Submitted to the Annals of Statistics
2017

BAHADUR REPRESENTATION AND SIMULTANEOUS INFERENCE FOR CURVE ESTIMATION IN TIME-VARYING MODELS

By Sayar Karmakar, Stefan Richter and Wei Biao Wu

University of Chicago and Heidelberg University

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.

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 [3] 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 [17], Fan and Zhang [18], Hoover et al. [23], Huang, Wu and Zhou [24], Lin and Ying [33], Ramsay and Silverman [41], Zhang, Lee and Song [51] among

Keywords and phrases: Time-varying models, Regression models, Auto-regressive processes, Bahadur representation, Simultaneous confidence band, Gaussian approximation

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, θ_0 and $\theta_i = \theta(i/n)$ are the regression coefficients. Here, θ_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. [23], Cai [7]) and Zhou and Wu [56] among others. Hypothesis testing is widely used to choose between model I and model II, see, for instance, Zhang and Wu [52], Zhang and Wu [53], Chow [9], Brown, Durbin and Evans [6], Nabeya and Tanaka [37], Leybourne and McCabe [30], Nyblom [38], Ploberger, Krämer and Kontrus [40], Andrews [2] and Lin and Teräsvirta [31]. Zhou and Wu [56] 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 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 [15] introduced the classical ARCH model and Bollerslev [5] 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 [45], Engle and Rangel [16] and Fryzlewicz, Sapatinas and Subba Rao [20]. Towards time-varying parameter models in the CH setting, numerous works discussed the CUSUM-type procedure, for instance, Kim, Cho and Lee [26] for testing change in parameters of GARCH(1,1). Kulperger et al. [29] 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 [25], Chen and Gupta [8], Lin et al. [32], Kokoszka et al. [27] 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 [36], 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. [21], 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 [13]. Since then, while several pointwise approaches were discussed in the tvARMA and tvARCH case (cf. Dahlhaus and Polonik [11], Dahlhaus and Subba Rao [13], Fryzlewicz, Sapatinas and Subba Rao [20]), pointwise theoretical results for estimation in tvGARCH processes were discussed in Rohan and Ramanathan [44] and Rohan [43] 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 [55] 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. (2.1)) 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, Richter and Wu [12].

The rest of the article is organized as follows. In Section 2, we introduce our framework, the functional dependence measure, the assumptions and the M-estimators of the parameter curves. Section 3 consists of the results about the Bahadur representation and the SCBs of the related contrasts. Section 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 5. Some summarized simulation studies and real data applications can be found in Section 6. We defer all the proofs to Supplement A.

2. Model assumptions and estimators.

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

$$Y_i = F_i(X_i, \theta(i/n)), \quad i = 1, \ldots, n, \tag{2.1}$$

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 θ :

$[0,1] \to \Theta \subset \mathbb{R}^{d_\Theta}$ 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 ζ_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

$$X_i = G_i(\mathcal{F}_i), \qquad Y_i = H_i(\mathcal{F}_i), \quad i = 1,\ldots,n, \tag{2.2}$$

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 (2.1) as it is only needed in an optional condition (2.13). Some more general formulations may still fit in the setting of this paper. There are some important special cases of (2.1):

(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 (2.1) 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, (2.1) 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. [14], Kowsar et al. [28] etc: for climatology, see Guanche, Mínguez and Méndez [22]; for risk management analysis see Taylor and Yu [47] etc.

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

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

$$L_{n,b_n}^c(t,\theta,\theta') := (nb_n)^{-1} \sum_{i=1}^n K_{b_n}(t - i/n)\ell(Z_i^c, \theta + \theta'(i/n - t)), \tag{2.3}$$

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

$$(\hat{\theta}_{b_n}(t), \widehat{\theta'}_{b_n}(t)) = \arg \min_{(\theta,\theta') \in \Theta \times \Theta'} L^c_{n,b_n}(t, \theta, \theta'), \qquad t \in [0, 1]. \tag{2.4}$$

In Examples 5.1 and 5.2, we discuss applications and choices of ℓ for general recursively defined locally stationary time series models and tvGARCH processes. In Example 5.3, we consider a time-varying logistic regression model with a Binomial likelihood function ℓ.

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 [48]. Assume that a stationary process Z_i has mean 0, $Z_i \in \mathcal{L}_q, q > 0$ and it admits the causal representation

$$Z_i = J(\zeta_i, \zeta_{i-1}, \ldots). \tag{2.5}$$

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

$$\delta_q^Z(i) = \|Z_i - Z_i^*\|_q, \tag{2.6}$$

where $\mathcal{F}_i^*$ is a coupled version of $\mathcal{F}_i$ with ζ_0 in $\mathcal{F}_i$ replaced by ζ_0^*,

$$\mathcal{F}^* = (\zeta_i, \zeta_{i-1}, \cdots, \zeta_1, \zeta_0^*, \zeta_{-1}, \zeta_{-2}, \cdots), \tag{2.7}$$

and $Z_i^* = J(\mathcal{F}_i^*)$. Note that $\delta_q^Z(i)$ measures the dependence of Z_i on ζ_0 in terms of the qth moment. The tail cumulative dependence measure $\Delta_q^Z(j)$ for $j \geq 0$ is defined as

$$\Delta_q^Z(j) = \sum_{i=j}^{\infty} \delta_q^Z(i). \tag{2.8}$$

2.4. *The class $\mathcal{H}(M_y, M_x, \chi, \bar{C})$.* To prove uniform convergence of L^c_{n,b_n} and its derivatives w.r.t. θ, we require ℓ to be Lipschitz continuous in direction of θ 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 ℓ and its derivatives to be in the class $\mathcal{H}(M_y, M_x, \chi, \bar{C})$ which is defined as follows: Let $\chi = (\chi_i)_{i=1,2,\ldots}$ 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'

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

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 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 $R_{M_y,M_x}(z) = 1 + |z|_{\hat{\chi},1}^M$.

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 ℓ are needed to prove results for the latter one which is postponed to Assumption 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 2.1. *Suppose that for some $r \ge 2$ and some $\gamma > 1$,*

- *(A1) (Smoothness in θ-direction) ℓ is twice continuously differentiable w.r.t. θ. It holds that $\ell, \nabla_\theta \ell, \nabla_\theta^2 \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$ for some $M_y, M_x \ge 1$, $\bar{C} > 0$ and $\chi = (\chi_i)_{i=1,2,\dots}$ with $\chi_i = O(i^{-(1+\gamma)})$.*
- *(A2) (Assumptions on unknown parameter curve) Θ is compact and for all $t \in [0,1]$, $\theta(t)$ lies in the interior of Θ. 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*

$$V(t) = \mathbb{E}\nabla_\theta^2 \ell(\tilde{Z}_0(t), \theta(t)), \tag{2.9}$$

$$I(t) = \mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) \cdot \nabla_\theta \ell(\tilde{Z}_0(t), \theta(t))^{\mathsf{T}}], \tag{2.10}$$

$$\Lambda(t) = \sum_{j\in\mathbb{Z}} \mathbb{E}[\nabla_\theta \ell(\tilde{Z}_0(t), \theta(t)) \cdot \nabla_\theta \ell(\tilde{Z}_j(t), \theta(t))^{\mathsf{T}}], \tag{2.11}$$

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* $n \in \mathbb{N}$*,* $i = 1, \dots, n$*,* $t, t' \in [0,1]$*,* $j \in \mathbb{N}$*:*

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

and

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

and <u>*either*</u>

$$\|Y_i - \tilde{Y}_i(i/n)\|_{rM} \le C_A n^{-1}, \qquad \|\tilde{Y}_0(t) - \tilde{Y}_0(t')\|_{rM} \le C_B|t - t'| \tag{2.12}$$

<u>*or*</u> *(with* χ *from (A1))*

$$\sup_{x \ne x'} \frac{\|F_i(x,\theta) - F_i(x',\theta)\|_{M_y}}{|x - x'|_{\chi,1}} < \infty. \tag{2.13}$$

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

(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> *(2.13)* <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 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 ℓ to be twice continuously differentiable w.r.t. θ as done in (A1) is a typical condition and is needed to use Taylor expansions. We additionally ask ℓ and its derivatives w.r.t. θ 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 ℓ. 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 (2.13) is dedicated to general linear models which may have discretely distributed observations Y_i and thus would not fulfill a condition like (2.12) 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 [10] and Dahlhaus and Subba Rao [13] and formalized in Dahlhaus, Richter and Wu [12] especially for processes with Markov structure.

ASSUMPTION 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 χ_l which is a stronger condition than the corresponding condition on $\nabla_\theta^2 \ell$ in (A1).

3. Main results.

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

$$A^{\otimes k} = A \otimes \ldots \otimes A \tag{3.1}$$

denote the k-fold Kronecker product.

THEOREM 3.1. *Fix $t \in (0,1)$. Let Assumption 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

$$\sqrt{nb_n}(\hat{\theta}_{b_n}(t) - \theta(t) - b_n^2 \frac{\mu_{K,2}}{2}\theta''(t)) \xrightarrow{d} N(0, \sigma_{K,0}^2 \cdot V(t)^{-1} I(t) V(t)^{-1}). \tag{3.2}$$

(iii) If additionally, Assumption 2.2 is fulfilled and $nb_n^9 \to 0$, then

$$\begin{aligned}
&\begin{pmatrix} \sqrt{nb_n}(\hat{\theta}_{b_n}(t) - \theta(t) - b_n^2 \frac{\mu_{K,2}}{2}\theta''(t)) \\ \sqrt{nb_n^3}(\widehat{\theta'}_{b_n}(t) - \theta'(t) - b_n^2 \frac{\mu_{K,4}}{2\mu_{K,2}} bias(t)) \end{pmatrix} \\
&\xrightarrow{d} \quad N\Big(0, \sigma_{K,0}^2 \begin{pmatrix} 1 & 0 \\ 0 & \mu_{K,2}^{-2} \end{pmatrix} \otimes \{V(t)^{-1} I(t) V(t)^{-1}\}\Big),
\end{aligned} \tag{3.3}$$

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 3.2. *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, Richter and Wu [12], 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 (About local constant estimation). *If instead of (2.3) and (2.4), 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 (3.2). If for instance twice differentiability of $t \mapsto \tilde{Z}_0(t)$ is assumed, one obtains

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

Note that the bias term changes significantly.

3.2. *A Bahadur representation for $\hat{\theta}_{b_n}$, $\widehat{\theta}'_{b_n}$.* 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 3.4(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 3.4(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 [56]. Similar results are obtained for $\widehat{\theta}'_{b_n}$ in Theorem 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 3.4 (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 2.1 be fulfilled.*

(i) It holds that

$$\sup_{t\in\mathcal{T}_n}\Big|V(t)\cdot\big\{\hat{\theta}_{b_n}(t)-\theta(t)\big\} - \nabla_\theta L^c_{n,b_n}(t,\theta(t),\theta'(t))\Big| = O_{\mathbb{P}}(\tau_n^{(1)}), \tag{3.4}$$

$$\sup_{t\in\mathcal{T}_n}\big|\nabla_\theta L^c_{n,b_n}(t,\theta(t),\theta'(t)) - b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) - (nb_n)^{-1}\sum_{i=1}^n K_{b_n}(i/n-t)h_i\big| = O_{\mathbb{P}}(\beta_n b_n^2 + b_n^3 + (nb_n)^{-1}). \tag{3.5}$$

(ii) If additionally Assumption 2.2 is fulfilled, then

$$\sup_{t\in\mathcal{T}_n}\Big|\mu_{K,2}V(t)\cdot b_n\big\{\widehat{\theta}'_{b_n}(t)-\theta'(t)\big\} - b_n^{-1}\nabla_{\theta'}L^c_{n,b_n}(t,\theta(t),\theta'(t))\Big| = O_{\mathbb{P}}(\tau_n^{(2)}), \tag{3.6}$$

$$\sup_{t\in\mathcal{T}_n}\big|b_n^{-1}\nabla_{\theta'}L^c_{n,b_n}(t,\theta(t),\theta'(t)) - b_n^3\frac{\mu_{K,4}}{2}V(t)\,bias(t) - (nb_n)^{-1}\sum_{i=1}^n K_{b_n}(i/n-t)\frac{(i/n-t)}{b_n}h_i\big| = O_{\mathbb{P}}(\beta_n b_n^2 + b_n^4 + (nb_n)^{-1}). \tag{3.7}$$

3.3. *Simultaneous confidence bands for $\hat{\theta}_{b_n}$, $\widehat{\theta}'_{b_n}$.* Based on the Bahadur result, we use results from Wu and Zhou [50] 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 3.5 (Theorem 1 and Corollary 2 from Wu and Zhou [50]). *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 3.6 (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 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*

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\sqrt{nb_n}}{\sigma_{K,0}} \sup_{t\in\mathcal{T}_n} \Big|\Sigma_C^{-1}(t)\Big\{\hat{\theta}_{b_n,C}(t) - \theta_C(t) - b_n^2\frac{\mu_{K,2}}{2}C^\mathsf{T}\theta''_C(t)\Big\}\Big| - B_K(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)), \tag{3.8}$$

(ii) If additionally, Assumption 2.2 is fulfilled and $nb_n^9\log(n) \to 0$, *then with* $\hat{K}(x) = K(x)x$,

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\sqrt{nb_n^3}\mu_{K,2}}{\sigma_{K,2}} \sup_{t\in\mathcal{T}_n} \Big|\Sigma_C^{-1}(t)\Big\{\hat{\theta}'_{b_n,C}(t) - \theta'_C(t) - b_n^2\frac{\mu_{K,4}}{2\mu_{K,2}}C^\mathsf{T}\mathit{bias}(t)\Big\}\Big| - B_{\hat{K}}(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)), \tag{3.9}$$

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

$$B_K(m^*) = \sqrt{2\log(m^*)} + \frac{\log(C_K) + (s/2 - 1/2)\log(\log(m^*)) - \log(2)}{\sqrt{2\log(m^*)}}, \tag{3.10}$$

with

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

REMARK 3.7. *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 (3.8) and (3.9), 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 4.

