_{M_y-1,M_x-1}(\hat{Y},\hat{X})\|_{qM/(M-1)} \\
& \qquad\qquad + \|R_{M_y-1,M_x-1}(\hat{Y},\hat{X}')\|_{qM/(M-1)} \big) \\
\leq\ & 2\bar{C}(1 + D(|\chi|_1 + 1))^{M-1} \big( \|\hat{Y} - \hat{Y}'\|_{qM} + \sum_{j=1}^{\infty} \chi_j \|\hat{X}_j - \hat{X}_j'\|_{qM} \big).
\end{aligned}
$$

Similarly, $R_{M_y,M_x}$ is a polynomial in $|x|_{\chi,1}$ and $|y|$ with (joint) degree at most $M$, thus

$$
\begin{aligned}
\|g(\hat{Z},\theta) - g(\hat{Z},\theta')\|_q &\leq \bar{C}|\theta - \theta'|_1 \|R_{M_y,M_x}(\hat{Y},\hat{X})\|_{qM} \\
&\leq \bar{C}(1 + D(|\chi|_1 + 1))^{M} |\theta - \theta'|_1.
\end{aligned}
$$

The proof of (3.61) is obvious from (3.59).

(ii) We first obtain (3.63) as before. The second summand has the upper bound

$$
2\bar{C}(1 + D(|\chi|_1 + 1))^{M-1} \sum_{j=1}^{\infty} \chi_j \|\hat{X}_j - \hat{X}_j'\|_{qM}.
$$

For the first summand in (3.63), notice that

$$
\begin{aligned}
& \big\| |\hat{Y} - \hat{Y}'| \cdot R_{M_y-1,M_x-1}(\hat{Y},\hat{X})\big\|_q \\
\leq\ & \sum_{k+l\leq M-1, 0\leq k\leq M_y-1, 0\leq l\leq M_x-1} \| |\hat{Y} - \hat{Y}'| \cdot |\hat{Y}|^k \cdot |\hat{X}|_{\chi,1}^l \|_q.
\end{aligned}
$$

By Hölder's inequality for conditional expectations,

$$
\begin{aligned}
& \big\| |\hat{Y} - \hat{Y}'| \cdot |\hat{Y}|^k |\hat{X}|_{\chi,1}^l \big\|_q \\
= \; & \mathbb{E}\big[\mathbb{E}[|\hat{Y} - \hat{Y}'|^q |\hat{Y}|^{qk} | \sigma(\hat{X}, \hat{X}')] \cdot |\hat{X}|_{\chi,1}^{ql}\big]^{1/q} \\
\leq \; & \mathbb{E}[\mathbb{E}[|\hat{Y} - \hat{Y}'|^{qM_y} | \sigma(\hat{X}, \hat{X}')]^{1/M_y} \mathbb{E}[|\hat{Y}|^{qkM_y/(M_y-1)} | \sigma(\hat{X}, \hat{X}')]^{(M_y-1)/M_y} |\hat{X}|_{\chi,1}^{ql}]^{1/q} \\
=: \; & \mathbb{E}[A_1 \cdot A_2 \cdot A_3]^{1/q}.
\end{aligned}
$$

By the additional condition, we have $A_1 \leq \tilde{D}^{1/M_y} |\hat{X} - \hat{X}'|_{\chi,1}^q$. By Hölder's inequality,

$$
\mathbb{E}[A_1 \cdot A_2 \cdot A_3]^{1/q} \leq \mathbb{E}[A_1^M]^{1/(qM)} \mathbb{E}[A_2^{M/k}]^{k/(qM)} \mathbb{E}[A_3^{M/(M-k-1)}]^{(M-k-1)/(qM)}.
$$

We have $\mathbb{E}[A_1^M]^{1/M} \leq \tilde{D}^{1/M_y} \| |\hat{X} - \hat{X}'|_{\chi,1} \|_{qM}^q$,

$$
\mathbb{E}[A_3^{M/(M-k-1)}]^{(M-k-1)/M} = \| |\hat{X}|_{\chi,1} \|_{Mql/(M-k-1)}^{ql} \leq \| |\hat{X}|_{\chi,1} \|_{qM}^{ql}
$$

and by Jensen's inequality for conditional expectations (note that $\frac{M_y-1}{k} \frac{M}{M_y} \geq 1$),

$$
\mathbb{E}[A_2^{M/k}]^{k/M} \leq \mathbb{E}[\mathbb{E}[|\hat{Y}|^{qkM_y/(M_y-1) \cdot \frac{M_y-1}{M_y} \frac{M}{k}} | \sigma(\hat{X}, \hat{X}')]]^{k/M} = \|\hat{Y}\|_{Mq}^{qk}.
$$

Putting the results together we obtain

$$
\| |\hat{Y} - \hat{Y}'| \cdot |\hat{Y}|^k |\hat{X}|_{\chi,1}^l \|_q \leq \tilde{D}^{1/(qM_y)} \sum_{i=1}^{\infty} \chi_i \|\hat{X}_i - \hat{X}_i'\|_{qM} \cdot D^k (|\chi|_1 D)^l,
$$

which leads to

$$
\big\| |\hat{Y} - \hat{Y}'| \cdot R_{M_y-1, M_x-1}(\hat{Y}, \hat{X}) \big\|_q \leq \tilde{D}^{1/(qM_y)} (1 + D(1 + |\chi|_1))^{M-1} \cdot \sum_{i=1}^{\infty} \chi_i \|\hat{X}_i - \hat{X}_i'\|_{qM},
$$

giving the result. □

**Lemma 3.7.2.** *Let $g \in \mathcal{H}(M'_y, M'_x, \chi, \bar{C})$ be continuously differentiable. Define $M' := \max\{M'_y, M'_x\}$. Let Assumption 3.2.1(A5) hold with some $r \geq 2$ and let Assumption 3.2.2(B3) hold. Suppose that*

- $\nabla_\theta g \in \mathcal{H}(M'_y, M'_x, \chi, \bar{C})$,
- *there exists $\tilde{\chi} = (\tilde{\chi}_i)_{i\in\mathbb{N}}$ such that for all $l \in \mathbb{N}$, $\partial_{x_l} g \in \mathcal{H}(M'_y - 1, M'_x - 1, \tilde{\chi}, \bar{C}\chi_l)$,*
- $\tilde{\theta} \in C^2[0,1]$.
- $\sup_{j\in\mathbb{N}_0} \sup_{t\in[0,1]} \|\partial_t \tilde{Z}_{0j}(t)\|_{M'} \leq D$ *with some $D > 0$.*

*Then*

*(i)* $\sup_{t\in[0,1]} \|\partial_t g(\tilde{Z}_0(t), \tilde{\theta}(t))\|_1 < \infty$,

*(ii)*
$$\sup_{t\neq t'} \frac{\|\partial_t g(\tilde{Z}_0(t), \tilde{\theta}(t)) - \partial_t g(\tilde{Z}_0(t'), \tilde{\theta}(t'))\|_1}{|t-t'|} < \infty.$$

*Proof of Lemma 3.7.2.* (i) Note that

$$\partial_t g(\tilde{Z}_0(t), \tilde{\theta}(t)) = \partial_z g(\tilde{Z}_0(t), \tilde{\theta}(t))\partial_t \tilde{Z}_0(t) + \nabla_\theta g(\tilde{Z}_0(t), \tilde{\theta}(t))\tilde{\theta}'(t). \tag{3.64}$$

By Lemma 3.7.1, it holds that for each $k = 1, \ldots, d_\Theta$, $\sup_t \|\nabla_{\theta_k} g(\tilde{Z}_0(t), \tilde{\theta}(t))\|_1 < \infty$. By Lemma 3.7.1(i) there exists a constant $C > 0$ such that for each $j \in \mathbb{N}_0$,

$$\|\partial_{z_j} g(\tilde{Z}_0(t), \tilde{\theta}(t))\|_{M'/(M'-1)} \leq C\bar{C}\hat{\chi}_j. \tag{3.65}$$

It follows that

$$\begin{aligned} \|\partial_z g(\tilde{Z}_0(t), \tilde{\theta}(t))\partial_t \tilde{Z}_0(t)\|_1 &\leq \sum_{j=0}^{\infty} \|\partial_{z_j} g(\tilde{Z}_0(t), \tilde{\theta}(t))\|_{M'/(M'-1)} \cdot \|\partial_t \tilde{Z}_{0j}(t)\|_{M'} \\ &\leq CD\bar{C}\sum_{j=0}^{\infty} \hat{\chi}_j < \infty, \end{aligned}$$

which shows the assertion.

(ii) Let $t, t' \in [0,1]$. From Lemma 3.7.1(i) we obtain with some constant $C > 0$, for each $k = 1, ..., d_\Theta$,

$$\begin{aligned} \|\nabla_{\theta_k} g(\tilde{Z}_0(t), \tilde{\theta}(t))\|_1 &\leq C, \\ \|\nabla_{\theta_k} g(\tilde{Z}_0(t), \tilde{\theta}(t)) - \nabla_{\theta_k} g(\tilde{Z}_0(t), \tilde{\theta}(t'))\|_1 &\leq C|\tilde{\theta}(t) - \tilde{\theta}(t')|_1. \end{aligned}$$

From Lemma 3.7.1(i) we obtain for each $k = 1, ..., d_\Theta$ (note that $rM \geq M'$ in Assumption 3.2.2):

$$\begin{aligned} &\|\nabla_{\theta_k} g(\tilde{Z}_0(t), \tilde{\theta}(t')) - \nabla_{\theta_k} g(\tilde{Z}_0(t'), \tilde{\theta}(t'))\|_1 \\ \leq\ & \bar{C}C\big(\|\tilde{Y}_0(t) - \tilde{Y}_0(t')\|_{M'} + \sum_{j=1}^{\infty} \chi_j \|\tilde{X}_0(t) - \tilde{X}_0(t')\|_{M'}\big) \\ \leq\ & \bar{C}CC_B|t - t'|(1 + |\chi|_1). \end{aligned} \tag{3.66}$$

or with Lemma 3.7.1(ii) the same bound with $1 + |\chi|_1$ replaced by $|\chi|_1$.
By Lipschitz continuity of $\tilde{\theta}$, $\tilde{\theta}'$, the above results imply that the second summand in (3.64) fulfills the assertion,

$$\sup_{t \neq t'} \frac{\|\nabla_\theta g(\tilde{Z}_0(t), \tilde{\theta}(t))\tilde{\theta}'(t) - \nabla_\theta g(\tilde{Z}_0(t'), \tilde{\theta}(t'))\tilde{\theta}'(t')\|_1}{|t - t'|} < \infty.$$

It remains to show the same for the first summand in (3.64). By (3.65) and $\|\partial_t \tilde{Z}_{0j}(t) - \partial_t \tilde{Z}_{0j}(t')\|_{M'} \leq C_B|t - t'|$ from Assumption 3.2.2(B3), we have

$$\|\partial_z g(\tilde{Z}_0(t), \tilde{\theta}(t))(\partial_t \tilde{Z}_0(t) - \partial_t \tilde{Z}_0(t'))\|_1 \leq C\bar{C}C_B|\chi|_1|t - t'|. \tag{3.67}$$

Similar as in (3.66), we see by Lemma 3.7.1(i) or (ii) that

$$\|\partial_{z_j} g(\tilde{Z}_0(t), \tilde{\theta}(t)) - \partial_{z_j} g(\tilde{Z}_0(t'), \tilde{\theta}(t))\|_{M'/(M'-1)} \leq \chi_j \bar{C}CC_B(1 + |\chi'|_1)|t - t'|. \tag{3.68}$$

Finally, by Lemma 3.7.1(i) and Lipschitz continuity of $\tilde{\theta}$, we have

$$\|\partial_{z_j} g(\tilde{Z}_0(t'), \tilde{\theta}(t)) - \partial_{z_j} g(\tilde{Z}_0(t'), \tilde{\theta}(t'))\|_{M'/(M'-1)} \leq \chi_j \bar{C} C |\tilde{\theta}(t) - \tilde{\theta}(t')|_1 = O(|t-t'|). \quad (3.69)$$

By Hölder's inequality, we conclude from (3.68) and (3.69) that

$$\|(\partial_z g(\tilde{Z}_0(t), \tilde{\theta}(t)) - \partial_z g(\tilde{Z}_0(t'), \tilde{\theta}(t')))\partial_t \tilde{Z}_0(t')\|_1 = O(|t - t'|),$$

which together with (3.67), finishes the proof. □

**Lemma 3.7.3.** *Suppose that Assumption 3.2.1(A5), (A7) hold with some $r \geq 2$. Let $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, where $\chi_i = O(i^{-(1+\gamma)})$. Assume that $\hat{K} : \mathbb{R} \to \mathbb{R}$ is bounded by $|\hat{K}|_\infty$ and has compact support $[-1, 1]$. Then it holds that*

*(i)* $\sup_{t\in[0,1]} \delta_r^{\sup_\theta |g(\tilde{Z}(t),\theta)|}(j) = O(j^{-(1+\gamma)})$.

*(ii) For $M_i(t, \eta, u) := \hat{K}_{b_n}(u - t) g(\tilde{Z}_i(u), \eta_1 + \eta_2(u - t)b_n^{-1})$, we have*

$$\sup_{u\in[0,1]} \sup_{t,\eta} \delta_r^{M(t,\eta,u)}(j) = O(j^{-(1+\gamma)}), \qquad \sup_{u\in[0,1]} \delta_r^{\sup_{t,\eta} |M(t,\eta,u)|}(j) = O(j^{-(1+\gamma)}).$$

*(iii) Let $d_u(t) = \theta(u) - \theta(t) - (u - t)\theta'(t)$ and $M_i^{(2)}(t, u) := \hat{K}_{b_n}(u - t)\{\int_0^1 g(\tilde{Z}_i(u), \theta(t) + sd_u(t))ds\} \cdot d_u(t)$. Then it holds for each component $i = 1, \ldots, k$, that*

$$\sup_{u\in[0,1]} \delta_r^{M^{(2)}(t,u)}(j) = O(b_n^2 j^{-(1+\gamma)}), \qquad \sup_{u\in[0,1]} \delta_r^{\sup_t |M^{(2)}(t,u)|}(j) = O(b_n^2 j^{-(1+\gamma)}).$$

*Proof.* (i) Let $\tilde{Z}_j(t)^*$ be a coupled version of $\tilde{Z}_j(t)$ where $\zeta_0$ is replaced by $\zeta_0^*$. By Lemma

3.7.1 we obtain in case (3.12) that with some constant $\tilde{C} > 0$:

$$\begin{aligned}
&\delta_q^{\sup_\theta |g(Z,\theta)|}(j) \\
= \ & \| \sup_\theta |g(\tilde{Z}_j(t),\theta)| - \sup_\theta |g(\tilde{Z}_j(t)^*,\theta)| \|_r \\
\leq \ & \| \sup_\theta |g(\tilde{Z}_j(t),\theta) - g(\tilde{Z}_j(t)^*,\theta)| \|_r \\
\leq \ & \tilde{C}\Big( \|\tilde{Y}_j(t) - \tilde{Y}_j(t)^*\|_{rM} + \sum_{i=1}^\infty \chi_i \|\tilde{X}_{j-i+1}(t) - \tilde{X}_{j-i+1}(t)^*\|_{rM}\Big) \\
\leq \ & \tilde{C}\Big( \delta_{rM}^{\tilde{Y}(t)}(j) + \sum_{i=1}^\infty \chi_i \delta_{rM}^{\tilde{X}(t)}(j-i+1)\Big),
\end{aligned} \tag{3.70}$$

and in case (3.13), similarly

$$\delta_r^{g(Z,\theta)}(j) \leq \tilde{C} \sum_{i=1}^\infty \chi_i \delta_{rM}^{\tilde{X}(t)}(j-i+1). \tag{3.71}$$

Note that if two sequences $a_i, b_i$ with $a_i = b_i = 0$ for $i < 0$ obey $a_i, b_i = O(i^{-(1+\gamma)})$ then the convolution $c_j = \sum_{i=1}^\infty a_i b_{j-i+1}$ still obeys $c_j = O(j^{-(1+\gamma)})$ due to

$$\begin{aligned}
|c_j| \ &\leq \sum_{i=1, i\geq (j+1)/2}^{j+1} |a_i| \cdot |b_{j-i+1}| + \sum_{i=1, |j-i| \geq (j+1)/2}^{j+1} |a_i||b_{j-i+1}| \\
&\leq \Big(\frac{j+1}{2}\Big)^{-(1+\gamma)} \sum_{i=1}^{j+1} |b_{j-i+1}| + \Big(\frac{j+1}{2}\Big)^{-(1+\gamma)} \sum_{i=1}^{j+1} |a_i| = O(j^{-(1+\gamma)}).
\end{aligned}$$

Together with Assumption (A7) and (3.70), (3.71), this shows $\sup_{t\in[0,1]} \delta_r^{g(\tilde{Z}(t),\theta)}(j) = O(j^{-(1+\gamma)})$.

The proof for (ii),(iii) is the same since

$$\begin{aligned}
\Big| \sup_{t,\eta} |M_i(t,\eta,u)| - \sup_{t,\eta} |M_i(t,\eta,u)^*| \Big| \ &\leq \ \sup_{t,\eta} |M_i(t,\eta,u) - M_i(t,\eta,u)^*| \\
&\leq \ |\hat{K}|_\infty \sup_\theta |g(\tilde{Z}_i(u),\theta) - g(\tilde{Z}_i(u)^*,\theta)|
\end{aligned}$$

and (since $|d_u(t)|_\infty \leq \sup_s |\theta''(s)|_\infty \cdot b_n^2$ if $|t-u| \leq b_n$), for each $l = 1, \dots, k$,

$$
\begin{aligned}
& \big| \sup_t |\tilde{M}_i^{(2)}(t,u)_l| - \sup_t |\tilde{M}_i^{(2)}(t,u)_l^*| \big| \leq \sup_t |M_i^{(2)}(t,u)_l - M_i^{(2)}(t,u)_l^*| \\
\leq\ & |\hat{K}|_\infty \sup_s |\theta''(s)|_\infty b_n^2 \\
& \quad \times \sup_t \int_0^1 |g(\tilde{Z}_i(u), \theta(t) + s d_u(t)) - g(\tilde{Z}_i(u)^*, \theta(t) + s d_u(t))| ds \\
\leq\ & |\hat{K}|_\infty \sup_s |\theta''(s)|_\infty b_n^2 \sup_{\theta \in \Theta} |g(\tilde{Z}_i(u), \theta) - g(\tilde{Z}_i(u)^*, \theta)|.
\end{aligned}
$$

□

For the proof of the following lemma, we will make use of the adjusted dependence measure $\| \cdot \|_{q,\alpha}$ which is defined as follows (cf. [85]): For some zero-mean random variable $Z$, let $\|Z\|_{q,\alpha} := \sup_{m \geq 0} (m+1)^\alpha \Delta_q^Z(m)$.

