Essays in Nonparametric Econometrics

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Michael Vogt

September 2011

Abteilungssprecher:	Prof. Dr. Martin Peitz
Referent:	Prof. Dr. Enno Mammen
Korreferent:	Prof. Oliver Linton, Ph.D.

Tag der Verteidigung:	26. Oktober 2011

Acknowledgements

First of all, I would like to thank my advisor Enno Mammen for all his support during my time in Mannheim. I have benefited a lot from his deep understanding of statistics. Indeed, I have always been (and still am) struck by his amazing intuition about statistical problems. I can hardly imagine receiving any better support.

I would also like to thank Oliver Linton. I experienced three very fruitful and stimulating months in London. The numerous discussions with him were very insightful and helped a lot in improving my thesis.

Kyusang Yu was an invaluable source of help when struggling with technical issues. I have profited a lot from his knowledge about additive models and statistics in general. With overwhelming probability, he could come up with good advice when I was stuck in a problem.

I am very indebted to Christopher Walsh. Working together with him on the first Ph.D. project made the start much easier for me. Not only could we share the frustration when stuck in a problem, we also had a lot of fun. Cheers, mate!

I have always appreciated the familiar atmosphere at the Chair of Statistics. I very much enjoyed working with my colleagues Christoph Breunig, Carsten Jentsch, Eun Ryung Lee, Christoph Nagel, Christoph Rothe and Melanie Schienle. Moreover, I was pleased to be part of the DFG/SNF Research Group FOR916. The conferences and lecture series of this group helped a lot in broadening my horizon in statistics. Finally, I would like to thank my fellow graduate students at the CDSE – in particular the class of 2007. They have contributed a lot towards making Mannheim a lively and pleasant environment for me.

I am especially grateful to my parents. They have always stuck by me whatever has happened. Thank you for that!

Mannheim, September 2011 *Michael Vogt*

Contents

Introduction

A wide range of financial and economic time series are likely to be nonstationary. Examples are return and volatility series as well as macroeconomic data such as GDP and interest rates. Moreover, it is often very restrictive to stipulate a parametric structure on the time series data at hand. Thus, in many financial and economic applications, we are faced with a non- or semiparametric estimation problem in a nonstationary time series setting.

To model the nonstationary behaviour of financial and economic time series, so-called locally stationary models have been proposed in recent years (see e.g. Dahlhaus & Subba Rao [8], Fryzlewicz et al. [15] or Linton & Hafner [28]). Local stationarity is a special kind of nonstationarity which was introduced in a series of papers by Dahlhaus (cf. [4], [5], and [6]). Intuitively speaking, a process is locally stationary if over short time spans, i.e. locally in time, it behaves approximately stationary. This intuitive concept can be turned into a rigorous definition in various different but related ways: A locally stationary process may be defined in terms of a time-varying spectral representation (cf. Dahlhaus [6]) or in terms of an MA(∞)-representation with time-varying coefficients (cf. Dahlhaus & Polonik [9]). Yet another way is to require that locally around each time point, the process can be approximated by a stationary process in a stochastic sense (cf. Dahlhaus & Subba Rao [8]).

Most of the locally stationary models suggested so far in the literature are of a parametric nature. Usually, parametric models are analyzed in which the coefficients are allowed to vary smoothly over time. The parametric form stipulated in these models is often ad hoc and not justified at all by a structural economic theory in the background. To avoid misleading conclusions under misspecification and to select an appropriate parametric model, non- and semiparametric approaches are required.

In this thesis, we study various non- and semiparametric estimation problems in a locally stationary time series setting. In particular, we provide asymptotic theory for a collection of non- and semiparametric models which have not been analyzed yet in the literature. The thesis consists of three chapters that are self-contained and can be read separately. Each chapter ends with an appendix that collects the proofs and technical details.

In *Chapter 1*, we introduce a nonparametric framework which is a natural extension of time series models with time-varying coefficients. Letting $Y_{t,T}$ and $X_{t,T}$ be random variables of dimension 1 and d, respectively, the model is given by

$$Y_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} \quad \text{for } t = 1, \ldots, T \tag{1}$$

with $\mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0$. Here, the model variables are locally stationary and the regression function is allowed to change smoothly over time. As usual in the literature on locally stationary processes, the function m depends on rescaled time $\frac{t}{T}$ rather than on real time t and the model variables form a triangular array rather than a sequence.[1] We introduce a kernel-based method to estimate the time-varying function and provide asymptotic theory for our estimates. Moreover, we show that the main conditions of the theory are satisfied for a large class of nonlinear autoregressive processes with a time-varying regression function. Finally, we examine structured models where the regression function splits up into time-varying additive components. As will be seen, estimation in these models does not suffer from the curse of dimensionality. The technical analysis is complemented by an application to index return data.

Chapter 2 studies a testing problem within the general framework (1). We are interested in the question whether the time-varying regression function m has the same shape at two different time points. This testing issue is not only interesting from a theoretical perspective but also from an applied point of view. In many applications, we want to find out whether the relationship between two variables is the same in two different economic situations, e.g. at a time point before a crisis and one during it. To tackle this kind of question, we propose a kernel-based L_2-test statistic. We derive the asymptotic distribution of the statistic both under the null and under local and fixed alternatives. To improve the small sample behaviour of the test, we set up a wild bootstrap procedure and derive the asymptotic properties thereof.

[1]Some remarks on the concept of rescaled time can be found in Section 3.2 of Chapter 3. A detailed discussion of the concept is given in Dahlhaus [5].

In *Chapter 3*, which is based on a joint paper with Christopher Walsh, we analyze a semiparametric multiplicative volatility model which splits up into a nonparametric part and a parametric GARCH component. The model is given by the equation

$$Y_{t,T} = \tau\Big(\frac{t}{T}, X_t\Big)\varepsilon_t \quad \text{for } t = 1, \dots, T. \tag{2}$$

Here, $Y_{t,T}$ are financial log-returns, $X_t = (X_t^1, \dots, X_t^d)$ is a vector of strictly stationary covariates, and τ is a nonparametric function of rescaled time and the variables X_t. Moreover, $\{\varepsilon_t\}$ is a strictly stationary GARCH process. Model (2) generalizes the simpler model

$$Y_{t,T} = \tau\Big(\frac{t}{T}\Big)\varepsilon_t,$$

where the function τ only depends on rescaled time. This simpler framework has for example been considered in Feng [13], a multivariate version has been analyzed in Linton & Hafner [28]. To avoid the curse of dimensionality, we impose some structural constraints on the nonparametric function τ in (2). In particular, the function is assumed to split up into multiplicative components according to

$$\tau\Big(\frac{t}{T}, X_t\Big) = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X_t^j).$$

We propose a two-step procedure to estimate the model. To estimate the multiplicative components of the τ-function, we extend the standard smooth backfitting procedure of Mammen et al. [29]. The GARCH parameters are estimated in a second step via a quasi-maximum likelihood based approach. Finally, the model is applied to S&P 500 return data using various interest rate spreads as covariates.

Chapter 1

Nonparametric Regression For Locally Stationary Time Series

1.1 Introduction

Classical time series analysis is based on the assumption of stationarity. However, many time series exhibit a nonstationary behaviour. Examples come from fields as diverse as finance, sound analysis and neuroscience.

One way to model nonstationary behaviour is provided by the theory of locally stationary processes developed in a series of papers by Dahlhaus (cf. [4], [5], and [6]). Intuitively speaking, a process is locally stationary if over short periods of time (i.e. locally in time) it behaves approximately stationary. So far, locally stationary models have been mainly considered within a parametric context. Usually, generalizations of classical parametric time series models are analyzed that allow the parameters to change smoothly over time.

There is a large amount of papers that deal with time series models with time-varying parameters. Dahlhaus et al. [7], for example, study wavelet estimation in autoregressive models with time-dependent coefficients. Chandler & Polonik [2] consider autoregressive processes with a time-varying variance and test for unimodality of the variance function. Dahlhaus & Subba Rao [8] analyze a class of ARCH models with time-varying parameters. They propose a kernel-based quasi-maximum likelihood method to estimate the parameter functions; a kernel-based normalized-least-squares method is suggested by Fryzlewicz et al. [15]. Linton & Hafner [28] provide estimation theory for a multivariate GARCH model with a time-varying unconditional variance. Finally, a diffusion process with a time-dependent drift and diffusion function is investigated in Koo & Linton [21].

In this chapter, we introduce a nonparametric framework which can be regarded as a natural extension of time series models with time-varying coefficients. In its most general form, the model is given by

$$Y_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} \quad \text{for } t = 1, \dots, T \tag{1.1}$$

with $\mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0$, where $Y_{t,T}$ and $X_{t,T}$ are random variables of dimension 1 and d, respectively. The model variables are assumed to be locally stationary and the regression function as a whole is allowed to change smoothly over time. As usual in the literature on locally stationary processes, the function m does not depend on real time t but rather on rescaled time $\frac{t}{T}$. This goes along with the model variables forming a triangular array instead of a sequence. Throughout the introduction, we stick to an intuitive concept of local stationarity. A technically rigorous definition is given in Section 1.2.

There is a wide range of interesting nonlinear time series models that fit into the general framework (1.1). An important example are nonparametric autoregressive models of the form

$$X_{t,T} = m\Big(\frac{t}{T}, X_{t-1,T}, \dots, X_{t-d,T}\Big) + \varepsilon_{t,T} \quad \text{for } t = 1, \dots, T \tag{1.2}$$

with $\mathbb{E}[\varepsilon_{t,T}|X_{t-1,T}, \dots, X_{t-d,T}] = 0$, which are analyzed in Section 1.3. As will be seen there, under certain conditions on the function m and the error terms $\varepsilon_{t,T}$, the process defined in (1.2) is locally stationary and strongly mixing. Independently to the present work, Kristensen [24] has developed results on local stationarity of the process given in (1.2) when the residuals $\varepsilon_{t,T}$ are i.i.d. In contrast to Kristensen, we do not restrict the residual process to be i.i.d. and also provide results on the mixing behaviour of the process.

In Section 1.4, we develop estimation theory for the nonparametric regression function in the general framework (1.1). As described there, the regression function is estimated by nonparametric kernel methods. We provide a complete asymptotic theory for our estimates. In particular, we derive uniform convergence rates and an asymptotic normality result. To do so, we split up the estimates into a variance part and a bias part. In order to control the variance part, we generalize results of Hansen [17] on uniform convergence rates for kernel estimates to our locally stationary setting. The locally stationary behaviour of the model variables also changes the asymptotic analysis of the bias part. In particular, it produces an additional bias term which can be regarded as measuring the deviation from stationarity.

Even though model (1.1) is theoretically interesting, it has an important drawback. Estimating the time-varying regression function in (1.1) suffers from an even more severe curse of dimensionality problem than in the standard strictly stationary setting with a time-invariant regression function. The reason is that in model (1.1), we fit a fully nonparametric function $m(u,\cdot)$ locally around *each* rescaled time point u. Compared to the standard case, this means that we additionally smooth in time direction and thus increase the dimensionality of the estimation problem by one. This makes the procedure even more data consuming than in the standard setting and thus infeasible in many applications.
In order to countervail this severe curse of dimensionality, we impose some structural constraints on the regression function in (1.1). In particular, we consider additive models of the form

$$Y_{t,T} = \sum_{j=1}^{d} m_j\Big(\frac{t}{T}, X^j_{t,T}\Big) + \varepsilon_{t,T} \quad \text{for } t = 1,\dots,T \tag{1.3}$$

with $\mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0$. In Section 1.5, we will show that the components of this model can be estimated with two-dimensional nonparametric convergence rates, no matter how large the dimension d. In order to do so, we extend the smooth backfitting approach of Mammen et al. [29] to our locally stationary setting.
To show the practical usefulness of our theory, we apply an additive volatility model with time-varying component functions to a sample of financial data in Section 1.6. The analysis makes visible how the component functions estimated at time points before and during the recent financial crisis differ from each other.

1.2 Local Stationarity

Heuristically speaking, the process $\{X_{t,T} : t = 1,\dots,T\}_{T=1}^{\infty}$ is locally stationary if it behaves approximately stationary locally in time. The next definition ensures this behaviour by requiring that for each rescaled time point u, there is a stationary process $\{X_t(u) : t \in \mathbb{Z}\}$ which approximates $\{X_{t,T}\}$ locally around u. This means that if $\frac{t}{T}$ is close to u, then $X_{t,T}$ is close to $X_t(u)$ at least in a stochastic sense.

Definition 1.1. *The process $\{X_{t,T}\}$ is locally stationary if for each time point $u \in [0,1]$ there exists an associated process $\{X_t(u)\}$ with the following two properties:*

(i) $\{X_t(u)\}$ is strictly stationary with density $f_{X_t(u)}$,

(ii) it holds that

$$\big\|X_{t,T} - X_t(u)\big\| \le \Big(\Big|\frac{t}{T} - u\Big| + \frac{1}{T}\Big) U_{t,T}(u) \quad a.s.,$$

where $\{U_{t,T}(u)\}$ is a process of positive variables satisfying $\mathbb{E}[(U_{t,T}(u))^\rho] < C$ for some $\rho > 0$ and $C < \infty$ independent of u, t, and T. $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^d$.

Since the ρ-th moments of the variables $U_{t,T}(u)$ are uniformly bounded by some $C < \infty$, it holds that $U_{t,T}(u) = O_p(1)$. As a consequence,

$$\big\|X_{t,T} - X_t(u)\big\| = O_p\Big(\Big|\frac{t}{T} - u\Big| + \frac{1}{T}\Big).$$

The constant ρ can be regarded as a measure of how well $X_{t,T}$ is approximated by $X_t(u)$: The larger ρ can be chosen, the less mass is contained in the tails of the distribution of $U_{t,T}(u)$. Thus, if ρ is large, then the bound $(|\frac{t}{T} - u| + \frac{1}{T})U_{t,T}(u)$ will take rather moderate values for most of the time. In this sense, the bound and thus the approximation of $X_{t,T}$ by $X_t(u)$ is getting better for larger ρ.

1.3 Locally Stationary Nonlinear AR Processes

In this section, we examine a large class of nonlinear autoregressive processes with a time-varying regression function that fit into the general framework (1.1). We show that these processes are locally stationary and strongly mixing under suitable conditions on the model components.

1.3.1 The tvNAR Process

We call an array $\{X_{t,T} : t \in \mathbb{Z}\}_{T=1}^{\infty}$ a time-varying nonlinear autoregressive (tv-NAR) process if $X_{t,T}$ behaves according to

$$X_{t,T} = m\Big(\frac{t}{T}, X_{t-1,T}, \dots, X_{t-d,T}\Big) + \sigma\Big(\frac{t}{T}, X_{t-1,T}, \dots, X_{t-d,T}\Big)\varepsilon_t. \qquad (1.4)$$

A tvNAR process is thus an autoregressive process of the form (1.2) with errors $\varepsilon_{t,T} = \sigma(\frac{t}{T}, X_{t-1,T}, \dots, X_{t-d,T})\varepsilon_t$. In the above definition, $m(u,x)$ and $\sigma(u,x)$ are smooth functions of rescaled time u and $x \in \mathbb{R}^d$. We stipulate that for $u \le 0$, $m(u,x) = m(0,x)$ and $\sigma(u,x) = \sigma(0,x)$. Analogously, we set $m(u,x) = m(1,x)$ and $\sigma(u,x) = \sigma(1,x)$ for $u \ge 1$. Furthermore, the variables ε_t are assumed to be i.i.d. with mean zero. For each $u \in \mathbb{R}$, we additionally define the associated process $\{X_t(u) : t \in \mathbb{Z}\}$ by

$$X_t(u) = m\big(u, X_{t-1}(u), \dots, X_{t-d}(u)\big) + \sigma\big(u, X_{t-1}(u), \dots, X_{t-d}(u)\big)\varepsilon_t, \qquad (1.5)$$

where the rescaled time argument of the functions m and σ is fixed at u.
As stipulated above, the conditional mean function m and the volatility function σ do not change over time for $t \leq 0$. Put differently,

$$X_{t,T} = m\big(0, X_{t-1,T}, \dots, X_{t-d,T}\big) + \sigma\big(0, X_{t-1,T}, \dots, X_{t-d,T}\big)\varepsilon_t \quad \text{for all } t \leq 0.$$

We can thus assume that $X_{t,T} = X_t(0)$ for $t \leq 0$. Consequently, if there exists a process $\{X_t(0)\}$ that satisfies the system of difference equations (1.5) for $u = 0$, then this immediately implies the existence of a tvNAR process $\{X_{t,T}\}$ satisfying (1.4). As will turn out, under appropriate conditions there exists a strictly stationary solution $\{X_t(u)\}$ to the system of equations (1.5) for each $u \in \mathbb{R}$, in particular for $u = 0$. We can thus take for granted that the tvNAR process $\{X_{t,T}\}$ defined by (1.4) exists.

1.3.2 Assumptions

We now list some conditions under which the tvNAR process is locally stationary and strongly mixing. To start with, the function m is supposed to satisfy the following conditions.

(M1) m is absolutely bounded by some constant $M < \infty$.

(M2) m is Lipschitz continuous with respect to rescaled time u, i.e. there exists a constant $L < \infty$ such that $|m(u,x) - m(u',x)| \leq L|u - u'|$ for all $x \in \mathbb{R}^d$.

(M3) m is continuously differentiable with respect to x. The partial derivatives $\partial_j m(u,x) := \frac{\partial}{\partial x_j} m(u,x)$ have the property that for some $K_1 < \infty$,

$$\sup_{u\in\mathbb{R}, \|x\|_\infty > K_1} |\partial_j m(u,x)| \leq \delta < 1.$$

An exact formula for the bound δ is given in (1.37) in Appendix A.

The function σ is required to fulfill analogous assumptions.

(Σ1) σ is bounded by some constant $\overline{\Sigma} < \infty$ from above and it is bounded away from zero by some constant $\underline{\Sigma} > 0$, i.e. $0 < \underline{\Sigma} \leq \sigma(u,x) \leq \overline{\Sigma} < \infty$ for all $u \in \mathbb{R}$ and $x \in \mathbb{R}^d$.

(Σ2) σ is Lipschitz continuous with respect to rescaled time u, i.e. $|\sigma(u,x) - \sigma(u',x)| \leq L|u-u'|$ for some $L < \infty$ and all $x \in \mathbb{R}^d$.

(Σ3) σ is continuously differentiable with respect to x. The partial derivatives $\partial_j \sigma(u,x) := \frac{\partial}{\partial x_j}\sigma(u,x)$ have the property that for some $K_1 < \infty$,

$$\sup_{u \in \mathbb{R}, \|x\|_\infty > K_1} |\partial_j \sigma(u,x)| \leq \delta < 1.$$

Finally, the error terms are required to have the following properties.

(E1) The variables ε_t are i.i.d. with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}|\varepsilon_t|^{1+\eta} < \infty$ for some $\eta > 0$. Moreover, they have an everywhere positive and continuous density f_ε.

(E2) The density f_ε is bounded and Lipschitz.

To show that the tvNAR process is strongly mixing, we additionally need the following condition on the densities of the error terms:

(E3) Let d_0, d_1 be any constants with $0 \leq d_0 \leq D_0 < \infty$ and $|d_1| \leq D_1 < \infty$. The density f_ε fulfills the condition

$$\int_{\mathbb{R}} |f_\varepsilon([1+d_0]x + d_1) - f_\varepsilon(x)| dx \leq C_{D_0,D_1}\big(d_0 + |d_1|\big)$$

with $C_{D_0,D_1} < \infty$ depending on the bounds D_0 and D_1.

We shortly give some remarks on the above assumptions:

- Conditions (M1) and (M3) together with (Σ1) and (Σ3) restrict the tvNAR process above all *outside* a large bounded set $\{x : \|x\| \leq K_1\}$. There, the functions m and σ are required to remain bounded and to be sufficiently flat. In a wide range of cases, the approximating processes $\{X_t(u)\}$ will exhibit a stable behaviour and will remain *within* a large bounded set for most of the time. The same will then also hold true for the process $\{X_{t,T}\}$. Therefore, the above conditions are not very severe. They only restrict the dynamics of the tvNAR process in a region to which it wanders very rarely.

- Our set of assumptions can be regarded as a strengthening of the assumptions needed to show geometric ergodicity of autoregressive processes of the form $X_t = m(X_{t-1},\ldots \ldots, X_{t-d}) + \sigma(X_{t-1},\ldots,X_{t-d})\varepsilon_t$. In particular, (M3) and (Σ3) are very close in spirit to assumptions from this context which require the mean and volatility functions m and σ not to grow too fast outside a large bounded set.

- Condition (M3) implies that the derivatives $\partial_j m(u,x)$ are absolutely bounded. Hence, there exists a constant $\Delta < \infty$ such that $|\partial_j m(u,x)| \leq \Delta$ for all $u \in \mathbb{R}$ and $x \in \mathbb{R}^d$. Similarly, (Σ3) implies that the derivatives $\partial_j \sigma(u,x)$ are absolutely bounded by some constant $\Delta < \infty$.

- As already noted, (E3) is only needed to prove that the tvNAR process is strongly mixing. It is for example fulfilled for the class of bounded densities f_ε whose first derivative f'_ε is bounded, satisfies $\int |x f'_\varepsilon(x)| dx < \infty$, and declines monotonically to zero for values $|x| > C$ for some constant $C > 0$. A proof for this can be found in the first subsection of Appendix A. (See also Section 3 in Fryzlewicz & Subba Rao [25] who work with assumptions closely related to (E3).)

1.3.3 Properties of the tvNAR Process

We now show that the tvNAR process is locally stationary and strongly mixing under the assumptions listed above. In addition, we show that the auxiliary processes $\{X_t(u)\}$ have densities that vary smoothly over rescaled time u. As will turn out, these three properties are central for the estimation theory developed in Sections 1.4 and 1.5. To formulate and prove the results, we repeatedly make use of the following notation: For any sequence of processes $\{Y_{t,T}, t \in \mathbb{Z}\}$ with $T = 1, 2, \ldots,$

$$Y_{t,T}^{t-k} := (Y_{t-k,T}, \ldots, Y_{t,T}) \quad \text{for } k > 0.$$

In particular, we let $X_{t,T}^{t-k} = (X_{t-k,T}, \ldots, X_{t,T})$ and $X_t^{t-k}(u) = (X_{t-k}(u), \ldots$ $\ldots, X_t(u))$ for $u \in \mathbb{R}$. The first theorem shows that the auxiliary process $\{X_t(u)\}$ is strictly stationary for each rescaled time point u.

Theorem 1.1. *Assume that (M1)–(M3), (Σ1)–(Σ3), and (E1) are fulfilled. Then*

(i) for each $u \in \mathbb{R}$, the process $\{X_t(u), t \in \mathbb{Z}\}$ has a strictly stationary solution with ε_t independent of $X_{t-k}(u)$ for $k < 0$,

(ii) the variables $X_{t-1}^{t-d}(u)$ have a density $f_{X_{t-1}^{t-d}(u)}$ w.r.t. Lebesgue measure,

(iii) the variables $X_{t-1,T}^{t-d}$ have densities $f_{X_{t-1,T}^{t-d}}$ w.r.t. Lebesgue measure.

The second result states that $\{X_t(u)\}$ locally approximates $\{X_{t,T}\}$ in the sense of Definition 1.1.

Theorem 1.2. *Assume that (M1)–(M3), (Σ1)–(Σ3), and (E1) are fulfilled. Then*

$$|X_{t,T} - X_t(u)| \leq \Big(\Big|\frac{t}{T} - u\Big| + \frac{1}{T}\Big) U_{t,T}(u), \tag{1.6}$$

where the variables $U_{t,T}(u)$ *satisfy the condition that* $\mathbb{E}[(U_{t,T}(u))^\rho] < C$ *for some* $\rho > 0$ *and* $C < \infty$ *independent of* u, t, *and* T.

Taken together, Theorems 1.1 and 1.2 show that the tvNAR process $\{X_{t,T}\}$ is locally stationary in the sense of Definition 1.1. As can be seen from the next result, the densities $f_{X_{t-1}^{t-d}(u)}$ change smoothly over time.

Theorem 1.3. *Let* $f(u,x) := f_{X_{t-1}^{t-d}(u)}(x)$ *be the density of* $X_{t-1}^{t-d}(u)$ *at* $x \in \mathbb{R}$. *If (M1)–(M3), (Σ1)–(Σ3), and (E1)–(E2) are fulfilled, then*

$$|f(u,x) - f(v,x)| \leq C_x |u-v|^p$$

with some constant $0 < p < 1$ *and* $C_x < \infty$ *continuously depending on* x.