4. Implementational issues. In this section, we discuss some issues which arise by implementing the procedure from Theorem 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.

4.1. *Bias correction.* There are several possible ways to eliminate the bias term in (3.8). 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 3.6 in practice. Here instead we propose a bias correction via the following jack-knife method: We define

$$\tilde{\theta}_{b_n}(t) := 2\hat{\theta}_{b_n/\sqrt{2}}(t) - \hat{\theta}_{b_n}(t). \tag{4.1}$$

Since the Bahadur representation from Theorem 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 3.6(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.

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 3.6. In Examples 5.1, 5.2 and 5.3 it can be seen that in many time series and independent regression models where the objective function ℓ 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 ℓ coincides with the true log conditional likelihood, one has even $V(t) = I(t)$. As it can be seen in Examples 5.1 and 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 ζ_0 which can be calculated by further assumptions on ζ_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

$$\Sigma_C(t) = C^\mathsf{T} V(t)^{-1} I(t) V(t)^{-1} C, \qquad \text{or} \tag{4.2}$$

$$\Sigma_C(t) = C^\mathsf{T} I(t)^{-1} C / c_0 \quad \text{with some known constant } c_0. \tag{4.3}$$

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

$$\hat{V}_{b_n}(t) := (nb_n)^{-1} \sum_{i=1}^n K_{b_n}(i/n - t) \nabla_\theta^2 \ell(Z_i^c, \hat{\theta}_{b_n}(t) + (i/n - t)\widehat{\theta}'_{b_n}(t)), \tag{4.4}$$

$$\begin{aligned} \hat{I}_{b_n}(t) := (nb_n)^{-1} \sum_{i=1}^n K_{b_n}(i/n - t) &\nabla_\theta \ell(Z_i^c, \hat{\theta}_{b_n}(t) + (i/n - t)\widehat{\theta}'_{b_n}(t)) \\ &\times \nabla_\theta \ell(Z_i^c, \hat{\theta}_{b_n}(t) + (i/n - t)\widehat{\theta}'_{b_n}(t))^\mathsf{T}. \end{aligned} \tag{4.5}$$

The convergence of these estimators is given in the next Proposition and easily follows from Lemma A.4 and Lemma A.5 in the appendix. Note that the following Proposition also holds if $\widehat{\theta}'_{b_n}$ in (4.4) and (4.5) is replaced by 0.

PROPOSITION 4.1. *Let Assumption 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 (4.2) or (4.3), 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 5) are kept by (4.2) or (4.3). Therefore, we only adopt a method from

Zhou and Wu [56] 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 $\tilde{\Phi}_i := \tilde{Q}_i \tilde{Q}_i^{\mathsf{T}}/(2m+1)$. Let τ_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 4.2. *Suppose that Assumption 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 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 τ_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 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 $q_1 < \min\{3/10, 1 - 2q_3\}$.

4.3. *Bandwidth selection.* Based on the asymptotic result (3.2) in Theorem 3.1 under Assumption 2.1, the MSE global optimal bandwidth choice reads

$$\hat{b}_n = n^{-1/5} \cdot \Big(\frac{\sigma_{K,0}^2 \int_0^1 \operatorname{tr}(V(t)^{-1} I(t) V(t)^{-1}) dt}{\mu_{K,2}^2 \int_0^1 |\theta''(t)|^2 dt}\Big)^{1/5}. \tag{4.6}$$

We therefore adapt a model-based cross validation method from Richter and Dahlhaus [42], 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 $j = 1, \dots, n$, define the leave-one-out local linear likelihood

$$L_{n,b_n,-j}^c(t,\theta,\theta') := (nb_n)^{-1} \sum_{i=1, i\neq j}^n K_{b_n}(i/n - t)\ell(Z_i^c, \theta + (i/n - t)\theta') \tag{4.7}$$

and the corresponding leave-one-out estimator

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

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

$$CV(b) := n^{-1} \sum_{i=1}^{n} \ell(Z_i^c, \hat{\theta}_{b_n,-i}(i/n)) w(i/n), \tag{4.8}$$

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). In Richter and Dahlhaus [42], 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 ℓ leads to estimators $\hat{\theta}_{b_n}$ which are not consistent. This motivates that a similar behavior should hold for the local constant version.

4.4. *Bootstrap method.* The SCB for $\theta_C(t)$ obtained in Theorem 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 (4.1). 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 4.3. *Suppose that Assumption 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*

$$\sup_{t \in \mathcal{T}_n} |\tilde{\theta}_C(t) - \theta_C(t) - W(t)| = O_{\mathbb{P}}\Big(\frac{n^{-\nu}}{\sqrt{nb_n}\log(n)^{1/2}}\Big), \tag{4.9}$$

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

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

The proof of Proposition 4.3 is immediate from the approximation rates (A.57), (A.58), (A.59) and (A.61) 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 (4.9) 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

$$\mu_{b_n}^{boot}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} V_i^* K_{b_n}(i/n - t), \tag{4.10}$$

where $V_1^*, V_2^*, \dots,$ 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 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 4.3 and compute $\tilde{\theta}_C(t)$ based on the jackknife-based estimator from 4.1.
- For $r = 1, \dots, N$ with some large N, generate n i.i.d. $N(0, I_{s\times s})$ random variables $V_1^*, \dots, V_n^*$ and compute $q_r = \sup_{t\in[0,1]} |\mu_{b_n}^{boot}(t)|$, where $\mu_{b_n}^{boot}(t)$ is computed according to (4.10).
- 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 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 $\mathcal{B}_s = \{x \in \mathbb{R}^s : |x| \leq 1\}$ is the unit ball in $\mathbb{R}^s$.

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

$$Y_i = \mu(X_i, \theta(i/n)) + \sigma(X_i, \theta(i/n))\zeta_i, \tag{5.1}$$

where $\theta = (\alpha_1, \dots, \alpha_k, \beta_0, \dots, \beta_l)^\mathsf{T}$ *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. ζ_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*

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

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*:*

$$\sup_{x\neq x'} \frac{|m_i(x) - m_i(x')|}{|x-x'|_{\kappa_{i\cdot},1}} \leq 1, \qquad \sup_{x\neq x'} \frac{|\sqrt{\nu_i(x)} - \sqrt{\nu_i(x')}|}{|x-x'|_{\rho_{i\cdot},1}} \leq 1. \tag{5.2}$$

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*

$$\sum_{j=1}^{p} \Big(\sup_{\theta\in\tilde{\Theta}} \sum_{i=1}^{k} |\alpha_i|\kappa_{ij} + \|\zeta_0\|_{4M} \cdot \sup_{\theta\in\tilde{\Theta}} \sum_{i=0}^{l} \sqrt{\beta_i}\rho_{ij}\Big) < 1. \tag{5.3}$$

4. Assumption 2.1 (A2) is valid with some $\Theta \subset \tilde{\Theta}$*.*

Then Assumption 2.1 is fulfilled for ℓ *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*

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

where I_d *denotes the d-dimensional identity matrix.*
If additionally, Assumption 2.2 (B1) is fulfilled and m_i*,* ν_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 2.2 is fulfilled for ℓ*.*

Note that in many special cases as the tvAR or the tvARCH processes, the restrictive conditions on the parameter space (5.3) can be relaxed by using matrix arguments, see also the proof techniques in Example 5.2.

In the following, we consider the tvGARCH model. This model was for instance studied in the stationary case in Francq and Zakoïan [19]. More recently, pointwise asymptotic results were obtained in Rohan and Ramanathan [44]. 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 (3.1).

EXAMPLE 5.2 (tvGARCH models). *For $i = 1, \ldots, n$, 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 $\theta = (\alpha_0, \ldots, \alpha_m, \beta_1, \ldots, \beta_l) : [0,1] \to \Theta \subset \mathbb{R}^{m+l+1}$. Let $f(\theta) = (\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_l)^\mathsf{T}$ and let $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^\mathsf{T}$ be the unit column vector with jth element being 1, $1 \leq j \leq l + m$. Define $M_i(\theta) = (f(\theta)\zeta_i^2, e_1, \ldots, e_{m-1}, f(\theta), e_{m+1}, \ldots, e_{m+l-1})^\mathsf{T}$. 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 2.1(A2) is fulfilled with $\Theta \subset \tilde{\Theta}$ and each component of $\theta(\cdot)$ is in $C^4[0,1]$,
(ii) ζ_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 2.1 and 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 5.1).

Let $q \in \mathbb{N}$. In the important GARCH(1,1) case, Ling [34] 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 [5]. 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

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

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 5.3 (Logistic regression). *Fix $m \in \mathbb{N}$. Assume that $\zeta_i = (\zeta_{i,0}, \zeta_{i,1}, \dots, \zeta_{i,m})^\mathsf{T}$, where $\zeta_{i,j}$, $i \in \mathbb{Z}$, $j = 0, \dots, m$ 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 $\mathcal{G}_i = (\dots, \zeta_{i-1,0}, \zeta_{i,0})$. For $i = 1, \dots, n$,*

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

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 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 $C_B > 0$,*
$$\|\tilde{X}_0(t) - \tilde{X}_0(t')\|_8 \le C_B|t - t'|.$$
4. *For each $t \in [0,1]$, $\mathbb{E}[\tilde{X}_0(t)\tilde{X}_0(t)^\mathsf{T}]$ is positive definite.*

Then Assumption 2.1 is fulfilled and $\Lambda(t) = I(t) = V(t)$.

Note that it is not possible to fulfill Assumption 2.2 in our setting of Example 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, Richter and Wu [12], that differentiability in L^1 should be enough to show the bias expansions for which Assumption 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.

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.

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 ζ_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 (4.6) 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 (4.5). We obtained the results given in Tables 1 and 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 1
Coverage probabilities of the SCB in (a) for $n = 2000$

	$\alpha = 90\%$			$\alpha = 95\%$		
b_n	α_0	α_1	$(\alpha_0, \alpha_1)^{\mathsf{T}}$	α_0	α_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

6.2. *Applications.* In this section, we consider a few real-data applications of our procedure. As mentioned in Section 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, Sapatinas and Subba Rao [20] proposed, ARCH models have good

TABLE 2
Coverage probabilities of the SCB in (b) for $n = 5000$

	$\alpha = 90\%$				$\alpha = 95\%$			
b_n	α_0	α_1	β_1	$(\alpha_0, \alpha_1, \beta_1)^\mathsf{T}$	α_0	α_1	β_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

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.

6.2.1. *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 [20] using tvARCH(p) models with $p = 0, 1, 2$. The authors suggested choosing $p = 1$ for USD-GBP exchange rates. We collect the same data from `www.federalreserve.gov/releases/h10/Hist/default1999.htm` and fit both tvARCH(1) and tvARCH(2) models. 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.

Based on Figure 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

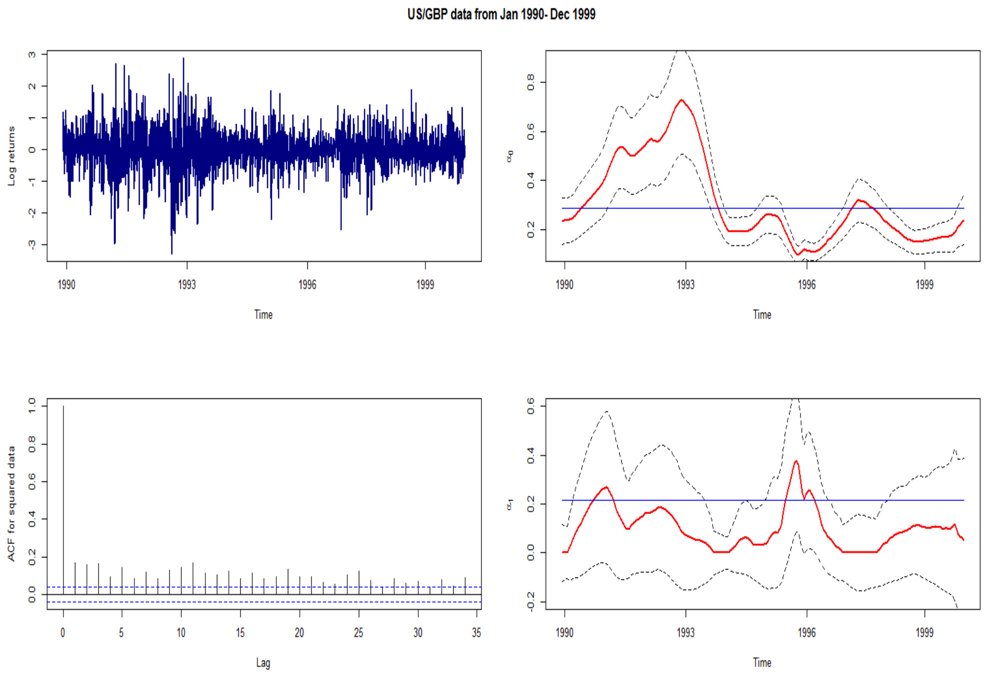

Fig 1. *Analysis of USD/GBP data from Jan 1990 to Dec 1999. Top left: Log-returns. Bottom left: ACF plot. Right panel: Estimates of the parameters $\alpha_0(\cdot)$, $\alpha_1(\cdot)$, respectively (red) with 95% SCBs (dashed) and estimates of the parameters assuming constancy (blue).*

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).