**Lemma 3.7.4.** *Let $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$. For $t \in (0,1)$ and $\eta \in E_n = \Theta \times (\Theta' \cdot b_n)$ and some continuous function $\hat{K}$ with bounded variation and compact support $[-1,1]$, define $\hat{K}_{b_n}(\cdot) := \hat{K}(\cdot / b_n)$ and*

$$
\begin{aligned}
G_n(t,\eta) \ :=\ & (nb_n)^{-1} \sum_{i=1}^n \hat{K}_{b_n}(i/n - t) \cdot \{ g(Z_i, \eta_1 + \eta_2 (i/n - t) b_n^{-1}) \\
& \qquad - \mathbb{E} g(Z_i, \eta_1 + \eta_2 (i/n - t) b_n^{-1}) \}.
\end{aligned}
$$

*Let $G_n^c(t,\eta)$, $\hat{G}_n(t,\eta)$ denote the same quantities but with $Z_i$ replaced by $Z_i^c$ or $\tilde{Z}_i(i/n)$, respectively.*

*(i) If Assumption 3.2.1(A5), 3.2.1(A6) hold with $r = 1$ and $\chi_j = O((j \log(j))^{-2})$, then*

$$
\| \sup_{t \in \mathcal{T}_n} \sup_{\eta \in E_n} |\hat{G}_n(t,\eta) - G_n^c(t,\eta)| \|_1 = O((nb_n)^{-1}).
$$

*(ii) Let Assumption 3.2.1(A5), (A7) hold with $r = 2$. If $nb_n \to \infty$, then for fixed $t \in [0,1]$,*

$$\sup_{\eta \in E_n} |\hat{G}_n(t,\eta)| = o_{\mathbb{P}}(1).$$

*If instead $nb_n^2 c_n^{-2d_\Theta} \to 0$, then*

$$\sup_{t \in \mathcal{T}_n} \sup_{\eta \in E_n} |\hat{G}_n(t,\eta)| = o_{\mathbb{P}}(c_n^{-1}).$$

*(iii) If Assumption 3.2.1(A5), (A7) hold with $r = 4$, then*

$$\sup_{t \in \mathcal{T}_n} \sup_{\eta \in E_n} |\hat{G}_n(t,\eta)| = O_{\mathbb{P}}(\beta_n),$$

*where $\beta_n = \log(n)^{1/2}(nb_n)^{-1/2} b_n^{-1/2}$.*

*Proof of Lemma 3.7.4.* (i) By Lemma 3.7.1(i),(ii) and by Assumption 3.2.1(A6), we obtain (independent of (3.12) or (3.13)) that for some $C > 0$:

$$\|\sup_{\theta \in \Theta} |g(Z_i,\theta) - g(Z_i^c,\theta)|\|_1 \;\leq\; C\sum_{j=0}^{\infty} \hat{\chi}_j \|Z_{ij} - Z_{ij}^c\|_M \leq 2C\sum_{j=i}^{\infty} \chi_j \|Z_{ij}\|_M \leq 2CD\sum_{j=i}^{\infty} \chi_j.$$

Similarly, we have in the case (3.12) that

$$\begin{aligned}\|\sup_{\theta \in \Theta} |g(Z_i,\theta) - g(\tilde{Z}_i(i/n),\theta)|\|_1 \;&\leq\; C\Big(\|Y_i - \tilde{Y}_i(i/n)\|_M + \sum_{j=1}^{\infty} \chi_j \|X_{ij} - \tilde{X}_{ij}(i/n)\|_M\Big)\\ &\leq\; CC_A|\chi|_1 n^{-1},\end{aligned}$$

while in the case (3.13) there exists $C_2 > 0$ such that

$$\|\sup_{\theta \in \Theta} |g(Z_i,\theta) - g(\tilde{Z}_i(i/n),\theta)|\|_1 \;\leq\; C_2\sum_{j=1}^{\infty} \chi_j \|X_{ij} - \tilde{X}_{ij}(i/n)\|_M \leq C_2 C_A |\chi|_1 n^{-1}.$$

Thus

$$\begin{aligned}
&\| \sup_{t\in\mathcal{T}_n} \sup_{\eta\in E_n} |G_n(t,\eta) - G_n^c(t,\eta)| \|_1 \\
\leq\ & |K|_\infty (nb_n)^{-1} \sum_{i=1}^n \| \sup_{\theta\in\Theta} |g(Z_i,\theta) - g(Z_i^c,\theta)| \|_1 \\
\leq\ & 2CD|K|_\infty (nb_n)^{-1} \sum_{i=1}^n \sum_{j=i}^\infty \chi_j + |K|_\infty (C\vee C_2)|\chi|_1 (nb_n)^{-1}.
\end{aligned}$$

Since $\chi_j = O((j\log(j))^{-2})$, $\sum_{i=1}^n \sum_{j=i}^\infty \chi_j = O(1)$ and the assertion is proved.

(ii) Fix $Q > 0$. Let $\iota > 0$. Let $E_n^{(\iota)}$ be a discretization of $E_n$ such that for each $\eta \in E_n$ one can find $\eta' \in E_n^{(\iota)}$ with $|\eta - \eta'|_1 \leq \iota$. Note that $\# E_n^{(\iota)}$ does not need to depend on $n$. Then

$$\begin{aligned}
\mathbb{P}\big( \sup_{\eta\in E_n} |\hat{G}_n(t,\eta)| > Q\big) \ \leq\ & \# E_n^{(\iota)} \sup_{\eta\in E_n} \mathbb{P}\big(|\hat{G}_n(t,\eta)| > Q/2\big) \\
& + \mathbb{P}( \sup_{|\eta-\eta'|_1\leq\iota} |\hat{G}_n(t,\eta) - \hat{G}_n(t,\eta')| > Q/2). \quad (3.72)
\end{aligned}$$

By Markov's inequality, we have

$$\mathbb{P}\big(|\hat{G}_n(t,\eta)| > Q/2\big) \leq \frac{\|\hat{G}_n(t,\eta)\|_2^2}{(Q/2)^2}.$$

The computation

$$\begin{aligned}
&\|\hat{G}_n(t,\eta)\|_2 \\
\leq\ & (nb_n)^{-1} \sum_{l=0}^\infty \Big\| \sum_{i=1}^n \hat{K}_{b_n}(i/n - t) P_{i-l} g(\tilde{Z}_i(i/n), \eta_1 + \eta_2(i/n - t)b_n^{-1}) \Big\|_2 \\
\leq\ & (nb_n)^{-1} \sum_{l=0}^\infty \Big( \sum_{i=1}^n \hat{K}_{b_n}(i/n - t)^2 \| P_{i-l} g(\tilde{Z}_i(i/n), \eta_1 + \eta_2(i/n - t)b_n^{-1}) \|_2^2 \Big)^{1/2} \\
\leq\ & (nb_n)^{-1/2} |\hat{K}|_\infty \sum_{l=0}^\infty \sup_{t\in[0,1]} \delta_2^{\sup_{\theta\in\Theta} |g(\tilde{Z}(t),\theta)|}(l) = O((nb_n)^{-1/2}), \quad (3.73)
\end{aligned}$$

shows that the first summand in (3.72) tends to zero. For the second summand, note that $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ implies that with some constant $\tilde{C} > 0$:

$$|g(z,\theta) - g(z,\theta')| \leq \tilde{C}|\theta - \theta'|_1 R_{M_y,M_x}(z).$$

Note that $\sup_t \|R_{M_y,M_x}(\tilde{Z}_0(t))\|_1 < \infty$ is bounded by using similar techniques as in the proof of Lemma 3.7.1. Thus

$$\begin{aligned}
&\mathbb{P}( \sup_{|\eta-\eta'|_1 \leq \iota} |\hat{G}_n(t,\eta) - \hat{G}_n(t,\eta')| > Q/2) \\
\leq\ & \mathbb{P}\Big((nb_n)^{-1} \sum_{i=1}^{n} \hat{K}_{b_n}(i/n - t)\{R_{M_y,M_x}(\tilde{Z}_i(i/n)) + \mathbb{E}R_{M_y,M_x}(\tilde{Z}_i(i/n))\} > \frac{Q}{2\iota}\Big) \\
\leq\ & \frac{4\iota}{Q}(nb_n)^{-1} \sum_{i=1}^{n} \hat{K}_{b_n}(i/n - t)\|R_{M_y,M_x}(\tilde{Z}_i(i/n))\|_1 \leq \frac{2\iota|K|_\infty}{Q} \sup_t \|R_{M_y,M_x}(\tilde{Z}_0(t))\|_1,
\end{aligned}$$

which can be made arbitrary small by choosing $\iota$ small enough. So we have shown that (3.72) tends to zero for $n \to \infty$.

For the second assertion, let $\mathcal{T}_n^{(\iota)} = \{i/\lfloor \iota^{-1} n^{1/2} \rfloor : i = 0, \ldots, \lfloor \iota^{-1} n^{1/2} \rfloor\}$ denote a discretization of $\mathcal{T}_n$. Let $E_n^{(\iota,c_n)}$ be a discretization of $E_n$ such that for each $\eta \in E_n$ one can find $\eta' \in E_n^{(\iota,c_n)}$ with $|\eta - \eta'|_1 \leq \iota c_n^{-1}$. Then $\#(E_n^{(\iota)} \times \mathcal{T}_n^{(\iota)}) = O(n^{1/2} c_n^{d_\Theta})$. Let $L_{\hat{K}}$ denote the Lipschitz constant of $\hat{K}$. We have with some constant $\tilde{C} > 0$:

$$\begin{aligned}
&|\hat{G}_n(t,\eta) - \hat{G}_n(t',\eta')| \\
\leq\ & \Big\{2(nb_n^2)^{-1}\tilde{C}(L_{\hat{K}} + R)|t - t'| + (nb_n)^{-1}|\eta - \eta'|\tilde{C}|\hat{K}|_\infty\Big\} \\
& \times \sum_{\substack{i=1 \\ |i/n-t| \leq b_n \text{ or } |i/n-t'| \leq b_n}}^{n} \{R_{M_y,M_x}(\tilde{Z}_i(i/n)) + \mathbb{E}R_{M_y,M_x}(\tilde{Z}_i(i/n))\} \\
\leq\ & \Big\{2(nb_n^2)^{-1}\tilde{C}(L_{\hat{K}} + R)|t - t'| + (nb_n)^{-1}|\eta - \eta'|_1\tilde{C}|\hat{K}|_\infty\Big\} \cdot \{W_n^{(1)} + nb_n W^{(2)}\},
\end{aligned}$$

with

$$W_n^{(1)} = \big|\sum_{i=1}^{n}\{R_{M_y,M_x}(\tilde{Z}_i(i/n)) - \mathbb{E}R_{M_y,M_x}(\tilde{Z}_i(i/n))\}\big|$$

and $W^{(2)} = \sup_{u\in[0,1]}\|R_{M_y,M_x}(\tilde{Z}_0(u))\|_1$. By using Lemma 3.7.3(i), we obtain

$$\|n^{-1/2}W_n^{(1)}\|_2 \le n^{-1/2}\sum_{l=0}^{\infty}\big\|\sum_{i=1}^{n}P_{i-l}R_{M_y,M_x}(\tilde{Z}_i(i/n))\big\|_2 \le \sum_{l=0}^{\infty}\sup_{u\in[0,1]}\delta_2^{\tilde{Z}(u)}(l).$$

We conclude that with some constant $\tilde{C}_2 > 0$,

$$\begin{aligned}
&\mathbb{P}\big(\sup_{|t-t'|\le\iota n^{-1/2},|\eta-\eta'|_1\le\iota}|\hat{G}_n(t,\eta) - \hat{G}_n(t,\eta')| > Qc_n^{-1}/2\big)\\
\le\ & \Big(\frac{2\tilde{C}_2\iota c_n}{Q}\Big)^2\cdot\big\{(n^{-3/2}b_n^{-2} + (nb_n)^{-1}c_n^{-1})(\|W_n^{(1)}\|_2 + nb_nW^{(2)})\big\}^2\\
\le\ & \Big(\frac{2\tilde{C}_2\iota}{Q}\Big)^2\cdot\{((nb_n^2)^{-1}c_n + (nb_n^2)^{-1/2})\|n^{-1/2}W_n^{(1)}\|_2 + ((nb_n^2)^{-1/2}c_n + 1)W^{(2)}\}^2,
\end{aligned}$$

which is arbitrary small if $\iota$ is chosen small enough due to $nb_n^2c_n^{-2}\to\infty$. By (3.73) and $\#(E_n^{(\iota)}\times\mathcal{T}_n^{(\iota)}) = O(n^{1/2}c_n^k)$, the assertion follows from

$$\begin{aligned}
&\mathbb{P}\big(\sup_{\eta\in E_n,t\in\mathcal{T}_n}|\hat{G}_n(t,\eta)| > Q\big)\\
\le\ & \#(E_n^{(\iota)}\times\mathcal{T}_n^{(\iota)})\sup_{\eta\in E_n,t\in\mathcal{T}_n}\mathbb{P}\big(|\hat{G}_n(t,\eta)| > Q/2\big)\\
&\quad+\mathbb{P}(\sup_{|\eta-\eta'|_1\le\iota,|t-t'|\le\iota n^{-1/2}}|\hat{G}_n(t,\eta) - \hat{G}_n(t,\eta')| > Q/2)
\end{aligned}$$

and $nb_n^2c_n^{-2d_\Theta}\to\infty$.

(iii) We use a chaining argument. Let $r = n^3$ and let $E_{n,r}$ be a discretization of $E_n$ such that for each $\eta\in E_n$ one can find $\eta'\in E_{n,r}$ with $|\eta-\eta'|\le r^{-1}$. Define $\mathcal{T}_{n,r} := \{i/r : i = 1,\ldots,r\}$ as a discretization of $\mathcal{T}_n$. Then $\#(E_{n,r}\times\mathcal{T}_{n,r}) = O(r^{2d_\Theta+1})$. For some constant

$Q > 0$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\eta\in E_n, t\in\mathcal{T}_n} |\hat{G}_n(t,\eta)| > Q\delta_n\Big)\\
\leq\ &\mathbb{P}\Big(\sup_{\eta\in E_{n,r}, t\in\mathcal{T}_n} |\hat{G}_n(t,\eta)| > Q\delta_n/2\Big)\\
&\quad+\mathbb{P}\Big(\sup_{|\eta-\eta'|\leq r^{-1}, |t-t'|\leq r^{-1}} |\hat{G}_n(t,\eta) - \hat{G}_n(t',\eta')| > Q\delta_n/2\Big). \qquad (3.74)
\end{aligned}$$

Let $\alpha = 1/2$. Let $M_i(t,\eta,u) := \hat{K}_{b_n}(u-t)g(\tilde{Z}_i(u), \eta_1 + \eta_2(u-t)b_n^{-1})$. By Assumption 3.2.1(A7) and Lemma 3.7.3(ii), we have $\sup_u \Delta_4^{\sup_{t,\eta}|M(t,\eta,u)|}(k) = O(k^{-\gamma})$. Thus

$$W_{4,\alpha} := \sup_{u\in[0,1]} \|\sup_{t,\eta} |M_i(t,\eta,u)|\|_{4,\alpha} = \sup_{m\geq 0}(m+1)^\alpha \sup_{u\in[0,1]}\sup_{t,\eta} \Delta_4^{\sup_{t,\eta}|M(t,\eta,u)|}(m) < \infty.$$

(independent of $n$) and

$$W_{2,\alpha} := \sup_{u\in[0,1]}\sup_{t,\eta} \|M_i(t,\eta,u)\|_{2,\alpha} = \sup_{m\geq 0}(m+1)^\alpha \sup_{u\in[0,1]}\sup_{t,\eta} \Delta_2^{M(t,\eta,u)}(m) < \infty$$

(independent of $n$). Note that $l = 1 \wedge \log\#(E_{n,r}\times\mathcal{T}_{n,r}) \leq 3(2d_\Theta+1)\log(n)$ and $Q\delta_n(nb_n) = Qn^{1/2}\log(n)^{1/2} \geq \sqrt{nl}W_{2,\alpha} + n^{1/q}l^{3/2}W_{4,\alpha} \gtrsim n^{1/2}\log(n)^{1/2} + n^{1/4}\log(n)^{3/2}$ for $Q$ large enough. By applying Theorem 6.2 of [85] (the proof therein also works for the functional dependence measure) with $q = 4$ and $\alpha = 1/2$ to $(M_i(t,\eta,i/n))_{t\in\mathcal{T}_{n,r},\eta\in E_{n,r}}$, we have with some constant $C_\alpha > 0$:

$$\begin{aligned}
&\mathbb{P}\big(\sup_{\eta'\in E_{n,r}, t'\in\mathcal{T}_{n,r}} |\hat{G}_n(t',\eta')| \geq Q\delta_n/2\big)\\
\leq\ &\frac{C_\alpha n\cdot l^2 W_{4,\alpha}^4}{(Q/2)^4(\delta_n(nb_n))^4} + C_\alpha \exp\Big(-\frac{C_\alpha (Q/2)^2(\delta_n(nb_n))^2}{nW_{2,\alpha}^2}\Big)\\
\lesssim\ &\frac{n\log(n)^2}{(nb_n)^2 b_n^{-2}\log(n)^2} + \exp\big(-\frac{(nb_n)b_n^{-1}\log(n)}{n}\big)\\
\lesssim\ &n^{-1} \to 0. \qquad (3.75)
\end{aligned}$$

Since $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ and $\hat{K}$ is Lipschitz with constant $L_{\hat{K}}$, it is easy to see that

$$\begin{aligned}
&|\hat{G}_n(t,\eta) - \hat{G}_n(t',\eta')| \\
\leq\ & C_M\{|\eta - \eta'|_1 + |t - t'|\} b_n^{-2} \cdot (L_{\hat{K}} + |\hat{K}|_\infty) \qquad (3.76) \\
& \times n^{-1} \sum_{i=1}^{n} \{R_{M_y,M_x}(\tilde{Z}_i(i/n)) + \mathbb{E} R_{M_y,M_x}(\tilde{Z}_i(i/n))\}, \qquad (3.77)
\end{aligned}$$

with some constant $C_M > 0$. Since $\|n^{-1}\sum_{i=1}^{n} R_{M_y,M_x}(\tilde{Z}_i(i/n))\|_1 = O(1)$ by Lemma 3.7.1, we conclude with Markov's inequality that

$$\mathbb{P}\big(\sup_{|\eta-\eta'|_1 \leq r^{-1}, |t-t'| \leq r^{-1}} |\hat{G}_n(t,\eta) - \hat{G}_n(t',\eta')| \geq C\delta_n/2\big) = O\Big(\frac{b_n^{-2} r^{-1}}{\delta_n}\Big). \qquad (3.78)$$