We finally characterize the mixing behaviour of the tvNAR process. We first give a quick reminder of the definitions of an α-mixing and β-mixing array. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{B}$ and $\mathcal{C}$ be subfields of $\mathcal{A}$. Define

$$\alpha(\mathcal{B},\mathcal{C}) = \sup_{B\in\mathcal{B}, C\in\mathcal{C}} |\mathbb{P}(B\cap C) - \mathbb{P}(B)\mathbb{P}(C)|$$
$$\beta(\mathcal{B},\mathcal{C}) = \mathbb{E}\sup_{C\in\mathcal{C}} |\mathbb{P}(C) - \mathbb{P}(C|\mathcal{B})|.$$

Moreover, for an array $\{Y_{t,T} : 1 \leq t \leq T\}$, define the coefficients

$$\alpha(k) = \sup_{t,T:1\leq t\leq T-k} \alpha\big(\sigma(Y_{s,T}, 1\leq s\leq t), \sigma(Y_{s,T}, t+k \leq s \leq T)\big)$$
$$\beta(k) = \sup_{t,T:1\leq t\leq T-k} \beta\big(\sigma(Y_{s,T}, 1\leq s\leq t), \sigma(Y_{s,T}, t+k \leq s \leq T)\big),$$

where $\sigma(Z)$ is the σ-field generated by Z. The array $\{Y_{t,T}\}$ is said to be α-mixing (or strongly mixing) if $\alpha(k) \to 0$ as $k \to \infty$. Similarly, it is called β-mixing if $\beta(k) \to 0$. Note that β-mixing implies α-mixing. The final result of this section shows that the tvNAR process is β-mixing with coefficients that converge exponentially fast to zero.

Theorem 1.4. *If (M1)–(M3), (Σ1)–(Σ3), and (E1)–(E3) are fulfilled, then the tvNAR process* $\{X_{t,T}\}$ *is geometrically* β*-mixing, i.e. there exists a positive constant* $\gamma < 1$ *such that* $\beta(k) \leq \gamma^k$.

The proofs of the above theorems can be found in Appendix A.

1.3.4 The Additive tvNAR Process

An interesting special case of the tvNAR process arises, when the functions m and σ split up into additive components. In this case, the process is defined by the difference equation

$$X_{t,T} = \sum_{j=1}^{d} m_j\Big(\frac{t}{T}, X_{t-j,T}\Big) + \Big(\sum_{j=1}^{d} \sigma_j\Big(\frac{t}{T}, X_{t-j,T}\Big)\Big)^{1/2} \varepsilon_t. \tag{1.7}$$

In this setting, we can replace the conditions (M1)–(M3) and (Σ1)–(Σ3) on the functions m and σ by analogous conditions on the additive component functions. Most importantly, (M3) (and analogously (Σ3)) can be replaced by

($\mathrm{M}_{\mathrm{add}}$3) $m_1, \ldots, m_d$ are continuously differentiable with respect to x. The partial derivatives $\partial m_j(u, x_j) := \frac{\partial}{\partial x_j} m_j(u, x_j)$ have the property that for some $K_1 < \infty$,

$$\sup_{u \in \mathbb{R}, |x_j| > K_1} |\partial m_j(u, x_j)| \leq \delta_{\mathrm{add}} < 1.$$

Here, δ_{add} is given by a similar expression as δ in (M3).

Inspecting the proofs of Theorems 1.1–1.4, it is straightforward to see that the theorems still hold true under these modified conditions.

1.4 Kernel Estimation in Locally Stationary Nonparametric Models

In this section, we consider kernel estimation in the general model (1.1),

$$Y_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} \quad \text{for } t = 1, \ldots, T,$$

where $\mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0$ and the covariates $X_{t,T} = (X_{t,T}^1, \ldots, X_{t,T}^d)$ are locally stationary. The next subsection introduces kernel estimates for the function m. In the subsequent subsections we provide asymptotic theory for these estimates. In particular, we derive results on uniform convergence rates in Subsections 1.4.3–1.4.5. The last subsection is devoted to results on asymptotic normality. The proofs are given in Appendix B.

1.4.1 Estimation Procedure

We restrict attention to local constant estimation. It is straightforward to extend the theory to local linear (or more generally local polynomial) estimation. The Nadaraya-Watson (NW) estimator for model (1.1) is given by

$$\hat{m}(u,x) = \frac{\sum_{t=1}^{T} K_h\big(u - \frac{t}{T}\big) \prod_{j=1}^{d} K_h\big(x^j - X_{t,T}^j\big) Y_{t,T}}{\sum_{t=1}^{T} K_h\big(u - \frac{t}{T}\big) \prod_{j=1}^{d} K_h\big(x^j - X_{t,T}^j\big)}. \tag{1.8}$$

Here, K denotes a one-dimensional kernel function and we use the notation $K_h(v) = K(\frac{v}{h})$. For convenience, we work with a product kernel $K^\times(u,x) = K(u) \prod_{j=1}^{d} K(x^j)$ and assume that the bandwidth h is the same in each direction. Our results can be easily modified to allow for non-product kernels and different bandwidths. Note that the above estimate differs from the NW estimator in the standard strictly stationary setting only in that there is an additional kernel in time direction.

1.4.2 Assumptions

The following three assumptions are central for our results:

(C1) The process $\{X_{t,T}\}$ is locally stationary in the sense of Definition 1.1. Thus, for each time point $u \in [0,1]$, there exists a strictly stationary process $\{X_t(u)\}$ with density $f(u,x) := f_{X_t(u)}(x)$ such that $\|X_{t,T} - X_t(u)\| \le (|\frac{t}{T} - u| + \frac{1}{T}) U_{t,T}(u)$ with $\mathbb{E}[(U_{t,T}(u))^\rho] \le C$ for some $\rho > 0$.

(C2) The densities $f_{X_t(u)}$ are smooth in u, i.e. $f(u,x) = f_{X_t(u)}(x)$ is a smooth function of u for each $x \in \mathbb{R}^d$. In particular, $f(u,x)$ is differentiable w.r.t. u for each $x \in \mathbb{R}^d$ and the derivative $\partial_0 f(u,x) := \frac{\partial}{\partial u} f(u,x)$ is continuous.

(C3) The array $\{X_{t,T}, \varepsilon_{t,T}\}$ is α-mixing.

In Section 1.3, we have seen that these three conditions are essentially fulfilled for the tvNAR process. Note that we do not necessarily need the densities $f_{X_t(u)}$ to be differentiable in time direction as assumed in (C2). We could also do with a lower degree of smoothness, e.g. continuity as shown for the tvNAR process, at the cost of having slower convergence rates for the bias part of the NW estimate. Furthermore, for the tvNAR process, (C3) is equivalent to $\{X_{t,T}^{t-d}\}$ being α-mixing. The latter condition is clearly fulfilled, as it is a direct consequence of Theorem 1.4.

In addition to the above three assumptions, we impose the following technical conditions on the model components:

(C4) $f(u,x)$ is partially differentiable w.r.t. x for each $u \in [0,1]$. The derivatives $\partial_j f(u,x) := \frac{\partial}{\partial x^j} f(u,x)$ are continuous for $j = 1, \ldots, d$.

(C5) $m(u,x)$ is twice continuously partially differentiable with first derivatives $\partial_j m(u,x)$ and second derivatives $\partial^2_{ij} m(u,x)$ for $i,j = 0, \ldots, d$.

The kernel K is assumed to satisfy the following condition:

(C6) K is symmetric about zero, bounded and has compact support, i.e. $K(v) = 0$ for all $|v| > C_1$ with some $C_1 < \infty$. Further, K is Lipschitz, i.e. $|K(v) - K(v')| \leq L|v - v'|$ for some $L < \infty$ and all $v, v' \in \mathbb{R}$.

Finally, note that throughout the chapter the bandwidth h is assumed to converge to zero at least at polynomial rate, i.e. there exists a small $\xi > 0$ such that $h \leq CT^{-\xi}$ for some constant $C > 0$.

1.4.3 Uniform Convergence Rates for Kernel Averages

As a first step in the asymptotic analysis of the NW estimate in model (1.1), we examine kernel averages of the general form

$$\hat{\psi}(u,x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X^j_{t,T}) W_{t,T} \tag{1.9}$$

and derive the uniform convergence rate of $\hat{\psi}(u,x) - \mathbb{E}\hat{\psi}(u,x)$. Later on we will make use of this result with $W_{t,T} = 1$ and $W_{t,T} = \varepsilon_{t,T}$ to calculate the uniform rate of the NW estimate.
We make the following assumptions on the components in (1.9):

(KA1) The array $\{X_{t,T}, W_{t,T}\}$ is α-mixing with mixing coefficients α that satisfy

$$\alpha(k) \leq Ak^{-\beta}$$

with $A < \infty$, where for some $s > 2$,

$$\mathbb{E}|W_{t,T}|^s \leq C \quad \text{and} \quad \beta > \frac{2s-2}{s-2}.$$

(KA2) The variables $X_{t,T}$ have densities $f_{X_{t,T}}$ with the following properties: For any compact set $S \subseteq \mathbb{R}^d$, there exist constants $B_0 = B_0(S)$, $B_1 = B_1(S)$, and $B_2 = B_2(S)$ such that

$$\sup_{t,T} \sup_{x \in S} f_{X_{t,T}}(x) \leq B_0 < \infty$$
$$\sup_{t,T} \sup_{x \in S} \mathbb{E}\big[|W_{t,T}|^s \big| X_{t,T} = x\big] f_{X_{t,T}}(x) \leq B_1 < \infty.$$

In addition,

$$\sup_{t,T} \sup_{l \geq 1} \sup_{x,x' \in S} \mathbb{E}\big[|W_{t,T}||W_{t+l,T}|\big| X_{t,T} = x, X_{t+l,T} = x'\big] \times f_{X_{t,T},X_{t+l,T}}(x,x') \leq B_2 < \infty,$$

where $f_{X_{t,T},X_{t+l,T}}$ is the joint density of $(X_{t,T}, X_{t+l,T})$.

The following theorem generalizes results of Hansen [17] for the strictly stationary case to our locally stationary setting. For related results, see Kristensen [23].

Theorem 1.5. *Let (KA1) and (KA2) be fulfilled and let the kernel K satisfy (C6). Assume that*

$$\beta > \frac{2 + s(1 + (d+1))}{s - 2} \tag{1.10}$$
$$\theta = \frac{\beta(1 - \frac{2}{s}) - \frac{2}{s} - 1 - (d+1)}{\beta + 3 - (d+1)}. \tag{1.11}$$

In addition, suppose that the bandwidth satisfies

$$\frac{\phi_T \log T}{T^\theta h^{d+1}} = o(1) \tag{1.12}$$

with ϕ_T slowly diverging to infinity (e.g. $\phi_T = \log \log T$). Then it holds that

$$\sup_{u \in [0,1], x \in S} \big|\hat{\psi}(u,x) - \mathbb{E}\hat{\psi}(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{T h^{d+1}}}\Big).$$

The convergence rate in the above theorem is identical to the rate obtained for a $(d+1)$-dimensional nonparametric problem in the standard strictly stationary setting. This reflects the fact that additionally smoothing in time direction, we essentially have a $(d+1)$-dimensional estimation problem in our case.

1.4.4 Uniform Convergence Rates for Density Estimates

Before we consider the NW estimates of model (1), we examine the asymptotic behaviour of density estimates in this model. Define

$$\hat{f}(u,x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j).$$

The following result shows that $\hat{f}(u,x)$ converges uniformly in (u,x) to the density $f(u,x)$ of $X_t(u)$.

Theorem 1.6. *Assume that (C1), (C2), (C4) and (C6) hold and that (KA1)–(KA2) are fulfilled for* $W_{t,T} = 1$. *Let* $\beta > 1 + (d+1)$ *and suppose that the bandwidth* h *satisfies*

$$\frac{\phi_T \log T}{T^{\theta} h^{d+1}} = o(1) \quad \textit{and} \quad \frac{1}{T^r h^{d+r}} = o(1)$$

with $\theta = \frac{\beta-1-(d+1)}{\beta+3-(d+1)}$, $\phi_T = \log\log T$, *and* $r = \min\{\rho, 1\}$, *where* ρ *has been introduced in (C1). Defining* $I_h = [C_1 h, 1 - C_1 h]$, *it then holds that*

$$\sup_{u \in I_h, x \in S} \big|\hat{f}(u,x) - f(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big) + O\Big(\frac{1}{T^r h^{d+r}}\Big) + o(h). \tag{1.13}$$

To derive the above result, we split up the difference $\hat{f}(u,x) - f(u,x)$ into a stochastic part and a bias part. The stochastic part is of the order $O_p(\sqrt{\log T/Th^{d+1}})$. As already noted in the previous subsection, this mirrors the fact that we essentially have to solve a $(d+1)$-dimensional estimation problem. The bias part converges at the rate $O(T^{-r}h^{-(d+r)}) + o(h)$. Thus, in contrast to the standard strictly stationary case, an additional bias component of order $T^{-r}h^{-(d+r)}$ shows up. As can be seen from the proof, this additional component results from replacing the variables $X_{t,T}$ by $X_t(\frac{t}{T})$ in the bias expression. It thus captures how far the variables $X_{t,T}$ are from their stationary approximation $X_t(\frac{t}{T})$. Put differently, it measures the deviation from stationarity.

Note that the additional bias term converges faster to zero for larger $r = \min\{\rho, 1\}$. This makes perfect sense if we recall from Section 1.2 that r measures how well $X_{t,T}$ is locally approximated by $X_t(\frac{t}{T})$: The larger r, the smaller the deviation of $X_{t,T}$ from its stationary approximation and thus the smaller the additional nonstationarity bias.

1.4.5 Uniform Convergence Rates for NW Estimates

Using the results from the previous subsections, we can derive the following theorem on the uniform convergence behaviour of the NW estimator in model (1).

Theorem 1.7. *Assume that (C1)–(C6) hold and that (KA1)–(KA2) are fulfilled for both $W_{t,T} = 1$ and $W_{t,T} = \varepsilon_{t,T}$. Let β and θ satisfy equations (1.10) and (1.11), and suppose that $\inf_{u\in[0,1],x\in S} f(u,x) > 0$. Moreover, assume that the bandwidth h satisfies*

$$\frac{\phi_T \log T}{T^{\theta} h^{d+1}} = o(1) \quad \text{and} \quad \frac{1}{T^r h^{d+r}} = o(1)$$

with $\phi_T = \log\log T$ and $r = \min\{\rho, 1\}$. Defining $I_h = [C_1 h, 1 - C_1 h]$, it then holds that

$$\sup_{u\in I_h, x\in S} \big|\hat{m}(u,x) - m(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{T h^{d+1}}} + \frac{1}{T^r h^d} + h^2\Big). \tag{1.14}$$

The convergence rate in (1.14) is composed of analogous terms as the rate of the kernel density estimator in Theorem 1.6. Note however that the additional nonstationarity bias is of the slightly different order $T^{-r}h^{-d}$. The reason is as follows: As already noted, the additional bias component results from replacing the variables $X_{t,T}$ by $X_t(\frac{t}{T})$ in the bias expression. Its order partly depends on the smoothness of the terms that show up in the bias. As will be seen in the proofs, this accounts for the slightly different order.

1.4.6 Asymptotic Normality

We conclude the asymptotic analysis of the NW estimator in model (1) with a result on asymptotic normality.

Theorem 1.8. *Assume that (C1)–(C6) hold and that (KA1)–(KA2) are fulfilled for both $W_{t,T} = 1$ and $W_{t,T} = \varepsilon_{t,T}$. Let $\beta \geq 4$ and $T^r h^{d+2} \to \infty$ with $r = \min\{\rho, 1\}$. Moreover, suppose that $f(u,x) > 0$ and that $\sigma^2(\frac{t}{T}, x) := \mathbb{E}[\varepsilon_{t,T}^2 | X_{t,T} = x]$ is continuous. Finally, let $r > \frac{d+2}{d+5}$ to ensure that the bandwidth h can be chosen to satisfy $Th^{d+5} \to 0$. Then*

$$\sqrt{Th^{d+1}}\big(\hat{m}(u,x) - m(u,x)\big) \xrightarrow{d} N(0, V_{u,x}).$$

Here, $V_{u,x} = \kappa_0^{d+1}\sigma^2(u,x)/f(u,x)$ with $\kappa_0 = \int K^2(\varphi)d\varphi$.

The above theorem parallels the asymptotic normality result for the standard strictly stationary setting. In particular, the variance expression $V_{u,x}$ is very similar to that for the standard case. By requiring that $T^r h^{d+2} \to \infty$, we make sure that the additional bias term which results from the nonstationarity of the model variables is asymptotically negligible.

1.5 Locally Stationary Additive Models

In this section, we put some structural constraints on the regression function m in the model

$$Y_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} \quad \text{with } \mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0.$$

In particular, we assume that for all rescaled time points $u \in [0,1]$ and all points x in a compact subset of $\mathbb{R}^d$, say $[0,1]^d$, the regression function can be split up into additive components according to

$$m(u,x) = m_0(u) + \sum_{j=1}^{d} m_j(u, x^j).$$

This means that for $x \in [0,1]^d$, we have the additive regression model

$$\mathbb{E}[Y_{t,T}|X_{t,T} = x] = m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j\Big(\frac{t}{T}, x^j\Big). \tag{1.15}$$

To identify the component functions in model (1.15), we introduce the density function

$$p(u,x) = \frac{I(x \in [0,1]^d) f(u,x)}{\mathbb{P}(X_0(u) \in [0,1]^d)}$$

together with the marginals $p_j(u,x^j) = \int p(u,x)dx^{-j}$, where as before $f(u,\cdot)$ is the density of the strictly stationary process $\{X_t(u)\}$. With these definitions at hand, we can impose the condition that

$$\int m_j(u,x^j)p_j(u,x^j)dx^j = 0 \tag{1.16}$$

for all $j = 1, \dots, d$ and all rescaled time points $u \in [0,1]^d$. Note that this normalization of the component functions varies over time in the sense that for each time point u, we integrate with respect to a different density.

For each rescaled time point u, the additive regression function $m(u,\cdot) = m_0(u) + \sum_{j=1}^d m_j(u,\cdot)$ can be characterized as the solution to an L_2-projection problem. To see this, let u be a fixed point in rescaled time and define $\mathcal{G}_{\text{add}}(p(u,\cdot))$ to be the class of functions $g : \mathbb{R}^d \to \mathbb{R}$ that are square integrable with respect to $p(u,\cdot)$ and that have an additive structure $g(x) = g_0 + g_1(x^1) + \ldots + g_d(x^d)$ for all $x \in [0,1]^d$ with $\int g_j(w^j) p_j(u, w^j) dw^j = 0$ for $j = 1, \ldots, d$. The regression function $m(u, \cdot)$ at time point u can ($p(u, \cdot)$ almost surely) be characterized by the projection equation

$$m(u, \cdot) = \min_{g \in \mathcal{G}_{\text{add}}(p(u,\cdot))} \int \big(m(u,w) - g(w)\big)^2 p(u,w) dw. \tag{1.17}$$

Note that $m(u, \cdot)$ trivially minimizes (1.17) as under the usual smoothness conditions, it belongs itself to $\mathcal{G}_{\text{add}}(p(u, \cdot))$.
We now define the smooth backfitting estimate

$$\tilde{m}(u, \cdot) = \tilde{m}_0(u) + \sum_{j=1}^{d} \tilde{m}_j(u, \cdot)$$

for some fixed $u \in [0,1]$ as the solution to an empirical version of the projection problem (1.17) with m and p replaced by kernel estimates $\hat{m}$ and $\hat{p}$. Choosing $\hat{m}$ as a $(d+1)$-dimensional NW estimate and $\hat{p}$ as a $(d+1)$-dimensional kernel density, the backfitting estimator $\tilde{m}(u, \cdot)$ of $m(u, \cdot)$ at time point u is given as

$$\tilde{m}(u, \cdot) = \min_{g \in \mathcal{G}_{\text{add}}(\hat{p}(u,\cdot))} \int \big(\hat{m}(u,w) - g(w)\big)^2 \hat{p}(u,w) dw, \tag{1.18}$$

where the minimization is done under the constraints

$$\int \tilde{m}_j(u, w^j) \hat{p}_j(u, w^j) dw^j = 0 \tag{1.19}$$

for all $j = 1, \ldots, d$. Note that (1.18) is a d-dimensional projection problem. In particular, rescaled time does not enter as an additional dimension. The projection is rather done separately for each time point $u \in [0,1]$. This means that we fit a smooth backfitting estimate to the data separately around each time point u.
By differentiation, we can show that the minimizer of (1.18) is characterized by the system of integral equations

$$\tilde{m}_j(u, x^j) = \hat{m}_j(u, x^j) - \sum_{k \neq j} \int \tilde{m}_k(u, x^k) \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j)} dx^k - \tilde{m}_0(u) \tag{1.20}$$

together with $\int \tilde{m}_j(u, w^j) \hat{p}_j(u, w^j) dw^j = 0$ for $j = 1 \ldots, d$. Here, $\hat{p}_j$ and $\hat{p}_{j,k}$ are kernel density estimates given by

$$\hat{p}_j(u, x^j) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) K_h(x^j, X_{t,T}^j) \tag{1.21}$$

$$\hat{p}_{j,k}(u, x^j, x^k) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) \\ \times K_h(x^j, X_{t,T}^j) K_h(x^k, X_{t,T}^k). \tag{1.22}$$

In these formulas,

$$T_{[0,1]^d} = \sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big) I(X_{t,T} \in [0,1]^d)$$

is the number of observations in the unit cube $[0,1]^d$, where only time points close to u are taken into account, and

$$K_h(v, w) = \frac{K_h(v - w)}{\int_0^1 K_h(s - w) ds}$$

is a modified kernel weight. These weights have the property that $\int_0^1 K_h(v, w) dv = 1$ for all v, which is needed to derive the asymptotic properties of the smooth backfitting estimators. Moreover, $\hat{m}_j$ is a Nadaraya-Watson smoother defined as

$$\hat{m}_j(u, x^j) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) \\ \times K_h(x^j, X_{t,T}^j) Y_{t,T} \big/ \hat{p}_j(u, x^j) \tag{1.23}$$

and the estimate $\tilde{m}_0(u)$ of the model constant at time point u is given by

$$\tilde{m}_0(u) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) Y_{t,T}. \tag{1.24}$$

We now summarize the assumptions needed to derive the asymptotic properties of the smooth backfitting estimates. First of all, the conditions of Section 1.4 must be satisfied for the kernel estimates (1.21)–(1.24). This is ensured by the following assumption.

(Add1) (C1)–(C6) are fulfilled together with (KA1)–(KA2) for $W_{t,T} = 1$ and $W_{t,T} = \varepsilon_{t,T}$. The parameters β and θ are such that

$$\beta > \max\Big\{4, \frac{2+3s}{s-2}\Big\} \quad \text{and} \quad \theta = \min\Big\{\frac{\beta - 4}{\beta}, \frac{\beta(1 - \frac{2}{s}) - \frac{2}{s} - 3}{\beta + 1}\Big\}$$

and $\inf_{u \in [0,1], x \in [0,1]^d} f(u, x) > 0$.

In addition, we need some restrictions on the admissible bandwidth. For convenience, we stipulate somewhat stronger conditions than in Section 1.4 to get rid of the additional nonstationarity bias from the very beginning.

(Add2) Let $\phi_T = \log\log T$ and $r = \min\{\rho, 1\}$ with ρ given in (C1). The bandwidth h satisfies $Th^5 \to \infty$. Moreover,

$$\frac{\phi_T \log T}{T^\theta h^2} = o(1) \quad \text{together with} \quad T^r h^3 \to \infty \text{ and } T^{\frac{r}{r+1}} h^2 \to \infty.$$

The condition $\frac{\phi_T \log T}{T^\theta h^2} = o(1)$ is already known from Section 1.4. The latter two restrictions ensure that $T^{-\frac{r}{r+1}} = o(h^2)$ and $T^{-r}h^{-1} = o(h^2)$. As will be seen in Appendix C, this implies that the additional nonstationarity bias is of smaller order than $O(h^2)$ and can thus be asymptotically neglected.
Under these assumptions, we can establish the following asymptotic results. Firstly, the smooth backfitting estimates uniformly converge to the true component functions at the two-dimensional rates no matter how large the dimension d of the full regression function.