6.2.2. *Real data application II: Merval index data.* In the empirical analysis of log-return for stock market data, however, as Palm [39] 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 [46], 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 2, the time series show significant lags in its ACF plot after squaring; indicating conditional heteroscedasticity.

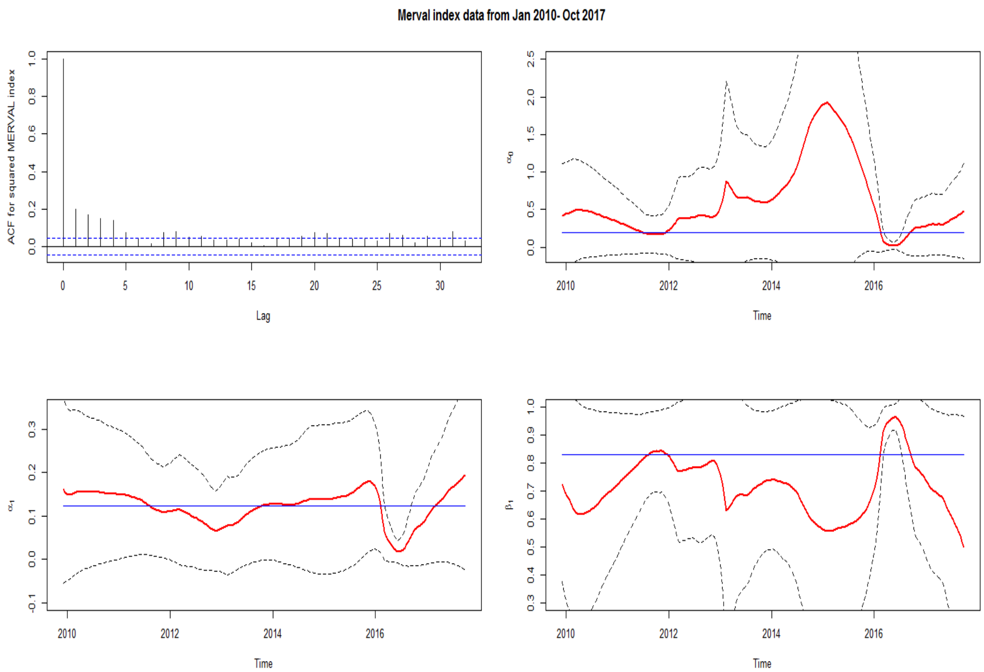

Fig 2. *Analysis of MERVAL index data from Jan 2010 to Oct 2017. Top left: ACF plot. Top right, bottom left, bottom right: Estimates of the parameters $\alpha_0(\cdot)$, $\alpha_1(\cdot)$ and $\beta_1(\cdot)$, respectively (red) with SCBs (dashed) and estimates of the parameters assuming constancy (blue).*

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 implications. 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.

References.

[1] Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets. *J. Econometrics* **135** 77–124. MR2328397

[2] Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[3] Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics* **79** 551–563.

[4] Billingsley, P. (1999). *Convergence of probability measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. MR1700749

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327. MR853051

[6] Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. 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. MR0378310

[7] Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. MR2328589

[8] Chen, J. and Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical association* **92** 739–747.

[9] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. MR0141193

[10] Dahlhaus, R. (2011). Locally Stationary Processes. *Handbook of Statistics*.

[11] Dahlhaus, R. and Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1–39. MR2546797

[12] Dahlhaus, R., Richter, S. and Wu, W. B. (2017). Towards a general theory for non-linear locally stationary processes. *ArXiv e-prints: 1704.02860*.

[13] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34** 1075–1114. MR2278352

[14] de Vries, S. O., Fidler, V., Kuipers, W. D. and Hunink, M. G. M. (1998). Fitting Multistate Transition Models with Autoregressive Logistic Regression: Supervised Exercise in Intermittent Claudication. *Medical Decision Making* **18** 52-60. PMID: 9456209.

[15] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** 987–1007. MR666121

[16] Engle, R. F. and Rangel, J. G. (2005). The spline garch model for unconditional volatility and its global macroeconomic causes.

[17] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497

[18] Fan, J. and Zhang, W. (2000). Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models. *Scandinavian Journal of Statistics* **27** 715–731.

[19] Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10** 605–637. MR2076065

[20] Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.* **36** 742–786. MR2396814

[21] Giacometti, R., Bertocchi, M., Rachev, S. T. and Fabozzi, F. J. (2012). A comparison of the Lee–Carter model and AR–ARCH model for forecasting mortality rates. *Insurance: Mathematics and Economics* **50** 85–93.

[22] Guanche, Y., Mínguez, R. and Méndez, F. J. (2014). Autoregressive logistic regression applied to atmospheric circulation patterns. *Climate dynamics* **42** 537–552.

[23] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699
[24] Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788. MR2087972
[25] James Chu, C.-S. (1995). Detecting parameter shift in GARCH models. *Econometric Reviews* **14** 241–266.
[26] Kim, S., Cho, S. and Lee, S. (2000). On the cusum test for parameter changes in GARCH (1, 1) models. *Communications in Statistics-Theory and Methods* **29** 445–462.
[27] Kokoszka, P., Leipus, R. et al. (2000). Change-point estimation in ARCH models. *Bernoulli* **6** 513–539.
[28] Kowsar, R., Keshtegar, B., A., M. M. and A., M. (2017). An autoregressive logistic model to predict the reciprocal effects of oviductal fluid components on in vitro spermophagy by neutrophils in cattle. *Scientific Reports* **7**.
[29] Kulperger, R., Yu, H. et al. (2005). High moment partial sum processes of residuals in GARCH models and their applications. *The Annals of Statistics* **33** 2395–2422.
[30] Leybourne, S. J. and McCabe, B. P. M. (1989). On the distribution of some test statistics for coefficient constancy. *Biometrika* **76** 169–177. MR991435
[31] Lin, C.-F. J. and Teräsvirta, T. (1999). Testing parameter constancy in linear models against stochastic stationary parameters. *J. Econometrics* **90** 193–213. MR1703341
[32] Lin, S.-J., Yang, J. et al. *Testing shifts in financial models with conditional heteroskedasticity: an empirical distribution function approach*.
[33] Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. With comments and a rejoinder by the authors. MR1952726
[34] Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *J. Appl. Probab.* **36** 688–705. MR1737046
[35] Ling, S. and McAleer, M. (2002). Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory* **18** 722–729. MR1906332
[36] Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *The Review of Economics and Statistics* **86** 378–390.
[37] Nabeya, S. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.* **16** 218–235. MR924867
[38] Nyblom, J. (1989). Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.* **84** 223–230. MR999682
[39] Palm, F. C. (1996). GARCH models of volatility. In *Statistical methods in finance. Handbook of Statist.* **14** 209–240. North-Holland, Amsterdam. MR1602132
[40] Ploberger, W., Krämer, W. and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *J. Econometrics* **40** 307–318. MR994952
[41] Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, second ed. *Springer Series in Statistics*. Springer, New York. MR2168993
[42] Richter, S. and Dahlhaus, R. (2017). Cross validation for locally stationary processes. *ArXiv e-prints: 1705.10046*.
[43] Rohan, N. (2013). A time varying GARCH (p, q) model and related statistical inference. *Statistics & Probability Letters* **83** 1983–1990.
[44] Rohan, N. and Ramanathan, T. V. (2013). Nonparametric estimation of a time-varying GARCH model. *J. Nonparametr. Stat.* **25** 33–52. MR3039969
[45] Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics and Statistics* **87** 503–522.
[46] Tagliafichi Ricardo, A. The Garch model and their application to the VaR.
[47] Taylor, J. W. and Yu, K. (2016). 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** 1069–1092.

[48] Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154 (electronic). MR2172215

[49] Wu, W. B. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic Processes and their Applications* **115** 939 - 958.

[50] Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528

[51] Zhang, W., Lee, S.-Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619

[52] Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029

[53] Zhang, T. and Wu, W. B. (2015). Time-varying nonlinear regression models: nonparametric estimation and model selection. *Ann. Statist.* **43** 741–768. MR3319142

[54] Zhang, D. and Wu, W. B. (2017). Gaussian Approximation for High Dimensional Time Series. *Ann. Statist.* **45** 1895–1919.

[55] Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444

[56] Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 513–531. MR2758526

SUPPLEMENTARY MATERIAL

Supplement A: Proofs
(doi: COMPLETED BY THE TYPESETTER; .pdf). This material contains the proofs of the results in the paper as well as the proofs of the examples.

APPENDIX A: PROOFS

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

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

and $\hat{L}_{n,b_n}^{\circ}$, L_{n,b_n}° similarly as $L_{n,b_n}^{\circ,c}$ 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 $L_{n,b_n}^{\circ}(t, \eta)$. In the proof of Theorem 3.1, it is shown that $L_{n,b_n}^{\circ}(t, \eta)$ 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 $\hat{\chi} = (1, \chi) \in \mathbb{R}^{\mathbb{N}_0}$.

A.1. Proofs of Section 3.

PROOF OF THEOREM 3.1. The proof is similar to the proof of Theorems 5.2 and 5.4 in [12].

(i) Fix $t \in (0, 1)$. By Lemma A.4(ii) applied to $\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we have

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

Applying Lemma A.5 to $\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we obtain

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

where $L^{\circ}(t, \eta) = \int_{-1}^{1} K(x)L(t, \eta_1 + \eta_2 x)dx$. By Lemma A.4(i), we have

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

and thus

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

By Lemma A.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 $L_{n,b_n}^{\circ,c}(t, \eta)$, standard arguments yield

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

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 (A.1), 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 $\nabla_\eta L_{n,b_n}^{\circ,c}(t, \hat{\eta}_{b_n}(t)) = 0$ with probability tending to 1. By a Taylor expansion we obtain

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

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 $V(t,\theta) := \mathbb{E}\nabla_\theta^2 \ell(\tilde{Z}_0(t), \theta)$. Since $g = \nabla_\theta^2 \ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we can use similar arguments as in (i) to obtain

$$\sup_{\eta \in E_n} |\nabla_\eta^2 L_{n,b_n}^{\circ,c}(t,\eta) - V^\circ(t,\eta)| = o_{\mathbb{P}}(1), \tag{A.3}$$

where

$$V^\circ(t,\eta) = \int_{-1}^1 K(x) \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix} \otimes V(t, \eta_1 + \eta_2 x) dx. \tag{A.4}$$

Let

$$V^\circ(t) := \begin{pmatrix} 1 & 0 \\ 0 & \mu_{K,2} \end{pmatrix} \otimes V(t). \tag{A.5}$$

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 (A.3), we conclude that

$$\nabla_\theta^2 L_{n,b_n}^{\circ,c}(t, \bar{\eta}(t)) = V^\circ(t, \bar{\eta}(t)) + o_{\mathbb{P}}(1) = V^\circ(t) + o_{\mathbb{P}}(1). \tag{A.6}$$

By Lemma A.4(i), we have

$$\|\nabla_\eta L_{n,b_n}^{\circ,c}(t, \eta_{b_n}(t)) - \nabla_\eta \hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t))\|_1 = O((nb_n)^{-1}). \tag{A.7}$$

With (A.2), (A.6) and (A.7) we obtain

$$\begin{aligned}
&\begin{pmatrix} \sqrt{nb_n}(\hat{\theta}_{b_n}(t) - \theta(t)) \\ \sqrt{nb_n^3}(\widehat{\theta}'_{b_n}(t) - \theta'(t)) \end{pmatrix} = \sqrt{nb_n}(\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)) \\
= \quad & -V^\circ(t)^{-1}\sqrt{nb_n}\nabla_\eta \hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t)) + o_{\mathbb{P}}(1) \\
= \quad & -V^\circ(t)^{-1}\sqrt{nb_n}\big\{\nabla_\eta \hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t))\big\} \\
& \qquad -V^\circ(t)^{-1} \begin{pmatrix} \sqrt{nb_n}\mathbb{E}\nabla_{\eta_1}\hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t)) \\ \sqrt{nb_n^3} b_n^{-1}\mathbb{E}\nabla_{\eta_2}\hat{L}_{n,b_n}^\circ(t, \eta_{b_n}(t)) \end{pmatrix} + o_{\mathbb{P}}(1).
\end{aligned} \tag{A.8}$$