We have $b_n^{-2} r^{-1} \delta_n^{-1} = b_n^{-2} n^{-3} (nb_n)^{1/2} b_n^{1/2} \log(n)^{-1/2} \to 0$. Inserting (3.75) and (3.78) into (3.74), we obtain the result. $\square$

**Lemma 3.7.5.** *Let $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$. Let $\hat{K}$ be some continuous function with bounded variation and compact support $[-1,1]$, and put $\hat{K}_{b_n}(\cdot) := \hat{K}(\cdot/b_n)$. Let Assumption 3.2.1(A5) be fulfilled with $r = 1$. Let*

$$\hat{B}_n(t,\eta) = (nb_n)^{-1} \sum_{i=1}^{n} \hat{K}_{b_n}(i/n - t) g(\tilde{Z}_i(i/n), \eta_1 + \eta_2(i/n - t) b_n^{-1}).$$

*Then we have*

$$\sup_{t \in \mathcal{T}_n} \sup_{\eta \in E_n} |\mathbb{E}\hat{B}_n(t,\eta) - \int \hat{K}(x) \mathbb{E} g(\tilde{Z}_0(t), \eta_1 + \eta_2 x) dx| = O((nb_n)^{-1} + b_n). \qquad (3.79)$$

*Proof of Lemma 3.7.5.* Let $\tilde{B}_n(t,\eta) := (nb_n)^{-1}\sum_{i=1}^{n} K_{b_n}(i/n - t) g(\tilde{Z}_i(t), \eta_1 + \eta_2(i/n - t) b_n^{-1})$. By Lemma 3.7.1(i), we have with some constant $\tilde{C} > 0$ that either in the case of (3.12),

$$\begin{aligned}
& \|g(\tilde{Z}_0(i/n), \eta_1 + \eta_2(i/n - t)b_n^{-1}) - g(\tilde{Z}_0(t), \eta_1 + \eta_2(i/n - t)b_n^{-1})\|_1 \\
\leq \; & \tilde{C}\Big(\|\tilde{Y}_0(i/n) - \tilde{Y}_0(t)\|_M + \sum_{i=1}^{\infty} \chi_i \|\tilde{X}_{-i}(i/n) - \tilde{X}_{-i}(t)\|_M\Big) \leq \tilde{C}C_B(1 + |\chi|_1)b_n
\end{aligned}$$

or in the case of (3.13),

$$\begin{aligned}
& \|g(\tilde{Z}_0(i/n), \eta_1 + \eta_2(i/n - t)b_n^{-1}) - g(\tilde{Z}_0(t), \eta_1 + \eta_2(i/n - t)b_n^{-1})\|_1 \\
\leq \; & \tilde{C}\sum_{i=1}^{\infty} \chi_i \|\tilde{X}_{-i}(i/n) - \tilde{X}_{-i}(t)\|_M \leq \tilde{C}C_B|\chi|_1 b_n.
\end{aligned}$$

Thus

$$\begin{aligned}
& \|\hat{B}_n(t, \eta) - \tilde{B}_n(t, \eta)\|_1 \\
\leq \; & (nb_n)^{-1} \sum_{i=1}^{n} |\hat{K}_{b_n}(i/n - t)| \\
& \qquad \times \|g(\tilde{Z}_i(i/n), \eta_1 + \eta_2(i/n - t)b_n^{-1}) - g(\tilde{Z}_i(t), \eta_1 + \eta_2(i/n - t)b_n^{-1})\|_1 \\
\leq \; & \tilde{C}|\hat{K}|_\infty C_B(1 + |\chi|_1)b_n.
\end{aligned}$$

Since $\hat{K}$ is of bounded variation and $\theta \mapsto \mathbb{E}g(\tilde{Z}_0(t), \theta)$ is Lipschitz continuous due to $g \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ and Lemma 3.7.1, a Riemannian sum argument yields

$$\begin{aligned}
\tilde{B}_n(t, \eta) &= (nb_n)^{-1} \sum_{i=1}^{n} \hat{K}_{b_n}(i/n - t)\mathbb{E}g(\tilde{Z}_0(t), \eta_1 + \eta_2(i/n - t)b_n^{-1}) \\
&= \int \hat{K}(x)\mathbb{E}g(\tilde{Z}_0(t), \eta_1 + \eta_2 x)dx + O((nb_n)^{-1}),
\end{aligned}$$

uniformly in $t \in \mathcal{T}_n$, $\eta \in E_n$. □

**Lemma 3.7.6.** *Let $\eta_{b_n}(t) = (\theta(t)^\intercal, b_n\theta'(t)^\intercal)^\intercal$. Let Assumption 3.2.1 hold with $r = 1$.*

*(i) Then uniformly in $t \in \mathcal{T}_n$,*

$$\mathbb{E}\nabla_{\eta_1}\hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) = b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) + O(b_n^3). \tag{3.80}$$

*(ii) If additionally Assumption 3.2.2 holds, then uniformly in $t \in \mathcal{T}_n$,*

$$\mathbb{E}\nabla_{\eta_2}\hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) = b_n^3\frac{\mu_{K,4}}{2}V(t)bias(t) + O(b_n^4),$$

*where $bias(t) = \frac{1}{3}\theta^{(3)}(t) + V(t)^{-1}\mathbb{E}[\partial_t\nabla_\theta^2\ell(\tilde{Z}_0(t),\theta(t))]\cdot\theta''(t)$, and the term $O(b_n^3)$ in (3.80) can be replaced by $O(b_n^4)$.*

*Proof of Lemma 3.7.6.* (i) Let $U_{i,n}(t) = (K_{b_n}(i/n-t), K_{b_n}(i/n-t)(i/n-t)b_n^{-1})^{\mathsf{T}}$. By a Taylor expansion of $\theta(i/n)$ around $t$, we have

$$\theta(i/n) = \theta(t) + \theta'(t)(i/n-t) + r_n(t),$$

where $r_n(t) = \theta''(t)\frac{(i/n-t)^2}{2} + \theta'''(\tilde{t})\frac{(i/n-t)^3}{6}$ and $\tilde{t}$ is between $t$ and $i/n$. We conclude that

$$\begin{aligned}
&\nabla_\eta\hat{L}^{\circ}_{n,b_n}(t,\eta_{b_n}(t)) - (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))\\
= \;& (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\Big\{\int_0^1\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n)+sr_n(t))ds\cdot r_n(t)\Big\}.
\end{aligned} \tag{3.81}$$

Since $\nabla_\theta^2\ell \in \mathcal{H}(M_y,M_x,\chi,\bar{C})$, we obtain with Lemma 3.7.1 for $|i/n-t| \le b_n$:

$$\|\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n)+sr_n(t)) - \nabla_\theta^2\ell(\tilde{Z}_i(t),\theta(t))\|_1 = O(b_n+n^{-1}). \tag{3.82}$$

Using (3.81), $\mathbb{E}\nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n)) = 0$ and (3.82), we obtain

$$\begin{aligned}
& \mathbb{E}\nabla_\eta \hat{L}^\circ_{n,b_n}(t, \eta_{b_n}(t)) \\
= \;& (nb_n)^{-1} \sum_{i=1}^n U_{i,n}(t) \otimes \big\{\mathbb{E}\nabla^2_\theta \ell(\tilde{Z}_i(t), \theta(t)) \cdot \theta''(t) \frac{(i/n-t)^2}{2}\big\} + O(b_n^3 + n^{-1}) \\
= \;& \begin{pmatrix} b_n^2 \frac{\mu_{K,2}}{2} V(t)\theta''(t) \\ 0 \end{pmatrix} + O(b_n^3 + n^{-1} + (nb_n)^{-1}).
\end{aligned} \tag{3.83}$$

Under Assumption 3.2.2, we have $r_n(t) = \theta''(t)\frac{(i/n-t)^2}{2} + \theta^{(3)}(t)\frac{(i/n-t)^3}{6} + \theta^{(4)}(\tilde{t})\frac{(i/n-t)^4}{24}$, where $\tilde{t}$ is between $t$ and $i/n$. We now use a more precise Taylor argument as in (3.81). We have

$$\begin{aligned}
& \nabla_\eta \hat{L}^\circ_{n,b_n}(t, \eta_{b_n}(t)) - (nb_n)^{-1} \sum_{i=1}^n U_{i,n}(t) \otimes \nabla_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n)) \\
= \;& (nb_n)^{-1} \sum_{i=1}^n U_{i,n}(t) \otimes \nabla^2_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n)) r_n(t) \qquad (3.84) \\
& + (nb_n)^{-1} \sum_{i=1}^n U_{i,n}(t) \otimes \big\{ \int_0^1 \nabla^2_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n) + s r_n(t)) \\
& \qquad - \nabla^2_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n)) ds \cdot r_n(t)\big\}.
\end{aligned}$$

Since $\nabla^2_\theta \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we have by Lemma 3.7.1:

$$\big\|\nabla^2_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n) + s r_n(t)) - \nabla^2_\theta \ell(\tilde{Z}_i(i/n), \theta(i/n))\big\|_1 = O(r_n(t)) = O(|i/n - t|^2). \tag{3.85}$$

This shows that the expectation of the second summand in (3.84) is $O(b_n^4)$. We now discuss the first term in (3.84). Put $v_i(t) := \nabla^2_\theta \ell(\tilde{Z}_i(t), \theta(t))$. By Assumption 3.2.2, $t \mapsto v_i(t)$ is continuously differentiable. By Taylor's expansion, $v_i(i/n) = v_i(t) + (i/n - t)\partial_t v_i(t) + (i/n -$

$t) \int_0^1 \partial_t v_i(t+s(i/n-t)) - \partial_t v_i(t)ds$. We have

$$\mathbb{E}\big[(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes v_i(t)r_n(t)\big] = \begin{pmatrix} b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) \\ b_n^3\frac{\mu_{K,4}}{6}V(t)\theta^{(3)}(t)\end{pmatrix} + O(n^{-1}+b_n^4), \tag{3.86}$$

since $K$ has bounded variation and $\int K(x)x^3dx = 0$ by symmetry. Similarly,

$$\begin{aligned}&\mathbb{E}\big[(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes \partial_t v_i(t)r_n(t)\big] \\ =\ & \begin{pmatrix} 0 \\ b_n^3\frac{\mu_{K,4}}{2}\mathbb{E}[\partial_t\nabla_\theta^2\ell(\tilde{Z}_0(t),\theta(t))]\theta''(t)\end{pmatrix} + O(n^{-1}+b_n^4).\end{aligned} \tag{3.87}$$

Finally, by Assumption 3.2.2 and since $\partial_z\nabla_\theta^2\ell, \nabla_\theta^3\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we have with Lemma 3.7.2 applied to $g = \nabla_\theta^2\ell$:

$$\|\partial_t v_i(t+s(i/n-t)) - \partial_t v_i(t)\|_1 = O(|i/n-t|). \tag{3.88}$$

The results (3.86), (3.87) and (3.88) imply

$$\begin{aligned}&\mathbb{E}\big[(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes \nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n))r_n(t)\big] \\ =\ & \begin{pmatrix} b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) \\ b_n^3\mu_{K,4}\cdot\big\{\frac{1}{6}V(t)\theta^{(3)}(t) + \frac{1}{2}\mathbb{E}[\partial_t\nabla_\theta^2\ell(\tilde{Z}_0(t),\theta(t))]\cdot\theta''(t)\big\}\end{pmatrix} + O(n^{-1}+b_n^4),\end{aligned}$$

which together with (3.84) gives the result. □

**Lemma 3.7.7.** *Let* $U_{i,n}(t) := K_{b_n}(i/n-t)\cdot(1,(i/n-t)b_n^{-1})^\mathsf{T}$. *Under the conditions of*

*Theorem 3.3.4, it holds that*

$$\sup_{t\in\mathcal{T}_n}\big|\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t)) \\ -(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))\big| = O_{\mathbb{P}}(\beta_n b_n^2).$$

*Proof.* Note that $\mathbb{E}\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n)) = 0$ by Assumption 3.2.1(A1) in connection with 3.2.1(A3). Put

$$\begin{aligned} &\Pi_n(t) \\ := \quad & (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\big\{[\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(t)+(i/n-t)\theta'(t)) - \nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))] \\ & -\mathbb{E}[\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(t)+(i/n-t)\theta'(t)) - \nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))]\big\}. \end{aligned}$$

We have to prove that $\sup_{t\in\mathcal{T}_n}\big|\Pi_n(t)\big| = O_{\mathbb{P}}(\delta_n b_n^2)$. Define $M_i(t,u) := \int_0^1 \nabla_\theta^2\ell(\tilde{Z}_i(u),\theta(t)+s(\theta(u)-\theta(t)-(u-t)\theta'(t)))ds$ and $M_i^{(2)}(t,u) = U_{i,n}(t)\otimes\big\{M_i(t,u)\{\theta(u)-\theta(t)-(u-t)\theta'(t)\}\big\}$. By a Taylor expansion of $\nabla_\theta\ell$ w.r.t. $\theta$, we have

$$\Pi_n(t) = (nb_n)^{-1}\sum_{i=1}^n (M_i^{(2)}(t,i/n) - \mathbb{E}M_i^{(2)}(t,i/n)).$$

We now apply a similar technique as in the proof of Lemma 3.7.4(iii), namely we use a chaining argument similar to (3.74) to prove

$$\mathbb{P}\big(\sup_{t\in\mathcal{T}_n}|\Pi_n(t)| > Q\beta_n b_n^2\big) \to 0,$$

for some $Q>0$ large enough. Since $\nabla_\theta^2\ell \in \mathcal{H}(M_y,M_x,\chi,\bar{C})$, there exists a constant $C>0$ such that

$$\sup_{\theta\in\Theta}|\nabla_\theta^2\ell(z,\theta)| \le CR_{M_y,M_x}(z)$$

and

$$\sup_{\theta \neq \theta'} \frac{|\nabla_\theta^2 \ell(z,\theta) - \nabla_\theta^2 \ell(z,\theta')|}{|\theta-\theta'|_1} \leq C R_{M_y,M_x}(z).$$

This implies

$$\begin{aligned} |M_i(t,i/n) - M_i(t',i/n)| &\leq 2C(|\theta(t)-\theta(t')|_1 + 2|\theta'(t)-\theta'(t')|_1 \\ &\qquad\qquad + |t-t'| \sup_s |\theta'(s)|_1) R_{M_y,M_x}(\tilde{Z}_i(i/n)) \\ &\leq \tilde{C}|t-t'| R_{M_y,M_x}(\tilde{Z}_i(i/n)), \end{aligned}$$

with some constant $\tilde{C} > 0$ due to Lipschitz continuity of $\theta(\cdot), \theta'(\cdot)$. Since additionally, $K$ and $x \mapsto K(x)x$ are Lipschitz continuous, we have $|M_i^{(2)}(t,i/n) - M_i^{(2)}(t',i/n)| \leq \tilde{C}_2 b_n^{-1}|t-t'| R_{M_y,M_x}(\tilde{Z}_i(i/n))$. This shows with some constant $\tilde{C}_3 > 0$ that

$$|\Pi_n(t) - \Pi_n(t')| \leq \tilde{C}_3 b_n^{-2} |t-t'| n^{-1} \sum_{i=1}^n \{R_{M_y,M_x}(\tilde{Z}_i(i/n)) + \mathbb{E}R_{M_y,M_x}(\tilde{Z}_i(i/n))\},$$

a similar result as in (3.77). Note that $\|R_{M_y,M_x}(\tilde{Z}_i(i/n))\|_1 \leq (1+D(|\chi|+1))^M$ by using similar arguments as in the proof of Lemma 3.7.1. By defining the discretization $\mathcal{T}_{n,r} := \{l/r : l = 1,\dots,r\}$ with $r = n^5$, we obtain for $Q > 0$ with Markov's inequality:

$$\mathbb{P}\big( \sup_{|t-t'| \leq r^{-1}} |\Pi_n(t) - \Pi_n(t')| > Q\beta_n b_n^2/2 \big) = O\big( \frac{b_n^{-2} r^{-1}}{\beta_n b_n^2} \big),$$

which converges to 0. Choose $\alpha = 1/2$. By Assumption 3.2.1(A7) and Lemma 3.7.3(iii), we have $\sup_u \Delta_4^{\sup_t |M^{(2)}(t,u)|}(k) = O(k^{-\gamma})$. Thus

$$\begin{aligned} W_{4,\alpha} &:= \sup_{u \in [0,1]} \sup_{t \in [0,1]} \| \sup_{t,\eta} |M_i^{(2)}(t,u)| \|_{4,\alpha} = \sup_{m \geq 0} (m+1)^\alpha \Delta_4^{\sup_t |M^{(2)}(t,u)|}(m) \\ &= O(b_n^2) \end{aligned} \tag{3.89}$$

(independent of $n$) and

$$\begin{aligned} W_{2,\alpha} &:= \sup_{t,u} \|M_i^{(2)}(t,u)\|_{2,\alpha} = \sup_{m\geq 0}(m+1)^{\alpha} \sup_{u\in[0,1]} \sup_{t} \Delta_2^{M^{(2)}(t,u)}(m) \\ &= O(b_n^2) \end{aligned} \tag{3.90}$$

(independent of $n$). We now apply Theorem 6.2 of [85] (the proof therein also works for the functional dependence measure) with $q = 4$, $\alpha = 1/2 > \frac{1}{4}$ to $(M_i^{(2)}(t, i/n))_{t\in\mathcal{T}_{n,r}}$, where $l = 1 \vee \#(\mathcal{T}_{n,t}) \leq 5\log(n)$. For $Q$ large enough, we obtain with some constant $C_\alpha$ only depending on $\alpha$:

$$\begin{aligned} & \mathbb{P}\big(\sup_{t'\in\mathcal{T}_r} |\Pi_n(t')| \geq Q\beta_n b_n^2/2\big) \\ \leq\ & \frac{C_\alpha n \cdot l^2 W_{4,\alpha}^4}{(Q/2)^4(\beta_n b_n^2(nb_n))^4} + C_\alpha \exp\Big(-\frac{C_\alpha (Q/2)^2(\beta_n b_n^2(nb_n))^2}{nW_{2,\alpha}^2}\Big) \\ \lesssim\ & \frac{n\log(n)^2}{(nb_n)^2 b_n^{-2}\log(n)^2} + \exp\big(-\frac{(nb_n)b_n^{-1}\log(n)}{n}\big) \\ \lesssim\ & n^{-1} \to 0, \end{aligned}$$

which finishes the proof. □

### *3.7.2 Proofs and Lemmas for the SCB*

From Lemma 1 in [91], we adopt the following result:

**Lemma 3.7.8.** *Suppose that $\hat{K} : \mathbb{R} \to \mathbb{R}$ has bounded variation and bounded support $[-1,1]$. Let $F_n(t) = \sum_{i=1}^n \hat{K}_{b_n}(t_i - t)V_i$, where $V_i, i \in \mathbb{Z}$ are i.i.d. $N(0, I_{k\times k})$. $b_n \to 0$ and $nb_n/\log^2(n) \to \infty$. Let $m^* = 1/b_n$. Then*

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{1}{\sigma_{\hat{K},0}\sqrt{nb_n}} \sup_{t\in\mathcal{T}_n} |F_n(t)| - B_{\hat{K}}(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp\{-2\exp(-u)\}. \tag{3.91}$$

*where $B_{\hat{K}}$ is defined in (3.23).*

The following lemma is an analogue of Lemma 2 in [91]. Since we use other Gaussian approximation rates from Theorem 3.3.5, we shortly state the proof for completeness.