Theorem 1.9. *Under (Add1) and (Add2), it holds that*

$$\sup_{u,x^j \in I_h} \left|\tilde{m}_j(u,x^j) - m_j(u,x^j)\right| = O_p\Big(\sqrt{\frac{\log T}{Th^2}} + h^2\Big) \tag{1.25}$$

with $I_h = [2C_1 h, 1 - 2C_1 h]$.

Secondly, the estimates are asymptotically normal if rescaled appropriately.

Theorem 1.10. *Suppose that (Add1) and (Add2) hold. In addition, let* $\theta > \frac{1}{3}$ *and* $r > \frac{1}{2}$, *which allows us to choose the bandwidth* h *such that* $T_{[0,1]^d} h^6 \to 0$. *Then for any* $u, x^1, \ldots, x^d \in (0,1)$,

$$\sqrt{T_{[0,1]^d} h^2} \begin{bmatrix} \tilde{m}_1(u,x^1) - m_1(u,x^1) \\ \vdots \\ \tilde{m}_d(u,x^d) - m_d(u,x^d) \end{bmatrix} \xrightarrow{d} N(0, V_{u,x}),$$

where $V_{u,x} = \operatorname{diag}(v_1(u,x^1), \ldots, v_d(u,x^d))$ *is a diagonal matrix with entries* $v_j(u,x^j) = \kappa_0^2 \sigma_j^2(u,x^j)/p_j(u,x^j)$ *and* $\kappa_0 = \int K^2(\varphi) d\varphi$.

The proof of Theorems 1.9 and 1.10 can be found in Appendix C.

1.6 Application

To illustrate our estimation theory, we apply it to a sample of NASDAQ Composite index data from the beginning of 2000 to the middle of 2011. For each day, our sample contains the return and the so-called high-low range. The latter is defined as the difference between the highest and lowest logarithmic price of a day. The range is a measure of daily volatility and has a long history in finance. It has been employed for example in the studies of Rogers & Satchell [38], Yang & Zhang [41], Alizadeh et al. [1], and Martens & van Dijk [31].

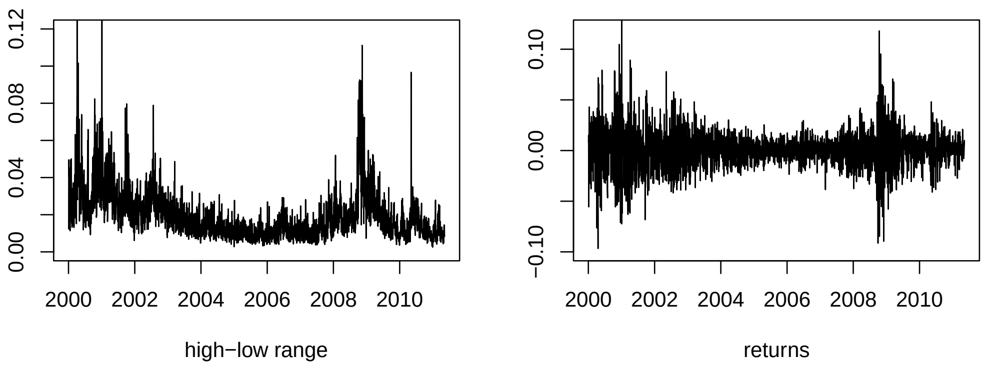

Figure 1.1: High-low range and returns of the NASDAQ Composite series.

In what follows, $y_{t,T}$ denotes the logarithm of the high-low range and $r_{t,T}$ is the daily return. With this notation, we define the model

$$y_{t,T} = m_0\Big(\frac{t}{T}\Big) + m_1\Big(\frac{t}{T}, y_{t-1,T}\Big) + m_2\Big(\frac{t}{T}, r_{t-1,T}\Big) + \varepsilon_{t,T}, \tag{1.26}$$

where $\mathbb{E}[\varepsilon_{t,T}|y_{t-1,T}, r_{t-1,T}] = 0$ and the functions m_1 and m_2 are normalized according to (1.16). Here, volatility is treated as an observed variable. We thus neglect the fact that the range only approximates the underlying true volatility. (1.26) can be regarded as a localized version of the model studied in Wu & Xiao [40].[1] It is very similar in structure to the volatility equation of a time-varying EGARCH(1,1) model. Clearly the conditional volatility in an EGARCH model is not identical with the daily range. However, following the argumentation in Wu & Xiao [40], if there is a relationship between daily range and conditional volatility, then the nonparametric fits of m_1 and m_2 may help in appropriately specifying the parametric form of time-varying EGARCH models.

[1] Wu & Xiao consider a model in which the component functions m_1 and m_2 do not depend on time and the first component m_1 is restricted to be linear. Moreover, implied volatility instead of the range is used as a daily volatility measure.

We fit model (1.26) locally around three different time points in our sample, choosing the bandwidth in time direction to span approximately one year and a half. As a result, we estimate the model for three different time periods, each spanning roughly three years. We include the period from 03/2000 to 03/2003 which corresponds to the aftermath of the technology bubble and the events of 9/11, the period from 11/2007 to 11/2010 which spans a great deal of the recent financial crisis, and an intermediate non-crisis period from 11/2003 to 11/2006.
The estimation results are shown in Figure 1.2. The solid, dashed and dotted lines are the nonparametric fits for the three different periods and the grey shaded areas are 95% pointwise confidence bands. The fits are normalized according to (1.19).

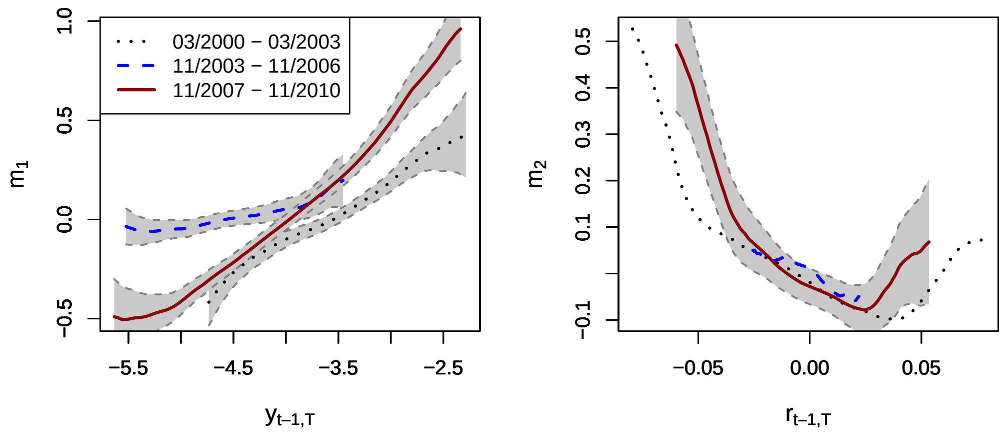

Figure 1.2: Estimation results for the additive model (1.26).

We have made several robustness checks. The first one concerns the choice of bandwidths. The bandwidth in time direction is handpicked rather than automatically selected. Given this, the bandwidths with respect to the two covariates are selected via a mean-squared error criterion. To check whether the estimation results are robust against different choices of bandwidth in time direction, we have gradually reduced the bandwidth to span only one year. This has virtually no effect on the fits. Moreover, we have smoothly varied the time points around which the model is estimated. As expected, this results in smooth changes of the nonparametric fits. In particular, shifting the time points only by a couple of months does not have major effects on the fits and preserves their qualitative form.
We now have a closer look at the estimation results in Figure 1.2.

- The estimates of m_1 are fairly linear. Interestingly, the fit for the financial crisis period (and presumably also the one for the period from 2000 to 2003) is much steeper than that for the intermediate non-crisis period from 2003 to

2006. This suggests that in more tense economic situations or crisis periods, today's volatility reacts more strongly to changes in yesterday's volatility. Put differently, the market is more sensitive to changes in volatility.

- The m_2-component is the news impact curve of the model. It captures how return shocks influence volatility. The estimates suggest that the overall form of the news impact curve is rather robust over time. Moreover, one can clearly see the asymmetric form of the curve which has been reported in numerous other studies before.

In the next step of our empirical analysis, we use the nonparametric fits of (1.26) as a guideline to set up a parametric model. We choose a specification with a linear m_1-component and a quadratic m_2-component that is flexible enough to allow for asymmetries. The model is given by

$$y_{t,T} = m_{0,\text{par}}\Big(\frac{t}{T}\Big) + m_{1,\text{par}}\Big(\frac{t}{T}, y_{t-1,T}\Big) + m_{2,\text{par}}\Big(\frac{t}{T}, r_{t-1,T}\Big) + \varepsilon_{t,T} \tag{1.27}$$

with

$$\begin{aligned} m_{1,\text{par}}\Big(\frac{t}{T}, y_{t-1,T}\Big) &= a_1\Big(\frac{t}{T}\Big) y_{t-1,T} \\ m_{2,\text{par}}\Big(\frac{t}{T}, r_{t-1,T}\Big) &= a_2\Big(\frac{t}{T}\Big) r_{t-1,T}^2 I(r_{t-1,T} < 0) + a_3\Big(\frac{t}{T}\Big) r_{t-1,T}^2 I(r_{t-1,T} \geq 0), \end{aligned}$$

where a_1, a_2 and a_3 are time-varying parameters. We estimate (1.27) locally around the same time points as the additive model (1.26) using the same bandwidth in time direction. The estimation is done by minimizing a least-squares criterion localized in time. Rather than reporting the estimates of the time-varying parameters in a table, we plot the fits of $m_{1,\text{par}}$ and $m_{2,\text{par}}$ in Figure 1.3.

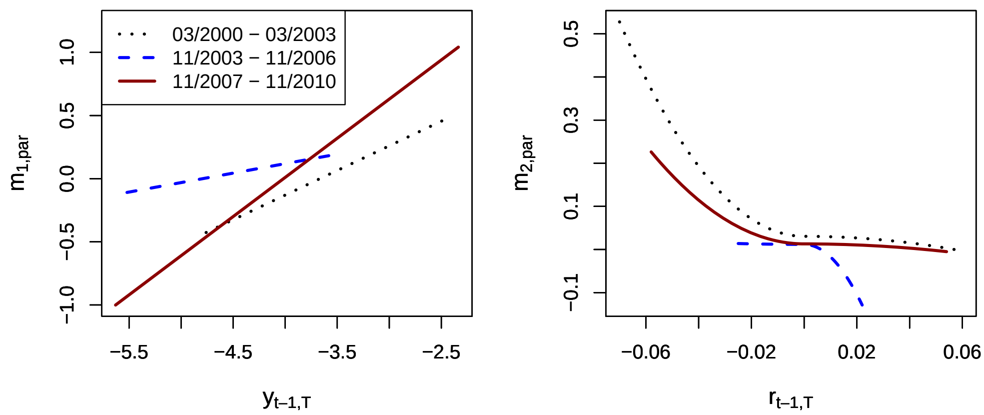

Figure 1.3: Estimation results for the parametric model (1.27).

The fits of $m_{1,\text{par}}$ give a very similar picture as their nonparametric counterparts. The estimates of $m_{2,\text{par}}$, however, do not. In particular, they suggest that the news impact curve in the intermediate non-crisis period from 2003 to 2006 substantially differs from the curves in the two crisis periods. Figure 1.4 makes visible the differences between the parametric and nonparametric fits of the news impact curve.

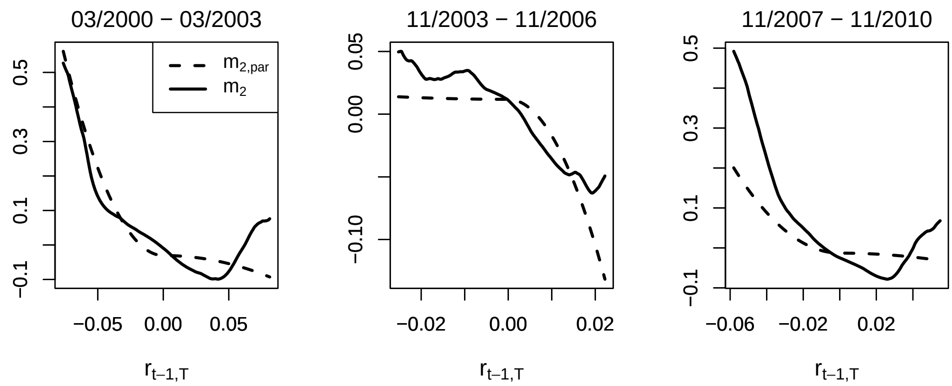

Figure 1.4: Comparison of the parametric function $m_{2,\text{par}}$ (dashed) and its nonparametric counterpart m_2 (solid).

As can be seen from Figure 1.4, the parametric estimates roughly capture the overall form of their nonparametric counterparts. However, they are not flexible enough to reproduce all important characteristics. In particular, the parametric estimate for the intermediate non-crisis period strongly exaggerates the slightly concave form of the corresponding nonparametric fit. This gives the impression that the news impact curve in the non-crisis period drastically differs from that in the two crisis periods.

The above considerations make visible an important shortcoming of the parametric analysis: If the parametric model is not flexible enough, then the fits may spuriously generate time-varying effects. Thus, the news impact curve may after all be much more robust over time than suggested by many parametric specifications.

1.7 Concluding Remarks

We have studied nonparametric regression models which are a natural extension of parametric time series models with time-varying coefficients. In these models, the regression function is allowed to vary smoothly over time and the model variables are locally stationary. We have developed a complete asymptotic theory for this framework. Moreover, we have shown that the main assumptions of the theory are satisfied for a large class of nonlinear autoregressive processes with time-varying regression function. Finally, we have introduced structured models where the regression function splits up into time-varying additive components. Importantly, estimation in these models does not suffer from the curse of dimensionality. This makes additive models a flexible option in many applications in which the fully nonparametric model is infeasible.

Appendix A

In this appendix, we prove the results on the tvNAR process from Section 1.3. Throughout the appendix, we use the symbol C to denote a universal real constant which may take a different value on each occurrence.

Preliminaries

Before we come to the proofs of the theorems, we state some auxiliary results needed for the arguments later on.

Linearization of the functions m and σ

Consider the function m. By the mean value theorem, it holds that

$$m(u,x) - m(u,x') = \sum_{j=1}^{d} \Delta_j^m(u,x,x')(x'_j - x_j)$$

with

$$\Delta_j^m(u,x,x') = \int_0^1 \partial_j m(u, x + s(x'-x))ds.$$

This allows us to write

$$\begin{aligned} &\Big| m\Big(\frac{t}{T}, X_{t-1,T}^{t-d}\Big) - m\big(u, X_{t-1}^{t-d}(u)\big)\Big| \\ &\le L\Big|\frac{t}{T} - u\Big| + \big|m\big(u, X_{t-1,T}^{t-d}\big) - m\big(u, X_{t-1}^{t-d}(u)\big)\big| \\ &\le L\Big|\frac{t}{T} - u\Big| + \sum_{j=1}^{d} \big|\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})\big|\big|X_{t-j,T} - X_{t-j}(u)\big|. \end{aligned} \tag{1.28}$$

The term $\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})$ has the property that

$$\big|\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})\big| \le \begin{cases} \Delta & \text{for } \|\varepsilon_{t-1}^{t-d}\|_\infty \le K_2 \\ \delta & \text{for } \|\varepsilon_{t-1}^{t-d}\|_\infty > K_2 \end{cases} \tag{1.29}$$

with $K_2 = (K_1 + M)/\underline{\Sigma}$ and $\Delta \geq \sup_{u,x} |\partial_j m(u,x)|$. This can be seen as follows: Using the abbreviations $m_{u,k} = m(u, X_{t-k-1}^{t-k-d}(u))$ and $m_{t,T,k} = m(\frac{t-k}{T}, X_{t-k-1,T}^{t-k-d})$ together with $\sigma_{u,k} = \sigma(u, X_{t-k-1}^{t-k-d}(u))$ and $\sigma_{t,T,k} = \sigma(\frac{t-k}{T}, X_{t-k-1,T}^{t-k-d})$, we obtain

$$\begin{aligned}
&\|X_{t-1}^{t-d}(u) + s(X_{t-1,T}^{t-d} - X_{t-1}^{t-d}(u))\|_\infty \\
&= \max_{k=1,\dots,d} |X_{t-k}(u) + s(X_{t-k,T} - X_{t-k}(u))| \\
&= \max_{k=1,\dots,d} |m_{u,k} + s(m_{t,T,k} - m_{u,k}) + \varepsilon_{t-k}(\sigma_{u,k} + s(\sigma_{t,T,k} - \sigma_{u,k}))| \\
&\geq \max_{k=1,\dots,d} \underline{\Sigma}|\varepsilon_{t-k}| - M, \qquad (1.30)
\end{aligned}$$

as $|m_{u,k} + s(m_{t,T,k} - m_{u,k})| \leq M$ and $|\sigma_{u,k} + s(\sigma_{t,T,k} - \sigma_{u,k})| \geq \underline{\Sigma} > 0$. Now assume that $\|\varepsilon_{t-1}^{t-d}\|_\infty > K_2$. In this case, (1.30) implies that $\|X_{t-1}^{t-d}(u) + s(X_{t-1,T}^{t-d} - X_{t-1}^{t-d}(u))\|_\infty > K_1$ for all $s \in [0,1]$. Hence, the region over which the integral in $\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})$ runs completely lies outside the area $[-K_1, K_1]^d$. Therefore, the integrand $\partial_j m$ is always smaller than δ in absolute value, which immediately implies that $|\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})| \leq \delta$. Now let $\|\varepsilon_{t-1}^{t-d}\|_\infty \leq K_2$. As $\sup_{u,x} |\partial_j m(u,x)| \leq \Delta < \infty$, the term $|\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})|$ is always bounded by Δ, in particular for $\|\varepsilon_{t-1}^{t-d}\|_\infty \leq K_2$.
From (1.29), it immediately follows that $\Delta_j^m(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})$ is absolutely bounded by

$$\Delta(\varepsilon_{t-1}^{t-d}) := \Delta I(\|\varepsilon_{t-1}^{t-d}\|_\infty \leq K_2) + \delta I(\|\varepsilon_{t-1}^{t-d}\|_\infty > K_2). \qquad (1.31)$$

As will turn out later on, this bound is particularly useful, as its stochastic behaviour is solely determined by the vector of residuals $(\varepsilon_{t-1}, \dots, \varepsilon_{t-d})$.
Finally, using analogous arguments for the function σ, we arrive at

$$\sigma(u,x) - \sigma(u,x') = \sum_{j=1}^d \Delta_j^\sigma(u,x,x')(x_j' - x_j)$$

with $\Delta_j^\sigma(u,x,x') = \int_0^1 \partial_j \sigma(u, x + s(x'-x))ds$. As before, we can bound the term $|\Delta_j^\sigma(u, X_{t-1}^{t-d}(u), X_{t-1,T}^{t-d})|$ by $\Delta(\varepsilon_{t-1}^{t-d})$.

Recursive formulas for $X_{t,T}$ and $X_t(u)$

For the proof of Theorem 1.4, we rewrite $X_{t,T}$ in a recursive fashion: Letting $x_{t-k_1}^{t-k_2}$ and $e_{t-k_1}^{t-k_2}$ be values of $X_{t-k_1}^{t-k_2}$ and $\varepsilon_{t-k_1}^{t-k_2}$, respectively, we recursively define the functions $m^{(i)}_{t,T}$ and $\sigma^{(i)}_{t,T}$ by

$$\begin{aligned} m^{(0)}_{t,T}\big(x^{t-d}_{t-1}\big) &= m\Big(\frac{t}{T}, x^{t-d}_{t-1}\Big) \\ \sigma^{(0)}_{t,T}\big(x^{t-d}_{t-1}\big) &= \sigma\Big(\frac{t}{T}, x^{t-d}_{t-1}\Big) \end{aligned}$$

and for $i \geq 1$ by

$$\begin{aligned} m^{(i+1)}_{t,T}\big(e^{t-i-1}_{t-1}, x^{t-i-1-d}_{t-i-2}\big) &= m^{(i)}_{t,T}\big(e^{t-i}_{t-1}, m^{(0)}_{t-i-1,T}(x^{t-i-1-d}_{t-i-2}) \\ &\qquad + \sigma^{(0)}_{t-i-1,T}(x^{t-i-1-d}_{t-i-2})e_{t-i-1}, x^{t-i-d}_{t-i-2}\big) \\ \sigma^{(i+1)}_{t,T}\big(e^{t-i-1}_{t-1}, x^{t-i-1-d}_{t-i-2}\big) &= \sigma^{(i)}_{t,T}\big(e^{t-i}_{t-1}, m^{(0)}_{t-i-1,T}(x^{t-i-1-d}_{t-i-2}) \\ &\qquad + \sigma^{(0)}_{t-i-1,T}(x^{t-i-1-d}_{t-i-2})e_{t-i-1}, x^{t-i-d}_{t-i-2}\big). \end{aligned}$$

With this definition, we can represent $X_{t,T}$ as

$$X_{t,T} = m^{(i)}_{t,T}\big(\varepsilon^{t-i}_{t-1}, X^{t-i-d}_{t-i-1,T}\big) + \sigma^{(i)}_{t,T}\big(\varepsilon^{t-i}_{t-1}, X^{t-i-d}_{t-i-1,T}\big)\varepsilon_t.$$

Moreover, for $i \geq d$ we can write

$$\begin{aligned} m^{(i)}_{t,T}\big(e^{t-i}_{t-1}, x^{t-i-d}_{t-i-1}\big) &= m\Big(\frac{t}{T}, m^{(i-1)}_{t-1,T}(e^{t-i}_{t-2}, x^{t-i-d}_{t-i-1}) + \sigma^{(i-1)}_{t-1,T}(e^{t-i}_{t-2}, x^{t-i-d}_{t-i-1})e_{t-1}, \dots \\ &\qquad \dots, m^{(i-d)}_{t-d,T}(e^{t-i}_{t-d-1}, x^{t-i-d}_{t-i-1}) + \sigma^{(i-d)}_{t-d,T}(e^{t-i}_{t-d-1}, x^{t-i-d}_{t-i-1})e_{t-d}\Big) \\ \sigma^{(i)}_{t,T}\big(e^{t-i}_{t-1}, x^{t-i-d}_{t-i-1}\big) &= \sigma\Big(\frac{t}{T}, m^{(i-1)}_{t-1,T}(e^{t-i}_{t-2}, x^{t-i-d}_{t-i-1}) + \sigma^{(i-1)}_{t-1,T}(e^{t-i}_{t-2}, x^{t-i-d}_{t-i-1})e_{t-1}, \dots \\ &\qquad \dots, m^{(i-d)}_{t-d,T}(e^{t-i}_{t-d-1}, x^{t-i-d}_{t-i-1}) + \sigma^{(i-d)}_{t-d,T}(e^{t-i}_{t-d-1}, x^{t-i-d}_{t-i-1})e_{t-d}\Big). \end{aligned}$$

Formulas for conditional densities

Throughout the appendix, the symbol $f_{V|W}$ is used to denote the density of V conditional on W. In particular, $f_{X_{t,T}|X^{t-r+1}_{t-1,T},\varepsilon^{-s}_{t-r},X^{-s-d}_{-s-1,T}}$ is the density of $X_{t,T}$ conditional on the variables $X^{t-r+1}_{t-1,T}$, ε^{-s}_{t-r}, and $X^{-s-d}_{-s-1,T}$ with $1 \leq r \leq d$ and $s > 0$. If the residuals ε_t have a density f_ε, then it can be shown that

$$f_{X_{t,T}|X^{t-r+1}_{t-1,T},\varepsilon^{-s}_{t-r},X^{-s-d}_{-s-1,T}}(x_t|x^{t-r+1}_{t-1}, e^{-s}_{t-r}, w) = \frac{1}{\sigma_{t,T}} f_\varepsilon\Big(\frac{x_t - m_{t,T}}{\sigma_{t,T}}\Big) \tag{1.32}$$

with x_t, x_{t-1}^{t-r+1}, e_{t-r}^{-s}, and w being values of $X_{t,T}$, $X_{t-1,T}^{t-r+1}$, ε_{t-r}^{-s}, and $X_{-s-1,T}^{-s-d}$, respectively, and

$$\begin{aligned} m_{t,T} &= m\Big(\frac{t}{T}, x_{t-1}^{t-r+1}, m_{t-r,T}^{(t-r+s)}(e_{t-r-1}^{-s}, w) + \sigma_{t-r,T}^{(t-r+s)}(e_{t-r-1}^{-s}, w)e_{t-r}, \dots \\ &\qquad \dots, m_{t-d,T}^{(t-d+s)}(e_{t-d-1}^{-s}, w) + \sigma_{t-d,T}^{(t-d+s)}(e_{t-d-1}^{-s}, w)e_{t-d}\Big) \\ \sigma_{t,T} &= \sigma\Big(\frac{t}{T}, x_{t-1}^{t-r+1}, m_{t-r,T}^{(t-r+s)}(e_{t-r-1}^{-s}, w) + \sigma_{t-r,T}^{(t-r+s)}(e_{t-r-1}^{-s}, w)e_{t-r}, \dots \\ &\qquad \dots, m_{t-d,T}^{(t-d+s)}(e_{t-d-1}^{-s}, w) + \sigma_{t-d,T}^{(t-d+s)}(e_{t-d-1}^{-s}, w)e_{t-d}\Big). \end{aligned}$$

Comments on assumption (E3)

We now show that (E3) is fulfilled for the class of bounded densities f_ε whose first derivative f'_ε is bounded, satisfies $\int |x f'_\varepsilon(x)| dx < \infty$ and declines monotonically to zero for values $|x| > R$ for some constant R.