By (A.8), it is enough to show the two stochastic convergences

$$b_n^{-1}\big\{\nabla_{\eta_2} L^\circ_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t))\big\} = o_{\mathbb{P}}(1), \tag{A.9}$$

$$b_n^{-1}\mathbb{E}\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t)) = o_{\mathbb{P}}(1). \tag{A.10}$$

Using (A.37) from the proof of Lemma A.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 (A.9) due to $nb_n^3 \to \infty$. Using the intermediate result (A.47) in the proof of A.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 (A.10) 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 (A.8). 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 [48], 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 $j = 1,\dots,2d_\Theta$, that

$$\begin{aligned}
&\mathrm{LIM}\ \|S_{n,b_n,m}(t)_j - (nb_n)^{1/2}\big\{\nabla_{\eta_j}\hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_{\eta_j}\hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t))\big\}\|_2\\
\leq\ & \mathrm{LIM}\ (nb_n)^{-1/2}\sum_{l=m}^\infty\Big\|\sum_{i=1}^n (U_{i,n}(t)\otimes P_{i-l}\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(t)+\theta'(t)(i/n-t)))_j\Big\|_2\\
=\ & \mathrm{LIM}\ (nb_n)^{-1/2}\Big(\sum_{i=1}^n\big\|(U_{i,n}(t)\otimes P_{i-l}\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(t)+\theta'(t)(i/n-t)))_j\big\|_2^2\Big)^{1/2}\\
\leq\ & |K|_\infty \mathrm{LIM}\sum_{l=m}^\infty\sup_{t\in[0,1]}\sup_{i,\theta}\delta_2^{\nabla_{\theta_i}\ell(\tilde{Z}(t),\theta)}(l) = 0,
\end{aligned} \tag{A.11}$$

by Lemma A.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 $j = 1,\dots,2d_\Theta$,

$$\|S_{n,b_n,m}(t)_j - \tilde{S}_{n,b_n,m}(t)_j\|_2 = O((nb_n)^{-1/2}). \tag{A.12}$$

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 A.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 [13]. 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 $\iota > 0$,

$$
\begin{aligned}
&\sum_{i=1}^n \mathbb{E}\big[M_i(t)_j^2 \mathbb{1}_{\{|M_i(t)_j| > \iota\}}\big] \\
\le\ & C(nb_n)^{-1} \sum_{l=0}^{m-1} \sum_{i=1}^n K_{b_n}(i/n-t)^2 \mathbb{E}\big[\tilde{M}_{ij,l}^2 \mathbb{1}_{\{|K|_\infty |\tilde{M}_{ij,l}| > \iota(nb_n)^{1/2}\}}\big].
\end{aligned} \tag{A.13}
$$

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 A.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 (A.13) is tending to 0. The proof for $j = d_\Theta + 1, \ldots, 2d_\Theta$ is similar. From Theorem 18.2 in Billingsley [4] and the Cramer-Wold device we obtain

$$
\tilde{S}_{n,b_n,m}(t) \xrightarrow{d} N\Big(0, \begin{pmatrix} \sigma_{K,0}^2 & 0 \\ 0 & \sigma_{K,2}^2 \end{pmatrix} \otimes \Sigma_m\Big). \tag{A.14}
$$

Using (A.11), (A.12), (A.14) and $\Sigma_m \to \Lambda(t)$ $(m \to \infty)$, we obtain

$$(nb_n)^{1/2}\{\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t)) - \mathbb{E}\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t))\} \xrightarrow{d} N\Big(0, \begin{pmatrix} \sigma^2_{K,0} & 0 \\ 0 & \sigma^2_{K,2}\end{pmatrix} \otimes \Lambda(t)\Big). \tag{A.15}$$

Using (A.15), the expansion (A.8) and Lemma A.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 3.4. (i),(ii) By Lemma A.4(i),(iii) and Lemma A.5 applied to $g = \ell$, we have that

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

i.e. $L^{\circ,c}_{n,b_n}(t,\eta)$ converges to $L^\circ(t,\eta)$ uniformly in t,η if $b_n = o(1)$ and $\beta_n = o(1)$. It was already seen in the proof of Theorem 3.1 that $L^\circ(t,\eta)$ is continuous w.r.t. η and uniquely minimized by $\eta = (\theta(t)^\mathsf{T}, 0)^\mathsf{T}$. Standard arguments give

$$\sup_{t\in\mathcal{T}_n} |\hat{\eta}_{b_n}(t) - \eta_n(t)| = o_{\mathbb{P}}(1). \tag{A.16}$$

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 $t \in \mathcal{T}_n$:

$$\hat{\eta}_{b_n}(t) - \eta_{b_n}(t) = -\big[V^\circ(t) + R_{n,b_n}(t)\big]^{-1} \cdot \nabla_\eta L^{\circ,c}_{n,b_n}(t,\eta_{b_n}(t)), \tag{A.17}$$

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 $|\bar{\eta}(t) - \eta_{b_n}(t)|_1 \le |\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)|_1$ and $V^\circ(t)$ is defined in (A.5).

By Lemma A.4(i),(iii) and Lemma A.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

$$\sup_{t\in\mathcal{T}_n}\sup_{\theta\in\Theta} |\nabla^2_\eta L^{\circ,c}_{n,b_n}(t,\eta) - V^\circ(t,\eta)| = O_{\mathbb{P}}(\beta_n + (nb_n)^{-1}) + O(b_n), \tag{A.18}$$

where $V^\circ(t,\eta)$ is defined in (A.4).

Note that $\mathbb{E}\nabla_\theta \ell(\tilde{Z}_0(t),\theta(t)) = 0$ by Assumption 2.1(A3) in connection with 2.1(A1). By Lemma A.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 A.1, we see that the assumptions of Lemma A.9 are fulfilled and thus:

$$\sup_{t\in\mathcal{T}_n} \Big|(nb_n)^{-1}\sum_{i=1}^n K_{b_n}(i/n - t)\nabla_\theta \ell(\tilde{Z}_i(i/n),\theta(i/n))\Big| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n)). \tag{A.19}$$

With Lemma A.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 A.6(i),(ii) and Lemma A.4(i):

$$\sup_{t\in\mathcal{T}_n} |\nabla_{\eta_j} L^{\circ,c}_{n,b_n}(t, \eta_{b_n}(t))| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n) + (nb_n)^{-1} + \beta_n b_n^2 + b_n^{1+j}), \tag{A.20}$$

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 A.1, the same holds for $\eta \mapsto V^\circ(t,\eta)$. We conclude that with some constant $C > 0$,

$$\sup_{t\in\mathcal{T}_n} |R_{n,b_n}(t)| \le \sup_{t\in\mathcal{T}_n}\sup_{\eta\in E_n} |\nabla^2_\eta L^{\circ,c}_{n,b_n}(t,\eta) - V^\circ(t,\eta)| + C\sup_{t\in\mathcal{T}_n} |\hat{\eta}_{b_n}(t) - \eta_{b_n}(t)|. \tag{A.21}$$

Inserting (A.20), (A.21) and (A.16) into (A.17), we obtain

$$\sup_{t\in\mathcal{T}_n} |\hat{\eta}_{b_n,j}(t) - \eta_{b_n,j}(t)| = O_{\mathbb{P}}((nb_n)^{-1/2}\log(n) + (nb_n)^{-1} + \beta_n b_n^2 + b_n^{1+j}), \tag{A.22}$$

where $j = 1, 2$. Inserting (A.22), (A.18) into (A.21), 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

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

and (A.20) we have (3.4) and (3.6). The second result (3.5) and (3.7) follows from Lemma A.4(i), Lemma A.7 and Lemma A.6. □

LEMMA A.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 $j \in \mathbb{N}$,*

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

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, χ and $\tilde{D}$ (only in (ii)) such that

(i)

$$\begin{aligned}
&\|\sup_{\theta\in\Theta} |g(\hat{Z},\theta) - g(\hat{Z}',\theta)|\|_q &\le\ & \bar{C}\cdot C\sum_{j=0}^{\infty} \chi_j \|\hat{Z}_j - \hat{Z}'_j\|_{qM}, &\text{(A.23)}\\
&\Big\|\sup_{\theta\ne\theta'} \frac{|g(\hat{Z},\theta) - g(\hat{Z},\theta')|}{|\theta-\theta'|_1}\Big\|_q &\le\ & \bar{C}\cdot C, &\text{(A.24)}\\
&\|\sup_{\theta\in\Theta} |g(\hat{Z},\theta)|\|_q &\le\ & \bar{C}\cdot C. &\text{(A.25)}
\end{aligned}$$

(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*

$$\|\sup_{\theta\in\Theta}|g(\hat{Z},\theta)-g(\hat{Z}',\theta)|\|_q \leq \bar{C}\cdot C\sum_{j=1}^{\infty}\chi_j\|\hat{X}_j-\hat{X}_j'\|_{qM}. \tag{A.26}$$

Proof of Lemma A.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:

$$\begin{aligned}
&\|\sup_{\theta\in\Theta}|g(\hat{Z},\theta)-g(\hat{Z}',\theta)|\|_q\\
\leq\ & \bar{C}\big\||\hat{Y}-\hat{Y}'|(R_{M_y-1,M_x-1}(\hat{Y},\hat{X})+R_{M_y-1,M_x-1}(\hat{Y}',\hat{X}))\big\|_q\\
&\quad+\big\||\hat{X}-\hat{X}'|_{\chi,1}\cdot(R_{M_y-1,M_x-1}(\hat{Y},\hat{X})+R_{M_y-1,M_x-1}(\hat{Y},\hat{X}'))\big\|_q\\
\leq\ & \bar{C}\|\hat{Y}-\hat{Y}'\|_{qM}\cdot\big(\|R_{M_y-1,M_x-1}(\hat{Y},\hat{X})\|_{qM/(M-1)}\\
&\qquad\qquad+\|R_{M_y-1,M_x-1}(\hat{Y}',\hat{X})\|_{qM/(M-1)}\big)\\
&\quad+\bar{C}\||\hat{X}-\hat{X}'|_{\chi,1}\|_{qM}\big(\|R_{M_y-1,M_x-1}(\hat{Y},\hat{X})\|_{qM/(M-1)}\\
&\qquad\qquad+\|R_{M_y-1,M_x-1}(\hat{Y},\hat{X}')\|_{qM/(M-1)}\big)\\
\leq\ & 2\bar{C}(1+D(|\chi|_1+1))^{M-1}\big(\|\hat{Y}-\hat{Y}'\|_{qM}+\sum_{j=1}^{\infty}\chi_j\|\hat{X}_j-\hat{X}_j'\|_{qM}\big).
\end{aligned} \tag{A.27}$$

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 (A.25) is obvious from (A.23).