**Lemma 3.7.9.** *Let the assumptions and notations from Theorem 3.3.5 hold. For some kernel function $\hat{K}:\mathbb{R}\to\mathbb{R}$ which has bounded variation and compact support $[-1,1]$, define*

$$D_h(t) := (nb_n)^{-1}\sum_{i=1}^{n}\hat{K}_{b_n}(i/n-t)\tilde{h}_i(i/n).$$

*Assume that $\Sigma_{\tilde{h}}(t)$ is Lipschitz-continuous and that its smallest eigenvalue is bounded away from 0 uniformly on $[0,1]$. Assume that $\log(n)^4(b_n n^{\frac{2\gamma-1}{1+4\gamma}})^{-1}\to 0$ and $b_n\log(n)^{3/2}\to 0$. Then*

$$\lim_{n\to\infty}\mathbb{P}\Big(\frac{\sqrt{nb_n}}{\sigma_{\hat{K},0}}\sup_{t\in\mathcal{T}_n}\Big|\Sigma_{\tilde{h}}^{-1}(t)D_{\tilde{h}}(t)\Big| - B_{\hat{K}}(m^*)\le\frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)), \quad (3.92)$$

*Proof of Lemma 3.7.9.* By Theorem 3.3.5 and summation-by-parts, there exist i.i.d. $V_i\sim N(0,I_{s\times s})$ such that

$$\sup_{t\in\mathcal{T}_n}|D_{\tilde{h}}(t)-\Xi(t)| = O_{\mathbb{P}}\Big(\frac{n^{\frac{1+\gamma}{1+4\gamma}}\log(n)^{3/2}}{nb_n}\Big) = O_{\mathbb{P}}\Big(\frac{\log(n)^2\big(b_n n^{\frac{2\gamma-1}{1+4\gamma}}\big)^{-1/2}}{(nb_n)^{1/2}\log(n)^{1/2}}\Big), \quad (3.93)$$

where $\Xi(t) = (nb_n)^{-1}\sum_{i=1}^n K_{b_n}(i/n-t)\Sigma_{\tilde{h}}(i/n)V_i$. Here, (3.93) is $o_{\mathbb{P}}((nb_n)^{-1/2}\log(n)^{-1/2})$ due to $\log(n)^4\big(b_n n^{\frac{2\gamma-1}{1+4\gamma}}\big)^{-1}\to 0$. Since $\Sigma_{\tilde{h}}(\cdot)$ is Lipschitz continuous by Lemma 3.7.1, we can use a standard chaining argument in $t$ (as it was done in Lemma 3.7.7 for $\Pi_n(t)$) and the fact that $(nb_n)^{-1}\sum_{i=1}^n(\Sigma_{\tilde{h}}(i/n)-\Sigma_{\tilde{h}}(t))K_{b_n}(i/n-t)V_i\sim N(0,v_n)$, with $|v_n|_\infty\le C\frac{b_n}{n}$

for some constant $C > 0$ to obtain

$$
\begin{aligned}
& \sup_{t \in \mathcal{T}_n} |\Xi(t) - (nb_n)^{-1} \Sigma_{\tilde{h}}(t) \sum_{i=1}^{n} K_{b_n}(i/n - t) V_i| \\
= & \sup_{t \in \mathcal{T}_n} \left| (nb_n)^{-1} \sum_{i=1}^{n} K_{b_n}(i/n - t)(\Sigma_{\tilde{h}}(i/n) - \Sigma_{\tilde{h}}(t)) V_i \right| \\
= & O_{\mathbb{P}}\Big( \frac{b_n \log(n)}{(nb_n)^{1/2}} \Big) = O_{\mathbb{P}}\Big( \frac{b_n \log(n)^{3/2}}{(nb_n)^{1/2} \log(n)^{1/2}} \Big), \qquad (3.94)
\end{aligned}
$$

which is $o_{\mathbb{P}}((nb_n)^{-1/2} \log(n)^{-1/2})$ due to $b_n \log(n)^{3/2} \to 0$. So the result follows from Lemma 3.7.8 in view of (3.93) and (3.94). $\square$

*Proof of Theorem 3.3.6.* Let $\tilde{h}_i(t) := \nabla_\theta \ell(\tilde{Z}_i(t), \theta(t))$ and $\hat{K}(x) = K(x)$ or $\hat{K}(x) = K(x)x$, respectively. Define

$$
\Omega_C(t) := (nb_n)^{-1} \sum_{i=1}^{n} \hat{K}_{b_n}(i/n - t) A_C(i/n)^{\mathsf{T}} \tilde{h}_i(i/n)
$$

and $D_{\tilde{h}}(t) = (nb_n)^{-1} \sum_{i=1}^{n} \hat{K}_{b_n}(i/n - t) \tilde{h}_i(i/n)$. Similar to the discussion of $\Pi_n(t)$ in the proof of Lemma 3.7.7 (note that the rates in (3.89) and (3.90) then change to $O(b_n)$ instead of $O(b_n^2)$), we can show that

$$
\sup_{t \in \mathcal{T}_n} |\Omega_C(t) - A_C(t)^{\mathsf{T}} \cdot D_{\tilde{h}}(t)| = O_{\mathbb{P}}(\beta_n b_n) = O_{\mathbb{P}}\Big( \frac{b_n^{1/2} \log(n)}{(nb_n)^{1/2} \log(n)^{1/2}} \Big), \qquad (3.95)
$$

which is $o_{\mathbb{P}}((nb_n)^{-1/2} \log(n)^{-1/2})$ since $b_n \log(n)^2 \to 0$.

$h_i := (A_C(i/n)^{\mathsf{T}} \tilde{h}_i(i/n))_{i=1,\ldots,n}$ is a locally stationary process with long-run variance $\Sigma_h^2(t) = \Sigma_C^2(t)$. By the result of Lemma 3.7.9, we have that

$$
\lim_{n \to \infty} \mathbb{P}\Big( \frac{\sqrt{nb_n}}{\sigma_{\hat{K},0}} \sup_{t \in \mathcal{T}_n} \left| \Sigma_C^{-1}(t) \Omega_C(t) \right| - B_{\hat{K}}(m^*) \leq \frac{u}{\sqrt{2 \log(m^*)}} \Big) = \exp(-2 \exp(-u)). \qquad (3.96)
$$

(i) By Theorem 3.3.4(i), we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}_n}\big|V(t)\{\hat\theta_{b_n}(t)-\theta(t)\}-b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t)-D_{\tilde h}(t)\big|\\
&= O_{\mathbb{P}}\big(b_n^3+(nb_n)^{-1}b_n^{-1/2}\log(n)^{3/2}+(nb_n)^{-1/2}b_n\log(n)\big)\\
&= O_{\mathbb{P}}\Big(\frac{(nb_n^7\log(n))^{1/2}+(nb_n^2\log(n)^{-4})^{-1/2}+b_n\log(n)^{3/2}}{(nb_n)^{1/2}\log(n)^{1/2}}\Big), \qquad (3.97)
\end{aligned}$$

which is $o_{\mathbb{P}}((nb_n)^{-1/2}\log(n)^{-1/2})$ since $nb_n^7\log(n)\to 0$, $nb_n^2\log(n)^{-4}\to\infty$ and $b_n\log(n)^2\to 0$. Together with (3.95) and (3.96) (with $\hat K = K$), this implies (3.21).

(ii) By Theorem 3.3.4(ii), we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}_n}\big|\mu_{K,2}V(t)b_n\{\widehat{\theta}'_{b_n}(t)-\theta'(t)\}-b_n^3\frac{\mu_{K,4}}{2}V(t)\mathrm{bias}(t)-D_{\tilde h}(t)\big|\\
&= O_{\mathbb{P}}\big(b_n^4+(nb_n)^{-1}b_n^{-1/2}\log(n)^{3/2}+(nb_n)^{-1/2}b_n\log(n)\big)\\
&= o_{\mathbb{P}}((nb_n)^{-1/2}\log(n)^{-1/2}),
\end{aligned}$$

as above. Together with (3.95) and (3.96) (with $\hat K(x) = K(x)x$), this implies (3.22). □

### 3.7.3 Proof of Section 3.4

*Proof of Proposition 3.4.1.* (i) By $nb_n^2\log(n)^{-2d_\Theta}\to\infty$, Lemma 3.7.4(i),(ii), Lemma 3.7.5 and the notation therein applied to $g=\nabla_\theta^2\ell\in\mathcal{H}(M_y,M_x,\chi,\bar C)$, it holds that

$$\begin{aligned}
\sup_{t\in\mathcal{T}_n}|\hat V_{b_n}(t)-V(t)| \leq\ & \sup_{t\in\mathcal{T}_n,\eta\in E_n}|G_n^c(t,\eta)-\hat G_n(t,\eta)| \qquad (3.98)\\
&+\sup_{t\in\mathcal{T}_n,\eta\in E_n}|\hat G_n(t,\eta)|\\
&+\sup_{t\in\mathcal{T}_n,\eta\in E_n}|\mathbb{E}\hat B_n(t,\eta)-V^\circ(t,\eta)|\\
&+\sup_{t\in\mathcal{T}_n}|V^\circ(t,\hat\eta_{b_n})-V(t)|\\
=\ & O_{\mathbb{P}}((nb_n)^{-1})+o_{\mathbb{P}}(\log(n)^{-1})+O(b_n)+\sup_{t\in\mathcal{T}_n}|V^\circ(t,\hat\eta_{b_n})-V(t)|.
\end{aligned}$$

By using Lemma 3.7.4(ii) instead of Lemma 3.7.4(iii), we obtain similar as in the proof of Theorem 3.3.4(i) that (3.52) that

$$\sup_{t\in\mathcal{T}_n}|\hat{\eta}_{b_n}(t)-\eta_n(t)| \leq O_{\mathbb{P}}(1)\sup_{t\in\mathcal{T}_n}|\nabla_\eta L_{n,b_n}^{\circ,c}(t,\eta_{b_n}(t))|.$$

Note that $\mathbb{E}\nabla_\theta\ell(\tilde{Z}_0(t),\theta(t)) = 0$. Using Lemma 3.7.4(i),(ii) and Lemma 3.7.5 with $g = \nabla_\theta\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we obtain

$$\begin{aligned}
\sup_{t\in\mathcal{T}_n}|\nabla_\eta L_{n,b_n}^{\circ,c}(t,\eta_{b_n}(t))| &\leq \sup_{t\in\mathcal{T}_n,\eta\in E_n}|G_n^c(t,\eta)-\hat{G}_n(t,\eta)| + \sup_{t\in\mathcal{T}_n,\eta\in E_n}|\hat{G}_n(t,\eta)| \\
&\quad + \sup_{t\in\mathcal{T}_n,\eta\in E_n}|\mathbb{E}\hat{B}_n(t,\eta)| \\
&= O_{\mathbb{P}}((nb_n)^{-1}) + o_{\mathbb{P}}(\log(n)^{-1}) + O(b_n)
\end{aligned}$$

and thus

$$\sup_{t\in\mathcal{T}_n}|\hat{\eta}_{b_n}(t)-\eta_n(t)| = O_{\mathbb{P}}((nb_n)^{-1}+b_n) + o_{\mathbb{P}}(\log(n)^{-1}).$$

Since $\eta \mapsto V^\circ(t,\eta)$ is Lipschitz continuous by Lemma 3.7.1, the result follows from (3.99) and $b_n\log(n) \to 0$.

(ii) follows similarly due to $\nabla_\theta\ell\cdot\nabla_\theta\ell^\intercal \in \mathcal{H}(2M_y, 2M_y, \chi, \tilde{C})$ with some $\tilde{C} > 0$.

□

To prove Theorem 3.4.2, we adopt the methods used in [91]. Let us first assume that $\theta(\cdot)$ and the stationary approximation $\tilde{Z}_i(t)$ is known. Define $D_i := \nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))$, $Q_i := \sum_{j=-m}^m D_{i+j}$ and $\Phi_i := Q_iQ_i^\intercal/(2m+1)$. Recall that $\tau_n$ is some bandwidth and $\gamma_n = \tau_n + (m+1)/n$. For $t \in \mathcal{I}_n = [\gamma_n, 1-\gamma_n] \subset (0,1)$, define

$$\hat{\Lambda}(t) := \frac{\sum_{i=1}^n K_{\tau_n}(i/n-t)\Phi_i}{\sum_{i=1}^n K_{\tau_n}(i/n-t)}.$$

Note that $\hat{\Lambda}(t)$ is always positive semi-definite. We have the following convergence result.

**Theorem 3.7.10.** *Suppose that Assumption 3.2.1 holds with $r = 4$. Assume that $m = m_n \to \infty$, $m = O(n^{1/3})$, $\tau_n \to 0$ and $n\tau_n \to \infty$. Then with $\rho = 1$,*

*(i) For fixed $t \in (0,1)$,*

$$\|\hat{\Lambda}(t) - \Lambda(t)\|_2 = O\Big(\sqrt{\frac{m}{n\tau_n}} + m^{-1} + \tau_n^{\rho}\Big).$$

*(ii) We have*

$$\|\sup_{t \in \mathcal{I}_n} |\hat{\Lambda}(t) - \Lambda(t)|\|_2 = O\Big(\sqrt{\frac{m}{n\tau_n^2}} + m^{-1} + \tau_n^{\rho}\Big).$$

*If additionally Assumption 3.2.2(B1),(B3) is fulfilled with $M' = 2M$ and $\nabla_\theta \ell$ is continuously differentiable with $\partial_{z_j} \nabla_\theta \ell \in \mathcal{H}(M_y - 1, M_x - 1, \chi', \hat{\chi}_j \bar{C})$ for all $j \in \mathbb{N}_0$, then one can choose $\rho = 2$.*

*Proof of Theorem 3.7.10.* Let $\tilde{D}_i(t) := \nabla_\theta \ell(\tilde{Z}_i(t), \theta(t))$. By Lemma 3.7.3(i) applied to $\nabla_\theta \ell$, it holds that $\sup_{t\in[0,1]} \delta_4^{\tilde{D}(t)}(l) = O(l^{-(1+\gamma)})$. By Lemma 3.7.1, $\sup_{t\in[0,1]} \|\tilde{D}_0(t)\|_4 < \infty$. It is easily seen by Lemma 3.7.1 applied to $\nabla_\theta \ell \nabla_\theta \ell^{\mathsf{T}} \in \mathcal{H}(2M_y, 2M_x, \chi)$ that $t \mapsto \Lambda(t)$ is Lipschitz-continuous. Thus $D_i = \tilde{D}_i(i/n)$ has the same properties as $L_i$ in [91]. The proof therefore is completely the same as the proof of Theorem 4 in [91] with a modified bias term $\rho = 1$ and is omitted.

Under the additional assumption, we have $g = \nabla_\theta \ell \nabla_\theta \ell^{\mathsf{T}} \in \mathcal{H}(2M_y, 2M_x, \chi, \bar{C}')$ and $\partial_{z_j}(\nabla_\theta \ell \nabla_\theta \ell) \in \mathcal{H}(2M_y - 1, 2M_y - 1, (\max\{\chi_i', \chi_i\})_{i\in\mathbb{N}}, \bar{C}'\chi_j)$ with some $\bar{C}' > 0$. Application of Lemma 3.7.2 to $g$ shows that $\Lambda(t)$ is continuously differentiable with Lipschitz continuous derivative. This shows that in this case, one can choose $\rho = 2$. □

*Proof of Theorem 3.4.2.* We follow the steps in the proof of Theorem 5 in [91]. Since $\nabla_\theta^2 \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we have

$$\sup_{i=1,\dots,n} \sup_{\theta\in\Theta} |\nabla_\theta^2 \ell(\tilde{Z}_i(i/n), \theta)|_\infty \le 2 \sup_{\theta\in\Theta} |\theta|_\infty \cdot \sup_{i=1,\dots,n} R_{M_y,M_x}(\tilde{Z}_i(i/n)).$$

Note that $\sup_{0\le t\le 1}\|\tilde{Z}_0(t)\|_{4M}<\infty$. By Lemma 3.7.1, we have $\sup_{i=1,\dots,n} R_{M_y,M_x}(\tilde{Z}_i(i/n)) = O_{\mathbb{P}}(n^{1/4})$ and thus

$$\sup_{i=1,\dots,n}\sup_{\theta\in\Theta}|\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta)|_\infty = O_{\mathbb{P}}(n^{1/4}). \tag{3.99}$$

Put $D_i^{\#}=\nabla_\theta\ell(\tilde{Z}_i(i/n),\hat{\theta}_{b_n}(i/n))$ and define $Q_i^{\#}$, $\Phi_i^{\#}$ and $\Lambda^{\#}(t)$ accordingly. Then we have

$$\begin{aligned}
&\|\sup_{t\in\mathcal{I}_n}|\tilde{\Lambda}(t)-\Lambda^{\#}(t)|\|_1\\
\le\ &\sup_{t\in\mathcal{I}_n}\Big(\sum_{i=1}^n K_{\tau_n}(i/n-t)\Big)^{-1}\cdot|K|_\infty\\
&\qquad\times\sum_{i=(m+1)n+\tau_n}^{n}\sup_{j=-m,\dots,m}\big\{\|\tilde{D}_{i+j}-D_{i+j}^{\#}\|_2\|\tilde{D}_{i+j}\|_2+\|\tilde{D}_{i+j}-D_{i+j}^{\#}\|_2\|D_{i+j}^{\#}\|_2\big\}.
\end{aligned}$$

By the results of Lemma 3.7.1 applied to $\nabla_\theta\ell\in\mathcal{H}(M_y,M_x,\chi,\bar{C})$, we obtain with some constant $\tilde{C}>0$ that

$$\sup_{i,j}\|\tilde{D}_{i+j}\|_2\le\tilde{C},\qquad \|\tilde{D}_{i+j}-D_{i+j}^{\#}\|_2\le\tilde{C}\big(n^{-1}+\sum_{l=i+j}^{\infty}\chi_l\big),$$

and thus $\|\sup_{t\in\mathcal{I}_n}\|\tilde{\Lambda}(t)-\Lambda^{\#}(t)|\|_1 = O((n\tau_n)^{-1})$.