Proof. W.l.o.g. assume that $R \gg D_0$, D_1 and let $d_1 \geq 0$. We write

$$\begin{aligned} I &:= \int_{-\infty}^{\infty} \big| f_\varepsilon([1+d_0]x + d_1) - f_\varepsilon(x) \big| dx \\ &= \int_{-\infty}^{-(R+d_1)} \dots + \int_{-(R+d_1)}^{-R} \dots + \int_{-R}^{R} \dots + \int_{R}^{\infty} \dots \\ &=: I_1 + I_2 + I_3 + I_4 \end{aligned}$$

and consider the terms $I_1, \dots, I_4$ one after the other. First,

$$I_1 = \int_{-\infty}^{-(R+d_1)} \big| f'_\varepsilon(\bar{x})(d_0 x + d_1) \big| dx,$$

where $\bar{x}$ is some intermediate point between x and $(1+d_0)x + d_1$. Note that for all $x \in (-\infty, -(R+d_1))$, it holds that $\bar{x} \leq -R$ and $\bar{x} \leq x + d_1$. Therefore,

$$\begin{aligned} I_1 &\leq \int_{-\infty}^{-(R+d_1)} |f'_\varepsilon(x+d_1)|(d_0|x| + d_1) dx \\ &= d_0 \int_{-\infty}^{-(R+d_1)} |x f'_\varepsilon(x+d_1)| dx + d_1 \int_{-\infty}^{-(R+d_1)} |f'_\varepsilon(x+d_1)| dx \end{aligned}$$

with

$$\begin{aligned} \int_{-\infty}^{-(R+d_1)} |x f'_\varepsilon(x+d_1)| dx &\leq \int_{-\infty}^{-R} |(y-d_1) f'_\varepsilon(y)| dy \\ &\leq \int_{-\infty}^{-R} |y f'_\varepsilon(y)| dy + d_1 \int_{-\infty}^{-R} |f'_\varepsilon(y)| dy \leq C(1+d_1) \end{aligned}$$

and

$$\int_{-\infty}^{-(R+d_1)} |f'_\varepsilon(x+d_1)|dx \le \int_{-\infty}^{-R} |f'_\varepsilon(y)|dy \le C.$$

From this, it is straightforward to see that $I_1 \le C_{D_0,D_1}(d_0+d_1)$ with some constant C_{D_0,D_1} only depending on D_0 and D_1. Analogous arguments yield that $I_4 \le C_{D_0,D_1}(d_0+d_1)$ as well. Furthermore, it trivially holds that $I_2 \le Cd_1$ and $I_3 \le C_{D_0,D_1}(d_0+d_1)$. This completes the proof. □

In the proof of Theorem 1.4, we will apply assumption (E3) to the following situation. Let σ, σ', m, m' be constants such that $-M \le m, m' \le M$ and $\underline{\Sigma} \le \sigma, \sigma' \le \overline{\Sigma}$ and assume w.l.o.g. that $\sigma \le \sigma'$. Then

$$\begin{aligned}
&\int_{-\infty}^{\infty} \Big| f_\varepsilon\Big(\frac{x+m}{\sigma}\Big) - f_\varepsilon\Big(\frac{x+m'}{\sigma'}\Big)\Big| dx \\
&= \int_{-\infty}^{\infty} \Big| f_\varepsilon\Big(\frac{\sigma'}{\sigma}y + \frac{m-m'}{\sigma}\Big) - f_\varepsilon(y)\Big| dy \\
&= \int_{-\infty}^{\infty} \Big| f_\varepsilon\Big(\Big[1+\frac{\sigma'-\sigma}{\sigma}\Big]y + \frac{m-m'}{\sigma}\Big) - f_\varepsilon(y)\Big| dy.
\end{aligned}$$

We can now apply (E3) with $d_0 = \frac{\sigma'-\sigma}{\sigma}$ and $d_1 = \frac{m-m'}{\sigma}$.

Proof of Theorem 1.1

(i) follows by standard arguments to be found for example in Chen & Chen [3]. (ii) immediately follows with the help of (1.32). For (iii), recall that $X_{t-1,T}^{t-d} = X_{t-1}^{t-d}(0)$ for $t \le 1$. This allows us to write the density of $X_{t-1,T}^{t-d}$ as

$$f_{X_{t-1,T}^{t-d}}(x) = \int f_{X_{t-1,T}^{t-d}|\varepsilon_{t-d-1}^1, X_0^{-d+1}(0)}(x|e,w) \prod_{i=1}^{t-d-1} f_\varepsilon(e_i) f_{X_0^{-d+1}(0)}(w)\; dedw,$$

where $e = e_{t-d-1}^1$ and the conditional density $f_{X_{t-1,T}^{t-d}|\varepsilon_{t-d-1}^1, X_0^{-d+1}(0)}$ can be expressed in terms of the error density f_ε with the help of (1.32). □

Proof of Theorem 1.2

We use the triangle inequality to get

$$\big|X_{t,T} - X_t(u)\big| \le \Big|X_{t,T} - X_t\Big(\frac{t}{T}\Big)\Big| + \Big|X_t\Big(\frac{t}{T}\Big) - X_t(u)\Big|$$

and bound the terms $|X_{t,T} - X_t(\frac{t}{T})|$ and $|X_t(\frac{t}{T}) - X_t(u)|$ separately. In what follows, we give a detailed exposition for the term $|X_t(\frac{t}{T}) - X_t(u)|$. The arguments for $|X_{t,T} - X_t(\frac{t}{T})|$ are very similar and shortly summarized at the end of the proof. To keep notation simple, we use the shorthands $\underline{X}_{t,T} = X_{t,T}^{t-d+1}$, $\underline{X}_t(u) = X_t^{t-d+1}(u)$ and $\underline{\varepsilon}_t = \varepsilon_t^{t-d+1}$. We proceed in several steps.

Backward Iteration

Using the smoothness conditions on m and σ, we can write

$$\begin{aligned}\Big|X_t\Big(\frac{t}{T}\Big) - X_t(u)\Big| &\leq C\Big|\frac{t}{T} - u\Big|\big(1 + |\varepsilon_t|\big) \\ &\quad + \sum_{j=1}^{d} \big(\Delta_j^m + \Delta_j^\sigma |\varepsilon_t|\big)\Big|X_{t-j}\Big(\frac{t}{T}\Big) - X_{t-j}(u)\Big| \end{aligned} \tag{1.33}$$

with $\Delta_j^m = |\Delta_j^m(u, \underline{X}_{t-1}(u), \underline{X}_{t-1}(\frac{t}{T}))|$ and $\Delta_j^\sigma = |\Delta_j^\sigma(u, \underline{X}_{t-1}(u), \underline{X}_{t-1}(\frac{t}{T}))|$ as introduced in (1.28). Iterating (1.33) yields

$$\Big|X_t\Big(\frac{t}{T}\Big) - X_t(u)\Big| \leq \Big|\frac{t}{T} - u\Big| V_{t,T,n}(u) + R_{t,T,n}(u)$$

with

$$\begin{aligned} V_{t,T,n}(u) &= C \sum_{r=0}^{n-1} \sum_{j_1,\ldots,j_r=1}^{d} \prod_{l=1}^{r} \big(\Delta_{j_l}^m + \Delta_{j_l}^\sigma |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\big(1 + |\varepsilon_{t-\sum_{k=0}^{r} j_k}|\big) \\ R_{t,T,n}(u) &= \sum_{j_1,\ldots,j_n=1}^{d} \prod_{l=1}^{n} \big(\Delta_{j_l}^m + \Delta_{j_l}^\sigma |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\Big|X_{t-\sum_{k=0}^{n} j_k}\Big(\frac{t}{T}\Big) - X_{t-\sum_{k=0}^{n} j_k}(u)\Big|, \end{aligned}$$

where $j_0 = 0$, $\Delta_{j_l}^m = |\Delta_{j_l}^m(u, \underline{X}_{t-\sum_{k=0}^{l-1} j_k - 1}(u), \underline{X}_{t-\sum_{k=0}^{l-1} j_k - 1}(\frac{t}{T}))|$ and $\Delta_{j_l}^\sigma$ is defined analogously. In what follows, we show that

(R) $R_{t,T,n}(u) \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad n \to \infty$

(V) $V_{t,T,n}(u) \leq V_{t,T}(u)$ with $\mathbb{E}[V_{t,T}(u)^\rho] \leq C$ for some $\rho > 0$.

These two claims imply that

$$\Big|X_t\Big(\frac{t}{T}\Big) - X_t(u)\Big| \leq \Big|\frac{t}{T} - u\Big| V_{t,T}(u) \quad \text{a.s.}$$

with variables $V_{t,T}(u)$ whose ρ-th moment is uniformly bounded by some constant C. Deriving an analogous result for the term $|X_{t,T} - X_t(\frac{t}{T})|$ completes the proof.

Proof of (R)

Define $\Delta_{j_l} := \Delta(\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1})$ with $\Delta(\underline{\varepsilon}_{t-s}) = \Delta I(\|\underline{\varepsilon}_{t-s}\|_\infty \leq K_2) + \delta I(\|\underline{\varepsilon}_{t-s}\|_\infty > K_2)$ as introduced in (1.31). As shown in the preliminaries section of the appendix,

$\Delta_{j_l}^m \leq \Delta_{j_l}$ and analogously $\Delta_{j_l}^\sigma \leq \Delta_{j_l}$. Using the boundedness of m and σ, this allows us to write

$$R_{t,T,n}(u) \leq R_{t,n} := C \sum_{j_1,\ldots,j_n=1}^{d} \prod_{l=1}^{n} \Delta_{j_l} \big(1 + |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\big(1 + |\varepsilon_{t-\sum_{k=0}^{n} j_k}|\big).$$

If the terms Δ_{j_l} were bounded by a sufficiently small constant, then it would be easy to show that $R_{t,n}$ is contracting and converges almost surely to zero as $n \to \infty$. In our case, however, the terms Δ_{j_l} may become rather large, depending on which values the variables $\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1}$ take. If too many of the terms $\Delta_{j_1}, \ldots, \Delta_{j_n}$ are large, then $R_{t,n}$ will explode rather than converge to zero.

In what follows, we show that this problematic case is asymptotically negligible in the sense that it almost surely does not occur. To do so, we exploit the fact that $\Delta_{j_l} \leq \delta$, whenever $\|\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1}\|_\infty > K_2$. Thus, the problematic case that too many terms $\Delta_{j_1}, \ldots, \Delta_{j_n}$ are large can only arise if $\|\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1}\|_\infty \leq K_2$ for too many indices $l = 1, \ldots, n$.

We now introduce indicator functions which allow us to distinguish formally between the problematic case with $\|\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1}\|_\infty \leq K_2$ for too many indices and the complementary unproblematic case. This is done as follows: The term $R_{t,n}$ depends on the residuals $\varepsilon_{t-1}, \ldots, \varepsilon_{t-nd}$. These can be split up into n blocks of d successive variables, i.e. into random vectors $\underline{\varepsilon}_{t-(l-1)d} = \varepsilon_{t-(l-1)d-1}^{t-ld}$ for $l = 1, \ldots, n$. For each block $l = 1, \ldots, n$, we define the indicator functions

$$I_{l,\leq} = I\big(|\varepsilon_{t-(l-1)d-i}| \leq K_2 \text{ for some } i = 1, \ldots, d\big)$$
$$I_{l,\leq}^c = I\big(|\varepsilon_{t-(l-1)d-i}| > K_2 \text{ for all } i = 1, \ldots, d\big),$$

where evidently $I_{l,\leq} + I_{l,\leq}^c = 1$. Additionally, let $n_d = \lceil \frac{n}{d} \rceil$ and define

$$I_n = I\Big(\sum_{l=1}^{n_d} I_{l,\leq} > \kappa n_d\Big) \quad \text{and} \quad I_n^c = I\Big(\sum_{l=1}^{n_d} I_{l,\leq}^c \geq (1-\kappa) n_d\Big),$$

where κ is a constant with $0 < \kappa < 1$ to be specified later on. Note that again $I_n + I_n^c = 1$ holds.

The two indicator functions I_n and I_n^c allow us to discriminate between the problematic and the unproblematic case. In particular, $I_n = 1$ represents the case in which $\|\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1}\|_\infty \leq K_2$ for too many indices, whereas $I_n^c = 1$ indicates the unproblematic case. More specifically, if $I_n^c = 1$, then at least $\lceil (1-\kappa) n_d \rceil$ among the n terms $\underline{\varepsilon}_{t-\sum_{k=0}^{l-1} j_k - 1}$ have a supremum norm larger than K_2. This can be seen as follows:

(1) If I_n^c equals one, then among the first n_d blocks of residuals there are at least $\lceil (1-\kappa)n_d \rceil$ blocks in which all elements are larger than K_2 in absolute value.

(2) Regardless of whether I_n^c equals zero or one, each vector $\underline{\varepsilon}_{t-\sum_{k=1}^{l-1} j_k - 1}$ either coincides with a block of residuals or it covers part of two successive blocks. Moreover, for any tuple $(j_1, \ldots, j_n)$ of indices, at least n_d among the n vectors $\underline{\varepsilon}_{t-\sum_{k=1}^{l-1} j_k - 1}$ have an element in common with one of the first n_d blocks of residuals. These vectors can be chosen such that a different block corresponds to each vector.

(3) Combining (1) and (2) yields that for any tuple $(j_1, \ldots, j_n)$ of indices, there are at least $\lceil (1-\kappa)n_d \rceil$ among the n terms $\underline{\varepsilon}_{t-\sum_{k=1}^{l-1} j_k - 1}$ that have a supremum norm larger than K_2 if $I_n^c = 1$. Hence, at least $\lceil (1-\kappa)n_d \rceil$ terms among $\Delta_{j_1}, \ldots, \Delta_{j_n}$ are bounded by δ if $I_n^c = 1$.

We now use the indicators I_n and I_n^c to decompose the variables $R_{t,n}$ into two parts:

$$\begin{aligned} R_{t,n} &= C I_n \sum_{j_1,\ldots,j_n=1}^{d} \prod_{l=1}^{n} \Delta_{j_l} \big(1 + |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\big(1 + |\varepsilon_{t-\sum_{k=0}^{n} j_k}|\big) \\ &\quad + C I_n^c \sum_{j_1,\ldots,j_n=1}^{d} \prod_{l=1}^{n} \Delta_{j_l} \big(1 + |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\big(1 + |\varepsilon_{t-\sum_{k=0}^{n} j_k}|\big) \\ &=: R_{t,n}^{(1)} + R_{t,n}^{(2)}. \end{aligned}$$

In order to handle the term $R_{t,n}^{(1)}$, we show that

$$\sum_{n=0}^{\infty} \mathbb{P}\Big(R_{t,n}^{(1)} > \phi_n\Big) < \infty, \tag{1.34}$$

where $\{\phi_n\}$ is any null sequence with $\phi_n > 0$ for all n. By the Borel-Cantelli lemma, this implies that $R_{t,n}^{(1)} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. To prove (1.34), we write

$$\begin{aligned} \mathbb{P}\Big(R_{t,n}^{(1)} > \phi_n\Big) \le \mathbb{P}\big(I_n > 0\big) &= \mathbb{P}\Big(\sum_{l=1}^{n_d} I_{l,\le} > \kappa n_d\Big) \\ &= \mathbb{P}\Big(\sum_{l=1}^{n_d} (I_{l,\le} - \mathbb{E}[I_{l,\le}]) > (\kappa - \mathbb{E}[I_{l,\le}]) n_d\Big) \\ &= \mathbb{P}\Big(\sum_{l=1}^{n_d} (I_{l,\le} - \mathbb{E}[I_{l,\le}]) > \kappa_0 n_d\Big) \end{aligned}$$

with $\kappa_0 := \kappa - \mathbb{E}[I_{l,\leq}]$. As the variables ε_t have an everywhere positive density by assumption, the expectation $\mathbb{E}[I_{l,\leq}]$ is strictly smaller than one. We can thus choose $0 < \kappa < 1$ slightly larger than $\mathbb{E}[I_{l,\leq}]$ to get that $0 < \kappa_0 < 1$. As the variables $I_{l,\leq} - \mathbb{E}[I_{l,\leq}]$ are independent for $l = 1, \dots, n$, we can apply Hoeffding's inequality to get

$$\mathbb{P}\Big(\sum_{l=1}^{n_d} (I_{l,\leq} - \mathbb{E}[I_{l,\leq}]) > \kappa_0 n_d \Big) \leq 2 \exp\Big(- \frac{\kappa_0^2 n_d}{2} \Big).$$

Putting everything together, we obtain

$$\mathbb{P}\Big(R_{t,n}^{(1)} > \phi_n \Big) \leq 2 \exp\Big(- \frac{\kappa_0^2 n_d}{2} \Big) \leq C \gamma^n \tag{1.35}$$

with some constant $\gamma < 1$. This shows (1.34).
The term $R_{t,n}^{(2)}$ is easier to handle. It is unequal to zero only if $I_n^c = 1$. Recalling (1)–(3) from above, we thus have the following: Whenever $R_{t,n}^{(2)} \neq 0$, for any tuple $(j_1, \dots, j_n)$ at least $\lceil (1-\kappa) n_d \rceil$ of the n terms $\Delta_{j_1}, \dots, \Delta_{j_n}$ are bounded by δ. Note that there are $\binom{n}{\lceil (1-\kappa) n_d \rceil}$ possibilities to pick $\lceil (1-\kappa) n_d \rceil$ out of n terms. Moreover, by Stirling's formula, it holds that

$$\binom{n}{\lceil (1-\kappa) n_d \rceil} \leq \binom{n}{n/2} \leq 2^n.$$

These considerations yield that

$$\begin{aligned}
\mathbb{E}[R_{t,n}^{(2)}] &= C \sum_{j_1,\dots,j_n=1}^{d} \mathbb{E}\Big[I_n^c \sum_{j_1,\dots,j_n=1}^{d} \prod_{l=1}^{n} \Delta_{j_l} \big(1 + |\varepsilon_{t - \sum_{k=0}^{l-1} j_k}|\big) \big(1 + |\varepsilon_{t - \sum_{k=0}^{n} j_k}|\big) \Big] \\
&\leq C \sum_{j_1,\dots,j_n=1}^{d} 2^n \big(\delta(1 + \mathbb{E}|\varepsilon_0|)\big)^{\lceil (1-\kappa) n_d \rceil} \big(\Delta(1 + \mathbb{E}|\varepsilon_0|)\big)^{n - \lceil (1-\kappa) n_d \rceil} \\
&\leq C \big[2d \, \delta^{\frac{1-\kappa}{d}} \Delta^{1 - \frac{1-\kappa}{d}} (1 + \mathbb{E}|\varepsilon_0|) \big]^n \leq \gamma^n,
\end{aligned} \tag{1.36}$$

where the constant γ can be chosen strictly smaller than one if

$$\delta < \Big(2d \Delta^{1 - \frac{1-\kappa}{d}} (1 + \mathbb{E}|\varepsilon_0|) \Big)^{-\frac{d}{1-\kappa}}. \tag{1.37}$$

Choosing $\phi_n = n^{-p}$ with some $p > 0$, this implies that

$$\mathbb{P}(R_{t,n}^{(2)} > \phi_n) \leq \frac{\mathbb{E}[R_{t,n}^{(2)}]}{\phi_n} \leq n^p \gamma^n$$

with some constant $\gamma < 1$. Thus, $\sum_{n=1}^{\infty} \mathbb{P}(R_{t,n}^{(2)} > \phi_n) < \infty$ and the Borel Cantelli lemma yields $R_{t,n}^{(2)} \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$. This completes the proof of (R).
Note that under our assumptions, an even stronger result than (R) holds. In particular, there exists $\rho > 0$ with

$$\mathbb{E}[(R_{t,T,n}(u))^{\rho}] \leq C\gamma^n \tag{1.38}$$

for some fixed $\gamma < 1$ and $C < \infty$. We show (1.38) for some $0 < \rho < 1$. For the proof, we again use the bound $R_{t,T,n}(u) \leq R_{t,n}$ and show that $\mathbb{E}[(R_{t,n})^{\rho}] \leq C\gamma^n$. In (1.36), we have already seen that $\mathbb{E}[R_{t,n}^{(2)}] \leq C\gamma^n$ with $\gamma < 1$. It thus remains to show that $\mathbb{E}[(R_{t,n}^{(1)})^{\rho}] \leq C\gamma^n$ for some $0 < \rho < 1$. Letting $\phi_n = (d\delta)^n$, we can write

$$E[(R_{t,n}^{(1)})^{\rho}] = E\big[(R_{t,n}^{(1)})^{\rho} I(R_{t,n}^{(1)} < \phi_n)\big] + E\big[(R_{t,n}^{(1)})^{\rho} I(R_{t,n}^{(1)} \geq \phi_n)\big].$$

First note that $E\big[(R_{t,n}^{(1)})^{\rho} I(R_{t,n}^{(1)} < \phi_n)\big] = 0$, as $R_{t,n}^{(1)}$ can become strictly smaller than ϕ_n only if $I_n = 0$. Furthermore, applying the Cauchy-Schwarz inequality together with (1.35) yields

$$\begin{aligned} E\big[(R_{t,n}^{(1)})^{\rho} I(R_{t,n}^{(1)} \geq \phi_n)\big] &\leq C\big(d\Delta(1+\mathbb{E}|\varepsilon_0|)\big)^{\rho n} \sqrt{\mathbb{P}(R_{t,n}^{(1)} \geq \phi_n)} \\ &\leq C\big(d\Delta(1+\mathbb{E}|\varepsilon_0|)\big)^{\rho n} \sqrt{C\gamma^n} \leq C\tilde{\gamma}^n \end{aligned}$$

with some constant $\tilde{\gamma} < 1$ if $\rho > 0$ is chosen sufficiently small. This shows (1.38).

Proof of (V)

We next turn to the variables $V_{t,T,n}(u)$. First note that $V_{t,T,n}(u) \leq V_{t,T}(u)$ with

$$V_{t,T}(u) = C \sum_{r=0}^{\infty} \sum_{j_1,\dots,j_r=1}^{d} \prod_{l=1}^{r} \big(\Delta_{j_l}^{m} + \Delta_{j_l}^{\sigma} |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\big(1 + |\varepsilon_{t-\sum_{k=0}^{r} j_k}|\big).$$

Similar to before, we have that $V_{t,T}(u) \leq V_t$ with

$$V_t = C \sum_{r=0}^{\infty} \sum_{j_1,\dots,j_r=1}^{d} \prod_{l=1}^{r} \Delta_{j_l} \big(1 + |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\big(1 + |\varepsilon_{t-\sum_{k=0}^{r} j_k}|\big) = \sum_{r=0}^{\infty} R_{t,r}.$$

Letting $\rho < 1$ and using the fact that $\mathbb{E}[(R_{t,r})^{\rho}] \leq C\gamma^r$, we obtain that

$$\mathbb{E}[V_t^{\rho}] \leq \sum_{r=0}^{\infty} \mathbb{E}[(R_{t,n})^{\rho}] \leq C \sum_{r=0}^{\infty} \gamma^r < \infty.$$

As a result, $\mathbb{E}[V_{t,T}^{\rho}(u)] \leq \mathbb{E}[V_t^{\rho}] < C$.