(ii) We first obtain (A.27) 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 (A.27), 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

$$\||\hat{Y}-\hat{Y}'|\cdot|\hat{Y}|^k|\hat{X}|_{\chi,1}^l\|_q\leq\tilde{D}^{1/(qM_y)}\sum_{i=1}^{\infty}\chi_i\|\hat{X}_i-\hat{X}_i'\|_{qM}\cdot D^k(|\chi|_1D)^l,$$

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 A.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 2.1(A5) hold with some $r \geq 2$ and let Assumption 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 A.2. (i) Note that

$$\partial_t g(\tilde{Z}_0(t), \tilde{\theta}(t)) = \partial_z g(\tilde{Z}_0(t), \tilde{\theta}(t))\partial_t \tilde{Z}_0(t) + \nabla_\theta g(\tilde{Z}_0(t), \tilde{\theta}(t))\tilde{\theta}'(t). \tag{A.28}$$

By Lemma A.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 A.1(i) there exists a constant $C > 0$ such that for each $j \in \mathbb{N}_0$,

$$\|\partial_{z_j} g(\tilde{Z}_0(t), \tilde{\theta}(t))\|_{M'/(M'-1)} \leq C\bar{C}\hat{\chi}_j. \tag{A.29}$$

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 A.1(i) we obtain with some constant $C > 0$, for each $k = 1, \ldots, d_\Theta$,

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

From Lemma A.1(i) we obtain for each $k = 1, \ldots, d_\Theta$ (note that $rM \geq M'$ in Assumption 2.2):

$$\begin{aligned}
&\|\nabla_{\theta_k} g(\tilde{Z}_0(t), \tilde{\theta}(t')) - \nabla_{\theta_k} g(\tilde{Z}_0(t'), \tilde{\theta}(t'))\|_1 \\
&\leq \bar{C}C\big(\|\tilde{Y}_0(t) - \tilde{Y}_0(t')\|_{M'} + \sum_{j=1}^{\infty} \chi_j \|\tilde{X}_0(t) - \tilde{X}_0(t')\|_{M'}\big) \\
&\leq \bar{C}CC_B|t-t'|(1 + |\chi|_1). \qquad \text{(A.30)}
\end{aligned}$$

or with Lemma A.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 (A.28) 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 (A.28). By (A.29) and $\|\partial_t \tilde{Z}_{0j}(t) - \partial_t \tilde{Z}_{0j}(t')\|_{M'} \leq C_B |t - t'|$ from Assumption 2.2(B3), we have

$$\|\partial_z g(\tilde{Z}_0(t), \tilde{\theta}(t))(\partial_t \tilde{Z}_0(t) - \partial_t \tilde{Z}_0(t'))\|_1 \leq C\bar{C}C_B |\chi|_1 |t - t'|. \tag{A.31}$$

Similar as in (A.30), we see by Lemma A.1(i) or (ii) that

$$\|\partial_{z_j} g(\tilde{Z}_0(t), \tilde{\theta}(t)) - \partial_{z_j} g(\tilde{Z}_0(t'), \tilde{\theta}(t))\|_{M'/(M'-1)} \leq \chi_j \bar{C}CC_B(1 + |\chi'|_1)|t - t'|. \tag{A.32}$$

Finally, by Lemma A.1(i) and Lipschitz continuity of $\tilde{\theta}$, we have

$$\|\partial_{z_j} g(\tilde{Z}_0(t'), \tilde{\theta}(t)) - \partial_{z_j} g(\tilde{Z}_0(t'), \tilde{\theta}(t'))\|_{M'/(M'-1)} \leq \chi_j \bar{C}C|\tilde{\theta}(t) - \tilde{\theta}(t')|_1 = O(|t - t'|). \tag{A.33}$$

By Hölder's inequality, we conclude from (A.32) and (A.33) 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 (A.31), finishes the proof. □

LEMMA A.3. *Suppose that Assumption 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 ζ_0 is replaced by ζ_0^*. By Lemma A.1 we obtain in case (2.12) that with some constant $\tilde{C} > 0$:

$$\begin{aligned}
&\delta_q^{\sup_\theta |g(Z,\theta)|}(j)\\
&= \|\sup_\theta |g(\tilde{Z}_j(t),\theta)| - \sup_\theta |g(\tilde{Z}_j(t)^*,\theta)|\|_r\\
&\leq \|\sup_\theta |g(\tilde{Z}_j(t),\theta) - g(\tilde{Z}_j(t)^*,\theta)|\|_r\\
&\leq \tilde{C}\Big(\|\tilde{Y}_j(t) - \tilde{Y}_j(t)^*\|_{rM} + \sum_{i=1}^\infty \chi_i \|\tilde{X}_{j-i+1}(t) - \tilde{X}_{j-i+1}(t)^*\|_{rM}\Big)\\
&\leq \tilde{C}\Big(\delta_{rM}^{\tilde{Y}(t)}(j) + \sum_{i=1}^\infty \chi_i \delta_{rM}^{\tilde{X}(t)}(j-i+1)\Big), \qquad \text{(A.34)}
\end{aligned}$$

and in case (2.13), similarly

$$\delta_r^{g(Z,\theta)}(j) \leq \tilde{C}\sum_{i=1}^\infty \chi_i \delta_{rM}^{\tilde{X}(t)}(j-i+1). \qquad \text{(A.35)}$$

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 (A.34), (A.35), 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 $|d_u(t)|_\infty \le \sup_s |\theta''(s)|_\infty \cdot b_n^2$ if $|t-u| \le b_n$), for each $l = 1, \dots, k$,

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

□

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. [54]): For some zero-mean random variable Z, let $\|Z\|_{q,\alpha} := \sup_{m \ge 0}(m+1)^\alpha \Delta_q^Z(m)$.

Lemma A.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*

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

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 2.1(A5), 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 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 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 A.4. (i) By Lemma A.1(i),(ii) and by Assumption 2.1(A6), we obtain (independent of (2.12) or (2.13)) 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 (2.12) 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 (2.13) there exists $C_2 > 0$ such that

$$\|\sup_{\theta\in\Theta}|g(Z_i,\theta) - g(\tilde{Z}_i(i/n),\theta)|\|_1 \leq C_2\sum_{j=1}^{\infty}\chi_j\|X_{ij} - \tilde{X}_{ij}(i/n)\|_M \leq C_2C_A|\chi|_1 n^{-1}.$$

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

$$\begin{aligned}\mathbb{P}\big(\sup_{\eta\in E_n}|\hat{G}_n(t,\eta)| > Q\big) &\leq \#E_n^{(\iota)}\sup_{\eta\in E_n}\mathbb{P}\big(|\hat{G}_n(t,\eta)| > Q/2\big)\\ &\quad+\mathbb{P}(\sup_{|\eta-\eta'|_1\leq\iota}|\hat{G}_n(t,\eta) - \hat{G}_n(t,\eta')| > Q/2).\end{aligned}\tag{A.36}$$

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

$$
\begin{aligned}
&\|\hat{G}_n(t,\eta)\|_2 \\
\leq\ & (nb_n)^{-1}\sum_{l=0}^{\infty}\Big\|\sum_{i=1}^{n}\hat{K}_{b_n}(i/n-t)P_{i-l}g(\tilde{Z}_i(i/n),\eta_1+\eta_2(i/n-t)b_n^{-1})\Big\|_2 \\
\leq\ & (nb_n)^{-1}\sum_{l=0}^{\infty}\Big(\sum_{i=1}^{n}\hat{K}_{b_n}(i/n-t)^2\|P_{i-l}g(\tilde{Z}_i(i/n),\eta_1+\eta_2(i/n-t)b_n^{-1})\|_2^2\Big)^{1/2} \\
\leq\ & (nb_n)^{-1/2}|\hat{K}|_\infty\sum_{l=0}^{\infty}\sup_{t\in[0,1]}\delta_2^{\sup_{\theta\in\Theta}|g(\tilde{Z}(t),\theta)|}(l)=O((nb_n)^{-1/2}),
\end{aligned}
\tag{A.37}
$$

shows that the first summand in (A.36) 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 $\tilde{C}>0$:

$$|g(z,\theta)-g(z,\theta')|\leq\tilde{C}|\theta-\theta'|_1R_{M_y,M_x}(z).$$

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 A.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 ι small enough. So we have shown that (A.36) tends to zero for $n\to\infty$.

For the second assertion, let $\mathcal{T}_n^{(\iota)}=\{i/\lfloor\iota^{-1}n^{1/2}\rfloor:i=0,\dots,\lfloor\iota^{-1}n^{1/2}\rfloor\}$ 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 $\tilde{C}>0$:

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

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 A.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 ι is chosen small enough due to $nb_n^2c_n^{-2}\to\infty$. By (A.37) and $\#(E_n^{(\iota)}\times\mathcal{T}_n^{(\iota)}) = O(n^{1/2}c_n^k)$, the assertion follows from

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

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 $\mathcal{T}_{n,r} := \{i/r : i = 1,\dots,r\}$ 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

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{\eta\in E_n,t\in\mathcal{T}_n}|\hat{G}_n(t,\eta)| > Q\delta_n\Big)\\
\le\;& \mathbb{P}\Big(\sup_{\eta\in E_{n,r},t\in\mathcal{T}_n}|\hat{G}_n(t,\eta)| > Q\delta_n/2\Big)\\
&+\mathbb{P}\Big(\sup_{|\eta-\eta'|\le r^{-1},|t-t'|\le r^{-1}}|\hat{G}_n(t,\eta)-\hat{G}_n(t',\eta')| > Q\delta_n/2\Big).
\end{aligned}\tag{A.38}$$

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 2.1(A7) and Lemma A.3(ii), we have $\sup_u \Delta_4^{\sup_{t,\eta}|M(t,\eta,u)|}(k) = O(k^{-\gamma})$. Thus

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

(independent of n) and

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

(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 [54] (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 $C_\alpha > 0$:

$$\begin{aligned}
&\mathbb{P}\big(\sup_{\eta'\in E_{n,r},t'\in\mathcal{T}_{n,r}} |\hat{G}_n(t',\eta')| \geq Q\delta_n/2\big) \\
\leq\ & \frac{C_\alpha n \cdot l^2 W_{4,\alpha}^4}{(Q/2)^4(\delta_n(nb_n))^4} + C_\alpha \exp\Big(-\frac{C_\alpha (Q/2)^2(\delta_n(nb_n))^2}{nW_{2,\alpha}^2}\Big) \\
\lesssim\ & \frac{n\log(n)^2}{(nb_n)^2 b_n^{-2}\log(n)^2} + \exp\big(-\frac{(nb_n)b_n^{-1}\log(n)}{n}\big) \\
\lesssim\ & n^{-1} \to 0. \qquad \text{(A.39)}
\end{aligned}$$

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

$$\begin{aligned}
&|\hat{G}_n(t,\eta) - \hat{G}_n(t',\eta')| \\
\leq\ & C_M\big\{|\eta-\eta'|_1 + |t-t'|\big\} b_n^{-2} \cdot (L_{\hat{K}} + |\hat{K}|_\infty) \qquad \text{(A.40)} \\
& \times n^{-1}\sum_{i=1}^n \{R_{M_y,M_x}(\tilde{Z}_i(i/n)) + \mathbb{E}R_{M_y,M_x}(\tilde{Z}_i(i/n))\}, \qquad \text{(A.41)}
\end{aligned}$$

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 A.1, we conclude with Markov's inequality that

$$\mathbb{P}\big(\sup_{|\eta-\eta'|_1\leq r^{-1},|t-t'|\leq r^{-1}} |\hat{G}_n(t,\eta) - \hat{G}_n(t',\eta')| \geq C\delta_n/2\big) = O\Big(\frac{b_n^{-2}r^{-1}}{\delta_n}\Big). \qquad \text{(A.42)}$$

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 (A.39) and (A.42) into (A.38), we obtain the result. □

LEMMA A.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 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

$$\sup_{t\in\mathcal{T}_n}\sup_{\eta\in E_n} |\mathbb{E}\hat{B}_n(t,\eta) - \int \hat{K}(x)\mathbb{E}g(\tilde{Z}_0(t), \eta_1 + \eta_2 x)dx| = O((nb_n)^{-1} + b_n). \tag{A.43}$$

PROOF OF LEMMA A.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 A.1(i), we have with some constant $\tilde{C} > 0$ that either in the case of (2.12),

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

or in the case of (2.13),

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

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 A.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 $t \in \mathcal{T}_n$, $\eta \in E_n$. $\square$

LEMMA A.6. *Let $\eta_{b_n}(t) = (\theta(t)^\top, b_n\theta'(t)^\top)^\top$. Let Assumption 2.1 hold with $r = 1$.*

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

$$\mathbb{E}\nabla_{\eta_1}\hat{L}^{\circ}_{n,b_n}(t, \eta_{b_n}(t)) = b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) + O(b_n^3). \tag{A.44}$$

(ii) If additionally Assumption 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 $bias(t) = \frac{1}{3}\theta^{(3)}(t) + V(t)^{-1}\mathbb{E}[\partial_t\nabla^2_\theta\ell(\tilde{Z}_0(t), \theta(t))] \cdot \theta''(t)$, and the term $O(b_n^3)$ in (A.44) can be replaced by $O(b_n^4)$.