Define $\beta_n' := (nb_n)^{-1/2}\log(n)+(nb_n)^{-1}+\beta_n b_n^2+b_n^2$. Then by (3.99) and the fact that $\hat{\theta}_{b_n}(i/n)-\theta(i/n)=O_{\mathbb{P}}(\beta_n')$ from (3.58), we have

$$\sup_{i/n\in\mathcal{I}_n}|D_i-\tilde{D}_i|\le\sup_{i/n\in\mathcal{I}_1}|\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\bar{\theta}(i/n))|\cdot|\hat{\theta}_{b_n}(i/n)-\theta(i/n)| = n^{1/4}\beta_n'. \tag{3.100}$$

Note that $Q_i/(2m+1)$ is the Nadaraya-Watson-type smoother of the series $D_i$ with the rectangle kernel and bandwidth $\tilde{b}_n=m/n$. By using (3.55) in this context, we obtain

$$\sup_{i/n\in\mathcal{I}_n}\frac{1}{2m+1}|Q_i| = O_{\mathbb{P}}((n\tilde{b}_n)^{-1/2}\log(n)) = O_{\mathbb{P}}(m^{1/2}\log(n)). \tag{3.101}$$

Comparing $\Phi_i$ and $\tilde{\Phi}_i$ we obtain

$$(2m+1)(\Phi_i - \tilde{\Phi}_i) = (Q_i - \tilde{Q}_i)Q_i^{\mathsf{T}} + Q_i(Q_i - \tilde{Q}_i)^{\mathsf{T}} - (Q_i - \tilde{Q}_i)(Q_i - \tilde{Q}_i)^{\mathsf{T}}.$$

By equations (3.100) and (3.101), we have $\sup_{i/n \in \mathcal{I}_1} |\Phi_i - \tilde{\Phi}_i| = O_{\mathbb{P}}(\omega_n)$. This implies

$$\sup_{i/n \in \mathcal{I}_n} |\hat{\Lambda}(i/n) - \tilde{\Lambda}(i/n)| = O_{\mathbb{P}}(\omega_n).$$

The results from Theorem 3.7.10 now imply the assertion. □

### *3.7.4 Proofs of the examples in Section 3.5*

*Proof of Example 3.5.1.* Let $q = 4M$. Let $\nu = (\nu_0, \dots, \nu_l)^{\mathsf{T}}$ and $m = (m_1, \dots, m_k)^{\mathsf{T}}$. As known from Proposition 4.3 and Lemma 4.4 in [13] the process $(Y_i)_{i=1,\dots,n}$ described by (3.34) exists and fulfills Assumption 3.2.1(A5), (A7) with $\delta_q^Y(i) = O(c^i)$ for some $0 < c < 1$ and $q \geq 1$ if the recursion function $G_\zeta(y,t) := \mu(y, \theta(t)) + \sigma(y, \theta(t))\zeta$ obeys

$$\Big\| \sup_{t \in [0,1]} \sup_{y \neq y'} \frac{|G_{\zeta_0}(y,t) - G_{\zeta_0}(y',t)|}{|y - y'|_{\chi,1}} \Big\|_q \leq 1 \tag{3.102}$$

and

$$\sup_{t \in [0,1]} \|C(\tilde{X}_t(t))\|_q < \infty, \quad C(y) := \sup_{t \neq t'} \frac{\|G_{\zeta_0}(y,t) - G_{\zeta_0}(y,t')\|_q}{|t - t'|}, \tag{3.103}$$

where $|z|_{\chi,1} := \sum_{i=1}^p |z_i|\chi_i$ for some $\chi = (\chi_i)_{i=1,\dots,p} \in \mathbb{R}^p_{\geq 0}$ with $|\chi|_1 = \sum_{i=1}^p \chi_i < 1$. Here, we can bound

$$|\mu(y,\theta) - \mu(y',\theta)| \leq \sum_{i=1}^k |\alpha_i| |y - y'|_{\kappa_{i\cdot},1} \leq |y - y'|_{\kappa^{(\mu)}(\alpha),1}, \tag{3.104}$$

where $\chi^{(\mu)}(\alpha) := \sum_{i=1}^{k} |\alpha_i|\kappa_{i\cdot}$. Furthermore,

$$\begin{aligned}
|\sigma(y,\theta)^2 - \sigma(y',\theta)^2| &\leq \sum_{i=0}^{l} \beta_i |\nu_i(y) - \nu_i(y')| \\
&\leq \sum_{i=0}^{l} \sqrt{\beta_i} |y-y'|_{\rho_{i\cdot},1} \cdot \big(\sqrt{\beta_i \nu_i(y)} + \sqrt{\beta_i \nu_i(y')}\big) \\
&\leq \sum_{i=0}^{l} \sqrt{\beta_i} |y-y'|_{\rho_{i\cdot},1} \cdot \big(\sigma(y,\theta) + \sigma(y',\theta)\big),
\end{aligned}$$

i.e.

$$|\sigma(y,\theta) - \sigma(y',\theta)| \leq |y-y'|_{\chi^{(\sigma)}(\beta),1}, \tag{3.105}$$

where $\chi^{(\sigma)}(\beta) := \sum_{i=1}^{l} \sqrt{\beta_i}\rho_{i\cdot}$. Define

$$\chi_j^{(\mu,max)} := \sup_t |\chi^{(\mu)}(\alpha(t))_j|, \qquad \chi_j^{(\sigma,max)} := \sup_t |\chi^{(\sigma)}(\beta(t))_j|.$$

Since $\theta(t) = (\alpha(t)^{\mathsf{T}}, \beta(t)^{\mathsf{T}})^{\mathsf{T}} \in \Theta$, we have that

$$\sum_{j=1}^{p} (\chi_j^{(\mu,max)} + \|\zeta_0\|_q \chi_j^{(\sigma,max)}) = \sum_{j=1}^{p} \big(\sup_t |\chi^{(\mu)}(\alpha(t))_j| + \|\zeta_0\|_q \sup_t |\chi^{(\sigma)}(\beta(t))_j|\big) < 1.$$

Define $\chi_j := \chi_j^{(\mu,max)} + \|\zeta_0\|_q \chi_j^{(\sigma,max)}$. Then we have for all $t, y \neq y'$:

$$|\mu(y,\theta(t)) - \mu(y',\theta(t))| + \|\zeta_0\|_q |\sigma(y,\theta(t)) - \sigma(y',\theta(t))| \leq |y-y'|_{\chi,1},$$

which implies (3.102). Proposition 4.3 from [13] now implies the existence of $Y_i$, the stationary approximation $\tilde{Y}_i(t)$ and $\sup_t \|\tilde{Y}_0(t)\|_q < \infty$. By Lipschitz continuity of $\theta$ with constant $L_\theta$, we have

$$|\mu(y,\theta(t)) - \mu(y,\theta(t'))| \leq L_\theta |t-t'| \sum_{i=1}^{k} |m_i(y)|, \tag{3.106}$$

and

$$\begin{aligned}|\sigma(y,\theta(t))^2-\sigma(y,\theta(t'))^2| &\le L_\theta|t-t'|\sum_{i=0}^{l}\sqrt{\nu_i(y)}\frac{1}{2\beta_{min}^{1/2}}\big(\sqrt{\beta_i(t)\nu_i(y)}+\sqrt{\beta_i(t')\nu_i(y)}\big)\\ &\le \frac{L_\theta}{2\beta_{min}^{1/2}}|t-t'|\sum_{i=0}^{l}\sqrt{\nu_i(y)}(\sigma(y,\theta(t))+\sigma(y,\theta(t'))),\end{aligned}$$

which shows that

$$|\sigma(y,\theta(t))-\sigma(y,\theta(t'))|\le\frac{L_\theta}{2\beta_{min}^{1/2}}\sum_{i=0}^{l}\sqrt{\nu_i(y)}. \tag{3.107}$$

Note that (3.35) implies

$$m_i(y),\sqrt{\nu_i(y)}\le C_1|y|_1+C_2,$$

with some constants $C_1,C_2>0$. By (3.106), (3.107), we have for $t\neq t'$

$$\begin{aligned}\|G_{\zeta_0}(y,t)-G_{\zeta_0}(y,t')\|_q &\le |\mu(y,\theta(t))-\mu(y,\theta(t'))|+\|\zeta_0\|_q|\sigma(y,\theta(t))-\sigma(y,\theta(t'))|\\ &\le C_3|t-t'|\big(1+|y|_1\big),\end{aligned}$$

with some constant $C_3>0$. Since $\sup_t\|\tilde{Y}_0(t)\|_q<\infty$, (3.103) follows.

We now inspect the properties of the function $\ell$. First note that the recursion of the stationary approximation,

$$\tilde{Y}_i(t)=\mu(\tilde{X}_i(t),\theta(t))+\sigma(\tilde{X}_i(t),\theta(t))\zeta_i,$$

implies $\mathbb{E}\tilde{Y}_0(t)=0$ and $\mathbb{E}\tilde{Y}_0(t)^2=\mathbb{E}\mu(\tilde{X}_0(t),\theta(t))^2+\mathbb{E}\sigma(\tilde{X}_0(t),\theta(t))^2\ge\beta_{min}\nu_{min}>0$. Furthermore, for $L(t,\theta):=\mathbb{E}\ell(\tilde{Z}_0(t),\theta)$ it holds that

$$\begin{aligned}L(t,\theta)-L(t,\theta(t)) &= \mathbb{E}\Big(\frac{\mu(\tilde{X}_0(t),\theta)-\mu(\tilde{X}_0(t),\theta(t))}{\sigma(\tilde{X}_0(t),\theta)}\Big)^2\\ &\quad+\mathbb{E}\Big[\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2}-\log\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2}-1\Big].\end{aligned}\tag{3.108}$$

In the following we use the notation $|x|_A^2 := x^\mathsf{T} A x$ for a weighted vector norm. Note that

$$\mathbb{E}\Big(\frac{\mu(\tilde{X}_0(t),\theta) - \mu(\tilde{X}_0(t),\theta(t))}{\sigma(\tilde{X}_0(t),\theta)}\Big)^2 \geq c_0|\alpha - \alpha(t)|^2_{M_1(t)}, \tag{3.109}$$

with $c_0 = (\max_{\theta\in\Theta} \max_i \theta_i^2)^{-1}$ and $M_1(t) := \mathbb{E}[\frac{m(\tilde{X}_0(t))m(\tilde{X}_0(t))^\mathsf{T}}{\mathbb{1}\nu(\tilde{X}(t))\nu(\tilde{X}(t))^\mathsf{T}\mathbb{1}}]$. If $M_1(t)$ was not positive definite, this would imply that there exists $v \in \mathbb{R}^k$ such that $v'M(t)v = 0$, which in turn would imply $v'\mu(\tilde{X}_0(t))\mu(\tilde{X}_0(t))v = 0$ a.s. and thus non-positive definiteness of $\mathbb{E}[\mu(\tilde{X}_0(t))\mu(\tilde{X}_0(t))^\mathsf{T}]$ which is a contradiction to the assumption.

By a Taylor expansion of $f(x) = x - \log(x) - 1$, we obtain

$$\begin{aligned}
& \mathbb{E}\Big[\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2} - \log\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2} - 1\Big] \\
\geq\ & \frac{1}{2}\mathbb{E}\Big[\frac{(\sigma(\tilde{X}_0(t),\theta)^2 - \sigma(\tilde{X}_0(t),\theta(t))^2)^2}{(\sigma(\tilde{X}_0(t),\theta)^2 - \sigma(\tilde{X}_0(t),\theta(t))^2)^2 + \sigma(\tilde{X}_0(t),\theta)^4}\Big] \\
\geq\ & \frac{c_0}{10}|\beta - \beta(t)|^2_{M_2(t)},
\end{aligned} \tag{3.110}$$

where $M_2(t) = \mathbb{E}[\frac{\nu(\tilde{X}_0(t))\nu(\tilde{X}_0(t))^\mathsf{T}}{\mathbb{1}\nu(\tilde{X}(t))\nu(\tilde{X}(t))^\mathsf{T}\mathbb{1}}]$ is positive definite by assumption (use a similar argumentation as above). By (3.108), (3.109) and (3.110) we conclude that $\theta \mapsto L(t,\theta)$ is uniquely minimized in $\theta = \theta(t)$. This shows 3.2.1(A3).

Omitting the arguments $z = (y, x)$ and $\theta$, we have

$$\ell = \frac{1}{2}\Big[\frac{(y - \langle\alpha, m\rangle)^2}{\langle\beta,\nu\rangle} + \log\langle\beta,\nu\rangle\Big], \tag{3.111}$$

$$\begin{aligned}
\nabla_\theta \ell &= -\frac{\nabla_\theta m}{\sigma}\Big(\frac{y-m}{\sigma}\Big) + \frac{\nabla_\theta(\sigma^2)}{2\sigma^2}\Big[1 - \Big(\frac{y-m}{\sigma}\Big)^2\Big] \\
&= \begin{pmatrix} -\frac{m}{\sigma}\Big(\frac{y-m}{\sigma}\Big) \\ \frac{\nu}{2\sigma^2}\big[1-\big(\frac{y-m}{\sigma}\big)^2\big] \end{pmatrix} = \begin{pmatrix} \frac{m}{\langle\beta,\nu\rangle}(y - \langle\alpha,m\rangle) \\ \frac{\nu}{2\langle\beta,\nu\rangle}\big(1 - \frac{(y-\langle\alpha,m\rangle)^2}{\langle\beta,\nu\rangle}\big) \end{pmatrix},
\end{aligned} \tag{3.112}$$

$$\begin{aligned}
\nabla^2_\theta \ell &= \frac{\nabla_\theta m \nabla_\theta m^\intercal}{\sigma^2} + \big(\frac{y-m}{\sigma}\big)\cdot\Big[\frac{\nabla_\theta m \nabla_\theta(\sigma^2)^\intercal + \nabla_\theta(\sigma^2)\nabla_\theta m^\intercal}{\sigma^3} - \frac{\nabla^2_\theta m}{\sigma}\Big] \\
&\quad + \frac{\nabla^2_\theta(\sigma^2)}{2\sigma^2}\Big[1 - \big(\frac{y-m}{\sigma}\big)^2\Big] + \frac{\nabla_\theta(\sigma^2)\nabla_\theta(\sigma^2)^\intercal}{2\sigma^4}\Big[2\Big(\frac{y-m}{\sigma}\Big)^2 - 1\Big] \\
&= \begin{pmatrix} \frac{mm^\intercal}{\sigma^2} & \frac{y-m}{\sigma^2}\cdot m\nu^\intercal \\ \frac{y-m}{\sigma^2}\cdot \nu m^\intercal & \frac{\nu\nu^\intercal}{2\sigma^4}\big[2\big(\frac{y-m}{\sigma}\big)^2 - 1\big] \end{pmatrix} \\
&= \begin{pmatrix} \frac{mm^\intercal}{\langle\beta,\nu\rangle} & \frac{y-\langle\alpha,m\rangle}{\langle\beta,\nu\rangle^2}\cdot m\nu^\intercal \\ \frac{y-\langle\alpha,m\rangle}{\langle\beta,\nu\rangle^2}\cdot \nu m^\intercal & \frac{\nu\nu^\intercal}{2\langle\beta,\nu\rangle^2}\big[2\frac{(y-\langle\alpha,m\rangle)^2}{\langle\beta,\nu\rangle} - 1\big] \end{pmatrix}.
\end{aligned} \tag{3.113}$$

Since $\zeta_1$ is independent of $\tilde{X}_0(t) \in \mathcal{F}_0$ and $\mathbb{E}\zeta_1 = 0$, $\mathbb{E}\zeta_1^2 = 1$, we conclude that

$$\mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t))|\mathcal{F}_{t-1}] = \mathbb{E}\Big[-\frac{\mu(\tilde{X}_j(t),\theta(t))}{\sigma(\tilde{X}_0(t),\theta(t))}\zeta_0 + \frac{\nu(\tilde{X}_0(t),\theta(t))}{2\sigma(\tilde{X}_j(t),\theta(t))^2}(1-\zeta_0^2)\big|\mathcal{F}_{t-1}\Big] = 0,$$

i.e. $\nabla_\theta \ell(\tilde{Z}_1(t), \theta(t))$ is a martingale difference sequence, showing that $V(t) = \Lambda(t)$. We furthermore have that (we omit the arguments $(\tilde{X}_0(t), \theta(t))$ of $\mu$, $\sigma$ in the following):

$$V(t) = \mathbb{E}\nabla^2_\theta \ell(\tilde{Z}_0(t), \theta(t)) = \begin{pmatrix} \mathbb{E}\big[\frac{mm^\intercal}{\langle\beta,\nu\rangle}\big] & 0 \\ 0 & \mathbb{E}\big[\frac{\nu\nu^\intercal}{2\langle\beta,\nu\rangle^2}\big] \end{pmatrix}.$$

With a similar argumentation as above, we conclude that $V(t)$ is positive definite (which then implies by continuity that the smallest eigenvalue of $V(t)$ is bounded away from 0 uniformly in $t$). By the martingale difference property, $I(t) = \Lambda(t)$. Omitting the arguments

$(\tilde{X}_0(t), \theta(t))$,

$$\begin{aligned}
I(t) &= \mathbb{E}[\nabla_\theta(\tilde{Z}_j(t), \theta(t))\nabla_\theta(\tilde{Z}_0(t), \theta(t))^\mathsf{T}] \\
&= \begin{pmatrix} \mathbb{E}\big[\frac{mm^\mathsf{T}}{\sigma^2}\big] & \mathbb{E}[\zeta_0^3] \cdot \mathbb{E}\big[\frac{m\nu^\mathsf{T}}{2\sigma^3}\big] \\ \mathbb{E}[\zeta_0^3] \cdot \mathbb{E}\big[\frac{\nu m^\mathsf{T}}{2\sigma^3}\big] & \frac{\mathbb{E}[\zeta_0^4]-1}{4} \cdot \mathbb{E}\big[\frac{\nu\nu^\mathsf{T}}{\sigma^4}\big] \end{pmatrix} \\
&= \mathbb{E}\Big[\frac{1}{\sigma^2}\begin{pmatrix} m \\ \frac{\mathbb{E}[\zeta_0^3]}{2\sigma}\nu \end{pmatrix}^\mathsf{T}\begin{pmatrix} m \\ \frac{\mathbb{E}[\zeta_0^3]}{2\sigma}\nu \end{pmatrix}\Big] + \begin{pmatrix} 0 & 0 \\ 0 & (\frac{\mathbb{E}[\zeta_0^4]-\mathbb{E}[\zeta_0^3]^2-1}{4})\mathbb{E}\big[\frac{\nu\nu^\mathsf{T}}{\sigma^4}\big] \end{pmatrix},
\end{aligned}$$

which is positive semidefinite since $\mathbb{E}[\zeta_0^3] = \mathbb{E}[\zeta_0(\zeta_0^2-1)] \le \mathbb{E}[\zeta_0^2]^{1/2}\mathbb{E}[(\zeta_0^2-1)^2]^{1/2} = (\mathbb{E}[\zeta_0^4]-1)^{1/2}$. Positive definiteness follows from the fact that $(v_1, v_2)^\mathsf{T} I(t)(v_1, v_2) = 0$ implies $\nu^\mathsf{T} v_2 = 0$ a.s. from the last summand and $v_1^\mathsf{T} m + \frac{\mathbb{E}[\zeta_0^3]}{2\sigma} v_2^\mathsf{T}\nu = 0$ a.s. from the first summand, i.e. $v_1^\mathsf{T} m = 0$ a.s. which leads to a contradiction to either the positive definiteness of $\mathbb{E}[\nu\nu^\mathsf{T}]$ or $\mathbb{E}[mm^\mathsf{T}]$. So we obtain that Assumption 3.2.1(A4) is fulfilled.