Outline of the arguments for $|X_{t,T} - X_t(\frac{t}{T})|$

Similary to before, we can derive the expansion

$$\Big| X_{t,T} - X_t\Big(\frac{t}{T}\Big)\Big| \leq \frac{C}{T} W_{t,T,n} + R_{t,T,n}$$

with

$$W_{t,T,n} = \sum_{r=1}^{n-1} \sum_{j_1,\ldots,j_r=1}^{d} \Big[\sum_{l=1}^{r} j_l \prod_{l=1}^{r} \big(\Delta_{j_l}^m + \Delta_{j_l}^\sigma |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\Big] \big(1 + |\varepsilon_{t-\sum_{k=0}^{r} j_k}|\big)$$

$$R_{t,T,n} = \sum_{j_1,\ldots,j_n=1}^{d} \prod_{l=1}^{n} \big(\Delta_{j_l}^m + \Delta_{j_l}^\sigma |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big) \Big| X_{t-\sum_{k=1}^{n} j_k,T} - X_{t-\sum_{k=1}^{n} j_k}\Big(\frac{t}{T}\Big)\Big|,$$

where now $\Delta_{j_l}^m = \Delta_{j_l}^m(\frac{t}{T}, \underline{X}_{t-\sum_{k=0}^{l-1} j_k - 1}(\frac{t}{T}), \underline{X}_{t-\sum_{k=0}^{l-1} j_k - 1,T})$ and $\Delta_{j_l}^\sigma$ is defined analogously. By the same arguments as above, we can show that

$$\Big| X_{t,T} - X_t\Big(\frac{t}{T}\Big)\Big| \leq \frac{C}{T} W_{t,T} \quad \text{a.s.}$$

with variables

$$W_{t,T} = \sum_{r=1}^{\infty} \sum_{j_1,\ldots,j_r=1}^{d} \Big[\sum_{l=1}^{r} j_l \prod_{l=1}^{r} \big(\Delta_{j_l}^m + \Delta_{j_l}^\sigma |\varepsilon_{t-\sum_{k=0}^{l-1} j_k}|\big)\Big] \big(1 + |\varepsilon_{t-\sum_{k=0}^{r} j_k}|\big),$$

whose ρ-th moments are uniformly bounded for some $\rho > 0$. □

Proof of Theorem 1.3

Throughout the proof, we use the following notation: x_j and y_j are values of the variables $X_{t-j}(u)$ and $X_{t-d-j}(u)$ for $j = 1, \ldots, d$. Moreover, we write $x = (x_1, \ldots, x_d)$ together with $y = (y_1, \ldots, y_d)$ and define

$$\begin{aligned} F_u &: \ \text{distribution function of } X_{t-d-1}^{t-2d}(u) \\ F_{u,v} &: \ \text{joint distribution function of } X_{t-d-1}^{t-2d}(u) \text{ and } X_{t-d-1}^{t-2d}(v) \\ f_u(x) &: \ \text{density of } X_{t-1}^{t-d}(u) \text{ at } x \\ f_u(x|y) &: \ \text{density of } X_{t-1}^{t-d}(u) \text{ at } x \text{ conditional on } X_{t-d-1}^{t-2d}(u) = y. \end{aligned}$$

In addition, we let $f_{u,j} = f_u(x_j|x_{j+1}, \ldots, x_d, y_1, \ldots, y_j)$ denote the conditional density of $X_{t-j}(u)$ given $X_{t-j-1}^{t-j-d}(u)$. Note that

$$f_u(x|y) = \prod_{j=1}^{d} f_u(x_j|x_{j+1}, \ldots, x_d, y_1, \ldots, y_j).$$

Moreover, the conditional densities $f_{u,j}$ can be expressed in terms of the error density according to

$$f_{u,j} = \frac{1}{\sigma_{u,j}} f_\varepsilon\Big(\frac{x_j - m_{u,j}}{\sigma_{u,j}}\Big), \tag{1.39}$$

where we have used the shorthands $m_{u,j} = m(u,(x_{j+1},\dots,x_d,y_1,\dots,y_j))$ and $\sigma_{u,j} = \sigma(u,(x_{j+1},\dots,x_d,y_1,\dots,y_j))$.
With this notation at hand, we can now analyze the term $|f_u(x) - f_v(x)|$. Letting $z = (z_1,\dots,z_d)$ be some value taken by the random vector $(X_{t-d-1}(v),\dots$ $\dots,X_{t-2d}(v))$, it holds that

$$\begin{aligned}
\big|f_u(x) - f_v(x)\big| &= \Big|\int_{\mathbb{R}^d} f_u(x|y)dF_u(y) - \int_{\mathbb{R}^d} f_v(x|z)dF_v(z)\Big| \\
&= \Big|\int_{\mathbb{R}^{2d}} \big[f_u(x|y) - f_v(x|z)\big]dF_{u,v}(y,z)\Big| \\
&= \Big|\int_{\mathbb{R}^{2d}} \Big[\sum_{k=1}^{d}\prod_{j=1}^{k-1} f_{v,j}\big[f_{u,k} - f_{v,k}\big]\prod_{j=k+1}^{d} f_{u,j}\Big]dF_{u,v}(y,z)\Big| \\
&\le \sum_{k=1}^{d}\int_{\mathbb{R}^{2d}} \big|f_{u,k} - f_{v,k}\big|dF_{u,v}(y,z) =: \sum_{k=1}^{d} Q^k_{u,v}(x),
\end{aligned}$$

where the third line is by a telescoping argument and the fourth line follows from the boundedness of f_ε. Furthermore, using the boundedness of m, σ, and f_ε yields

$$\begin{aligned}
Q^k_{u,v}(x) &= \int_{\mathbb{R}^{2d}} \Big|\frac{1}{\sigma_{u,k}} f_\varepsilon\Big(\frac{x_k - m_{u,k}}{\sigma_{u,k}}\Big) - \frac{1}{\sigma_{v,k}} f_\varepsilon\Big(\frac{x_k - m_{v,k}}{\sigma_{v,k}}\Big)\Big|dF_{u,v}(y,z) \\
&\le C\int_{\mathbb{R}^{2d}} \Big|f_\varepsilon\Big(\frac{x_k - m_{u,k}}{\sigma_{u,k}}\Big) - f_\varepsilon\Big(\frac{x_k - m_{v,k}}{\sigma_{v,k}}\Big)\Big|dF_{u,v}(y,z) \\
&\quad + C\int_{\mathbb{R}^{2d}} \big|\sigma_{u,k} - \sigma_{v,k}\big|dF_{u,v}(y,z) \\
&=: Q^{k,1}_{u,v}(x) + Q^{k,2}_{u,v}(x).
\end{aligned}$$

Exploiting the Lipschitz continuity of f_ε together with the smoothness conditions on m and σ, we obtain

$$\begin{aligned}
Q^{k,1}_{u,v}(x) &\le C(1+|x_k|)\int \big(|u-v| + |y_1 - z_1| + \ldots + |y_k - z_k|\big)dF_{u,v}(y,z) \\
&= C(1+|x_k|)\Big(|u-v| + \sum_{j=1}^{k}\mathbb{E}\big|X_{t-d-j}(u) - X_{t-d-j}(v)\big|\Big)
\end{aligned}$$

and analogously

$$Q_{u,v}^{k,2}(x) \leq C\Big(|u-v| + \sum_{j=1}^{k} \mathbb{E}\big|X_{t-d-j}(u) - X_{t-d-j}(v)\big|\Big).$$

As an intermediate result, we have thus shown that

$$\big|f_u(x) - f_v(x)\big| \leq C(1 + \|x\|_1)\Big(|u-v| + \mathbb{E}\big|X_t(u) - X_t(v)\big|\Big), \tag{1.40}$$

where $C < \infty$ is some sufficiently large constant and $\|\cdot\|_1$ denotes the usual l_1-norm for $\mathbb{R}^d$-valued vectors.
In the remainder of the proof, we derive a bound for the expression $\mathbb{E}|X_t(u) - X_t(v)|$. As shown in the proof of Theorem 1.2,

$$\big|X_t(u) - X_t(v)\big| \leq |u-v| U_t(u,v)$$

with random variables $U_t(u,v)$ having the property that $\mathbb{E}[U_t(u,v)^\rho] < C$ for some $\rho > 0$. Letting q be a constant with $0 < q < \rho$, we arrive at

$$\begin{aligned}\mathbb{E}\big|X_t(u) - X_t(v)\big| &= \mathbb{E}\Big[\big|X_t(u) - X_t(v)\big| I\Big(U_t(u,v) \leq \frac{C}{|u-v|^q}\Big)\Big] \\ &\quad + \mathbb{E}\Big[\big|X_t(u) - X_t(v)\big| I\Big(U_t(u,v) > \frac{C}{|u-v|^q}\Big)\Big] \\ &=: E_1(u,v) + E_2(u,v)\end{aligned}$$

with

$$E_1(u,v) \leq |u-v| \; \mathbb{E}\Big[U_t(u,v) I\Big(U_t(u,v) \leq \frac{C}{|u-v|^q}\Big)\Big] \leq C|u-v|^{1-q}. \tag{1.41}$$

Moreover, since $|X_t(u) - X_t(v)| \leq C(1 + |\varepsilon_t|)$ and

$$\begin{aligned}\mathbb{E}\Big[I\Big(U_t(u,v) > \frac{C}{|u-v|^q}\Big)\Big] &\leq \mathbb{E}\Big[\Big(\frac{U_t(u,v)}{C|u-v|^{-q}}\Big)^\rho I\Big(U_t(u,v) > \frac{C}{|u-v|^q}\Big)\Big] \\ &\leq C|u-v|^{q\rho},\end{aligned}$$

we can apply Hölder's inequality to get

$$E_2(u,v) \leq C|u-v|^r \tag{1.42}$$

for some $r > 0$. Plugging (1.41) and (1.42) into (1.40) completes the proof. □

Proof of Theorem 1.4

To start with, note that the process $\{X_{t,T}\}$ is d-Markovian. This implies that

$$\beta(k) = \sup_{T\in\mathbb{Z}} \sup_{t\in\mathbb{Z}} \beta\big(\sigma(X_{t-k,T}^{t-k-d+1}), \sigma(X_{t+d-1,T}^{t})\big)$$

with

$$\beta\big(\sigma(X_{t-k,T}^{t-k-d+1}), \sigma(X_{t+d-1,T}^{t})\big) = \mathbb{E} \sup_{C\in\sigma(X_{t+d-1,T}^{t})} \big|\mathbb{P}(C) - \mathbb{P}(C|\sigma(X_{t-k,T}^{t-k-d+1})\big|.$$

In the following, we bound the expression $|\mathbb{P}(C) - \mathbb{P}(C|\sigma(X_{t-k,T}^{t-k-d+1})|$ for arbitrary sets $C \in \sigma(X_{t+d-1,T}^{t})$. As will be seen, this provides us with a bound for the mixing coefficients $\beta(k)$ of the process $\{X_{t,T}\}$.
We use the following notation: Throughout the proof, we let x_{t+j}, x_{t+j-1}^{t}, e, and z be values of $X_{t+j,T}$, $X_{t+j-1,T}^{t}$, $\varepsilon_{t-1}^{t-k+1}$ and $X_{t-k,T}^{t-k-d+1}$, respectively. Moreover, we use the shorthand

$$f_{X_{t+j,T}}^{\text{cond}}(z) := f_{X_{t+j,T}|X_{t+j-1,T}^{t},\varepsilon_{t-1}^{t-k+1},X_{t-k,T}^{t-k-d+1}}(x_{t+j}|x_{t+j-1}^{t}, e, z).$$

Finally, note that by (1.32), the above conditional density can be expressed in terms of the error density f_ε according to

$$f_{X_{t+j,T}}^{\text{cond}}(z) = \frac{1}{\sigma_{t,T,j}(z)} f_\varepsilon\Big(\frac{x_{t+j} - m_{t,T,j}(z)}{\sigma_{t,T,j}(z)}\Big) \tag{1.43}$$

with

$$\begin{aligned} m_{t,T,j}(z) = m\Big(&\frac{t+j}{T}, x_{t+j-1}^{t}, m_{t-1,T}^{(k-2)}(e_{t-2}^{t-k+1}, z) + \sigma_{t-1,T}^{(k-2)}(e_{t-2}^{t-k+1}, z)e_{t-1}, \ldots \\ &\ldots, m_{t+j-d,T}^{(k-j+d-1)}(e_{t+j-d-1}^{t-k+1}, z) + \sigma_{t+j-d,T}^{(k-j+d-1)}(e_{t+j-d-1}^{t-k+1}, z)e_{t+j-d}\Big) \\ \sigma_{t,T,j}(z) = \sigma\Big(&\frac{t+j}{T}, x_{t+j-1}^{t}, m_{t-1,T}^{(k-2)}(e_{t-2}^{t-k+1}, z) + \sigma_{t-1,T}^{(k-2)}(e_{t-2}^{t-k+1}, z)e_{t-1}, \ldots \\ &\ldots, m_{t+j-d,T}^{(k-j+d-1)}(e_{t+j-d-1}^{t-k+1}, z) + \sigma_{t+j-d,T}^{(k-j+d-1)}(e_{t+j-d-1}^{t-k+1}, z)e_{t+j-d}\Big). \end{aligned}$$

The recursively defined functions $m_{t-1,T}^{(k-2)}$, $\sigma_{t-1,T}^{(k-2)}, \ldots$ were introduced in the preliminaries section of the appendix. With this notation at hand, we can write

$$\begin{aligned} \mathbb{P}(C) &= \int I(x\in C) f_{X_{t+d-1,T}^{t}}(x)dx \\ &= \int I(x\in C) f_{X_{t+d-1,T}^{t}|\varepsilon_{t-1}^{t-k+1},X_{t-k,T}^{t-k-d+1}}(x|e,z) \prod_{l=1}^{k-1} f_\varepsilon(e_{t-l}) f_{X_{t-k,T}^{t-k-d+1}}(z)dedzdx \\ &= \int I(x\in C) \prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z) \prod_{l=1}^{k-1} f_\varepsilon(e_{t-l}) f_{X_{t-k,T}^{t-k-d+1}}(z)dedzdx \end{aligned}$$

and similarly

$$
\begin{aligned}
&\mathbb{P}(C|\sigma(X_{t-k,T}^{t-k-d+1})) \\
&= \mathbb{E}\big[I(X_{t+d-1,T}^{t} \in C)|X_{t-k,T}^{t-k-d+1}\big] \\
&= \mathbb{E}\Big[\mathbb{E}\big[I(X_{t+d-1,T}^{t} \in C)|\varepsilon_{t-1}^{t-k+1}, X_{t-k,T}^{t-k-d+1}\big]\Big|X_{t-k,T}^{t-k-d+1}\Big] \\
&= \int I(x \in C) f_{X_{t+d-1,T}^{t}|\varepsilon_{t-1}^{t-k+1}, X_{t-k,T}^{t-k-d+1}}(x|e, X_{t-k,T}^{t-k-d+1}) \prod_{l=1}^{k-1} f_{\varepsilon}(e_{t-l}) de dx \\
&= \int I(x \in C) \prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(X_{t-k,T}^{t-k-d+1}) \prod_{l=1}^{k-1} f_{\varepsilon}(e_{t-l}) de dx.
\end{aligned}
$$

Using the shorthand $\underline{X} = X_{t-k,T}^{t-k-d+1}$, we thus obtain

$$
\begin{aligned}
&\big|\mathbb{P}(C) - \mathbb{P}(C|\sigma(\underline{X}))\big| \\
&= \Big|\int I(x \in C)\Big[\prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z) - \prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(\underline{X})\Big] \prod_{l=1}^{k-1} f_{\varepsilon}(e_{t-l}) f_{\underline{X}}(z) de dz dx\Big| \\
&\leq \int \underbrace{\Big[\int \Big|\prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z) - \prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(\underline{X})\Big| dx\Big]}_{=:(*)} \prod_{l=1}^{k-1} f_{\varepsilon}(e_{t-l}) f_{\underline{X}}(z) de dz.
\end{aligned}
$$

We next consider $(*)$ more closely. First note that by a telescoping argument

$$
\begin{aligned}
&\prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z) - \prod_{j=0}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(\underline{X}) \\
&= \sum_{i=0}^{d-1} \prod_{j=0}^{i-1} f_{X_{t+j,T}}^{\text{cond}}(\underline{X}) \big[f_{X_{t+i,T}}^{\text{cond}}(z) - f_{X_{t+i,T}}^{\text{cond}}(\underline{X})\big] \prod_{j=i+1}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z).
\end{aligned}
$$

Using this together with Fubini's theorem, we obtain that

$$
\begin{aligned}
(*) &\leq \sum_{i=0}^{d-1} \int \Big[\prod_{j=0}^{i-1} f_{X_{t+j,T}}^{\text{cond}}(\underline{X}) \big|f_{X_{t+i,T}}^{\text{cond}}(z) - f_{X_{t+i,T}}^{\text{cond}}(\underline{X})\big| \prod_{j=i+1}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z)\Big] dx \\
&= \sum_{i=0}^{d-1} \int \Big[\int \Big[\int \prod_{j=i+1}^{d-1} f_{X_{t+j,T}}^{\text{cond}}(z) dx_{t+d-1} \dots d_{t+i+1}\Big] \\
&\qquad \times \big|f_{X_{t+i,T}}^{\text{cond}}(z) - f_{X_{t+i,T}}^{\text{cond}}(\underline{X})\big| dx_{t+i}\Big] \prod_{j=0}^{i-1} f_{X_{t+j,T}}^{\text{cond}}(\underline{X}) dx_{t+i-1} \dots dx_t
\end{aligned}
$$

$$\leq \sum_{i=0}^{d-1} \int \underbrace{\Big[\int \big|f^{\text{cond}}_{X_{t+i,T}}(z) - f^{\text{cond}}_{X_{t+i,T}}(\underline{X})\big| dx_{t+i}\Big]}_{=:(**)} \prod_{j=0}^{i-1} f^{\text{cond}}_{X_{t+j,T}}(\underline{X}) dx_{t+i-1} \dots dx_t, \tag{1.44}$$

where the last inequality uses the fact that $\int \prod_{j=i+1}^{d-1} f^{\text{cond}}_{X_{t+j,T}}(z) dx_{t+d-1} \dots d_{t+i+1}$ is a conditional probability and thus almost surely bounded by one. Applying (1.43) together with (E3), it is straightforward to see that

$$\begin{aligned}
(**) &= \int \Big| \frac{1}{\sigma_{t,T,i}(z)} f_\varepsilon\Big(\frac{x_{t+i} - m_{t,T,i}(z)}{\sigma_{t,T,i}(z)}\Big) - \frac{1}{\sigma_{t,T,i}(\underline{X})} f_\varepsilon\Big(\frac{x_{t+i} - m_{t,T,i}(\underline{X})}{\sigma_{t,T,i}(\underline{X})}\Big) \Big| dx_{t+i} \\
&\leq C\Big(\big|m_{t,T,i}(z) - m_{t,T,i}(\underline{X})\big| + \big|\sigma_{t,T,i}(z) - \sigma_{t,T,i}(\underline{X})\big|\Big) \\
&\leq C(2M + 2\overline{\Sigma})\Big(\big|m_{t,T,i}(z) - m_{t,T,i}(\underline{X})\big| + \big|\sigma_{t,T,i}(z) - \sigma_{t,T,i}(\underline{X})\big|\Big)^p,
\end{aligned} \tag{1.45}$$

where p is some constant with $0 < p < 1$. Iterating backwards $n < \lfloor \frac{k}{d} \rfloor$ times in the same way as in Theorem 1.2, we can further show that

$$\begin{aligned}
&\big|m_{t,T,i}(z) - m_{t,T,i}(\underline{X})\big| + \big|\sigma_{t,T,i}(z) - \sigma_{t,T,i}(\underline{X})\big| \\
&\qquad \leq C \sum_{j_1=1}^{d-i} \sum_{j_2,\dots,j_n=1}^{d} \prod_{m=1}^{n} \Delta_{j_m} (1 + |e_{t-\sum_{l=1}^{m-1} j_l}|)(1 + |e_{t-\sum_{l=1}^{n} j_l}|),
\end{aligned} \tag{1.46}$$

where $\Delta_{j_m} = \Delta(\underline{e}_{t-\sum_{l=1}^{m-1} j_l - 1})$ as defined in (1.31). In particular, note that Δ_{j_m} only depends on the residual values $e_{t-\sum_{l=1}^{m-1} j_l - 1}, \dots, e_{t-\sum_{l=1}^{m-1} j_l - d}$. Plugging (1.46) into the bound (1.45) for $(**)$ and inserting this into the bound (1.44) for $(*)$, we arrive at

$$(*) \leq C\Big(\sum_{j_1,\dots,j_n=1}^{d} \prod_{m=1}^{n} \Delta_{j_m} (1 + |e_{t-\sum_{l=1}^{m-1} j_l}|)(1 + |e_{t-\sum_{l=1}^{n} j_l}|)\Big)^p.$$

As a consequence,

$$\begin{aligned}
&\big|\mathbb{P}(C) - \mathbb{P}(C|\sigma(\underline{X}))\big| \\
&\leq C \int \Big(\sum_{j_1,\dots,j_n=1}^{d} \prod_{m=1}^{n} \Delta_{j_m} (1 + |e_{t-\sum_{l=1}^{m-1} j_l}|)(1 + |e_{t-\sum_{l=1}^{n} j_l}|)\Big)^p \prod_{l=1}^{k-1} f_\varepsilon(e_{t-l}) de \\
&= C\mathbb{E}\Big[\Big(\sum_{j_1,\dots,j_n=1}^{d} \prod_{m=1}^{n} \Delta_{j_m} (1 + |\varepsilon_{t-\sum_{l=1}^{m-1} j_l}|)(1 + |\varepsilon_{t-\sum_{l=1}^{n} j_l}|)\Big)^p\Big].
\end{aligned}$$

Using the arguments from Theorem 1.2, we can show that for $p > 0$ sufficiently small,

$$\mathbb{E}\Big[\Big(\sum_{j_1,\ldots,j_n=1}^{d}\prod_{m=1}^{n}\Delta_{j_m}(1+|\varepsilon_{t-\sum_{l=1}^{m-1}j_l}|)(1+|\varepsilon_{t-\sum_{l=1}^{n}j_l}|)\Big)^{\rho}\Big] \le \gamma^n$$

with some positive constant $\gamma < 1$. Choosing $n = \lfloor \frac{k}{2d} \rfloor$ for instance, we thus obtain that

$$\big|\mathbb{P}(C) - \mathbb{P}(C|\mathcal{B})\big| \le C\gamma^{\lfloor \frac{k}{2d} \rfloor} \le \tilde{\gamma}^k$$

for some constant $\tilde{\gamma} < 1$. This immediately implies that $\beta(k) \le \tilde{\gamma}^k$. □

Appendix B

In this appendix, we prove the results of Section 1.4. As in Appendix A, C denotes a universal real constant which may take a different value on each occurrence.

Auxiliary Results

Before we come to the proofs of the main theorems, we state some auxiliary results that are needed later on. The first two lemmas describe the asymptotic behaviour of Riemann sums that frequently show up throughout the appendix. The proofs are straightforward and thus omitted.

Lemma B1. *Suppose the kernel K satisfies (C6) and let $I_h = [C_1 h, 1 - C_1 h]$. Then for $k = 0, 1, 2$,*

$$\sup_{u \in I_h}\Big|\frac{1}{Th}\sum_{t=1}^{T}K_h\Big(u - \frac{t}{T}\Big)\Big(\frac{u - \frac{t}{T}}{h}\Big)^k - \int_0^1 \frac{1}{h}K_h(u-\varphi)\Big(\frac{u-\varphi}{h}\Big)^k d\varphi\Big| = O\Big(\frac{1}{Th^2}\Big).$$

Lemma B2. *Suppose K satisfies (C6) and let $g : [0,1] \times \mathbb{R}^d \to \mathbb{R}$, $(u,x) \mapsto g(u,x)$ be continuously differentiable w.r.t. u. Then for any compact subset $S \subseteq \mathbb{R}^d$,*

$$\sup_{u \in I_h, x \in S}\Big|\frac{1}{Th}\sum_{t=1}^{T}K_h\Big(u - \frac{t}{T}\Big)g\Big(\frac{t}{T}, x\Big) - g(u,x)\Big| = O\Big(\frac{1}{Th^2}\Big) + o(h).$$

The next lemma is needed for the proof of Theorem 1.5.