PROOF OF LEMMA A.6. (i) Let $U_{i,n}(t) = (K_{b_n}(i/n-t), K_{b_n}(i/n-t)(i/n-t)b_n^{-1})^\top$. 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

$$\begin{aligned}
&\nabla_\eta\hat{L}^{\circ}_{n,b_n}(t, \eta_{b_n}(t)) - (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t) \otimes \nabla_\theta\ell(\tilde{Z}_i(i/n), \theta(i/n)) \\
&= (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t) \otimes \Big\{\int_0^1 \nabla^2_\theta\ell(\tilde{Z}_i(i/n), \theta(i/n) + sr_n(t))ds \cdot r_n(t)\Big\}.
\end{aligned} \tag{A.45}$$

Since $\nabla^2_\theta\ell \in \mathcal{H}(M_y, M_x, \chi, \bar{C})$, we obtain with Lemma A.1 for $|i/n-t| \le b_n$:

$$\|\nabla^2_\theta\ell(\tilde{Z}_i(i/n), \theta(i/n) + sr_n(t)) - \nabla^2_\theta\ell(\tilde{Z}_i(t), \theta(t))\|_1 = O(b_n + n^{-1}). \tag{A.46}$$

Using (A.45), $\mathbb{E}\nabla_\theta\ell(\tilde{Z}_i(i/n), \theta(i/n)) = 0$ and (A.46), we obtain

$$\begin{aligned}
&\mathbb{E}\nabla_\eta\hat{L}^{\circ}_{n,b_n}(t, \eta_{b_n}(t)) \\
&= (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t) \otimes \Big\{\mathbb{E}\nabla^2_\theta\ell(\tilde{Z}_i(t), \theta(t)) \cdot \theta''(t)\frac{(i/n-t)^2}{2}\Big\} + O(b_n^3 + n^{-1}) \\
&= \begin{pmatrix} b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) \\ 0 \end{pmatrix} + O(b_n^3 + n^{-1} + (nb_n)^{-1}).
\end{aligned} \tag{A.47}$$

Under Assumption 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 (A.45). We have

$$\begin{aligned}
&\nabla_\eta \hat{L}^\circ_{n,b_n}(t,\eta_{b_n}(t)) - (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\nabla_\theta \ell(\tilde{Z}_i(i/n),\theta(i/n))\\
&= (nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes \nabla_\theta^2 \ell(\tilde{Z}_i(i/n),\theta(i/n)) r_n(t)\\
&\quad +(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\Big\{\int_0^1 \nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n)+sr_n(t))\\
&\qquad\qquad -\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n))ds\cdot r_n(t)\Big\}.
\end{aligned}\tag{A.48}$$

Since $\nabla_\theta^2\ell\in\mathcal{H}(M_y,M_x,\chi,\bar{C})$, we have by Lemma A.1:

$$\big\|\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n)+sr_n(t))-\nabla_\theta^2\ell(\tilde{Z}_i(i/n),\theta(i/n))\big\|_1 = O(r_n(t)) = O(|i/n-t|^2).\tag{A.49}$$

This shows that the expectation of the second summand in (A.48) is $O(b_n^4)$. We now discuss the first term in (A.48). Put $v_i(t) := \nabla_\theta^2\ell(\tilde{Z}_i(t),\theta(t))$. By Assumption 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

$$\mathbb{E}\big[(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes v_i(t) r_n(t)\big] = \begin{pmatrix} b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t)\\ b_n^3\frac{\mu_{K,4}}{6}V(t)\theta^{(3)}(t)\end{pmatrix} + O(n^{-1}+b_n^4),\tag{A.50}$$

since K has bounded variation and $\int K(x)x^3dx = 0$ by symmetry. Similarly,

$$\begin{aligned}
&\mathbb{E}\big[(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes \partial_t v_i(t) r_n(t)\big]\\
&= \begin{pmatrix} 0\\ b_n^3\frac{\mu_{K,4}}{2}\mathbb{E}[\partial_t\nabla_\theta^2\ell(\tilde{Z}_0(t),\theta(t))]\theta''(t)\end{pmatrix} + O(n^{-1}+b_n^4).
\end{aligned}\tag{A.51}$$

Finally, by Assumption 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 A.2 applied to $g=\nabla_\theta^2\ell$:

$$\|\partial_t v_i(t+s(i/n-t)) - \partial_t v_i(t)\|_1 = O(|i/n-t|).\tag{A.52}$$

The results (A.50), (A.51) and (A.52) 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 (A.48) gives the result. □

LEMMA A.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 3.4, 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 2.1(A1) in connection with 2.1(A3). Put

$$\begin{aligned} &\Pi_n(t) \\ := \ &(nb_n)^{-1}\sum_{i=1}^n U_{i,n}(t)\otimes\big\{[\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(t)+(i/n-t)\theta'(t)) - \nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))] \\ &\qquad -\mathbb{E}[\nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(t)+(i/n-t)\theta'(t)) - \nabla_\theta\ell(\tilde{Z}_i(i/n),\theta(i/n))]\big\}. \end{aligned}$$

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. θ, 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 A.4(iii), namely we use a chaining argument similar to (A.38) 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

$$\sup_{\theta\in\Theta}|\nabla_\theta^2 \ell(z,\theta)| \le C R_{M_y,M_x}(z)$$

and

$$\sup_{\theta\neq\theta'}\frac{|\nabla_\theta^2\ell(z,\theta)-\nabla_\theta^2\ell(z,\theta')|}{|\theta-\theta'|_1} \le CR_{M_y,M_x}(z).$$

This implies

$$\begin{aligned} |M_i(t,i/n) - M_i(t',i/n)| \ &\le\ 2C(|\theta(t)-\theta(t')|_1 + 2|\theta'(t)-\theta'(t')|_1 \\ &\qquad + |t-t'|\sup_s|\theta'(s)|_1)R_{M_y,M_x}(\tilde{Z}_i(i/n)) \\ &\le\ \tilde{C}|t-t'|R_{M_y,M_x}(\tilde{Z}_i(i/n)), \end{aligned}$$

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 $|M_i^{(2)}(t, i/n) - M_i^{(2)}(t', i/n)| \le \tilde{C}_2 b_n^{-1}|t - t'|R_{M_y,M_x}(\tilde{Z}_i(i/n))$. This shows with some constant $\tilde{C}_3 > 0$ that

$$|\Pi_n(t) - \Pi_n(t')| \le \tilde{C}_3 b_n^{-2}|t - t'|n^{-1}\sum_{i=1}^{n}\{R_{M_y,M_x}(\tilde{Z}_i(i/n)) + \mathbb{E}R_{M_y,M_x}(\tilde{Z}_i(i/n))\},$$

a similar result as in (A.41). Note that $\|R_{M_y,M_x}(\tilde{Z}_i(i/n))\|_1 \le (1 + D(|\chi| + 1))^M$ by using similar arguments as in the proof of Lemma A.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:

$$\mathbb{P}\big(\sup_{|t-t'|\le r^{-1}} |\Pi_n(t) - \Pi_n(t')| > Q\beta_n b_n^2/2\big) = O\big(\frac{b_n^{-2}r^{-1}}{\beta_n b_n^2}\big),$$

which converges to 0. Choose $\alpha = 1/2$. By Assumption 2.1(A7) and Lemma A.3(iii), we have $\sup_u \Delta_4^{\sup_t |M^{(2)}(t,u)|}(k) = O(k^{-\gamma})$. Thus

$$\begin{aligned} W_{4,\alpha} &:= \sup_{u\in[0,1]}\sup_{t\in[0,1]}\|\sup_{t,\eta}|M_i^{(2)}(t,u)|\|_{4,\alpha} = \sup_{m\ge 0}(m+1)^{\alpha}\Delta_4^{\sup_t |M^{(2)}(t,u)|}(m) \\ &= O(b_n^2) \end{aligned} \tag{A.53}$$

(independent of n) and

$$\begin{aligned} W_{2,\alpha} &:= \sup_{t,u}\|M_i^{(2)}(t,u)\|_{2,\alpha} = \sup_{m\ge 0}(m+1)^{\alpha}\sup_{u\in[0,1]}\sup_t \Delta_2^{M^{(2)}(t,u)}(m) \\ &= O(b_n^2) \end{aligned} \tag{A.54}$$

(independent of n). We now apply Theorem 6.2 of [54] (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 $l = 1 \vee \#(\mathcal{T}_{n,t}) \le 5\log(n)$. For Q large enough, we obtain with some constant C_α only depending on α:

$$\begin{aligned} &\mathbb{P}\big(\sup_{t'\in\mathcal{T}_r}|\Pi_n(t')| \ge Q\beta_n b_n^2/2\big) \\ \le\ & \frac{C_\alpha n\cdot l^2 W_{4,\alpha}^4}{(Q/2)^4(\beta_n b_n^2(nb_n))^4} + C_\alpha \exp\Big(-\frac{C_\alpha (Q/2)^2(\beta_n b_n^2(nb_n))^2}{nW_{2,\alpha}^2}\Big) \\ \lesssim\ & \frac{n\log(n)^2}{(nb_n)^2 b_n^{-2}\log(n)^2} + \exp\big(-\frac{(nb_n)b_n^{-1}\log(n)}{n}\big) \\ \lesssim\ & n^{-1} \to 0, \end{aligned}$$

which finishes the proof. □

A.2. Proofs and Lemmas for the SCB. From Lemma 1 in [56], we adopt the following result:

LEMMA A.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*

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{1}{\sigma_{\hat{K},0}\sqrt{nb_n}} \sup_{t\in\mathcal{T}_n} |F_n(t)| - B_{\hat{K}}(m^*) \le \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp\{-2\exp(-u)\}. \tag{A.55}$$

where $B_{\hat{K}}$ is defined in (3.10).

The following lemma is an analogue of Lemma 2 in [56]. Since we use other Gaussian approximation rates from Theorem 3.5, we shortly state the proof for completeness.

LEMMA A.9. *Let the assumptions and notations from Theorem 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

$$\lim_{n\to\infty} \mathbb{P}\Big(\frac{\sqrt{nb_n}}{\sigma_{\hat{K},0}} \sup_{t\in\mathcal{T}_n} \Big|\Sigma_{\tilde{h}}^{-1}(t)D_{\tilde{h}}(t)\Big| - B_{\hat{K}}(m^*) \le \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)), \tag{A.56}$$

PROOF OF LEMMA A.9. By Theorem 3.5 and summation-by-parts, there exist i.i.d. $V_i \sim N(0, I_{s\times s})$ such that

$$\sup_{t\in\mathcal{T}_n} |D_{\tilde{h}}(t) - \Xi(t)| = O_{\mathbb{P}}\Big(\frac{n^{\frac{1+\gamma}{1+4\gamma}}\log(n)^{3/2}}{nb_n}\Big) = O_{\mathbb{P}}\Big(\frac{\log(n)^2\big(b_n n^{\frac{2\gamma-1}{1+4\gamma}}\big)^{-1/2}}{(nb_n)^{1/2}\log(n)^{1/2}}\Big), \tag{A.57}$$

where $\Xi(t) = (nb_n)^{-1}\sum_{i=1}^{n} K_{b_n}(i/n-t)\Sigma_{\tilde{h}}(i/n)V_i$. Here, (A.57) 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 A.1, we can use a standard chaining argument in t (as it was done in Lemma A.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

$$\begin{aligned}
&\sup_{t\in\mathcal{T}_n} |\Xi(t) - (nb_n)^{-1}\Sigma_{\tilde{h}}(t)\sum_{i=1}^{n} K_{b_n}(i/n-t)V_i| \\
&= \sup_{t\in\mathcal{T}_n} \big|(nb_n)^{-1}\sum_{i=1}^{n} K_{b_n}(i/n-t)(\Sigma_{\tilde{h}}(i/n) - \Sigma_{\tilde{h}}(t))V_i\big| \\
&= O_{\mathbb{P}}\Big(\frac{b_n\log(n)}{(nb_n)^{1/2}}\Big) = O_{\mathbb{P}}\Big(\frac{b_n\log(n)^{3/2}}{(nb_n)^{1/2}\log(n)^{1/2}}\Big), \qquad \text{(A.58)}
\end{aligned}$$

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 A.8 in view of (A.57) and (A.58). □

PROOF OF THEOREM 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 A.7 (note that the rates in (A.53) and (A.54) then change to $O(b_n)$ instead of $O(b_n^2)$), we can show that

$$\sup_{t\in\mathcal{T}_n} |\Omega_C(t) - A_C(t)^{\mathsf{T}}\cdot D_{\tilde{h}}(t)| = O_{\mathbb{P}}(\beta_n b_n) = O_{\mathbb{P}}\Big(\frac{b_n^{1/2}\log(n)}{(nb_n)^{1/2}\log(n)^{1/2}}\Big), \qquad \text{(A.59)}$$

which is $o_{\mathbb{P}}((nb_n)^{-1/2}\log(n)^{-1/2})$ since $b_n\log(n)^2 \to 0$.

$h_i := (A_C(i/n)^{\mathsf{T}}\tilde{h}_i(i/n))_{i=1,\dots,n}$ is a locally stationary process with long-run variance $\Sigma_h^2(t) = \Sigma_C^2(t)$. By the result of Lemma A.9, we have that

$$\lim_{n\to\infty}\mathbb{P}\Big(\frac{\sqrt{nb_n}}{\sigma_{\hat{K},0}}\sup_{t\in\mathcal{T}_n}\big|\Sigma_C^{-1}(t)\Omega_C(t)\big| - B_{\hat{K}}(m^*) \le \frac{u}{\sqrt{2\log(m^*)}}\Big) = \exp(-2\exp(-u)). \qquad \text{(A.60)}$$

(i) By Theorem 3.4(i), we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}_n}\big|V(t)\{\hat{\theta}_{b_n}(t) - \theta(t)\} - b_n^2\frac{\mu_{K,2}}{2}V(t)\theta''(t) - D_{\tilde{h}}(t)\big| \\
&= O_{\mathbb{P}}\big(b_n^3 + (nb_n)^{-1}b_n^{-1/2}\log(n)^{3/2} + (nb_n)^{-1/2}b_n\log(n)\big) \\
&= O_{\mathbb{P}}\Big(\frac{(nb_n^7\log(n))^{1/2} + (nb_n^2\log(n)^{-4})^{-1/2} + b_n\log(n)^{3/2}}{(nb_n)^{1/2}\log(n)^{1/2}}\Big), \qquad \text{(A.61)}
\end{aligned}$$

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 (A.59) and (A.60) (with $\hat{K} = K$), this implies (3.8).