A careful inspection of (3.111), (3.112) and (3.113) shows that $\ell, \nabla_\theta\ell, \nabla_\theta^2\ell \in \mathcal{H}(2, 3, \tilde{\chi}, \tilde{C})$ with some $\tilde{C} > 0$ and $\tilde{\chi} = (1, \ldots, 1, 0, 0, \ldots)$ consisting of $\max\{k, l\}$ ones followed by zeros, which shows Assumption 3.2.1(A1). In the special case $\mu(x, \theta) \equiv 0$, it seems as if no direct improvement of the value $M$ is possible. In the special case of $\sigma(x, \theta)^2 \equiv \beta_0$, we have

$$\begin{aligned}
\ell &= \frac{1}{2}\Big[\frac{(y - \langle\alpha, m\rangle)^2}{\beta_0} + \log\beta_0\Big], \\
\nabla_\theta\ell &= \begin{pmatrix} \frac{m}{\beta_0}(y - \langle\alpha, m\rangle) \\ \frac{1}{2\beta_0}(1 - \frac{(y-\langle\alpha,m\rangle)^2}{\beta_0}) \end{pmatrix}, \\
\nabla_\theta^2\ell &= \begin{pmatrix} \frac{mm^\mathsf{T}}{\beta_0} & \frac{y-\langle\alpha,m\rangle}{\beta_0^2} m \\ \frac{y-\langle\alpha,m\rangle}{\beta_0^2} m^\mathsf{T} & \frac{1}{2\beta_0^2}\big[2\frac{(y-\langle\alpha,m\rangle)^2}{\beta_0} - 1\big] \end{pmatrix},
\end{aligned}$$

which implies that $\ell, \nabla_\theta\ell, \nabla_\theta^2\ell \in \mathcal{H}(2, 2, \tilde{\chi}, \tilde{C})$.

Now suppose that Assumption 3.2.2(B1) is fulfilled. We use results from Section 4 in [13] to show that the first derivative process $\partial_t\tilde{Y}_i(t)$ exists and fulfills a Lipschitz condition. By assumption, with some constant $C > 0$,

$$
\begin{aligned}
& |\partial_{x_j} G_{\zeta_0}(x,t) - \partial_{x_j} G_{\zeta_0}(x',t)| \\
\leq\ & |\langle \alpha(t), \partial_{x_j} m(x) - \partial_{x_j} m(x')\rangle| \\
& \quad + \Big| \frac{\langle \beta(t), \partial_{x_j}\nu(x)\rangle}{2\langle \beta(t), \nu(x)\rangle^{1/2}} - \frac{\langle \beta(t), \partial_{x_j}\nu(x')\rangle}{2\langle \beta(t), \nu(x')\rangle^{1/2}} \Big| \cdot |\zeta_0| \\
\leq\ & |\alpha(t)|_\infty C|x - x'|_1 \\
& \quad + |\zeta_0| \cdot \Big( \frac{1}{2\beta_{min}^{1/2}} |\langle \beta(t), \partial_{x_j}\nu(x) - \partial_{x_j}\nu(x')\rangle| \\
& + \frac{|\langle \beta(t), \partial_{x_j}\nu(x')\rangle|}{2} \frac{|\langle \beta(t), \nu(x) - \nu(x')\rangle|}{\langle \beta(t), \nu(x)\rangle^{1/2}\langle \beta(t), \nu(x')\rangle^{1/2}(\langle \beta(t), \nu(x)\rangle^{1/2} + \langle \beta(t), \nu(x')\rangle^{1/2})} \Big).
\end{aligned}
$$

By assumption,

$$
|\langle \beta(t), \partial_{x_j}\nu(x) - \partial_{x_j}\nu(x')\rangle| \leq C|\beta(t)|_\infty |x - x'|_1.
$$

Furthermore,

$$
\begin{aligned}
|\langle \beta(t), \nu(x) - \nu(x')\rangle| &\leq \sum_{i=1}^{l} \beta_i(t) |\sqrt{\nu_i(x)} - \sqrt{\nu_i(x')}| \cdot |\sqrt{\nu_i(x)} + \sqrt{\nu_i(x')}| \\
&\leq |\beta(t)|_\infty |x - x'|_1 \sum_{i=1}^{l} \big(\sqrt{\nu_i(x)} + \sqrt{\nu_i(x')}\big).
\end{aligned}
$$

Since each component of $\beta(t)$ is lower bounded by $\beta_{min}$ and therefore $\langle \beta(t), \nu(x)\rangle^{1/2} \geq \beta_{min}\sqrt{\nu_i(x)}$ for each $i$, we conclude that

$$
\frac{|\langle \beta(t), \nu(x) - \nu(x')\rangle|}{\langle \beta(t), \nu(x)\rangle^{1/2}(\langle \beta(t), \nu(x)\rangle^{1/2} + \langle \beta(t), \nu(x')\rangle^{1/2})} \leq C|x - x'|_1,
$$

with some constant $C > 0$. Finally, let $e_j$ be the $j$-th unit vector in $\mathbb{R}^k$. Notice that for all $i$,

$$
\frac{|\partial_{x_j}\nu_i(x)|}{\nu_i(x)^{1/2}} = 2|\partial_{x_j}\sqrt{\nu_i(x)}| \leq 2\lim_{h\to 0} \frac{|\sqrt{\nu_i(x)} - \sqrt{\nu_i(x + he_j)}|}{|h|} \leq \lim_{h\to 0} \frac{|he_j|_1}{|h|} \leq 1.
$$

This shows that $\frac{|\langle\beta(t),\partial_{x_j}\nu(x')\rangle|}{\langle\beta(t),\nu(x')\rangle^{1/2}}$ is bounded and we obtain that for some constant $> 0$,

$$|\partial_{x_j} G_{\zeta_0}(x,t) - \partial_{x_j} G_{\zeta_0}(x',t)| \leq C(1+|\zeta_0|)|x-x'|_1.$$

With similar but simpler arguments we obtain that

$$|\partial_t G_{\zeta_0}(x,t) - \partial_t G_{\zeta_0}(x',t)| \leq C(1+|\zeta_0|)|x-x'|_1$$

and

$$\frac{|\partial_{x_j} G_{\zeta_0}(x,t) - \partial_{x_j} G_{\zeta_0}(x,t')|}{|t-t'|\cdot|x|_1} \leq C(1+|\zeta_0|), \qquad \frac{|\partial_t G_{\zeta_0}(x,t) - \partial_t G_{\zeta_0}(x,t')|}{|t-t'|\cdot|x|_1} \leq C(1+|\zeta_0|).$$

By Theorem 4.8 and Proposition 4.11 in [13], we obtain 3.2.2(B3) with $M_2 = 2M$. Finally, straightforward calculations show that each component of $\partial_{x_j}\nabla_\theta^2\ell$ and $\nabla_\theta^3\ell$ in $\mathcal{H}(M_2,\tilde{\chi})$ with $\tilde{\chi} = (1,\dots,1,0,\dots,0)$ consisting of $p$ ones. This shows Assumption 3.2.2(B2).

□

*Proof of Example 3.5.2.* Let $q = 4M$, and $M = 4$. Fix $t \in [0,1]$. Consider the recursion of the corresponding stationary approximation

$$\begin{aligned} \tilde{Y}_i(t) &= \tilde{\sigma}_i(t)^2\zeta_i^2, \\ \tilde{\sigma}_i(t)^2 &= \alpha_0(t) + \sum_{j=1}^m \alpha_j(t)\tilde{Y}_{i-j}(t) + \sum_{j=1}^l \beta_j(t)\tilde{\sigma}_{i-j}(t)^2. \end{aligned} \tag{3.114}$$

Define

$$\begin{aligned} \tilde{P}_i(t) &:= (\tilde{Y}_i(t),\dots,\tilde{Y}_{i-m+1}(t),\tilde{\sigma}_i(t)^2,\dots,\tilde{\sigma}_{i-l+1}(t)^2)^\mathsf{T}, \\ a_i(t) &:= (\alpha_0(t)\zeta_i^2,0,\dots,0,\alpha_0(t),0,\dots,0)^\mathsf{T}. \end{aligned}$$

For brevity, let $M_i(t) = M_i(\theta(t))$. Following Section 3.1 in [78], the model (3.114) admits the representation

$$\tilde{P}_i(t) = M_i(t)\tilde{P}_{i-1}(t) + a_i(t). \tag{3.115}$$

In Theorem 2.1 [48], it was shown that a necessary and sufficient condition for $\tilde{Y}_i(t) = H(t, \mathcal{F}_i)$ to exist and having $q$-th moments is

$$\lambda_{max}(\mathbb{E}[M_0(t)^{\otimes q}]) < 1.$$

It is easy to see in their proofs, that the condition $\sup_{t\in[0,1]} \lambda_{max}(\mathbb{E}[M_0(t)^{\otimes q}]) < 1$ then implies $\sup_{t\in[0,1]} \|\tilde{Y}_i(t)\|_q \le D$ with some $D > 0$.

(3.115) implies the explicit representation

$$\tilde{P}_i(t) = \sum_{k=0}^{\infty}\Big(\prod_{j=0}^{k-1} M_{i-j}(t)\Big)a_{i-k}(t). \tag{3.116}$$

We therefore have for $t, t' \in [0,1]$:

$$\begin{aligned}\|\tilde{P}_i(t) - \tilde{P}_i(t')\|_q \;\le\; & \sum_{k=0}^{\infty}\sum_{l=0}^{k-1}\Big(\prod_{0\le j<l} \|M_{i-j}(t)\|_q\Big)\|M_{i-l}(t) - M_{i-l}(t')\|_q \\ & \qquad \times\Big(\prod_{l<j\le k-1} \|M_{i-j}(t)\|_q\Big)\cdot \|a_{i-k}(t)\|_q \\ & + \sum_{k=0}^{\infty}\Big(\prod_{j=0}^{k-1} \|M_{i-j}(t)\|_q\Big)\|a_{i-k}(t) - a_{i-k}(t')\|_q. \end{aligned} \tag{3.117}$$

Note that $\sup_{t\in[0,1]} \lambda_{max}(\|M_0(t)\|_q) < 1$. By Lipschitz continuity of $\theta(\cdot)$, we have $\|a_0(t) - a_0(t')\|_q = |\alpha_0(t) - \alpha_0(t')|(\|\zeta_0^2\|_q, 0, \ldots, 0, 1, 0, \ldots, 0)^{\mathsf{T}} = O(|t-t'|)$ and $\|M_0(t) - M_0(t')\|_q = (\|\zeta_0^2\|_q|f(\theta(t)) - f(\theta(t'))|, 0, \ldots, 0, |f(\theta(t)) - f(\theta(t'))|, 0, \ldots, 0)^{\mathsf{T}} = O(|t-t'|)$. We conclude from the first component of (3.117), that for all $t, t' \in [0,1]$:

$$\|\tilde{Y}_i(t) - \tilde{Y}_i(t')\|_q \le C \cdot |t - t'|, \tag{3.118}$$

with some constant $C > 0$.

Put $P_i = (Y_i, \ldots, Y_{i-m+1}, \sigma_i^2, \ldots, \sigma_{i-l+1}^2)^\top$. Similarly to (3.115), we have

$$P_i = M_i(i/n)P_{i-1} + a_i(i/n), \quad i = 1, \ldots, n. \tag{3.119}$$

Note that $i$ iterations of (3.119) lead to $P_0 = \tilde{P}_0(0)$, thus existence of $Y_i$ follows from existence of $\tilde{Y}_i(0)$. We have

$$\begin{aligned} \|P_i - \tilde{P}_i(i/n)\|_q &\leq \|M_i(i/n)\|_q \|P_{i-1} - \tilde{P}_{i-1}(i/n)\|_q \\ &\leq \|M_0(i/n)\|_q \|P_{i-1} - \tilde{P}_{i-1}((i-1)/n)\|_q \\ &\qquad + \|M_0(i/n)\|_q \|\tilde{P}_0(i/n) - \tilde{P}_0((i-1)/n)\|_q. \end{aligned}$$

Iteration of this inequality leads to

$$\|P_i - \tilde{P}_i(i/n)\|_q \leq \sum_{k=1}^{i} \Big( \prod_{j=0}^{k} \|M_0((i-j)/n)\|_q \Big) \|\tilde{P}_0((i-k)/n) - \tilde{P}_0((i-k-1)/n)\|_q.$$

Due to $\sup_{t\in[0,1]} \lambda_{max}(\|M_0(t)\|_q) < 1$ and (3.118), we conclude from the first component that $\|Y_i - \tilde{Y}_i(i/n)\|_q = O(n^{-1})$. This shows Assumption 3.2.1(A5).

Note that $P_i = J_i(\mathcal{F}_i)$ and $\tilde{P}_i(t) = J(t, \mathcal{F}_i)$ with some measurable functions $J_i$, $J(t, \cdot)$. Following [75] (Example 9 and Proposition 1 therein), define $\hat{P}_i(t) := \tilde{P}_i(t) - \tilde{P}_i(t)^*$. Then

$$\hat{P}_i(t)^{\otimes q} = M_i(t)^{\otimes q} \hat{P}_{i-1}(t)^{\otimes q},$$

which implies $\sup_{t\in[0,1]} \delta_q^{\tilde{Y}(t)}(k) = \|\tilde{Y}_i(t) - \tilde{Y}_i(t)^*\|_q = O(c^k)$ with some $c \in (0,1)$ uniformly in $t$ due to $\sup_{t\in[0,1]} \lambda_{max}(\mathbb{E}M_i(t)^{\otimes q}) < 1$. The same argumentation can be done for $Y_i$ which yields $\sup_{n\in\mathbb{N}} \delta_q^Y(k) = O(c^k)$. This shows Assumption 3.2.1(A7).

Let $\Sigma(x,\theta) := (\sigma(x,\theta)^2, \ldots, \sigma(x_{(l-1)\to}, \theta)^2)^\top$ and $A(x,\theta) := (\alpha_0 + \sum_{j=1}^m \alpha_j x_j, \ldots, \alpha_0 +$

$\sum_{j=1}^{m} \alpha_j x_{j+l-1})^\mathsf{T}$, and

$$B(\theta) = \begin{pmatrix} \beta_1 & \dots & \dots & \dots & \beta_l \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

As said in Theorem 2.1 in [48], $\lambda_{max}(\mathbb{E}M_0(\theta)^{\otimes q})$ is a necessary and sufficient condition for the corresponding GARCH process with parameters $\theta$ to have 8-th moments. As a consequence, $\lambda_{max}(\mathbb{E}M_0(\theta)) < 1$ which by Proposition 1 in [28] implies $\lambda_{max}(B(\theta)) < 1$. It holds that

$$\Sigma(x,\theta) = A(x,\theta) + B(\theta) \cdot \Sigma(x_{1\to},\theta). \tag{3.120}$$

Put $R_{x,\theta}(\Sigma) := A(x,\theta) + B(\theta)\cdot\Sigma$. For fixed $x \in \mathbb{R}^{\mathbb{N}}$, $R_{x,\theta}$ is a contraction in the space of continuous functions $C(\Theta)$ on $\Theta$ due to $\lambda_{max}(B(\theta)) < 1$. By Banach's fix point theorem, we conclude that

$$\sigma(x,\theta)^2 = \sum_{k=0}^{\infty} \big(B(\theta)^k A(x_{k\to},\theta)\big)_1. \tag{3.121}$$

Since $A(0,\theta) = (\alpha_0,\dots,\alpha_0)^\mathsf{T}$, we have

$$\sigma(0,\theta)^2 = \beta_0 \sum_{j=1}^{l} \Big(\sum_{k=0}^{\infty} B(\theta)^k\Big)_{1j} = \beta_0 \sum_{j=1}^{l} \big((1-B(\theta))^{-1}\big)_{1j}.$$

Notice that $|(A(x,\theta)-A(x',\theta))_i| \le \sum_{j=1}^{m} \beta_j |x_{j+i-1} - x'_{j+i-1}|$. By (3.121), we conclude that

$$\begin{aligned} |\sigma(x,\theta)^2 - \sigma(x',\theta)^2| &\le \Big|\sum_{k=0}^{\infty} \big(B(\theta)^k\{A(x_{k\to},\theta) - A(x'_{k\to},\theta)\}\big)_1\Big| \\ &\le \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=0}^{\infty} (B(\theta)^k)_{1i}\beta_j |x_{k+j+i-1} - x'_{k+j+i-1}| \\ &\le |x-x'|_{\tilde{\chi},1}, \end{aligned} \tag{3.122}$$

with some sequence $\tilde{\chi} = (\tilde{\chi}_i)_{i\in\mathbb{N}}$ satisfying $\tilde{\chi}_i = O(c^i)$ with $0 < c < 1$. Due to the explicit representation (3.121) with geometrically decaying summands, it is easy to see that $\sigma(x,\theta)^2$ is four times continuously differentiable w.r.t. $\theta$. Note that

$$\begin{aligned}
\partial_{\alpha_\nu}(\sigma(x,\theta)^2) &= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{\infty}(kB(\theta)^{k-1}\partial_{\alpha_\nu}B(\theta))_{1i}\cdot \beta_j x_{k+j+i-1},\\
\partial_{\beta_\nu}(\sigma(x,\theta)^2) &= \sum_{i=1}^{l}\sum_{k=0}^{\infty}(B(\theta)^k)_{1i}x_{k+\nu+i-1},
\end{aligned}$$

which shows that $\nabla_{\theta_\nu}(\sigma(x,\theta)^2) \le |x|_{\tilde{\chi},1}$ with some $\tilde{\chi} = (\tilde{\chi}_i)_{i\in\mathbb{N}}$ with $\tilde{\chi}_i = O(c^i)$. Using similar arguments as in (3.122), one can show that for each component, it holds that $|\nabla_\theta^k(\sigma(x,\theta)^2) - \nabla_\theta^k(\sigma(x',\theta)^2)| \le |x-x'|_{\tilde{\chi},1}$ ($k = 1,2,3,4$) and $|\nabla_\theta^k(\sigma(x,\theta)^2)| \le |x|_{\tilde{\chi},1}$ with some geometrically decaying $\tilde{\chi}$. Due to a Taylor expansion, we have

$$|\sigma(x,\theta)^2 - \sigma(x,\theta')^2| \le |\theta-\theta'|_1 \cdot \sup_{\theta\in\Theta}|\nabla_\theta(\sigma(x,\theta)^2)|_\infty \le |\theta-\theta'|_1\cdot|x|_{\tilde{\chi},1}, \tag{3.123}$$

with some geometrically decaying $\tilde{\chi}$. Similar arguments hold for higher order derivatives $\nabla_\theta^k(\sigma(x,\theta)^2)$, $k = 1,2,3$. Since $\sigma(x,\theta)^2 \ge \beta_0 \ge \beta_{min}$ uniformly in $\theta$ and

$$\begin{aligned}
\ell(y,x,\theta) &= \frac{1}{2}\Big(\frac{y}{\sigma(x,\theta)^2} + \log(\sigma(x,\theta)^2)\Big),\\
\nabla_\theta\ell(y,x,\theta) &= \frac{\nabla_\theta(\sigma(x,\theta)^2)}{2\sigma(x,\theta)^2}\Big(1 - \frac{y}{\sigma(x,\theta)^2}\Big),\\
\nabla_\theta^2\ell(y,x,\theta) &= \Big[-\frac{\nabla_\theta(\sigma(x,\theta)^2)\nabla_\theta(\sigma(x,\theta)^2)^{\mathsf{T}}}{2\sigma(x,\theta)^4} + \frac{\nabla_\theta^2(\sigma(x,\theta)^2)}{2\sigma(x,\theta)^2}\Big]\Big(1-\frac{y}{\sigma(x,\theta)^2}\Big)\\
&\quad + \frac{\nabla_\theta(\sigma(x,\theta)^2)\nabla_\theta(\sigma(x,\theta)^2)^{\mathsf{T}}}{2\sigma(x,\theta)^4}\cdot\frac{y}{\sigma(x,\theta)^2},
\end{aligned}$$

it is easy to see from the results (3.122) and (3.123) and similar results for their derivatives w.r.t. $\theta$ that $\ell \in \mathcal{H}(1,2,\chi,\tilde{C})$, $\nabla_\theta\ell \in \mathcal{H}(1,3,\chi,\tilde{C})$, $\nabla_\theta^2\ell \in \mathcal{H}(1,4,\chi,\tilde{C})$ with some geometrically decaying $\chi$ and some $\tilde{C} > 0$, i.e. Assumption 3.2.1(A1) is fulfilled with $M_y = 1$ and $M_x = 4$.