Lemma B3. *Assume that (KA1) and (KA2) hold and that the kernel satisfies (C6). Then for any compact set $S \subseteq \mathbb{R}^d$, there exists a constant $\Theta = \Theta(S) < \infty$ such that for T sufficiently large,*

$$Var(\hat{\psi}(u,x)) \leq \frac{\Theta}{Th^{d+1}}$$

uniformly for $u \in [0,1]$ and $x \in S$.

Proof. Throughout the proof, let $u \in [0,1]$ and $x \in S$. Moreover, define $S^{\bullet} = \{x \in \mathbb{R}^d : \|x - S\|_\infty \leq C_1\}$ with $\|x - S\|_\infty := \min_{y \in S} \|x - y\|_\infty$ and write

$$\hat{\psi}(u,x) - \mathbb{E}[\hat{\psi}(u,x)] = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} \big(Z_{t,T}(u,x) - \mathbb{E}[Z_{t,T}(u,x)]\big)$$

with

$$Z_{t,T} = Z_{t,T}(u,x) = K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) W_{t,T}.$$

With this notation,

$$\begin{aligned} Th^{d+1}\mathrm{Var}(\hat{\psi}(u,x)) &= \frac{1}{Th^{d+1}} \sum_{t=1}^{T} \mathrm{Var}(Z_{t,T}) + \frac{2}{Th^{d+1}} \sum_{l=1}^{T-1} \sum_{t=1}^{T-l} \mathrm{Cov}(Z_{t,T}, Z_{t+l,T}) \\ &=: V_1(u,x) + V_2(u,x). \end{aligned}$$

Following the arguments in Hansen [17], we first derive some preliminary bounds using (KA2):

($*_1$) For $1 \leq r \leq s$,

$$\begin{aligned} \mathbb{E}[|W_{t,T}|^r | X_{t,T} = x] f_{X_{t,T}}(x) &\leq \big(\mathbb{E}[|W_{t,T}|^s | X_{t,T} = x]\big)^{\frac{r}{s}} f_{X_{t,T}}(x) \\ &\leq \big(\mathbb{E}[|W_{t,T}|^s | X_{t,T} = x] f_{X_{t,T}}(x)\big)^{\frac{r}{s}} f_{X_{t,T}}(x)^{\frac{s-r}{s}} \\ &\leq B_1(S) B_0(S). \end{aligned}$$

($*_2$) For $1 \leq r \leq s$,

$$\begin{aligned} \mathbb{E}[|Z_{t,T}|^r] &= K_h^r\Big(u - \frac{t}{T}\Big) \mathbb{E}\Big[\prod_{j=1}^{d} K_h^r(x^j - X_{t,T}^j) \mathbb{E}[|W_{t,T}|^r | X_{t,T}]\Big] \\ &= K_h^r\Big(u - \frac{t}{T}\Big) \int_{\mathbb{R}^d} \prod_{j=1}^{d} K_h^r(x^j - w^j) \mathbb{E}[|W_{t,T}|^r | X_{t,T} = w] f_{X_{t,T}}(w) dw \\ &= h^d K_h^r\Big(u - \frac{t}{T}\Big) \int_{\mathbb{R}^d} K^r(\varphi) \underbrace{\mathbb{E}[|W_{t,T}|^r | X_{t,T} = x - h\varphi] f_{X_{t,T}}(x - h\varphi)}_{\leq B_1(S^{\bullet}) B_0(S^{\bullet}) \text{ by } (*_1) \text{ for } T \text{ sufficiently large}} d\varphi \\ &\leq C h^d K_h^r\Big(u - \frac{t}{T}\Big). \end{aligned}$$

$(*_3)$ For $l \geq 1$,

$$\begin{aligned}
&\mathbb{E}[Z_{t,T} Z_{t+l,T}] \\
&= K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) \mathbb{E}\Big[\prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) \prod_{j=1}^{d} K_h(x^j - X_{t+l,T}^j) \\
&\qquad \times \mathbb{E}\big[|W_{t,T}||W_{t+l,T}|\big| X_{t,T}, X_{t+l,T}\big]\Big] \\
&= K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} K_h(x^j - v^j) \prod_{j=1}^{d} K_h(x^j - w^j) \\
&\qquad \times \underbrace{\mathbb{E}\big[|W_{t,T}||W_{t+l,T}|\big| X_{t,T} = v, X_{t+l,T} = w\big] f_{X_{t,T}, X_{t+l,T}}(v, w)}_{\leq B_2(S^\bullet) \text{ for } T \text{ sufficiently large}} dv dw \\
&\leq C K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) h^{2d}.
\end{aligned}$$

We now bound the covariances $\text{Cov}(Z_{t,T}, Z_{t+l,T})$: Let $\tau_T = \lceil (\log T)^{-1} h^{-d} \rceil$ and distinguish between two cases:

$(*_4)$ Let $l \leq \tau_T$. Then by $(*_2)$ and $(*_3)$,

$$\begin{aligned}
\big|\text{Cov}(Z_{t,T}, Z_{t+l,T})\big| &= \big|\mathbb{E}\big[(Z_{t,T} - \mathbb{E}[Z_{t,T}])(Z_{t+l,T} - \mathbb{E}[Z_{t+l,T}])\big]\big| \\
&\leq \mathbb{E}|Z_{t,T} Z_{t+l,T}| + \mathbb{E}|Z_{t,T}| \mathbb{E}|Z_{t+l,T}| \\
&\leq C K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) h^{2d}.
\end{aligned}$$

$(*_5)$ Let $l \geq \tau_T + 1$. Then by Davydov's inequality and $(*_2)$,

$$\begin{aligned}
\big|\text{Cov}(Z_{t,T}, Z_{t+l,T})\big| &\leq C \alpha(l)^{1-\frac{2}{s}} \big(\mathbb{E}|Z_{t,T}|^s \mathbb{E}|Z_{t+l,T}|^s\big)^{\frac{1}{s}} \\
&\leq C \alpha(l)^{1-\frac{2}{s}} K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) h^{\frac{2d}{s}} \\
&\leq C l^{-[(2-\frac{2}{s}) + \delta(1-\frac{2}{s})]} K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) h^{\frac{2d}{s}}
\end{aligned}$$

with some constant $\delta > 0$. The last inequality follows from the assumption that $\beta > \frac{2s-2}{s-2}$. This means that there exists $\delta > 0$ with $\beta = \frac{2s-2}{s-2} + \delta$. Thus, $\alpha(l)^{1-\frac{2}{s}} \leq l^{-\beta(1-\frac{2}{s})} \leq l^{-[(2-\frac{2}{s}) + \delta(1-\frac{2}{s})]}$.

We are now in a position to bound V_1 and V_2. Using $(*_2)$, we obtain

$$V_1(u, x) \leq \frac{1}{Th^{d+1}} \sum_{t=1}^{T} \mathbb{E}[Z_{t,T}^2] \leq \frac{C}{Th} \sum_{t=1}^{T} K_h^2\Big(u - \frac{t}{T}\Big) \leq C$$

uniformly in u and x. Further, applying $(*_4)$ and $(*_5)$ yields

$$\begin{aligned}
V_2(u,x) &= \frac{2}{Th^{d+1}} \sum_{l=1}^{T-1} \sum_{t=1}^{T-l} \mathrm{Cov}(Z_{t,T}, Z_{t+l,T}) \\
&= \frac{2}{Th^{d+1}} \Big(\sum_{l=1}^{\tau_T} \sum_{t=1}^{T-l} \mathrm{Cov}(Z_{t,T}, Z_{t+l,T}) + \sum_{l=\tau_T+1}^{T-1} \sum_{t=1}^{T-l} \mathrm{Cov}(Z_{t,T}, Z_{t+l,T}) \Big) \\
&\leq \frac{C}{Th^{d+1}} \sum_{l=1}^{\tau_T} \sum_{t=1}^{T-l} K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) h^{2d} \\
&\quad + \frac{C}{Th^{d+1}} \sum_{l=\tau_T+1}^{T-1} \sum_{t=1}^{T-l} K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big) h^{\frac{2d}{s}} l^{-[(2-\frac{2}{s})+\delta(1-\frac{2}{s})]} \\
&=: V_{2,1}(u,x) + V_{2,2}(u,x),
\end{aligned}$$

where

$$\begin{aligned}
|V_{2,1}(u,x)| &\leq C \frac{h^{2d}}{h^d} \sum_{l=1}^{\tau_T} \underbrace{\frac{1}{Th} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big)}_{\leq C \text{ uniformly in } u \text{ and } l} \\
&\leq C \tau_T h^d = (\log T)^{-1} \to 0.
\end{aligned}$$

In order to bound $V_{2,2}(u,x)$, we make use of the fact that for $\eta > 1$ and $k \geq 1$, $\sum_{j=k+1}^{\infty} j^{-\eta} \leq \int_k^{\infty} x^{-\eta} dx = \frac{k^{1-\eta}}{\eta - 1}$. This implies that

$$\begin{aligned}
\sum_{l=\tau_T+1}^{T-1} l^{-[(2-\frac{2}{s})+\delta(1-\frac{2}{s})]} &\leq \frac{\tau_T^{1-[(2-\frac{2}{s})+\delta(1-\frac{2}{s})]}}{[(2-\frac{2}{s})+\delta(1-\frac{2}{s})]-1} \\
&\leq C h^{d-\frac{2d}{s}} \underbrace{(\log T)^{-(1-[(2-\frac{2}{s})+\delta(1-\frac{2}{s})])} h^{\delta d(1-\frac{2}{s})}}_{=: q_T \to 0}.
\end{aligned}$$

Using this, we obtain

$$\begin{aligned}
|V_{2,2}(u,x)| &\leq C \frac{h^{\frac{2d}{s}}}{h^d} \sum_{l=\tau_T+1}^{T-1} l^{-[(2-\frac{2}{s})+\delta(1-\frac{2}{s})]} \underbrace{\frac{1}{Th} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) K_h\Big(u - \frac{t+l}{T}\Big)}_{\leq C \text{ uniformly in } u \text{ and } l} \\
&\leq C q_T \to 0.
\end{aligned}$$

□

Proof of Theorem 1.5

The proof extends Theorem 2 of Hansen [17]. Define

$$B = \{(u,x) \in \mathbb{R}^{d+1} : u \in [0,1], x \in S\} \quad \text{and} \quad \tau_T = \rho_T T^{\frac{1}{s}}$$

with ρ_T slowly diverging to infinity as $T \to \infty$. To simplify the calculations in later parts of the proof, we choose $\rho_T = (\log T)^{\frac{1}{1+\beta}} \phi_T^{(1+\frac{\beta-d}{2})\frac{1}{1+\beta}}$ with $\phi_T = \log\log T$. Defining

$$\hat{\psi}_1(u,x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) W_{t,T} I(|W_{t,T}| \le \tau_T)$$

$$\hat{\psi}_2(u,x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) W_{t,T} I(|W_{t,T}| > \tau_T),$$

we can write

$$\hat{\psi}(u,x) - \mathbb{E}[\hat{\psi}(u,x)] = \big(\hat{\psi}_1(u,x) - \mathbb{E}[\hat{\psi}_1(u,x)]\big) + \big(\hat{\psi}_2(u,x) - \mathbb{E}[\hat{\psi}_2(u,x)]\big).$$

In what follows, we analyze the two terms on the right-hand side. We proceed in several steps.

Step 1: Truncation

With $B = [0,1] \times S$ and $a_T = \sqrt{\frac{\log T}{Th^{d+1}}}$, it holds that

$$\begin{aligned}
\mathbb{P}\Big(\sup_{(u,x)\in B} |\hat{\psi}_2(u,x)| > Ca_T\Big) &\le P\big(|W_{t,T}| > \tau_T \text{ for some } 1 \le t \le T\big) \\
&\le \sum_{t=1}^{T} \mathbb{P}(|W_{t,T}| > \tau_T) \le \sum_{t=1}^{T} \mathbb{E}\Big[\frac{|W_{t,T}|^s}{\tau_T^s} I(|W_{t,T}| > \tau_T)\Big] \\
&\le \tau_T^{-s} \sum_{t=1}^{T} \mathbb{E}|W_{t,T}|^s \le CT\tau_T^{-s} = \rho_T^{-s} \to 0
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\big|\hat{\psi}_2(u,x)\big| &\le \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \int_{\mathbb{R}^d} \prod_{j=1}^{d} K_h(x^j - w^j) \\
&\qquad \times \mathbb{E}\big[|W_{t,T}| I(|W_{t,T}| > \tau_T)\big| X_{t,T} = w\big] f_{X_{t,T}}(w) dw \\
&= \frac{1}{Th} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \int_{\mathbb{R}^d} \prod_{j=1}^{d} K(\varphi^j)
\end{aligned}$$

$$
\begin{aligned}
&\times \mathbb{E}\big[|W_{t,T}| I(|W_{t,T}| > \tau_T)\big| X_{t,T} = x - h\varphi\big] f_{X_{t,T}}(x - h\varphi) d\varphi \\
&\le \frac{1}{Th} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \frac{1}{\tau_T^{s-1}} \int_{\mathbb{R}^d} \prod_{j=1}^{d} K(\varphi^j) \\
&\quad \times \underbrace{\mathbb{E}\big[|W_{t,T}|^s I(|W_{t,T}| > \tau_T)\big| X_{t,T} = x - h\varphi\big] f_{X_{t,T}}(x - h\varphi)}_{\le B_1} d\varphi \\
&\le \frac{C}{\tau_T^{s-1}} \underbrace{\frac{1}{Th} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big)}_{\le C \text{ uniformly in u}} \le \frac{C}{\tau_T^{s-1}} = C \rho_T^{-(s-1)} T^{-\frac{s-1}{s}} \le C a_T
\end{aligned}
$$

with a constant C that does not depend on (u, x). Hence,

$$
\sup_{(u,x)\in B} \big|\hat{\psi}_2(u, x) - \mathbb{E}\hat{\psi}_2(u, x)\big| = O_p(a_T).
$$
[2]

Step 2: Discretization

We cover the region B with $N \le C h^{-(d+1)} a_T^{-(d+1)}$ balls $B_n = \{(u, x) \in \mathbb{R}^{d+1} : \|(u, x) - (u_n, x_n)\|_\infty \le a_T h\}$ and use (u_n, x_n) to denote the midpoint of B_n. Now let $K^*(v) = C \prod_{j=0}^{d} I(|v^j| \le 2C_1)$ for $v \in \mathbb{R}^{d+1}$ and note that for $(u, x) \in B_n$,

$$
\begin{aligned}
&\Big|K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) - K_h\Big(u_n - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x_n^j - X_{t,T}^j)\Big| \\
&\le a_T K_h^*\Big(u_n - \frac{t}{T}, x_n - X_{t,T}\Big)
\end{aligned}
$$

with $K_h^*(v) = K^*(\frac{v}{h})$. Defining

$$
\tilde{\psi}_1(u, x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h^*\Big(u - \frac{t}{T}, x - X_{t,T}\Big) |W_{t,T}| I(|W_{t,T}| \le \tau_T)
$$

and noting that $\mathbb{E}|\tilde{\psi}_1(u, x)| \le C < \infty$, we thus obtain

$$
\begin{aligned}
&\sup_{(u,x)\in B_n} \big|\hat{\psi}_1(u, x) - \mathbb{E}\hat{\psi}_1(u, x)\big| \\
&\le \big|\hat{\psi}_1(u_n, x_n) - \mathbb{E}\hat{\psi}_1(u_n, x_n)\big| + a_T\Big(\big|\tilde{\psi}_1(u_n, x_n)\big| + \mathbb{E}\big|\tilde{\psi}_1(u_n, x_n)\big|\Big) \\
&\le \big|\hat{\psi}_1(u_n, x_n) - \mathbb{E}\hat{\psi}_1(u_n, x_n)\big| + \big|\tilde{\psi}_1(u_n, x_n) - \mathbb{E}\tilde{\psi}_1(u_n, x_n)\big| + 2a_T M
\end{aligned}
$$

[2]Hansen [17] uses the more slowly diverging truncation sequence $\tau_T = a_T^{-1/(s-1)}$. He shows that with this choice of τ_T, it holds that $|\hat{\psi}_2(u, x) - \mathbb{E}\hat{\psi}_2(u, x)| = O_p(a_T)$. It is however not clear at all whether $\sup_{u,x} |\hat{\psi}_2(u, x) - \mathbb{E}\hat{\psi}_2(u, x)| = O_p(a_T)$ in his case, which is needed for the proof. To ensure uniform convergence, we have set $\tau_T = \rho_T T^{1/s}$.

for any $M > \mathbb{E}|\tilde{\psi}_1(u_n,x_n)|$. As a consequence,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{(u,x)\in B}\big|\hat{\psi}_1(u,x)-\mathbb{E}\hat{\psi}_1(u,x)\big| > 4Ma_T\Big)\\
&\leq N\max_{1\leq n\leq N}\mathbb{P}\Big(\sup_{(u,x)\in B_n}\big|\hat{\psi}_1(u,x)-\mathbb{E}\hat{\psi}_1(u,x)\big| > 4Ma_T\Big)\\
&\leq N\max_{1\leq n\leq N}\mathbb{P}\Big(\big|\hat{\psi}_1(u_n,x_n)-\mathbb{E}\hat{\psi}_1(u_n,x_n)\big| > Ma_T\Big) \qquad \text{(A)}\\
&\quad + N\max_{1\leq n\leq N}\mathbb{P}\Big(\big|\tilde{\psi}_1(u_n,x_n)-\mathbb{E}\tilde{\psi}_1(u_n,x_n)\big| > Ma_T\Big). \qquad \text{(B)}
\end{aligned}$$

As the terms (A) and (B) can be bounded in the same way, we restrict attention to (A) in what follows. We use the notation

$$\hat{\psi}_1(u,x)-\mathbb{E}\hat{\psi}_1(u,x) = \frac{1}{Th^{d+1}}\sum_{t=1}^{T} Z_{t,T}(u,x)$$

with $Z_{t,T}(u,x) = K_h(u-\frac{t}{T})\{\prod_{j=1}^d K_h(x^j-X^j_{t,T})W_{t,T}I(|W_{t,T}|\leq\tau_T)-\mathbb{E}[\prod_{j=1}^d K_h(x^j-X^j_{t,T})W_{t,T}I(|W_{t,T}|\leq\tau_T)]\}$. Note that for each fixed (u,x), the array $\{Z_{t,T}(u,x)\}$ is α-mixing with mixing coefficients α_T^Z satisfying $\alpha_T^Z(k)\leq\alpha(k)$.

Step 3: Bounding (A)

We now bound (A) with the help of an exponential inequality by Liebscher (see Theorem 2.1 in [27]).

Lemma (Liebscher)**.** *Let $Z_{t,T}$ be a zero-mean triangular array such that $|Z_{t,T}|\leq b_T$ with strong mixing coefficients $\alpha(k)$. Then for any $\varepsilon>0$ and $S_T\leq T$ with $\varepsilon > 4S_Tb_T$,*

$$\mathbb{P}\Big(\Big|\sum_{t=1}^{T} Z_{t,T}\Big| > \varepsilon\Big) \leq 4\exp\Big(-\frac{\varepsilon^2}{64\sigma^2_{S_T,T}\frac{T}{S_T}+\frac{8}{3}\varepsilon b_T S_T}\Big)+4\frac{T}{S_T}\alpha(S_T),$$

where $\sigma^2_{S_T,T}=\sup_{0\leq j\leq T-1}\mathbb{E}[(\sum_{t=j+1}^{\min\{j+S_T,T\}} Z_{t,T})^2]$.

We apply this exponential inequality as follows to our situation:

- As we are interested in bounding the term

$$\begin{aligned}
\mathbb{P}\Big(\big|\hat{\psi}_1(u,x)-\mathbb{E}\hat{\psi}_1(u,x)\big| > Ma_T\Big) &= \mathbb{P}\Big(\Big|\frac{1}{Th^{d+1}}\sum_{t=1}^{T} Z_{t,T}(u,x)\Big| > Ma_T\Big)\\
&= \mathbb{P}\Big(\Big|\sum_{t=1}^{T} Z_{t,T}(u,x)\Big| > Ma_TTh^{d+1}\Big),
\end{aligned}$$

  we choose $\varepsilon = Ma_TTh^{d+1}$.

- As $|W_{t,T}|I(|W_{t,T}| \leq \tau_T) \leq \tau_T$ and $K_h(u - \frac{t}{T}) \prod_{j=1}^d K_h(x^j - X_{t,T}^j) \leq \bar{C}$, we have that
$$|Z_{t,T}(u,x)| \leq 2\tau_T \bar{C} =: b_T.$$
- By Lemma B3, $\sigma_{S_T,T}^2 \leq \Theta S_T h^{d+1}$ with a constant Θ independent of (u,x).
- It remains to choose S_T in a suitable way. The exponential inequality of Liebscher reads
$$\mathbb{P}\Big(\Big|\sum_{t=1}^T Z_{t,T}(u,x)\Big| > \varepsilon\Big) \leq \underbrace{4\exp\Big(-\frac{\varepsilon^2}{64\sigma_{S_T,T}^2 \frac{T}{S_T} + \frac{8}{3}\varepsilon b_T S_T}\Big)}_{=:(*)} + 4\frac{T}{S_T}\alpha(S_T)$$
with
$$\begin{aligned}(*) &= \exp\Big(-\frac{M^2 a_T^2 T^2 h^{2(d+1)}}{64\Theta S_T h^{d+1}\frac{T}{S_T} + \frac{16\bar{C}}{3}\varepsilon\tau_T S_T}\Big)\\ &= \exp\Big(-\frac{M^2 T h^{d+1}\log T}{64\Theta T h^{d+1} + \frac{16\bar{C}}{3}\varepsilon\tau_T S_T}\Big).\end{aligned}$$
If we choose $S_T = a_T^{-1}\tau_T^{-1}$, then the second term in the denominator becomes
$$\frac{16\bar{C}}{3}\varepsilon\tau_T S_T = \frac{16\bar{C}}{3} M a_T T h^{d+1}\tau_T S_T = \frac{16\bar{C}}{3} M T h^{d+1}$$
and therefore
$$(*) = \exp\Big(-\frac{M^2\log T}{64\theta + \frac{16}{3}M\bar{C}}\Big) = T^{-\frac{M^2}{64\theta + \frac{16}{3}M\bar{C}}}.$$
Hence, we choose $S_T = a_T^{-1}\tau_T^{-1}$.

It is straightforward to see that with these choices, the conditions needed to apply the exponential inequality of Liebscher are fulfilled. For any fixed (u,x) and T sufficiently large, we now get

$$\begin{aligned}&\mathbb{P}\Big(\big|\hat{\psi}_1(u,x) - \mathbb{E}\hat{\psi}_1(u,x)\big| > M a_T\Big)\\ &\leq 4\exp\Big(-\frac{\varepsilon^2}{64\Theta S_T h^{d+1}\frac{T}{S_T} + \frac{8}{3}\varepsilon S_T b_T}\Big) + 4\frac{T}{S_T}\alpha(S_T)\\ &\leq 4\exp\Big(-\frac{M^2\log T}{64\Theta + \frac{16}{3}M\bar{C}}\Big) + 4\frac{T}{S_T}A S_T^{-\beta}\\ &\leq 4\exp\Big(-\frac{M\log T}{64 + 6\bar{C}}\Big) + 4ATS_T^{-1-\beta}\\ &= 4T^{-\frac{M}{64+6\bar{C}}} + 4ATS_T^{-1-\beta},\end{aligned}$$

where we have chosen $M > \Theta$ to get the last inequality. Recalling that $N \leq Ch^{-(d+1)} a_T^{-(d+1)}$, it follows that

$$\mathbb{P}\Big(\sup_{(u,x)\in B} \big|\hat{\psi}_1(u,x) - \mathbb{E}\hat{\psi}_1(u,x)\big| > 4Ma_T \Big) \leq O(R_{1T}) + O(R_{2T})$$

with

$$R_{1T} = h^{-(d+1)} a_T^{-(d+1)} T^{-\frac{M}{64+6C}}$$
$$R_{2T} = h^{-(d+1)} a_T^{-(d+1)} T S_T^{-1-\beta}.$$

As $\frac{\phi_T \log T}{T^\theta h^{d+1}} = o(1)$ by assumption, we obtain

$$R_{1T} = h^{-(d+1)} a_T^{-(d+1)} T^{-\frac{M}{64+6C}} = o\Big(\frac{T^\theta}{\phi_T \log T}\Big)\Big(\frac{Th^{d+1}}{\log T}\Big)^{\frac{d+1}{2}} T^{-\frac{M}{64+6C}} \leq T^{-\eta}$$

for some small $\eta > 0$, if we choose M large enough. Furthermore,

$$\begin{aligned} R_{2T} &= h^{-(d+1)} a_T^{-(d+1)} T (a_T \tau_T)^{1+\beta} \\ &= \Big(\frac{\phi_T \log T}{h^{d+1}}\Big)^{1+\frac{\beta-d}{2}} T^{1-\frac{\beta-d}{2}+\frac{1+\beta}{s}} \\ &= o\big(T^{\theta(1+\frac{\beta-d}{2})+1-\frac{\beta-d}{2}+\frac{1+\beta}{s}}\big) \end{aligned}$$

By our assumptions on θ and β, it holds that $R_{2T} = o(1)$. This shows the result.