(ii) By Theorem 3.4(ii), we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}_n}\left|\mu_{K,2}V(t)b_n\{\widehat{\theta}'_{b_n}(t)-\theta'(t)\}-b_n^3\frac{\mu_{K,4}}{2}V(t)\mathrm{bias}(t)-D_{\tilde{h}}(t)\right|\\
=\ &O_{\mathbb{P}}\big(b_n^4+(nb_n)^{-1}b_n^{-1/2}\log(n)^{3/2}+(nb_n)^{-1/2}b_n\log(n)\big)\\
=\ &o_{\mathbb{P}}((nb_n)^{-1/2}\log(n)^{-1/2}),
\end{aligned}$$

as above. Together with (A.59) and (A.60) (with $\hat{K}(x)=K(x)x$), this implies (3.9). □

A.3. Proof of Section 4.

Proof of Proposition 4.1. (i) By $nb_n^2\log(n)^{-2d_\Theta}\to\infty$, Lemma A.4(i),(ii), Lemma A.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

$$\begin{aligned}
\sup_{t\in\mathcal{T}_n}|\hat{V}_{b_n}(t)-V(t)| \leq\ &\sup_{t\in\mathcal{T}_n,\eta\in E_n}|G_n^c(t,\eta)-\hat{G}_n(t,\eta)|\\
&+\sup_{t\in\mathcal{T}_n,\eta\in E_n}|\hat{G}_n(t,\eta)|\\
&+\sup_{t\in\mathcal{T}_n,\eta\in E_n}|\mathbb{E}\hat{B}_n(t,\eta)-V^\circ(t,\eta)|\\
&+\sup_{t\in\mathcal{T}_n}|V^\circ(t,\hat{\eta}_{b_n})-V(t)|\\
=\ &O_{\mathbb{P}}((nb_n)^{-1})+o_{\mathbb{P}}(\log(n)^{-1})+O(b_n)+\sup_{t\in\mathcal{T}_n}|V^\circ(t,\hat{\eta}_{b_n})-V(t)|.
\end{aligned}\tag{A.62}$$

By using Lemma A.4(ii) instead of Lemma A.4(iii), we obtain similar as in the proof of Theorem 3.4(i) that (A.16) 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 A.4(i),(ii) and Lemma A.5 with $g=\nabla_\theta\ell\in\mathcal{H}(M_y,M_x,\chi,\bar{C})$, we obtain

$$\begin{aligned}
\sup_{t\in\mathcal{T}_n}|\nabla_\eta L_{n,b_n}^{\circ,c}(t,\eta_{b_n}(t))| \leq\ &\sup_{t\in\mathcal{T}_n,\eta\in E_n}|G_n^c(t,\eta)-\hat{G}_n(t,\eta)|+\sup_{t\in\mathcal{T}_n,\eta\in E_n}|\hat{G}_n(t,\eta)|\\
&+\sup_{t\in\mathcal{T}_n,\eta\in E_n}|\mathbb{E}\hat{B}_n(t,\eta)|\\
=\ &O_{\mathbb{P}}((nb_n)^{-1})+o_{\mathbb{P}}(\log(n)^{-1})+O(b_n)
\end{aligned}$$

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 A.1, the result follows from (A.62) and $b_n\log(n)\to 0$.

(ii) follows similarly due to $\nabla_\theta\ell\cdot\nabla_\theta\ell^\mathsf{T}\in\mathcal{H}(2M_y,2M_y,\chi,\tilde{C})$ with some $\tilde{C}>0$.

□

To prove Theorem 4.2, we adopt the methods used in [56]. 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 $\Phi_i := Q_i Q_i^{\mathsf{T}}/(2m+1)$. Recall that τ_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 A.10. *Suppose that Assumption 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 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 A.10. Let $\tilde{D}_i(t) := \nabla_\theta \ell(\tilde{Z}_i(t), \theta(t))$. By Lemma A.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 A.1, $\sup_{t\in[0,1]} \|\tilde{D}_0(t)\|_4 < \infty$. It is easily seen by Lemma A.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 [56]. The proof therefore is completely the same as the proof of Theorem 4 in [56] 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 A.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 4.2. We follow the steps in the proof of Theorem 5 in [56]. 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 A.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

$$\sup_{i=1,\dots,n}\sup_{\theta\in\Theta}|\nabla^2_\theta \ell(\tilde{Z}_i(i/n),\theta)|_\infty = O_{\mathbb{P}}(n^{1/4}). \tag{A.63}$$

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

$$\begin{aligned}
&\|\sup_{t\in\mathcal{I}_n}|\tilde{\Lambda}(t)-\Lambda^{\#}(t)|\|_1\\
\le\;& \sup_{t\in\mathcal{I}_n}\Big(\sum_{i=1}^n K_{\tau_n}(i/n-t)\Big)^{-1}\cdot|K|_\infty\\
&\quad\times\sum_{i=(m+1)n+\tau_n}^{n}\sup_{j=-m,\dots,m}\big\{\|\tilde{D}_{i+j}-D^{\#}_{i+j}\|_2\|\tilde{D}_{i+j}\|_2+\|\tilde{D}_{i+j}-D^{\#}_{i+j}\|_2\|D^{\#}_{i+j}\|_2\big\}.
\end{aligned}$$

By the results of Lemma A.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 (A.63) and the fact that $\hat{\theta}_{b_n}(i/n)-\theta(i/n)=O_{\mathbb{P}}(\beta_n')$ from (A.22), we have

$$\sup_{i/n\in\mathcal{I}_n}|D_i-\tilde{D}_i|\le\sup_{i/n\in\mathcal{I}_1}|\nabla^2_\theta\ell(\tilde{Z}_i(i/n),\bar{\theta}(i/n))|\cdot|\hat{\theta}_{b_n}(i/n)-\theta(i/n)|=n^{1/4}\beta_n'. \tag{A.64}$$

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 (A.19) in this context, we obtain

$$\sup_{i/n\in\mathcal{I}_n}\frac{1}{2m+1}|Q_i|=O_{\mathbb{P}}((n\tilde{b}_n)^{-1/2}\log(n))=O_{\mathbb{P}}(m^{1/2}\log(n)). \tag{A.65}$$

Comparing Φ_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 (A.64) and (A.65), 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 A.10 now imply the assertion. □

A.4. Proofs of the examples in Section 5.

PROOF OF EXAMPLE 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 [12] the process $(Y_i)_{i=1,\dots,n}$ described by (5.1) exists and fulfills Assumption 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

$$\Big\| \sup_{t\in[0,1]} \sup_{y\neq y'} \frac{|G_{\zeta_0}(y,t) - G_{\zeta_0}(y',t)|}{|y-y'|_{\chi,1}} \Big\|_q \leq 1 \tag{A.66}$$

and

$$\sup_{t\in[0,1]} \|C(\tilde{X}_t(t))\|_q < \infty, \quad C(y) := \sup_{t\neq t'} \frac{\|G_{\zeta_0}(y,t) - G_{\zeta_0}(y,t')\|_q}{|t-t'|}, \tag{A.67}$$

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

$$|\mu(y,\theta) - \mu(y',\theta)| \leq \sum_{i=1}^k |\alpha_i||y-y'|_{\kappa_{i\cdot},1} \leq |y-y'|_{\kappa^{(\mu)}(\alpha),1}, \tag{A.68}$$

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.

$$|\sigma(y,\theta) - \sigma(y',\theta)| \leq |y-y'|_{\chi^{(\sigma)}(\beta),1}, \tag{A.69}$$

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 (A.66). Proposition 4.3 from [12] 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 θ with constant L_θ, we have

$$|\mu(y,\theta(t)) - \mu(y,\theta(t'))| \leq L_\theta |t-t'| \sum_{i=1}^{k} |m_i(y)|, \tag{A.70}$$

and

$$\begin{aligned} |\sigma(y,\theta(t))^2 - \sigma(y,\theta(t'))^2| &\leq L_\theta|t-t'| \sum_{i=0}^{l} \sqrt{\nu_i(y)} \frac{1}{2\beta_{min}^{1/2}} \big(\sqrt{\beta_i(t)\nu_i(y)} + \sqrt{\beta_i(t')\nu_i(y)}\big) \\ &\leq \frac{L_\theta}{2\beta_{min}^{1/2}} |t-t'| \sum_{i=0}^{l} \sqrt{\nu_i(y)} (\sigma(y,\theta(t)) + \sigma(y,\theta(t'))), \end{aligned}$$

which shows that

$$|\sigma(y,\theta(t)) - \sigma(y,\theta(t'))| \leq \frac{L_\theta}{2\beta_{min}^{1/2}} \sum_{i=0}^{l} \sqrt{\nu_i(y)}. \tag{A.71}$$

Note that (5.2) implies

$$m_i(y), \sqrt{\nu_i(y)} \leq C_1|y|_1 + C_2,$$

with some constants $C_1, C_2 > 0$. By (A.70), (A.71), we have for $t \neq t'$

$$\begin{aligned} \|G_{\zeta_0}(y,t) - G_{\zeta_0}(y,t')\|_q &\leq |\mu(y,\theta(t)) - \mu(y,\theta(t'))| + \|\zeta_0\|_q |\sigma(y,\theta(t)) - \sigma(y,\theta(t'))| \\ &\leq C_3|t-t'|\big(1+|y|_1\big), \end{aligned}$$

with some constant $C_3 > 0$. Since $\sup_t \|\tilde{Y}_0(t)\|_q < \infty$, (A.67) follows.

We now inspect the properties of the function ℓ. 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 $\mathbb{E}\tilde{Y}_0(t)^2 = \mathbb{E}\mu(\tilde{X}_0(t),\theta(t))^2 + \mathbb{E}\sigma(\tilde{X}_0(t),\theta(t))^2 \geq \beta_{min}\nu_{min} > 0$. Furthermore, for $L(t,\theta) := \mathbb{E}\ell(\tilde{Z}_0(t),\theta)$ it holds that

$$\begin{aligned} L(t,\theta) - L(t,\theta(t)) &= \mathbb{E}\Big(\frac{\mu(\tilde{X}_0(t),\theta) - \mu(\tilde{X}_0(t),\theta(t))}{\sigma(\tilde{X}_0(t),\theta)}\Big)^2 \\ &\quad + \mathbb{E}\Big[\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2} - \log\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2} - 1\Big]. \end{aligned} \tag{A.72}$$

In the following we use the notation $|x|_A^2 := x^\mathsf{T} A x$ for a weighted vector norm. Note that

$$\mathbb{E}\Big(\frac{\mu(\tilde{X}_0(t),\theta)-\mu(\tilde{X}_0(t),\theta(t))}{\sigma(\tilde{X}_0(t),\theta)}\Big)^2 \geq c_0|\alpha-\alpha(t)|^2_{M_1(t)}, \tag{A.73}$$

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

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2} - \log\frac{\sigma(\tilde{X}_0(t),\theta(t))^2}{\sigma(\tilde{X}_0(t),\theta)^2} - 1\Big]\\
\geq\ & \frac{1}{2}\mathbb{E}\Big[\frac{(\sigma(\tilde{X}_0(t),\theta)^2-\sigma(\tilde{X}_0(t),\theta(t))^2)^2}{(\sigma(\tilde{X}_0(t),\theta)^2-\sigma(\tilde{X}_0(t),\theta(t))^2)^2+\sigma(\tilde{X}_0(t),\theta)^4}\Big]\\
\geq\ & \frac{c_0}{10}|\beta-\beta(t)|^2_{M_2(t)}, \qquad \text{(A.74)}
\end{aligned}$$

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 (A.72), (A.73) and (A.74) we conclude that $\theta\mapsto L(t,\theta)$ is uniquely minimized in $\theta=\theta(t)$. This shows 2.1(A3).