It was shown in the proof of Theorem 2.1 in [28], that $\theta \mapsto L(t,\theta) = \mathbb{E}\ell(\tilde{Z}_0(t),\theta)$ is uniquely minimized in $\theta = \theta(t)$, which shows Assumption 3.2.1(A3). As in the proof of Example 3.5.1, we obtain that

$$V(t) = \mathbb{E}\Big[\frac{\nabla_\theta(\sigma(\tilde{X}_0(t),\theta(t))^2)\nabla_\theta(\sigma(\tilde{X}_0(t),\theta(t))^2)^\intercal}{2\sigma(\tilde{X}_0(t),\theta(t))^4}\Big] = I(t)\frac{2}{\mathbb{E}\zeta_0^4 - 1}.$$

Furthermore,

$$\nabla_\theta \ell(\tilde{Z}_i(t),\theta(t)) = \frac{\nabla_\theta(\sigma(\tilde{X}_i(t),\theta(t))^2)}{2\sigma(\tilde{X}_i(t),\theta(t))^2}\{1-\zeta_i^2\},$$

which shows that $\nabla_\theta \ell(\tilde{Z}_i(t),\theta(t))$ is a martingale difference sequence w.r.t. $\mathcal{F}_i$. Thus $\Lambda(t) = I(t)$. It was shown in the proof of Theorem 2.2 in [28] that $V(t)$ is positive definite for each $t \in [0,1]$. By continuity, we conclude that Assumption 3.2.1(A4) is fulfilled.

Regarding Assumption 3.2.2, notice that from the explicit representation (3.116) and the eigenvalue restriction $\sup_{t\in[0,1]} \lambda_{max}(\|M_0(t)\|_q) < 1$ it follows that the following sequence is geometrically decaying uniformly in $t$, thus

$$\begin{aligned}\partial_t \tilde{P}_i(t) &= \sum_{k=0}^{\infty}\sum_{l=0}^{k-1}\Big(\prod_{0\le j<l} M_{i-j}(t)\Big)\partial_t M_{i-l}(t)\Big(\prod_{l<j\le k-1} M_{i-j}(t)\Big)a_{i-k}(t)\\ &\quad + \sum_{k=0}^{\infty}\Big(\prod_{j=0}^{k-1} M_{i-j}(t)\Big)\partial_t a_{i-k}(t)\end{aligned}$$

exists a.s. and has $q$-th moments, so does its first component $\partial_t \tilde{Y}_t(t)$. Similar arguments that were used to prove (3.118) can be applied here and yield for $t,t' \in [0,1]$:

$$\|\partial_t \tilde{Y}_i(t) - \partial_t \tilde{Y}_i(t')\|_q \le C' \cdot |t-t'|,$$

with some constant $C' > 0$, i.e. Assumption 3.2.2(B3) is shown.

From (3.121) and $\sup_{\theta\in\Theta} \lambda_{max}(B(\theta)) < 1$, it follows that $x_i \mapsto \sigma(x,\theta)$ is differentiable for all $i \in \mathbb{N}$ and

$$\partial_{x_i}(\sigma(x,\theta)^2) = \sum_{k=0}^{\infty} \big(B(\theta)^k \partial_{x_i} A(x_{k\to},\theta)\big)_1.$$

Let $M_y' = 1$, $M_x' = 5$. It follows that $z_i \mapsto \nabla_\theta^2 \ell(z,\theta)$ is differentiable and $\partial_{z_i}\nabla_\theta^2 \in \mathcal{H}(1,5,\chi,\tilde{C})$ for all $i \in \mathbb{N}_0$ with some geometrically decaying $\chi$ and some $\tilde{C} > 0$. It is also easy to see that $\nabla_\theta^3 \ell \in \mathcal{H}(1,5,\chi,\tilde{C})$ with some geometrically decaying $\chi$. This shows Assumption 3.2.2(B2). □

*Proof of Example 3.5.3.* Define $\tilde{Y}_i(t) := \sum_{j=1}^m \mathbb{1}_{\{\zeta_{i,j} \le \pi(\tilde{X}_i(t)^\mathsf{T}\theta(t))\}}$. Put $M_y = 1$. The model follows (3.1) with $F_i(x,\theta) = \sum_{j=1}^m \mathbb{1}_{\{\xi_{i,j} \le \pi(x^\mathsf{T}\theta(i/n))\}}$. We have

$$\begin{aligned}
&\|F_i(x,\theta) - F_i(x',\theta)\|_1 \\
\le\ & \sum_{j=1}^m \|\mathbb{1}_{\{\zeta_{i,j} \le \pi(x^\mathsf{T}\theta(t))\}} - \mathbb{1}_{\{\zeta_{i,j} \le \pi(x^\mathsf{T}\theta(t'))\}}\|_1 \\
\le\ & m\{\mathbb{P}(\pi(x^\mathsf{T}\theta) \le \xi_{i,1} \le \pi((x')^\mathsf{T}\theta)) + \mathbb{P}(\pi((x')^\mathsf{T}\theta) \le \xi_{i,1} \le \pi(x^\mathsf{T}\theta))\} \\
\le\ & 2m|\pi(x^\mathsf{T}\theta) - \pi((x')^\mathsf{T}\theta)|.
\end{aligned}$$

Since $|\partial_w \pi(w)| \le \frac{1}{4}$, we conclude that

$$\|F_i(x,\theta) - F_i(x',\theta)\|_1 \le \frac{m}{2} \sup_j \sup_{\theta\in\Theta} |\theta_j| \cdot |x - x'|_1,$$

i.e. (3.13) and thus Assumption 3.2.1(A5), (A7) is fulfilled.

Note that for fixed $c \in (0,1)$, $f(w) = \log(1+e^w) - c\cdot w$ is strongly convex with minimum at $w_0$ defined by $c = \frac{e^{w_0}}{1+e^{w_0}}$. It holds that

$$L(t,\theta) := \mathbb{E}\ell(\tilde{Y}_0(t),\tilde{X}_0(t),\theta) = m\cdot\mathbb{E}\Big[\log\Big(1+\exp\big(\tilde{X}_0(t)^\mathsf{T}\theta\big)\Big) - \pi(\tilde{X}_0(t)^\mathsf{T}\theta(t))\cdot\tilde{X}_0(t)^\mathsf{T}\theta\Big].$$

By a Taylor expansion of $f$ around $w_0$, we obtain $f(w) = f(w_0) + \frac{1}{2}(w-w_0)^2\partial_w f(\tilde{w})$. Since $f''(w) = \frac{e^w}{(1+e^w)^2}$ is increasing for $w < 0$ and decreasing for $w > 0$, $f''(\tilde{w}) \ge \min\{\pi(w),\pi(w_0)\}$. In the following we use the notation $|x|_A^2 := x^\mathsf{T} A x$ for a weighted vector norm. We obtain that

$$L(t,\theta) - L(t,\theta(t)) \geq |\theta - \theta(t)|^2_{\tilde{V}(t,\theta)},$$

with $\tilde{V}(t,\theta) = \mathbb{E}\big[\min\{\pi(\tilde{X}_0(t)^\mathsf{T}\theta), \pi(\tilde{X}_0(t)^\mathsf{T}\theta(t))\}\tilde{X}_0\tilde{X}_0(t)^\mathsf{T}\big]$. If $\tilde{V}(t,\theta)$ was not positive definite for one $\theta$, there would exist $v \in \mathbb{R}^p$ such that $v'\tilde{V}(t,\theta)v = 0$ which would imply that either $v'\tilde{X}_0(t) = 0$ a.s. or $\min\{\pi(\tilde{X}_0(t),\theta(t)), \pi(\tilde{X}_0(t),\theta)\} = 0$ a.s.. But it holds $\pi(\tilde{X}_0(t),\theta) \in (0,1)$ a.s. since $\sup_{j=1,\ldots,p} |\tilde{X}_{0j}(t)| < \infty$ a.s. and $\Theta$ is compact. Furthermore, $v'\tilde{X}_0(t) = 0$ a.s. is a contradication to the positive definiteness of $\mathbb{E}[\tilde{X}_0(t)\tilde{X}_0(t)^\mathsf{T}]$. Thus $\tilde{V}(t,\theta)$ is positive definite for each $\theta$ and we conclude that $L(t,\theta)$ is uniquely minimized by $\theta = \theta(t)$. This shows Assumption 3.2.1(A3). We furthermore have

$$\begin{aligned}
\nabla_\theta \ell(z,\theta) &= m\pi(x^\mathsf{T}\theta)x - yx, \\
\nabla^2_\theta \ell(z,\theta) &= m\frac{\exp(x^\mathsf{T}\theta)}{(1+\exp(x^\mathsf{T}\theta))^2} \cdot xx^\mathsf{T}.
\end{aligned}$$

It is easy to see that $\ell \in \mathcal{H}(1,1,\chi,\tilde{C})$, $\nabla_\theta \ell \in \mathcal{H}(1,2,\chi,\tilde{C})$ and $\nabla_\theta \ell \in \mathcal{H}(1,2,\chi,\tilde{C})$ with some $\tilde{C} > 0$ and $\chi = (1,\ldots,1,0,0,\ldots)$, a vector with $p$ ones followed from zeros, i.e. Assumption 3.2.1(A1).

Since $\tilde{Y}_i(t)$ given $\tilde{X}_i(t)$ is binomial distributed with parameters $(m, \pi(\tilde{X}_0(t)^\mathsf{T}\theta(t)))$, we have

$$\mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t),\theta(t))|\tilde{X}_0(t)] = m\pi(\tilde{X}_0(t)^\mathsf{T}\theta)\tilde{X}_0(t) - m\pi(\tilde{X}_0(t)^\mathsf{T}\theta)\tilde{X}_0(t) = 0.$$

Furthermore, $(\nabla_\theta \ell(\tilde{Z}_i(t),\theta(t)))_i$ is an uncorrelated sequence, thus we have $\Lambda(t) = I(t)$. Here,

$$V(t) = \mathbb{E}\nabla^2_\theta \ell(\tilde{Z}_0(t),\theta(t)) = m\mathbb{E}\Big[\frac{\pi(\tilde{X}_0(t)^\mathsf{T}\theta(t))}{1+\exp(\tilde{X}_0(t)^\mathsf{T}\theta(t))} \cdot \tilde{X}_0(t)^\mathsf{T}\tilde{X}_0(t)\Big],$$

which is positive definite by a similar argumentation as it was done for $\tilde{V}(t,\theta)$. Finally, since

$\tilde{Y}_0(t)$ is binomial distributed with parameters $(m, \pi(\tilde{X}_0(t)^\mathsf{T}\theta(t)))$ given $\tilde{X}_0(t)$, we obtain

$$\begin{aligned} I(t) &= \mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) \nabla_\theta \ell(\tilde{Z}_0(t), \theta(t))^\mathsf{T}] \\ &= \mathbb{E}\big[\big(m\pi(\tilde{X}_0(t)^\mathsf{T}\theta(t)) - \tilde{Y}_0(t)\big)^2 \tilde{X}_0(t)\tilde{X}_0(t)^\mathsf{T}\big] \\ &= m\mathbb{E}[\pi(\tilde{X}_0(t)^\mathsf{T}\theta(t))(1-\pi(\tilde{X}_0(t)^\mathsf{T}\theta(t)))\tilde{X}_0(t)\tilde{X}_0(t)^\mathsf{T}] = V(t), \end{aligned}$$

and thus its positive definiteness and Assumption 3.2.1(A4).

□

**Comment 3.7.11.**

**Lemma 3.7.12** (Fisher information ARCH)**.** *Consider the ARCH model* $X_t^2 = (\alpha_0 + \alpha_1 X_{t-1}^2)\zeta_t^2$ *with* $\mathbb{E}\zeta_0 = 0$, $\mathbb{E}\zeta_0^2 = 1$ *and* $\mathbb{E}\zeta_4 = \mu_4$*. Then*

$$V(\theta) = \mathbb{E}\Big[\frac{1}{2(\alpha_0 + \alpha_1 X_0^2)^2} \begin{pmatrix} 1 & X_0^2 \\ X_0^2 & X_0^4 \end{pmatrix}\Big].$$

*It holds that*

$$\mathbb{E}\begin{pmatrix} 1 & X_0^2 \\ X_0^2 & X_0^4 \end{pmatrix} = \frac{1}{1-\alpha_1}\begin{pmatrix} 1-\alpha_1 & \alpha_0 \\ \alpha_0 & \frac{\mu_4\alpha_0^2}{1-\mu_4\alpha_1^2}(1+\alpha_1) \end{pmatrix}.$$

*Proof of Lemma 3.7.12.* By recursion,

$$\mathbb{E}X_0^2 = (\alpha_0 + \alpha_1 \mathbb{E}X_0^2)\mathbb{E}\zeta_0^2 = \alpha_0 + \alpha_1\mathbb{E}X_0^2,$$

which implies $\mathbb{E}X_0^2 = \frac{\alpha_0}{1-\alpha_1}$. By recursion,

$$\begin{aligned} \mathbb{E}X_0^4 &= \mathbb{E}(\alpha_0 + \alpha_1 X_0^2)^2\mathbb{E}\zeta_0^4 = (\alpha_0^2 + 2\alpha_0\alpha_1\mathbb{E}X_0^2 + \alpha_1^2\mathbb{E}X_0^4)\mu_4 \\ &= \mu_4\alpha_0^2(1 + 2\frac{\alpha_1}{1-\alpha_1}) + \mu_4\alpha_1^2\mathbb{E}X_0^4 \end{aligned}$$

which implies $\mathbb{E}X_0^4 = \frac{\mu_4\alpha_0^2}{1-\mu_4\alpha_1^2}\frac{1+\alpha_1}{1-\alpha_1}$. □

**Lemma 3.7.13** (Fisher information GARCH)**.** *Consider the GARCH model* $X_t^2 = \sigma_t^2 \zeta_0^2$, $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2$. *Then*

$$V(\theta) = \mathbb{E}\Big[\frac{1}{2\sigma_0^4}\nabla_\theta(\sigma_0^2)\nabla_\theta(\sigma_0^2)\Big] = \mathbb{E}\Big[\frac{1}{2\sigma_0^4}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\Big]$$

# CHAPTER 4

# SUMMARY AND FUTURE WORKS

This thesis is a discussion of different tools needed to perform simultaneous inference on models where coefficients change over time in smooth or piecewise continuous fashion. It starts with the discussion of achieving the optimal rate for a Gaussian approximation, often used as an invariance principle to construct confidence interval. This theme was explored in conjunction with the Gaussian extreme value theory and Bahadur expansion to explore the nontrivial simultaneous confidence intervals for conditional heteroscedastic models such as ARMA-GARCH etc.

For the research direction of chapter 1, it is difficult to achieve better result than $n^{1/p} \log n$. However, a few possible directions remain to be explored. One such is to see if such a principle can be obtained for strongly dependent series. One can start with the simple linear long-range dependent series and the insights can lead to more general results. Another interesting direction is to see what happens if one let the dimension grow. Obviously one cannot expect $d \gg \sqrt{n}$ but for smaller $d$ it can be an interesting direction to see if one can achieve such sharp result where $d$ can also grow.

The simultaneous confidence band construction can also give rise to a few possible future directions. One of them can be to include piecewise continuous models such as Thresholded Autoregression. Another interesting exploration that can be done is to see if one can lessen the moment conditions using our sharper Gaussian approximation from Chapter 1. Also, since the parameter function is changing over time, it is natural to ask if this facilitates some Bayesian treatment and if we can replace simultaneous confidence interval by Bayesian credible interval with some proper choice of prior. See Sarris [67] or Stock and Watson [69].

# REFERENCES


[1] Elena Andreou and Eric Ghysels. Monitoring disruptions in financial markets. *J. Econometrics*, 135(1-2):77–124, 2006. ISSN 0304-4076. doi: 10.1016/j.jeconom.2005.07.023. URL http://dx.doi.org/10.1016/j.jeconom.2005.07.023.

[2] Donald W. K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993. ISSN 0012-9682. doi: 10.2307/2951764. URL http://dx.doi.org/10.2307/2951764.

[3] Alexander Aue, Siegfried Hörmann, Lajos Horváth, and Matthew Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37(6B):4046–4087, 2009.