□

Proof of Theorem 1.6

We split up the term $\hat{f}(u,x) - f(u,x)$ into a variance part $\hat{f}(u,x) - \mathbb{E}\hat{f}(u,x)$ and a bias part $\mathbb{E}\hat{f}(u,x) - f(u,x)$. For the variance part, we immediately obtain

$$\sup_{u\in[0,1],x\in S} \big|\hat{f}(u,x) - \mathbb{E}\hat{f}(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big)$$

by Theorem 1.5. The rate of the bias part can be derived as follows: As the kernel K is bounded, we can use a telescoping argument to get that

$$\begin{aligned} &\Big| \prod_{j=1}^d K_h(x^j - X_{t,T}^j) - \prod_{j=1}^d K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big| \\ &= \Big| \sum_{k=1}^d \prod_{j=1}^{k-1} K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\big[K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\tfrac{t}{T})\big)\big] \prod_{j=k+1}^d K_h(x^j - X_{t,T}^j)\Big| \\ &\leq C \sum_{k=1}^d \big|K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\tfrac{t}{T})\big)\big|. \end{aligned}$$

Moreover, exploiting again the boundedness of K, there exists a constant $C < \infty$ with $|K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\frac{t}{T})\big)| \le C|K_h(x^k - X_{t,T}^k) - K_h(x^k - X_t^k(\frac{t}{T}))|^r$ for $r = \min\{\rho, 1\}$. Hence,

$$\begin{aligned}
&\Big|\prod_{j=1}^d K_h(x^j - X_{t,T}^j) - \prod_{j=1}^d K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big| \\
&\qquad \le C\sum_{k=1}^d \big|K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\tfrac{t}{T})\big)\big|^r. \qquad (1.47)
\end{aligned}$$

Using (1.47), we obtain

$$\begin{aligned}
&\big|\mathbb{E}\hat{f}(u,x) - f(u,x)\big| \\
&\le \Big|\frac{1}{Th^{d+1}}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big)\mathbb{E}\Big[\prod_{j=1}^d K_h(x^j - X_{t,T}^j) - \prod_{j=1}^d K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big]\Big| \\
&\quad + \Big|\frac{1}{Th^{d+1}}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big)\mathbb{E}\Big[\prod_{j=1}^d K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big] - f(u,x)\Big| \\
&\le \frac{C}{Th^{d+1}}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big)\sum_{k=1}^d \mathbb{E}\big|K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\tfrac{t}{T})\big)\big|^r \\
&\quad + \Big|\frac{1}{Th^{d+1}}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big)\int\prod_{j=1}^d K_h(x^j - w^j) f\Big(\frac{t}{T}, w\Big)dw - f(u,x)\Big| \\
&=: B_1(u,x) + B_2(u,x).
\end{aligned}$$

Since K is Lipschitz, $|X_{t,T}^k - X_t^k(\frac{t}{T})| \le \frac{C}{T}U_{t,T}(\frac{t}{T})$, and $U_{t,T}(\frac{t}{T})$ has finite r-th moment, it holds that

$$B_1(u,x) \le \frac{C}{Th^{d+1}}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big)\sum_{k=1}^d \mathbb{E}\Big|\frac{1}{Th}U_{t,T}(\tfrac{t}{T})\Big|^r \le \frac{C}{T^r h^{d+r}}$$

uniformly for u and x. By the smoothness conditions on f,

$$B_2(u,x) = \Big|\frac{1}{Th}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big) f\Big(\frac{t}{T}, x\Big) - f(u,x)\Big| + o(h)$$

uniformly in u and x. Moreover,

$$\Big|\frac{1}{Th}\sum_{t=1}^T K_h\Big(u - \frac{t}{T}\Big) f\Big(\frac{t}{T}, x\Big) - f(u,x)\Big| = O\Big(\frac{1}{Th^2}\Big) + o(h)$$

uniformly in u and x by Lemma B2. Hence,

$$\sup_{u\in I_h, x\in S} \left|\mathbb{E}\hat{f}(u,x) - f(u,x)\right| = o(h) + O\Big(\frac{1}{T^r h^{d+r}}\Big).$$

□

Proof of Theorem 1.7

We write

$$\begin{aligned}\hat{m}(u,x) - m(u,x) &= \frac{\hat{g}^V(u,x)}{\hat{f}(u,x)} + \frac{\hat{g}^B(u,x)}{\hat{f}(u,x)} - m(u,x) \\ &= \frac{1}{\hat{f}(u,x)}\big(\hat{g}^V(u,x) + \hat{g}^B(u,x) - m(u,x)\hat{f}(u,x)\big)\end{aligned}$$

with

$$\begin{aligned}\hat{g}^V(u,x) &= \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)\varepsilon_{t,T} \\ \hat{g}^B(u,x) &= \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) m\Big(\frac{t}{T}, X_{t,T}\Big).\end{aligned}$$

(a) By Theorem 1.5 with $W_{t,T} = \varepsilon_{t,T}$,

$$\sup_{u\in[0,1], x\in S} \left|\hat{g}^V(u,x)\right| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big).$$

(b) It holds that

$$\begin{aligned}\sup_{u\in[0,1], x\in S} \Big|\hat{g}^B(u,x) &- m(u,x)\hat{f}(u,x) \\ &- \mathbb{E}[\hat{g}^B(u,x) - m(u,x)\hat{f}(u,x)]\Big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big).\end{aligned}$$

This follows by applying Theorem 1.5 to the term $\hat{g}^B(u,x) - m(u,x)\hat{f}(u,x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)\{m(\frac{t}{T}, X_{t,T}) - m(u,x)\}$.

(c) It holds that

$$\begin{aligned}&\sup_{u\in I_h, x\in S} \left|\mathbb{E}[\hat{g}^B(u,x) - m(u,x)\hat{f}(u,x)]\right| \\ &= h^2 \frac{\kappa_2}{2} \sum_{i=0}^{d} \Big(2\partial_i m(u,x)\partial_i f(u,x) + \partial_{i,i}^2 m(u,x) f(u,x)\Big) + O\Big(\frac{1}{T^r h^d}\Big) + o(h^2)\end{aligned}$$

with $r = \min\{\rho, 1\}$. To show this, let $\bar{K} : \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function with support $[-2C_1, 2C_1]$ (or more generally with support $[-qC_1, qC_1]$ for some $q > 1$). Assume that $\bar{K}(x) = 1$ for all $x \in [-C_1, C_1]$ and write $\bar{K}_h(x) = \bar{K}(\frac{x}{h})$. Then

$$\begin{aligned}
\mathbb{E}[\hat{g}^B(u,x) - m(u,x)\hat{f}(u,x)] &= \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \mathbb{E}\Big[\prod_{j=1}^{d} \bar{K}_h(x_j - X_{t,T}^j) \\
&\qquad \times \prod_{j=1}^{d} K_h(x_j - X_{t,T}^j)\Big\{m\Big(\frac{t}{T}, X_{t,T}\Big) - m(u,x)\Big\}\Big] \\
&=: Q_1(u,x) + Q_2(u,x) + Q_3(u,x) + Q_4(u,x)
\end{aligned}$$

with

$$Q_i(u,x) = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) q_i(u,x)$$

and

$$\begin{aligned}
q_1(u,x) &= \mathbb{E}\Big[\prod_{j=1}^{d} \bar{K}_h(x_j - X_{t,T}^j)\Big\{\prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) \\
&\qquad - \prod_{j=1}^{d} K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big\}\Big\{m\Big(\frac{t}{T}, X_{t,T}\Big) - m(u,x)\Big\}\Big] \\
q_2(u,x) &= \mathbb{E}\Big[\prod_{j=1}^{d} \bar{K}_h(x_j - X_{t,T}^j) \prod_{j=1}^{d} K_h\big(x^j - X_t^j(\tfrac{t}{T})\big) \\
&\qquad \times \Big\{m\Big(\frac{t}{T}, X_{t,T}\Big) - m\Big(\frac{t}{T}, X_t(\tfrac{t}{T})\Big)\Big\}\Big] \\
q_3(u,x) &= \mathbb{E}\Big[\Big\{\prod_{j=1}^{d} \bar{K}_h(x^j - X_{t,T}^j) - \prod_{j=1}^{d} \bar{K}_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big\} \\
&\qquad \times \prod_{j=1}^{d} K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big\{m\Big(\frac{t}{T}, X_t(\tfrac{t}{T})\Big) - m(u,x)\Big\} \\
q_4(u,x) &= \mathbb{E}\Big[\prod_{j=1}^{d} K_h\big(x^j - X_t^j(\tfrac{t}{T})\big)\Big\{m\Big(\frac{t}{T}, X_t(\tfrac{t}{T})\Big) - m(u,x)\Big\}\Big].
\end{aligned}$$

We first consider $Q_1(u,x)$. Using (1.47), we obtain

$$|Q_1(u,x)| \le \frac{C}{Th^{d+1}} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \mathbb{E}\Big[\sum_{k=1}^{d} \big|K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\tfrac{t}{T})\big)\big|^r$$

$$\times \prod_{j=1}^{d} \bar{K}_h(x_j - X_{t,T}^j)\Big\{m\Big(\frac{t}{T}, X_{t,T}\Big) - m(u,x)\Big\}\Big]$$

with $r = \min\{\rho, 1\}$. The term $\prod_{j=1}^{d} \bar{K}_h(x_j - X_{t,T}^j)\{m(\frac{t}{T}, X_{t,T}) - m(u,x)\}$ in the above expression can be bounded by Ch. Since K is Lipschitz, $|X_{t,T}^k - X_t^k(\frac{t}{T})| \leq \frac{C}{T} U_{t,T}(\frac{t}{T})$, and the variables $U_{t,T}(\frac{t}{T})$ have finite r-th moment, we can infer that

$$\begin{aligned} |Q_1(u,x)| &\leq \frac{C}{Th^d} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big)\mathbb{E}\Big[\sum_{k=1}^{d} \big|K_h(x^k - X_{t,T}^k) - K_h\big(x^k - X_t^k(\tfrac{t}{T})\big)\big|^r\Big] \\ &\leq \frac{C}{Th^d} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big)\mathbb{E}\Big[\sum_{k=1}^{d} \Big|\frac{1}{Th} U_{t,T}(\tfrac{t}{T})\Big|^r\Big] \leq \frac{C}{T^r h^{d-1+r}} \end{aligned}$$

uniformly in u and x. We next turn to $Q_2(u,x)$. Note that the expression in the expectation of $q_2(u,x)$ is non-zero only if $X_{t,T} \in [x^j - 2C_1 h, x^j + 2C_1 h]_{j=1}^d$ and $X_t(\frac{t}{T}) \in [x^j - C_1 h, x^j + C_1 h]_{j=1}^d$. As m is continuous, this implies that $|m(\frac{t}{T}, X_{t,T}) - m(\frac{t}{T}, X_t(\frac{t}{T}))| \leq C$ for some constant $C < \infty$, whenever the expression in the expectation is non-zero. This allows us to use the bound

$$\Big|m\Big(\frac{t}{T}, X_{t,T}\Big) - m\Big(\frac{t}{T}, X_t(\tfrac{t}{T})\Big)\Big| \leq C\Big|m\Big(\frac{t}{T}, X_{t,T}\Big) - m\Big(\frac{t}{T}, X_t(\tfrac{t}{T})\Big)\Big|^r$$

with $r = \min\{\rho, 1\}$ and some constant $C < \infty$. We thus arrive at

$$\begin{aligned} |q_2(u,x)| &\leq C\mathbb{E}\Big[\prod_{j=1}^{d} \bar{K}_h(x_j - X_{t,T}^j) \prod_{j=1}^{d} K_h\big(x^j - X_t^j(\tfrac{t}{T})\big) \\ &\qquad \times \Big|m\Big(\frac{t}{T}, X_{t,T}\Big) - m\Big(\frac{t}{T}, X_t(\tfrac{t}{T})\Big)\Big|^r\Big] \\ &\leq C\mathbb{E}\Big[\Big(\sum_{j=1}^{d} |X_{t,T}^j - X_t^j(\tfrac{t}{T})|\Big)^r\Big] \\ &\leq C\mathbb{E}\Big[\Big(\frac{1}{T} U_{t,T}(\tfrac{t}{T})\Big)^r\Big] \leq \frac{C}{T^r} \end{aligned}$$

uniformly in u and x. As a result, $\sup_{u,x} |Q_2(u,x)| \leq \frac{C}{T^r h^d}$. Using analogous arguments as for $Q_1(u,x)$, we can further show that $\sup_{u,x} |Q_3(u,x)| \leq \frac{C}{T^r h^{d-1+r}}$. Finally, applying Lemmas B1 and B2 and exploiting the smoothness conditions on m and f, we obtain that

$$Q_4(u,x) = h^2 \frac{\kappa_2}{2} \sum_{i=0}^{d} \Big(2\partial_i m(u,x)\partial_i f(u,x) + \partial_{i,i}^2 m(u,x) f(u,x)\Big) + o(h^2)$$

uniformly in u and x. Combining the results on $Q_1(u,x),\dots,Q_4(u,x)$ completes the proof.

Using the intermediate results (a)–(c), we now obtain that

$$\begin{aligned}
&\sup_{u\in I_h, x\in S}\big|\hat{m}(u,x)-m(u,x)\big| \\
&\qquad\le \sup \frac{1}{|\hat{f}(u,x)|}\Big(\sup\big|\hat{g}^V(u,x)\big| + \sup\big|\hat{g}^B(u,x)-m(u,x)\hat{f}(u,x)\big|\Big) \\
&\qquad= \sup \frac{1}{\hat{f}(u,x)}\; O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}} + \frac{1}{T^r h^d} + h^2\Big).
\end{aligned}$$

with $r = \min\{\rho, 1\}$. Moreover, since $\sup \hat{f}(u,x)^{-1} = O_p(1)$ (which immediately follows from Theorem 1.6 and the assumption that $f(u,x) > 0$), we finally arrive at

$$\sup_{u\in I_h, x\in S}\big|\hat{m}(u,x)-m(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}} + \frac{1}{T^r h^d} + h^2\Big).$$

□

Proof of Theorem 1.8

With $\hat{g}^V(u,x)$ and $\hat{g}^B(u,x)$ as in the proof of Theorem 1.7, we let

$$\begin{aligned}
\sqrt{Th^{d+1}}\big(\hat{m}(u,x)-m(u,x)\big) &= \sqrt{Th^{d+1}}\Big(\frac{\hat{g}^V(u,x)}{\hat{f}(u,x)} + \frac{\hat{g}^B(u,x)}{\hat{f}(u,x)} - m(u,x)\Big) \\
&= \frac{\sqrt{Th^{d+1}}}{\hat{f}(u,x)}\big(\hat{g}^V(u,x) + \hat{g}^B(u,x) - m(u,x)\hat{f}(u,x)\big)
\end{aligned}$$

and use the shorthands

$$\begin{aligned}
B(u,x) &= \sqrt{Th^{d+1}}\big(\hat{g}^B(u,x) - m(u,x)\hat{f}(u,x)\big) \\
V(u,x) &= \sqrt{Th^{d+1}}\hat{g}^V(u,x).
\end{aligned}$$

In what follows, we refer to $B(u,x)$ as the bias part and to $V(u,x)$ as the stochastic part.
The bias part vanishes asymptotically, i.e. $B(u,x) = o_p(1)$. This follows from (c) of Theorem 1.7 and the fact that $B(u,x) - \mathbb{E}[B(u,x)] = o_p(1)$. In order to prove the latter, it suffices to show that $\text{Var}(B(u,x)) = o(1)$, which can be achieved by arguments similar to those for Lemma B3.

The stochastic part is asymptotically normal. In particular,

$$V(u,x) \xrightarrow{d} N(0, \kappa_0^{d+1}\sigma^2(u,x)f(u,x)) \tag{1.48}$$

with $\kappa_0 = \int K^2(\varphi)d\varphi$. The proof proceeds by the usual blocking argument. Decomposing $V(u,x)$ alternately into big blocks and small blocks, we can neglect the small blocks and exploit the mixing conditions to replace the big blocks by independent random variables. This allows us to apply a Lindeberg theorem to get the result. We omit the details, as the proof is very similar to that for the standard strictly stationary setting. We however shortly comment on how to calculate the variance of $V(u,x)$. First, by the same steps as in Lemma B3,

$$\begin{aligned}\operatorname{Var}(V(u,x)) &= \operatorname{Var}\Big(\frac{1}{\sqrt{Th^{d+1}}}\sum_{t=1}^{T} K_h\Big(u-\frac{t}{T}\Big)\prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)\varepsilon_{t,T}\Big) \\ &= \frac{1}{Th^{d+1}}\sum_{t=1}^{T} K_h^2\Big(u-\frac{t}{T}\Big)\mathbb{E}\Big[\prod_{j=1}^{d} K_h^2(x^j - X_{t,T}^j)\varepsilon_{t,T}^2\Big] + o(1).\end{aligned}$$

Moreover, by similar steps as for (d) in Theorem 1.7,

$$\frac{1}{Th^{d+1}}\sum_{t=1}^{T} K_h^2\Big(u-\frac{t}{T}\Big)\mathbb{E}\Big[\prod_{j=1}^{d} K_h^2(x^j - X_{t,T}^j)\varepsilon_{t,T}^2\Big] = \kappa_0^{d+1}\sigma^2(u,x)f(u,x) + o(1)$$

with $\kappa_0 = \int K^2(\varphi)d\varphi$. Hence,

$$\operatorname{Var}\big(V(u,x)\big) = \kappa_0^{d+1}\sigma^2(u,x)f(u,x) + o(1).$$

As $\hat{f}(u,x) - f(u,x) = o_p(1)$ and $\frac{1}{\hat{f}(u,x)} = O_p(1)$, we can now combine (1.48) with the fact that $B(u,x) = o_p(1)$ to arrive at

$$\begin{aligned}\sqrt{Th^{d+1}}\big(\hat{m}(u,x) - m(u,x)\big) &= \frac{1}{\hat{f}(u,x)}\big(B(u,x) + V(u,x)\big) \\ &= \frac{V(u,x)}{f(u,x)} + o_p(1) \xrightarrow{d} N(0, V_{u,x}).\end{aligned}$$

This completes the proof. □

Appendix C

In this appendix, we prove the results concerning the smooth backfitting estimates of Section 1.5. Throughout the appendix, conditions (Add1) and (Add2) are assumed to be satisfied. Moreover, C is used to denote a universal real constant which may take a different value on each occurrence.

Auxiliary Results

Before we come to the proof of Theorems 1.9 and 1.10, we provide results on uniform convergence rates for the kernel smoothers that are used as pilot estimates in the smooth backfitting procedure. We start with an auxiliary lemma which is needed to derive the various rates.

Lemma C1. *Define $T_0 = \mathbb{E}[T_{[0,1]^d}]$. Then uniformly for $u \in I_h$,*

$$\frac{T_0}{T} = \mathbb{P}(X_0(u) \in [0,1]^d) + O(T^{-\frac{\rho}{1+\rho}}) + o(h) \tag{1.49}$$

with ρ defined in assumption (C1) and

$$\frac{T_{[0,1]^d} - T_0}{T_0} = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big). \tag{1.50}$$

Proof. We first show (1.49). Let $U_{t,T} := U_{t,T}(\frac{t}{T})$ for short and recall that $\|X_{t,T} - X_t(\frac{t}{T})\| \leq \frac{1}{T}U_{t,T}$ almost surely with $\mathbb{E}[U_{t,T}^\rho] \leq C$ for some $\rho > 0$. It holds that

$$\mathbb{E}\big[I(X_{t,T} \in [0,1]^d)\big] = \mathbb{E}\big[I\big(X_{t,T} \in [0,1]^d, \|X_{t,T} - X_t(\tfrac{t}{T})\| \leq \tfrac{1}{T}U_{t,T}\big)\big]$$
$$\begin{cases} \geq \mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[\tfrac{C}{T}U_{t,T}, 1 - \tfrac{C}{T}U_{t,T}\big]^d\big)\big] \\ \leq \mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[-\tfrac{C}{T}U_{t,T}, 1 + \tfrac{C}{T}U_{t,T}\big]^d\big)\big] \end{cases}$$

for some sufficiently large $C < \infty$. Hence, with

$$B_L = \frac{1}{T}\sum_{t=1}^T K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[\tfrac{C}{T}U_{t,T}, 1 - \tfrac{C}{T}U_{t,T}\big]^d\big)\big]$$
$$B_U = \frac{1}{T}\sum_{t=1}^T K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[-\tfrac{C}{T}U_{t,T}, 1 + \tfrac{C}{T}U_{t,T}\big]^d\big)\big],$$

we obtain

$$B_L \leq \frac{T_0}{T} \leq B_U.$$

Now letting $q < 1$, it holds that

$$\begin{aligned} B_U &= \frac{1}{T}\sum_{t=1}^T K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[-\tfrac{C}{T}U_{t,T}, 1 + \tfrac{C}{T}U_{t,T}\big]^d, U_{t,T} \leq T^q\big)\big] \\ &\quad + \frac{1}{T}\sum_{t=1}^T K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[-\tfrac{C}{T}U_{t,T}, 1 + \tfrac{C}{T}U_{t,T}\big]^d, U_{t,T} > T^q\big)\big] \\ &=: B_U^{(1)} + B_U^{(2)}, \end{aligned}$$

where

$$\begin{aligned}
B_U^{(1)} &\leq \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\big[I\big(X_t(\tfrac{t}{T}) \in \big[-\tfrac{C}{T^{1-q}}, 1+\tfrac{C}{T^{1-q}}\big]^d\big)\big] \\
&= \int I\big(x \in \big[-\tfrac{C}{T^{1-q}}, 1+\tfrac{C}{T^{1-q}}\big]^d\big) \underbrace{\frac{1}{T}\sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big) f\Big(\frac{t}{T}, x\Big)}_{=f(u,x)+o(h) \text{ by Lemma } B2} dx \\
&= \int I\big(x \in \big[-\tfrac{C}{T^{1-q}}, 1+\tfrac{C}{T^{1-q}}\big]^d\big) f(u,x)dx + o(h) \\
&= \int I\big(x \in [0,1]^d\big) f(u,x)dx + O\Big(\frac{1}{T^{1-q}}\Big) + o(h)
\end{aligned}$$

and

$$\begin{aligned}
B_U^{(2)} &\leq \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\big[I(U_{t,T} > T^q)\big] \\
&\leq \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big)\mathbb{E}\Big[\Big(\frac{U_{t,T}}{T^q}\Big)^\rho\Big] \leq \frac{C}{T^{q\rho}}
\end{aligned}$$

uniformly for $u \in I_h$. Setting $q = \frac{1}{1+\rho}$, we arrive at

$$B_U \leq \int I(x \in [0,1]^d) f(u,x)dx + O(T^{-\frac{\rho}{1+\rho}}) + o(h) \tag{1.51}$$

uniformly in u. By similar arguments, we can show that

$$B_L \geq \int I(x \in [0,1]^d) f(u,x)dx + O(T^{-\frac{\rho}{1+\rho}}) + o(h). \tag{1.52}$$

Combining (1.51) and (1.52) yields (1.49), since $\int I(x \in [0,1]^d) f(u,x)dx = \mathbb{P}(X_0(u) \in [0,1]^d)$. Equation (1.50) now follows immediately:

$$\frac{T_{[0,1]^d} - T_0}{T_0} = \frac{T}{T_0} \cdot \frac{1}{T}(T_{[0,1]^d} - T_0) = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big)$$

uniformly in u, as

$$\begin{aligned}
\frac{1}{T}(T_{[0,1]^d} - T_0) &= \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big)\big(I(X_{t,T} \in [0,1]^d) - \mathbb{E}[I(X_{t,T} \in [0,1]^d)]\big) \\
&= O_p\Big(\sqrt{\frac{\log T}{Th}}\Big)
\end{aligned}$$

uniformly for $u \in I_h$ by Theorem 1.5 and $\frac{T_0}{T} = O_p(1)$ uniformly in u by (1.49). □

We now examine the convergence behaviour of the pilot estimates of the smooth backfitting procedure. We first consider the kernel density estimates defined in (1.21) and (1.22).