Omitting the arguments $z=(y,x)$ and θ, we have

$$\begin{aligned}
\ell &= \frac{1}{2}\Big[\frac{(y-\langle\alpha,m\rangle)^2}{\langle\beta,\nu\rangle} + \log\langle\beta,\nu\rangle\Big], && \text{(A.75)}\\
\nabla_\theta\ell &= -\frac{\nabla_\theta m}{\sigma}\Big(\frac{y-m}{\sigma}\Big) + \frac{\nabla_\theta(\sigma^2)}{2\sigma^2}\Big[1-\Big(\frac{y-m}{\sigma}\Big)^2\Big]\\
&= \begin{pmatrix} -\frac{m}{\sigma}\Big(\frac{y-m}{\sigma}\Big)\\ \frac{\nu}{2\sigma^2}\big[1-\big(\frac{y-m}{\sigma}\big)^2\big]\end{pmatrix} = \begin{pmatrix}\frac{m}{\langle\beta,\nu\rangle}(y-\langle\alpha,m\rangle)\\ \frac{\nu}{2\langle\beta,\nu\rangle}\big(1-\frac{(y-\langle\alpha,m\rangle)^2}{\langle\beta,\nu\rangle}\big)\end{pmatrix}, && \text{(A.76)}\\
\nabla_\theta^2\ell &= \frac{\nabla_\theta m\nabla_\theta m^\mathsf{T}}{\sigma^2} + \big(\frac{y-m}{\sigma}\big)\cdot\Big[\frac{\nabla_\theta m\nabla_\theta(\sigma^2)^\mathsf{T}+\nabla_\theta(\sigma^2)\nabla_\theta m^\mathsf{T}}{\sigma^3} - \frac{\nabla_\theta^2 m}{\sigma}\Big]\\
&\qquad + \frac{\nabla_\theta^2(\sigma^2)}{2\sigma^2}\Big[1-\big(\frac{y-m}{\sigma}\big)^2\Big] + \frac{\nabla_\theta(\sigma^2)\nabla_\theta(\sigma^2)^\mathsf{T}}{2\sigma^4}\Big[2\Big(\frac{y-m}{\sigma}\Big)^2-1\Big]\\
&= \begin{pmatrix}\frac{mm^\mathsf{T}}{\sigma^2} & \frac{y-m}{\sigma^2}\cdot m\nu^\mathsf{T}\\ \frac{y-m}{\sigma^2}\cdot\nu m^\mathsf{T} & \frac{\nu\nu^\mathsf{T}}{2\sigma^4}\big[2\big(\frac{y-m}{\sigma}\big)^2-1\big]\end{pmatrix}\\
&= \begin{pmatrix}\frac{mm^\mathsf{T}}{\langle\beta,\nu\rangle} & \frac{y-\langle\alpha,m\rangle}{\langle\beta,\nu\rangle^2}\cdot m\nu^\mathsf{T}\\ \frac{y-\langle\alpha,m\rangle}{\langle\beta,\nu\rangle^2}\cdot\nu m^\mathsf{T} & \frac{\nu\nu^\mathsf{T}}{2\langle\beta,\nu\rangle^2}\big[2\frac{(y-\langle\alpha,m\rangle)^2}{\langle\beta,\nu\rangle}-1\big]\end{pmatrix}. && \text{(A.77)}
\end{aligned}$$

Since ζ_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 μ, σ in the following):

$$V(t) = \mathbb{E}\nabla_\theta^2 \ell(\tilde{Z}_0(t), \theta(t)) = \begin{pmatrix} \mathbb{E}\big[\frac{mm^\mathsf{T}}{\langle \beta, \nu\rangle}\big] & 0 \\ 0 & \mathbb{E}\big[\frac{\nu\nu^\mathsf{T}}{2\langle \beta, \nu\rangle^2}\big]\end{pmatrix}.$$

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 2.1(A4) is fulfilled.

A careful inspection of (A.75), (A.76) and (A.77) 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 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 2.2(B1) is fulfilled. We use results from Section 4 in [12] 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,

$$\begin{aligned}
|\langle \beta(t), \nu(x) - \nu(x')\rangle| &\leq \sum_{i=1}^{l} \beta_i(t) \big|\sqrt{\nu_i(x)} - \sqrt{\nu_i(x')}\big| \cdot \big|\sqrt{\nu_i(x)} + \sqrt{\nu_i(x')}\big| \\
&\leq |\beta(t)|_\infty |x-x'|_1 \sum_{i=1}^{l} \big(\sqrt{\nu_i(x)} + \sqrt{\nu_i(x')}\big).
\end{aligned}$$

Since each component of $\beta(t)$ is lower bounded by β_{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,

$$|\partial_{x_j} G_{\zeta_0}(x,t) - \partial_{x_j}(x',t)| \leq C(1+|\zeta_0|)|x-x'|_1.$$

With similar but simpler arguments we obtain that

$$|\partial_t G_{\zeta_0}(x,t) - \partial_t G_{\zeta_0}(x',t)| \le C(1+|\zeta_0|)|x-x'|_1$$

and

$$\frac{|\partial_{x_j} G_{\zeta_0}(x,t) - \partial_{x_j} G_{\zeta_0}(x,t')|}{|t-t'|\cdot|x|_1} \le C(1+|\zeta_0|), \qquad \frac{|\partial_t G_{\zeta_0}(x,t) - \partial_t G_{\zeta_0}(x,t')|}{|t-t'|\cdot|x|_1} \le C(1+|\zeta_0|).$$

By Theorem 4.8 and Proposition 4.11 in [12], we obtain 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 2.2(B2). □

Proof of Example 5.2. Let $q = 4M$, and $M = 4$. Fix $t \in [0,1]$. Consider the recursion of the corresponding stationary approximation

$$\begin{aligned} \tilde{Y}_i(t) &= \tilde{\sigma}_i(t)^2\zeta_i^2, \\ \tilde{\sigma}_i(t)^2 &= \alpha_0(t) + \sum_{j=1}^m \alpha_j(t)\tilde{Y}_{i-j}(t) + \sum_{j=1}^l \beta_j(t)\tilde{\sigma}_{i-j}(t)^2. \end{aligned} \tag{A.78}$$

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 [49], the model (A.78) admits the representation

$$\tilde{P}_i(t) = M_i(t)\tilde{P}_{i-1}(t) + a_i(t). \tag{A.79}$$

In Theorem 2.1 [35], 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 $D > 0$.

(A.79) implies the explicit representation

$$\tilde{P}_i(t) = \sum_{k=0}^\infty \Big(\prod_{j=0}^{k-1} M_{i-j}(t)\Big) a_{i-k}(t). \tag{A.80}$$

We therefore have for $t, t' \in [0,1]$:

$$\begin{aligned}\|\tilde{P}_i(t) - \tilde{P}_i(t')\|_q &\leq \sum_{k=0}^{\infty}\sum_{l=0}^{k-1}\Big(\prod_{0\leq j<l}\|M_{i-j}(t)\|_q\Big)\|M_{i-l}(t) - M_{i-l}(t')\|_q \\ &\qquad\qquad \times\Big(\prod_{l<j\leq k-1}\|M_{i-j}(t)\|_q\Big)\cdot\|a_{i-k}(t)\|_q \\ &\quad + \sum_{k=0}^{\infty}\Big(\prod_{j=0}^{k-1}\|M_{i-j}(t)\|_q\Big)\|a_{i-k}(t) - a_{i-k}(t')\|_q. \end{aligned} \tag{A.81}$$

Note that $\sup_{t\in[0,1]}\lambda_{max}(\|M_0(t)\|_q) < 1$. By Lipschitz continuity of $\theta(\cdot)$, we have $\|a_0(t) - a_0(t')\|_q = |\alpha_0(t) - \alpha_0(t')|(\|\zeta_0^2\|_q, 0, \dots, 0, 1, 0, \dots, 0)^{\mathsf{T}} = O(|t-t'|)$ and $\|M_0(t) - M_0(t')\|_q = (\|\zeta_0^2\|_q|f(\theta(t)) - f(\theta(t'))|, 0, \dots, 0, |f(\theta(t)) - f(\theta(t'))|, 0, \dots, 0)^{\mathsf{T}} = O(|t-t'|)$. We conclude from the first component of (A.81), that for all $t, t' \in [0,1]$:

$$\|\tilde{Y}_i(t) - \tilde{Y}_i(t')\|_q \leq C\cdot|t-t'|, \tag{A.82}$$

with some constant $C > 0$.

Put $P_i = (Y_i, \dots, Y_{i-m+1}, \sigma_i^2, \dots, \sigma_{i-l+1}^2)^{\mathsf{T}}$. Similarly to (A.79), we have

$$P_i = M_i(i/n)P_{i-1} + a_i(i/n), \quad i = 1, \dots, n. \tag{A.83}$$

Note that i iterations of (A.83) 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 (A.82), we conclude from the first component that $\|Y_i - \tilde{Y}_i(i/n)\|_q = O(n^{-1})$. This shows Assumption 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 [48] (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 2.1(A7).

Let $\Sigma(x,\theta) := (\sigma(x,\theta)^2, \ldots, \sigma(x_{(l-1)\to},\theta)^2)^\intercal$ and $A(x,\theta) := (\alpha_0 + \sum_{j=1}^m \alpha_j x_j, \ldots, \alpha_0 + \sum_{j=1}^m \alpha_j x_{j+l-1})^\intercal$, and

$$B(\theta) = \begin{pmatrix} \beta_1 & \ldots & \ldots & \ldots & \beta_l \\ 1 & 0 & \ldots & \ldots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}.$$

As said in Theorem 2.1 in [35], $\lambda_{max}(\mathbb{E}M_0(\theta)^{\otimes q})$ is a necessary and sufficient condition for the corresponding GARCH process with parameters θ to have 8-th moments. As a consequence, $\lambda_{max}(\mathbb{E}M_0(\theta)) < 1$ which by Proposition 1 in [19] implies $\lambda_{max}(B(\theta)) < 1$. It holds that

$$\Sigma(x,\theta) = A(x,\theta) + B(\theta)\cdot\Sigma(x_{1\to},\theta). \tag{A.84}$$

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 Θ due to $\lambda_{max}(B(\theta)) < 1$. By Banach's fix point theorem, we conclude that

$$\sigma(x,\theta)^2 = \sum_{k=0}^{\infty} \big(B(\theta)^k A(x_{k\to},\theta)\big)_1. \tag{A.85}$$

Since $A(0,\theta) = (\alpha_0, \ldots, \alpha_0)^\intercal$, 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 (A.85), we conclude that

$$\begin{aligned} |\sigma(x,\theta)^2 - \sigma(x',\theta)^2| &\le \Big|\sum_{k=0}^{\infty} \big(B(\theta)^k\{A(x_{k\to},\theta) - A(x'_{k\to},\theta)\}\big)_1\Big| \\ &\le \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=0}^{\infty} (B(\theta)^k)_{1i}\beta_j |x_{k+j+i-1} - x'_{k+j+i-1}| \\ &\le |x - x'|_{\tilde{\chi},1}, \end{aligned} \tag{A.86}$$

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 (A.85) with geometrically decaying summands, it is easy to see that $\sigma(x,\theta)^2$

is four times continuously differentiable w.r.t. θ. 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 (A.86), 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

$$|\sigma(x,\theta)^2 - \sigma(x,\theta')^2| \le |\theta-\theta'|_1\cdot\sup_{\theta\in\Theta}|\nabla_\theta(\sigma(x,\theta)^2)|_\infty \le |\theta-\theta'|_1\cdot|x|_{\tilde{\chi},1}, \tag{A.87}$$

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 θ 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 (A.86) and (A.87) and similar results for their derivatives w.r.t. θ 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 χ and some $\tilde{C} > 0$, i.e. Assumption 2.1(A1) is fulfilled with $M_y = 1$ and $M_x = 4$.

It was shown in the proof of Theorem 2.1 in [19], that $\theta \mapsto L(t,\theta) = \mathbb{E}\ell(\tilde{Z}_0(t),\theta)$ is uniquely minimized in $\theta = \theta(t)$, which shows Assumption 2.1(A3). As in the proof of Example 5.1, we obtain that

$$V(t) = \mathbb{E}\Big[\frac{\nabla_\theta(\sigma(\tilde{X}_0(t),\theta(t))^2)\nabla_\theta(\sigma(\tilde{X}_0(t),\theta(t))^2)^{\mathsf{T}}}{2\sigma(\tilde{X}_0(t),\theta(t))^4}\Big] = I(t)\frac{2}{\mathbb{E}\zeta_0^4-1}.$$

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 [19] that $V(t)$ is positive definite for each $t \in [0,1]$. By continuity, we conclude that Assumption 2.1(A4) is fulfilled.

Regarding Assumption 2.2, notice that from the explicit representation (A.80) 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

$$\begin{aligned}\partial_t \tilde{P}_i(t) &= \sum_{k=0}^{\infty}\sum_{l=0}^{k-1}\Big(\prod_{0\le j<l} M_{i-j}(t)\Big)\partial_t M_{i-l}(t)\Big(\prod_{l<j\le k-1} M_{i-j}(t)\Big)a_{i-k}(t)\\ &\qquad + \sum_{k=0}^{\infty}\Big(\prod_{j=0}^{k-1} M_{i-j}(t)\Big)\partial_t a_{i-k}(t)\end{aligned}$$

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 (A.82) 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 2.2(B3) is shown.

From (A.85) 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 χ 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 χ. This shows Assumption 2.2(B2). □

PROOF OF EXAMPLE 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 (2.1) 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. (2.13) and thus Assumption 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 $w > 0$, $f''(\tilde{w}) \geq \min\{\pi(w), \pi(w_0)\}$. In the following we use the notation $|x|_A^2 := x^\mathsf{T}Ax$ 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 θ, 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 $\sup_{j=1,\dots,p}|\tilde{X}_{0j}(t)| < \infty$ a.s. and Θ 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 θ and we conclude that $L(t,\theta)$ is uniquely minimized by $\theta = \theta(t)$. This shows Assumption 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 $\chi = (1,\dots,1,0,0,\dots)$, a vector with p ones followed from zeros, i.e. Assumption 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 2.1(A4).

$\square$

Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg, Germany.
E-mail: stefan.richter@iwr.uni-heidelberg.de

Department of Statistics
University of Chicago
5747 S. Ellis Avenue
Chicago, IL 60637, USA.
E-mail: sayarkarmakar@uchicago.edu,
E-mail: wbwu@galton.uchicago.edu