[4] Jushan Bai. Estimation of a change point in multiple regression models. *The Review of Economics and Statistics*, 79(4):551–563, 1997.

[5] István Berkes, Weidong Liu, and Wei Biao Wu. Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.*, 42(2):794–817, 2014. ISSN 0091-1798. doi: 10.1214/13-AOP850. URL http://dx.doi.org/10.1214/13-AOP850.

[6] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. doi: 10.1002/9780470316962. URL http://dx.doi.org/10.1002/9780470316962. A Wiley-Interscience Publication.

[7] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31(3):307–327, 1986. ISSN 0304-4076. doi: 10.1016/0304-4076(86)90063-1. URL http://dx.doi.org/10.1016/0304-4076(86)90063-1.

[8] R. L. Brown, James Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B*, 37:149–192, 1975. ISSN 0035-9246. URL http://links.jstor.org/sici?sici=0035-9246(1975)37:2<149:TFTTCO>2.0.CO;2-7origin=MSN. With discussion by D. R. Cox, P. R. Fisk, Maurice Kendall, M. B. Priestley, Peter C. Young, G. Phillips, T. W. Anderson, A. F. M. Smith, M. R. B. Clarke, A. C. Harvey, Agnes M. Herzberg, M. C. Hutchison, Mohsin S. Khan, J. A. Nelder, Richard E. Quant, T. Subba Rao, H. Tong and W. G. Gilchrist and with reply by J. Durbin and J. M. Evans.

[9] Zongwu Cai. Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics*, 136(1):163–188, 2007. ISSN 0304-4076. doi: 10.1016/j.jeconom.2005.08.004. URL http://dx.doi.org/10.1016/j.jeconom.2005.08.004.

[10] Jie Chen and Arjun K Gupta. Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical association*, 92(438):739–747, 1997.

[11] Gregory C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28:591–605, 1960. ISSN 0012-9682. doi: 10.2307/1910133. URL http://dx.doi.org/10.2307/1910133.

[12] R. Dahlhaus. Locally Stationary Processes. *Handbook of Statistics*, 09 2011.

[13] R. Dahlhaus, S. Richter, and W. B. Wu. Towards a general theory for non-linear locally stationary processes. *ArXiv e-prints: 1704.02860*, 04 2017.

[14] Rainer Dahlhaus and Wolfgang Polonik. Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39, 2009. ISSN 1350-7265. doi: 10.3150/08-BEJ137. URL http://dx.doi.org/10.3150/08-BEJ137.

[15] Rainer Dahlhaus and Suhasini Subba Rao. Statistical inference for time-varying ARCH processes. *Ann. Statist.*, 34(3):1075–1114, 2006. ISSN 0090-5364. doi: 10.1214/009053606000000227. URL http://dx.doi.org/10.1214/009053606000000227.

[16] Sybolt O. de Vries, Vaclav Fidler, Wietze D. Kuipers, and Maria G.M. Hunink. Fitting multistate transition models with autoregressive logistic regression: Supervised exercise in intermittent claudication. *Medical Decision Making*, 18(1):52–60, 1998. doi: 10.1177/0272989X9801800112. URL https://doi.org/10.1177/0272989X9801800112. PMID: 9456209.

[17] Monroe D. Donsker. Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23:277–281, 1952. ISSN 0003-4851.

[18] J. L. Doob. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, 20:393–403, 1949. ISSN 0003-4851.

[19] Ernst Eberlein. On strong invariance principles under dependence assumptions. *Ann. Probab.*, 14(1):260–270, 1986. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(198601)14:1<260:OSIPUD>2.0.CO;2-Oorigin=MSN.

[20] Uwe Einmahl. A useful estimate in the multidimensional invariance principle. *Probab. Theory Related Fields*, 76(1):81–101, 1987. ISSN 0178-8051. doi: 10.1007/BF00390277. URL http://dx.doi.org/10.1007/BF00390277.

[21] Uwe Einmahl. Strong invariance principles for partial sums of independent random vectors. *Ann. Probab.*, 15(4):1419–1440, 1987. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(198710)15:4<1419:SIPFPS>2.0.CO;2-Yorigin=MSN.

[22] Uwe Einmahl. Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.*, 28(1):20–68, 1989. ISSN 0047-259X. doi: 10.1016/0047-259X(89)90097-3. URL http://dx.doi.org/10.1016/0047-259X(89)90097-3.

[23] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007, 1982. ISSN 0012-9682. doi: 10.2307/1912773. URL `http://dx.doi.org/10.2307/1912773`.

[24] Robert F Engle and J Gonzalo Rangel. The spline garch model for unconditional volatility and its global macroeconomic causes. 2005.

[25] P. Erdös and M. Kac. On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.*, 52:292–302, 1946. ISSN 0002-9904.

[26] Jianqing Fan and Wenyang Zhang. Statistical estimation in varying coefficient models. *Ann. Statist.*, 27(5):1491–1518, 1999. ISSN 0090-5364. doi: 10.1214/aos/1017939139. URL `http://dx.doi.org/10.1214/aos/1017939139`.

[27] Jianqing Fan and Wenyang Zhang. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27(4):715–731, 2000.

[28] Christian Francq and Jean-Michel Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4): 605–637, 2004. ISSN 1350-7265. doi: 10.3150/bj/1093265632. URL `http://dx.doi.org/10.3150/bj/1093265632`.

[29] Piotr Fryzlewicz, Theofanis Sapatinas, and Suhasini Subba Rao. Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.*, 36(2):742–786, 2008. ISSN 0090-5364. doi: 10.1214/07-AOS510. URL `http://dx.doi.org/10.1214/07-AOS510`.

[30] Rosella Giacometti, Marida Bertocchi, Svetlozar T Rachev, and Frank J Fabozzi. A comparison of the lee–carter model and ar–arch model for forecasting mortality rates. *Insurance: Mathematics and Economics*, 50(1):85–93, 2012.

[31] F. Götze and A. Yu. Zaitsev. Bounds for the rate of strong approximation in the multidimensional invariance principle. *Teor. Veroyatn. Primen.*, 53(1): 100–123, 2008. ISSN 0040-361X. doi: 10.1137/S0040585X9798350X. URL `http://dx.doi.org/10.1137/S0040585X9798350X`.

[32] Yanira Guanche, Roberto Mínguez, and Fernando J Méndez. Autoregressive logistic regression applied to atmospheric circulation patterns. *Climate dynamics*, 42(1-2):537–552, 2014.

[33] Donald R. Hoover, John A. Rice, Colin O. Wu, and Li-Ping Yang. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822, 1998. ISSN 0006-3444. doi: 10.1093/biomet/85.4.809. URL `http://dx.doi.org/10.1093/biomet/85.4.809`.

[34] Jianhua Z. Huang, Colin O. Wu, and Lan Zhou. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica*, 14(3): 763–788, 2004. ISSN 1017-0405.

[35] Chia-Shang James Chu. Detecting parameter shift in garch models. *Econometric Reviews*, 14(2):241–266, 1995.

[36] Soohwa Kim, Sinsup Cho, and Sangyeol Lee. On the cusum test for parameter changes in garch (1, 1) models. *Communications in Statistics-Theory and Methods*, 29(2):445–462, 2000.

[37] Piotr Kokoszka, Remigijus Leipus, et al. Change-point estimation in arch models. *Bernoulli*, 6(3):513–539, 2000.

[38] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32:111–131, 1975.

[39] J. Komlós, P. Major, and G. Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34(1): 33–58, 1976.

[40] Rasoul Kowsar, Behrooz Keshtegar, Marey Mohamed. A., and Miyamoto A. An autoregressive logistic model to predict the reciprocal effects of oviductal fluid components on in vitro spermophagy by neutrophils in cattle. *Scientific Reports*, 7(4482), 2017. doi: 10.1038/s41598-017-04841-z. URL `https://www.nature.com/articles/s41598-017-04841-z`.

[41] Reg Kulperger, Hao Yu, et al. High moment partial sum processes of residuals in garch models and their applications. *The Annals of Statistics*, 33(5):2395–2422, 2005.

[42] S. J. Leybourne and B. P. M. McCabe. On the distribution of some test statistics for coefficient constancy. *Biometrika*, 76(1):169–177, 1989. ISSN 0006-3444. doi: 10.1093/biomet/76.1.169. URL `http://dx.doi.org/10.1093/biomet/76.1.169`.

[43] Chien-Fu Jeff Lin and Timo Teräsvirta. Testing parameter constancy in linear models against stochastic stationary parameters. *J. Econometrics*, 90(2): 193–213, 1999. ISSN 0304-4076. doi: 10.1016/S0304-4076(98)00041-4. URL `http://dx.doi.org/10.1016/S0304-4076(98)00041-4`.

[44] D. Y. Lin and Z. Ying. Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.*, 96(453):103–126, 2001. ISSN 0162-1459. doi: 10.1198/016214501750333018. URL `http://dx.doi.org/10.1198/016214501750333018`. With comments and a rejoinder by the authors.

[45] Shinn-Juh Lin, Jian Yang, et al. *Testing shifts in financial models with conditional heteroskedasticity: an empirical distribution function approach.*

[46] Georg Lindgren. Extreme values and crossings for the x 2-process and other functions of multidimensional gaussian processes, by reliability applications. *Advances in Applied Probability*, 12(3):746–774, 1980.

[47] Shiqing Ling. On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.*, 36(3):688–705, 1999. ISSN 0021-9002.

[48] Shiqing Ling and Michael McAleer. Necessary and sufficient moment conditions for the GARCH$(r, s)$ and asymmetric power GARCH$(r, s)$ models. *Econometric Theory*, 18(3):722–729, 2002. ISSN 0266-4666. doi: 10.1017/S0266466602183071. URL http://dx.doi.org/10.1017/S0266466602183071.

[49] Weidong Liu and Zhengyan Lin. Strong approximation for a class of stationary processes. *Stochastic Process. Appl.*, 119(1):249–280, 2009. ISSN 0304-4149. doi: 10.1016/j.spa.2008.01.012. URL http://dx.doi.org/10.1016/j.spa.2008.01.012.

[50] Weidong Liu and Wei Biao Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26(4):1218–1245, 2010. ISSN 0266-4666. doi: 10.1017/S026646660999051X. URL http://dx.doi.org/10.1017/S026646660999051X.

[51] Weidong Liu, Han Xiao, and Wei Biao Wu. Probability and moment inequalities under dependence. *Statist. Sinica*, 23(3):1257–1272, 2013. ISSN 1017-0405.

[52] Helmut Lütkepohl. *New introduction to multiple time series analysis*. Springer-Verlag, Berlin, 2005. ISBN 3-540-40172-5. doi: 10.1007/978-3-540-27752-1. URL http://dx.doi.org/10.1007/978-3-540-27752-1.

[53] Thomas Mikosch and Cătălin Stărică. Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *The Review of Economics and Statistics*, 86(1):378–390, 2004.

[54] Ditlev Monrad and Walter Philipp. The problem of embedding vector-valued martingales in a gaussian process. *Theory of Probability & Its Applications*, 35(2):374–377, 1991.

[55] Seiji Nabeya and Katsuto Tanaka. Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.*, 16(1):218–235, 1988. ISSN 0090-5364. doi: 10.1214/aos/1176350701. URL http://dx.doi.org/10.1214/aos/1176350701.

[56] S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.*, 7(5):745–789, 1979. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(197910)7:5<745:LDOSOI>2.0.CO;2-Sorigin=MSN.

[57] Jukka Nyblom. Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.*, 84(405):223–230, 1989. ISSN 0162-1459. URL http://links.jstor.org/sici?sici=0162-1459(198903)84:405<223:TFTCOP>2.0.CO;2-Worigin=MSN.

[58] F. C. Palm. GARCH models of volatility. In *Statistical methods in finance*, volume 14 of *Handbook of Statist.*, pages 209–240. North-Holland, Amsterdam, 1996. doi: 10.1016/S0169-7161(96)14009-8. URL `http://dx.doi.org/10.1016/S0169-7161(96)14009-8`.

[59] Werner Ploberger, Walter Krämer, and Karl Kontrus. A new test for structural stability in the linear regression model. *J. Econometrics*, 40(2):307–318, 1989. ISSN 0304-4076. doi: 10.1016/0304-4076(89)90087-0. URL `http://dx.doi.org/10.1016/0304-4076(89)90087-0`.

[60] M. B. Priestley. *Nonlinear and nonstationary time series analysis.* Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1988. ISBN 0-12-564910-X.

[61] Yu. V. Prohorov. Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen.*, 1:177–238, 1956. ISSN 0040-361x.

[62] J. O. Ramsay and B. W. Silverman. *Functional data analysis.* Springer Series in Statistics. Springer, New York, second edition, 2005. ISBN 978-0387-40080-8; 0-387-40080-X.

[63] S. Richter and R. Dahlhaus. Cross validation for locally stationary processes. *ArXiv e-prints: 1705.10046*, 05 2017.

[64] Neelabh Rohan. A time varying garch (p, q) model and related statistical inference. *Statistics & Probability Letters*, 83(9):1983–1990, 2013.

[65] Neelabh Rohan and T. V. Ramanathan. Nonparametric estimation of a time-varying GARCH model. *J. Nonparametr. Stat.*, 25(1):33–52, 2013. ISSN 1048-5252. doi: 10.1080/10485252.2012.728600. URL `http://dx.doi.org/10.1080/10485252.2012.728600`.

[66] A. I. Sakhanenko. Estimates in the invariance principle in terms of truncated power moments. *Sibirsk. Mat. Zh.*, 47(6):1355–1371, 2006. ISSN 0037-4474. doi: 10.1007/s11202-006-0119-1. URL `http://dx.doi.org/10.1007/s11202-006-0119-1`.

[67] Alexander H Sarris. A bayesian approach to estimation of time-varying regression coefficients. In *Annals of Economic and Social Measurement, Volume 2, number 4*, pages 501–523. NBER, 1973.

[68] Cătălin Stărică and Clive Granger. Nonstationarities in stock returns. *The Review of Economics and Statistics*, 87(3):503–522, 2005.

[69] James H Stock and Mark W Watson. Median unbiased estimation of coefficient variance in a time-varying parameter model. *Journal of the American Statistical Association*, 93 (441):349–358, 1998.

[70] A Tagliafichi Ricardo. The garch model and their application to the var.

[71] James W Taylor and Keming Yu. Using auto-regressive logit models to forecast the exceedance probability for financial risk management. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(4):1069–1092, 2016.

[72] L. Trapani, G. Urga, and C Kao. Testing for instability in covariance structures. *Bernoulli*, page To Appear, 2017.

[73] Ruey S. Tsay. *Analysis of financial time series*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, third edition, 2010. ISBN 978-0-470-41435-4. doi: 10.1002/9780470644560. URL `http://dx.doi.org/10.1002/9780470644560`.

[74] N. Wiener. *Nonlinear Problems in Random Theory*. Wiley, New York, 1958.

[75] Wei Biao Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005. ISSN 1091-6490. doi: 10.1073/pnas.0506715102. URL `http://dx.doi.org/10.1073/pnas.0506715102`.

[76] Wei Biao Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154 (electronic), 2005. ISSN 1091-6490. doi: 10.1073/pnas.0506715102. URL `http://dx.doi.org/10.1073/pnas.0506715102`.

[77] Wei Biao Wu. Strong invariance principles for dependent random variables. *Ann. Probab.*, 35(6):2294–2320, 2007. ISSN 0091-1798. doi: 10.1214/009117907000000060. URL `http://dx.doi.org/10.1214/009117907000000060`.

[78] Wei Biao Wu and Wanli Min. On linear processes with dependent innovations. *Stochastic Processes and their Applications*, 115(6):939 – 958, 2005. ISSN 0304-4149. doi: http://dx.doi.org/10.1016/j.spa.2005.01.001. URL `http://www.sciencedirect.com/science/article/pii/S0304414905000025`.

[79] Wei-Biao Wu and Ying Nian Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.*, 10(1):352–379, 2016. ISSN 1935-7524. doi: 10.1214/16-EJS1108. URL `http://dx.doi.org/10.1214/16-EJS1108`.

[80] Wei Biao Wu and Zhibiao Zhao. Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(3):391–410, 2007. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2007.00594.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2007.00594.x`.

[81] Wei Biao Wu and Zhou Zhou. Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*, 21(3):1397–1413, 2011. ISSN 1017-0405. doi: 10.5705/ss.2008.223. URL `http://dx.doi.org/10.5705/ss.2008.223`.

[82] A. Yu. Zaitsev. Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I. *Teor. Veroyatnost. i Primenen.*, 45(4):718–738, 2000. ISSN 0040-361X. doi: 10.1137/S0040585X97978555. URL `http://dx.doi.org/10.1137/S0040585X97978555`.

[83] A. Yu. Zaitsev. Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. III. *Teor. Veroyatnost. i Primenen.*, 46(4):744–769, 2001. ISSN 0040-361X. doi: 10.1137/S0040585X97979305. URL `http://dx.doi.org/10.1137/S0040585X97979305`.

[84] A. Yu. Zaitsev. Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. II. *Teor. Veroyatnost. i Primenen.*, 46(3):535–561, 2001. ISSN 0040-361X. doi: 10.1137/S0040585X97979123. URL `http://dx.doi.org/10.1137/S0040585X97979123`.

[85] D. Zhang and W. B. Wu. Gaussian Approximation for High Dimensional Time Series. *Ann. Statist.*, 45(5):1895–1919, 2017.

[86] Ting Zhang and Wei Biao Wu. Inference of time-varying regression models. *Ann. Statist.*, 40(3):1376–1402, 2012. ISSN 0090-5364. doi: 10.1214/12-AOS1010. URL `http://dx.doi.org/10.1214/12-AOS1010`.

[87] Ting Zhang and Wei Biao Wu. Time-varying nonlinear regression models: nonparametric estimation and model selection. *Ann. Statist.*, 43(2):741–768, 2015. ISSN 0090-5364. doi: 10.1214/14-AOS1299. URL `http://dx.doi.org/10.1214/14-AOS1299`.

[88] Wenyang Zhang, Sik-Yum Lee, and Xinyuan Song. Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.*, 82(1):166–188, 2002. ISSN 0047-259X. doi: 10.1006/jmva.2001.2012. URL `http://dx.doi.org/10.1006/jmva.2001.2012`.

[89] Zhou Zhou and Wei Biao Wu. Local linear quantile estimation for nonstationary time series. *Ann. Statist.*, 37(5B):2696–2729, 2009. ISSN 0090-5364. doi: 10.1214/08-AOS636. URL `http://dx.doi.org/10.1214/08-AOS636`.

[90] Zhou Zhou and Wei Biao Wu. Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(4):513–531, 2010. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2010.00743.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2010.00743.x`.

[91] Zhou Zhou and Wei Biao Wu. Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(4):513–531, 2010. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2010.00743.x. URL `http://dx.doi.org/10.1111/j.1467-9868.2010.00743.x`.