Lemma C2. *It holds that*

$$\sup_{u,x^j \in I_h} \big|\hat{p}_j(u,x^j) - p_j(u,x^j)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^2}} + \frac{1}{T^r h^{d+r}}\Big) + o(h) \quad (1.53)$$

$$\sup_{u \in I_h,\, x^j \in [0,1]} \big|\hat{p}_j(u,x^j) - \kappa_0(x^j) p_j(u,x^j)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^2}} + \frac{1}{T^r h^{d+r}} + h\Big) \quad (1.54)$$

$$\sup_{u,x^j,x^k \in I_h} \big|\hat{p}_{j,k}(u,x^j,x^k) - p_{j,k}(u,x^j,x^k)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^3}} + \frac{1}{T^r h^{d+r}}\Big) + o(h) \quad (1.55)$$

and

$$\sup_{u \in I_h,\, x^j,x^k \in [0,1]} \big|\hat{p}_{j,k}(u,x^j,x^k) - \kappa_0(x^j)\kappa_0(x^k) p_{j,k}(u,x^j,x^k)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^3}} + \frac{1}{T^r h^{d+r}} + h\Big) \quad (1.56)$$

with $r = \min\{\rho, 1\}$ *and* $\kappa_0(w) = \int K_h(w,v)dv$.

Proof. We only consider the term $\hat{p}_j$, the proof for $\hat{p}_{j,k}$ being analogous. Defining $T_0 = \mathbb{E}[T_{[0,1]^d}]$ and

$$\check{p}_j(u,x^j) = \frac{1}{T_0} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) K_h(x^j, X_{t,T}^j),$$

we obtain that

$$\begin{aligned} \hat{p}_j(u,x^j) &= \frac{T_0}{T_{[0,1]^d} - T_0 + T_0} \check{p}_j(u,x^j) \\ &= \Big[1 + \frac{T_{[0,1]^d} - T_0}{T_0}\Big]^{-1} \check{p}_j(u,x^j) \\ &= \Big[1 - \frac{T_{[0,1]^d} - T_0}{T_0} + O_p\Big(\Big(\frac{T_{[0,1]^d} - T_0}{T_0}\Big)^2\Big)\Big] \check{p}_j(u,x^j). \end{aligned}$$

Using (1.50) from Lemma C1, this implies that

$$\hat{p}_j(u,x^j) = \check{p}_j(u,x^j) + O_p\Big(\sqrt{\frac{\log T}{Th}}\Big)$$

uniformly for $u \in I_h$ and $x^j \in [0,1]$. Applying the proving strategy of Theorem 1.6 to $\check{p}_j(u,x^j)$ completes the proof of (1.53) and (1.54). □

We next examine the Nadaraya-Watson smoother $\hat{m}_j$. To this purpose, we decompose it into a variance part $\hat{m}_j^A$ and a bias part $\hat{m}_j^B$. The decomposition is given by $\hat{m}_j(u,x^j) = \hat{m}_j^A(u,x^j) + \hat{m}_j^B(u,x^j)$ with

$$\hat{m}_j^A(u,x^j) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) K_h(x^j, X_{t,T}^j) \varepsilon_{t,T} / \hat{p}_j(u,x^j) \tag{1.57}$$

$$\begin{aligned}\hat{m}_j^B(u,x^j) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) K_h(x^j, X_{t,T}^j) \\ \times \Big(m_0\Big(\frac{t}{T}\Big) + \sum_{k=1}^{d} m_k\Big(\frac{t}{T}, X_{t,T}^k\Big)\Big) / \hat{p}_j(u,x^j).\end{aligned} \tag{1.58}$$

For the variance part $\hat{m}_j^A$, we have

Lemma C3. *It holds that*

$$\sup_{u,x^j \in [0,1]} \big|\hat{m}_j^A(u,x^j)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^2}}\Big). \tag{1.59}$$

Proof. Replacing the occurrences of $T_{[0,1]^d}$ in (1.57) by $T_0 = \mathbb{E}[T_{[0,1]^d}]$ and then applying Theorem 1.5 gives the result. □

For the bias part, we have the following expansion:

Lemma C4. *It holds that*

$$\sup_{u,x^j \in I_h} \big|\hat{m}_j^B(u,x^j) - \hat{\mu}_{T,j}(u,x^j)\big| = o_p(h^2) \tag{1.60}$$

$$\sup_{u \in I_h,\, x^j \in I_h^c} \big|\hat{m}_j^B(u,x^j) - \hat{\mu}_{T,j}(u,x^j)\big| = O_p(h^2) \tag{1.61}$$

with $I_h^c = [0,1] \setminus I_h$ *and*

$$\begin{aligned}\hat{\mu}_{T,j}(u,x^j) = \alpha_{T,0}(u) + \alpha_{T,j}(u,x^j) + \sum_{k \neq j} \int \alpha_{T,k}(u,x^k) \frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)} dx^k \\ + h^2 \int \beta(u,x) \frac{p(u,x)}{p_j(u,x^j)} dx^{-j}.\end{aligned}$$

Here,

$$\alpha_{T,0}(u) = m_0(u) + h\kappa_1(u)\partial_u m_0(u) + \frac{h^2}{2}\kappa_2(u)\partial^2_{uu} m_0(u)$$

$$\alpha_{T,k}(u,x^k) = m_k(u,x^k) + h\Big[\kappa_1(u)\partial_u m_k(u,x^k) + \frac{\kappa_0(u)\kappa_1(x^k)}{\kappa_0(x^k)}\partial_{x^k} m_k(u,x^k)\Big]$$

$$\beta(u,x) = \kappa_2 \partial_u m_0(u)\partial_u \log p(u,x) + \kappa_2 \sum_{k=1}^{d}\Big\{\partial_u m_k(u,x^k)\partial_u \log p(u,x) + \frac{1}{2}\partial^2_{uu} m_k(u,x^k) + \partial_{x^k} m_k(u,x^k)\partial_{x^k}\log p(u,x) + \frac{1}{2}\partial^2_{x^k x^k} m_k(u,x^k)\Big\}$$

with $\kappa_2 = \int w^2 K(w)dw$ *and* $\kappa_l(v) = \int w^l K_h(v,w)dw$ *for* $l = 0,1,2$.

Proof. By definition

$$\hat{m}_j^B(u,x^j) = \hat{m}_j^{B,0}(u,x^j) + \sum_{k=1}^{d}\hat{m}_j^{B,k}(u,x^j)$$

with

$$\hat{m}_j^{B,0}(u,x^j) = \frac{1}{T_{[0,1]^d}}\sum_{t=1}^{T} I(X_{t,T}\in[0,1]^d)K_h\Big(u,\frac{t}{T}\Big)K_h(x^j,X^j_{t,T})m_0\Big(\frac{t}{T}\Big)/\hat{p}_j(u,x^j)$$

$$\hat{m}_j^{B,k}(u,x^j) = \frac{1}{T_{[0,1]^d}}\sum_{t=1}^{T} I(X_{t,T}\in[0,1]^d)K_h\Big(u,\frac{t}{T}\Big)K_h(x^j,X^j_{t,T})m_k\Big(\frac{t}{T},X^k_{t,T}\Big)/\hat{p}_j(u,x^j)$$

for $k = 1,\dots,d$. We show that

$$\hat{m}_j^{B,0}(u,x^j) = m_0(u) + h\kappa_1(u)\partial_u m_0(u) + h^2\big[\kappa_2(u)\partial_u m_0(u)\frac{\partial_u p_j(u,x^j)}{p_j(u,x^j)} + \frac{1}{2}\kappa_2(u)\partial^2_{uu}m_0(u)\big] + R^0_T(u,x^j) \tag{1.62}$$

with $\sup_{u,x^j\in I_h}|R^0_T(u,x^j)| = o_p(h^2)$ and $\sup_{u\in I_h,\, x^j\in I_h^c}|R^0_T(u,x^j)| = O_p(h^2)$,

$$\begin{aligned}\hat{m}_j^{B,j}(u,x^j) = {} & m_j(u,x^j) \\ & + h\big[\kappa_1(u)\partial_u m_j(u,x^j) + \frac{\kappa_0(u)\kappa_1(x^j)}{\kappa_0(x^j)}\partial_{x^j} m_j(u,x^j)\big] \\ & + h^2\big[\kappa_2(u)\partial_u m_j(u,x^j)\frac{\partial_u p_j(u,x^j)}{p_j(u,x^j)} + \frac{1}{2}\kappa_2(u)\partial^2_{uu} m_j(u,x^j) \\ & \quad + \frac{\kappa_0(u)\kappa_2(x^j)}{\kappa_0(x^j)}\partial_{x^j} m_j(u,x^j)\frac{\partial_{x^j} p_j(u,x^j)}{p_j(u,x^j)} \\ & \quad + \frac{1}{2}\frac{\kappa_0(u)\kappa_2(x^j)}{\kappa_0(x^j)}\partial^2_{x^j x^j} m_j(u,x^j)\big] \\ & + R^j_T(u,x^j),\end{aligned} \tag{1.63}$$

where R_T^j is of the same uniform order as R_T^0, and for $k \neq j$,

$$\begin{aligned}\hat{m}_j^{B,k}(u,x^j) = &\int m_k(u,x^k)\frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)}dx^k + h\int \big[\kappa_1(u)\partial_u m_k(u,x^k) \\ &\quad + \frac{\kappa_0(u)\kappa_1(x^k)}{\kappa_0(x^k)}\partial_{x^k} m_k(u,x^k)\big]\frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)}dx^k \\ &+ h^2\big[\kappa_2(u)\int \kappa_0(x^k)\partial_u m_k(u,x^k)\frac{\partial_u p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)}dx^k \\ &\quad + \kappa_0(u)\int \kappa_2(x^k)\partial_{x^k} m_k(u,x^k)\frac{\partial_{x^k} p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)}dx^k \\ &\quad + \kappa_2(u)\int \kappa_0(x^k)\frac{1}{2}\partial^2_{uu} m_k(u,x^k)\frac{p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)}dx^k \\ &\quad + \kappa_0(u)\int \kappa_2(x^k)\frac{1}{2}\partial^2_{x^k x^k} m_k(u,x^k)\frac{p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)}dx^k\big] \\ &+ R_T^k(u,x^j), \end{aligned} \tag{1.64}$$

where again R_T^k is of the same uniform order as R_T^0. Combining (1.62)–(1.64) completes the proof.

We only give the proof of (1.64), as this is the most complicated term: Recalling that $\int K_h(x^k, X_{t,T}^k)dx^k = 1$, a second-order Taylor expansion of $m_k(\frac{t}{T}, X_{t,T}^k)$ around (u,x^k) yields

$$\begin{aligned}\hat{m}_j^{B,k}(u,x^j) = &\int \frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)} m_k(u,x^k)dx^k \\ &+ \frac{1}{T_{[0,1]^d}}\sum_{t=1}^{T}\big(V_{t,T}^k(u,x^j) + W_{t,T}^k(u,x^j)\big)/\hat{p}_j(u,x^j) + o_p(h^2)\end{aligned}$$

uniformly for $u \in I_h$ and $x^j \in [0,1]$ with

$$\begin{aligned}V_{t,T}^k(u,x^j) = &I(X_{t,T} \in [0,1]^d)K_h\Big(u,\frac{t}{T}\Big)K_h(x^j,X_{t,T}^j)\int K_h(x^k,X_{t,T}^k) \\ &\times\Big[\partial_u m_k(u,x^k)\Big(\frac{t}{T}-u\Big) + \partial_{x^k} m_k(u,x^k)(X_{t,T}^k - x^k)\Big]dx^k \\ W_{t,T}^k(u,x^j) = &I(X_{t,T} \in [0,1]^d)K_h\Big(u,\frac{t}{T}\Big)K_h(x^j,X_{t,T}^j)\int K_h(x^k,X_{t,T}^k) \\ &\times\Big[\frac{1}{2}\partial^2_{uu} m_k(u,x^k)\Big(\frac{t}{T}-u\Big)^2 + \partial^2_{ux^k} m_k(u,x^k)\Big(\frac{t}{T}-u\Big)(X_{t,T}^k - x^k) \\ &+ \frac{1}{2}\partial^2_{x^k x^k} m_k(u,x^k)(X_{t,T}^k - x^k)^2\Big]dx^k.\end{aligned}$$

We now have a closer look at the expectations of $V^k_{t,T}(u,x^j)$ and $W^k_{t,T}(u,x^j)$. First, note that

$$\mathbb{E}[V^k_{t,T}(u,x^j)] = \mathbb{E}\Big[I\big(X_t(\tfrac{t}{T}) \in [0,1]^d\big)K_h\Big(u,\frac{t}{T}\Big)K_h\big(x^j,X^j_t(\tfrac{t}{T})\big)\int K_h\big(x^k,X^k_t(\tfrac{t}{T})\big) \times \Big\{\partial_u m_k(u,x^k)\Big(\frac{t}{T}-u\Big) + \partial_{x^k} m_k(u,x^k)\big(X^k_t(\tfrac{t}{T})-x^k\big)\Big\}dx^k\Big] + O\Big(\frac{1}{T^{\frac{r}{r+1}}}+\frac{1}{T^r h}\Big) \tag{1.65}$$

with $r = \min\{\rho,1\}$ uniformly for $u \in I_h$ and $x^j \in [0,1]$. This is shown by successively replacing the occurrences of $X_{t,T}$ in $\mathbb{E}[V^k_{t,T}(u,x^j)]$ by $X_t(\frac{t}{T})$. In order to replace the occurrence in the indicator function $I(X_{t,T} \in [0,1]^d)$, similar arguments as in Lemma C1 can be used. For replacing the occurrences in $K_h(x^j,X^j_{t,T})$ and $K_h(x^k,X^k_{t,T})$, we exploit the Lipschitz continuity of K and use arguments similar to those in part (c) of the proof of Theorem 1.7. With (1.65), we can now write

$$\frac{1}{T_{[0,1]^d}}\sum_{t=1}^T \mathbb{E}[V^k_{t,T}(u,x^j)] = \frac{1}{T_{[0,1]^d}}\sum_{t=1}^T K_h\Big(u,\frac{t}{T}\Big)\int K_h(x^j,w^j)K_h(x^k,w^k) \times \Big[\partial_u m_k(u,x^k)\Big(\frac{t}{T}-u\Big) + \partial_{x^k} m_k(u,x^k)(w^k-x^k)\Big] \times \Big(\int I(w\in[0,1]^d)f\Big(\frac{t}{T},w\Big)dw^{-j,k}\Big)dw^j dw^k dx^k + O\Big(\frac{1}{T^{\frac{r}{r+1}}}+\frac{1}{T^r h}\Big)$$

uniformly for $u \in I_h$ and $x^j \in [0,1]$, where $w^{-j,k}$ denotes all but the j-th and k-th component of the vector w. Noting that $O(T^{-\frac{r}{r+1}} + \frac{1}{T^r h}) = o(h^2)$ by (Add2), using a first-order Taylor expansion of $f(\frac{t}{T},w)$ and recalling the definition of the density p, we can infer that

$$\begin{aligned}
&\frac{1}{T_{[0,1]^d}}\sum_{t=1}^T \mathbb{E}[V^k_{t,T}(u,x^j)] \\
&= \frac{T}{T_{[0,1]^d}}P(X_0(u)\in[0,1]^d) \\
&\quad\times\Big\{\frac{1}{T}\sum_{t=1}^T K_h\Big(u,\frac{t}{T}\Big)\Big(\frac{t}{T}-u\Big)\int \kappa_0(x^j)\kappa_0(x^k)\partial_u m_k(u,x^k)p_{j,k}(u,x^j,x^k)dx^k \\
&\qquad+\frac{1}{T}\sum_{t=1}^T K_h\Big(u,\frac{t}{T}\Big)\int h\kappa_0(x^j)\kappa_1(x^k)\partial_{x^k} m_k(u,x^k)p_{j,k}(u,x^j,x^k)dx^k \\
&\qquad+\frac{1}{T}\sum_{t=1}^T K_h\Big(u,\frac{t}{T}\Big)\Big(\frac{t}{T}-u\Big)^2\int \kappa_0(x^j)\kappa_0(x^k)\partial_u m_k(u,x^k)\partial_u p_{j,k}(u,x^j,x^k)dx^k
\end{aligned}$$

$$+ \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u, \frac{t}{T}\Big) \int h^2 \kappa_0(x^j)\kappa_2(x^k)\partial_{x^k} m_k(u, x^k)\partial_{x^k} p_{j,k}(u, x^j, x^k) dx^k\Big\}$$

$$+ o_p(h^2)$$

uniformly for $u \in I_h$ and $x^j \in [0,1]$. Combining the two claims of Lemma C1, it holds that

$$\frac{T}{T_{[0,1]^d}} P(X_0(u) \in [0,1]^d) = 1 + O\Big(\sqrt{\frac{\log T}{Th}}\Big) + O(T^{-\frac{\rho}{1+\rho}}) + o(h)$$

uniformly in u. We can thus use Lemmas B1 and B2 from Appendix B to get

$$\begin{aligned}
&\frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} \mathbb{E}[V_{t,T}^k(u, x^j)] \\
&= h\Big[\kappa_1(u)\kappa_0(x^j) \int \kappa_0(x^k)\partial_u m_k(u, x^k) p_{j,k}(u, x^j, x^k) dx^k \\
&\qquad + \kappa_0(u)\kappa_0(x^j) \int \kappa_1(x^k)\partial_{x^k} m_k(u, x^k) p_{j,k}(u, x^j, x^k) dx^k\Big] \\
&\quad + h^2\Big[\kappa_2(u)\kappa_0(x^j) \int \kappa_0(x^k)\partial_u m_k(u, x^k)\partial_u p_{j,k}(u, x^j, x^k) dx^k \\
&\qquad + \kappa_0(u)\kappa_0(x^j) \int \kappa_2(x^k)\partial_{x^k} m_k(u, x^k)\partial_{x^k} p_{j,k}(u, x^j, x^k) dx^k\Big] \\
&\quad + R_T^V(u, x^j)
\end{aligned} \tag{1.66}$$

with $\sup_{u,x^j \in I_h} |R_T^V(u, x^j)| = o(h^2)$ and $\sup_{u \in I_h, x^j \in I_h^c} |R_T^V(u, x^j)| = O(h^2)$. Exploiting the fact that $\kappa_1(u) = 0$ for all $u \in I_h$ and that

$$\int \partial_{x^k} m_k(u, x^k)\Big[\frac{1}{\kappa_0(x^k)}\hat{p}_{j,k}(u, x^j, x^k) - \kappa_0(x^j) p_{j,k}(u, x^j, x^k)\Big] h\kappa_1(x^k) dx^k = O_p(h^2)$$

uniformly for $u \in I_h$ and $x^j \in [0,1]$, we can rewrite (1.66) as

$$\begin{aligned}
&\frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} \mathbb{E}[V_{t,T}^k(u, x^j)] \\
&= h\Big[\kappa_1(u) \int \partial_u m_k(u, x^k)\hat{p}_{j,k}(u, x^j, x^k) dx^k \\
&\qquad + \kappa_0(u) \int \frac{\kappa_1(x^k)}{\kappa_0(x^k)}\partial_{x^k} m_k(u, x^k)\hat{p}_{j,k}(u, x^j, x^k) dx^k\Big] \\
&\quad + h^2\Big[\kappa_2(u)\kappa_0(x^j) \int \kappa_0(x^k)\partial_u m_k(u, x^k)\partial_u p_{j,k}(u, x^j, x^k) dx^k \\
&\qquad + \kappa_0(u)\kappa_0(x^j) \int \kappa_2(x^k)\partial_{x^k} m_k(u, x^k)\partial_{x^k} p_{j,k}(u, x^j, x^k) dx^k\Big] \\
&\quad + \tilde{R}_T^V(u, x^j),
\end{aligned} \tag{1.67}$$

where $\tilde{R}^V_T(u,x^j)$ is of the same uniform order as $R^V_T(u,x^j)$. Using analogous arguments as above, we can further show that

$$\begin{aligned}
&\frac{1}{T_{[0,1]^d}}\sum_{t=1}^T \mathbb{E}[W^k_{t,T}(u,x^j)] \\
&= \frac{h^2}{2}\Big[\kappa_2(u)\kappa_0(x^j)\int \kappa_0(x^k)\partial^2_{uu}m_k(u,x^k)p_{j,k}(u,x^j,x^k)dx^k \\
&\qquad + \kappa_0(u)\kappa_0(x^j)\int \kappa_2(x^k)\partial^2_{x^kx^k}m_k(u,x^k)p_{j,k}(u,x^j,x^k)dx^k\Big] \\
&\quad + R^W_T(u,x^j) \qquad (1.68)
\end{aligned}$$

with $\sup_{u,x^j\in I_h}|R^W_T(u,x^j)| = o(h^2)$ and $\sup_{u\in I_h,x^j\in I^c_h}|R^W_T(u,x^j)| = O(h^2)$. Finally, applying the same proving strategy as in Theorem 1.5, one can show that

$$\sup_{u\in I_h,\,x^j\in[0,1]}\Big|\frac{1}{T_{[0,1]^d}}\sum_{t=1}^T\big(V^k_{t,T}(u,x^j)-\mathbb{E}[V^k_{t,T}(u,x^j)]\big)\Big| = o_p(h^2)$$

$$\sup_{u\in I_h,\,x^j\in[0,1]}\Big|\frac{1}{T_{[0,1]^d}}\sum_{t=1}^T\big(W^k_{t,T}(u,x^j)-\mathbb{E}[W^k_{t,T}(u,x^j)]\big)\Big| = o_p(h^2).$$

Therefore,

$$\begin{aligned}
\hat{m}^{B,k}_j(u,x^j) &= \int m_k(u,x^k)\frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)}dx^k \\
&\quad + \frac{1}{T_{[0,1]^d}}\sum_{t=1}^T\big(\mathbb{E}[V^k_{t,T}(u,x^j)]+\mathbb{E}[W^k_{t,T}(u,x^j)]\big)/\hat{p}_j(u,x^j) + o_p(h^2)
\end{aligned}$$

uniformly for $u\in I_h$ and $x^j\in[0,1]$. Plugging (1.67) and (1.68) into the above expression and using the fact that $\hat{p}_j(u,x^j)$ converges uniformly to $\kappa_0(x^j)p_j(u,x^j)$ yields (1.64). □

We finally state a result on the convergence behaviour of the term $\tilde{m}_0(u)$.

Lemma C5. *It holds that*

$$\sup_{u\in I_h}\big|\tilde{m}_0(u)-m_0(u)\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}+h^2\Big). \qquad (1.69)$$

Proof. The claim can be shown by replacing the term $T_{[0,1]^d}$ by $T_0=\mathbb{E}[T_{[0,1]^d}]$ in the expression for $\tilde{m}_0(u)$ and then using arguments from Theorem 1.7. □

Proof of Theorems 1.9 and 1.10

To prove Theorems 1.9 and 1.10, it suffices to show that the high-level conditions (A1)–(A6), (A8), and (A9) of Mammen et al. [29] are satisfied. This allows us to apply their Theorems 1–3, which imply the result. As will be seen, the high-level conditions are satisfied uniformly for $u \in I_h$ rather than only pointwise. For this reason, we can infer that the convergence rates in (1.25) hold uniformly over $u \in I_h$ rather than only pointwise. In what follows, we formulate the high-level conditions and show that they are fulfilled in our setting.

(A1) For all $j \neq k$, it holds that

$$\int \frac{p_{j,k}^2(u, x^j, x^k)}{p_k(u, x^k) p_j(u, x^j)} dx^j dx^k < \infty$$

uniformly for $u \in I_h$.

This condition follows immediately from the assumptions on the density $f(u, x)$. These imply that $p_j(u, x^j) \geq c > 0$ and $p_{j,k}(u, x^j, x^k) \leq C < \infty$ for all $u \in [0, 1]$ and x^j, $x^k \in [0, 1]$ with some appropriately chosen constants c and C.

(A2) For all $j \neq k$, it holds that

$$\int \Big[\frac{\hat{p}_j(u, x^j) - p_j(u, x^j)}{p_j(u, x^j)}\Big]^2 p_j(u, x^j) dx^j = o_p(1)$$

$$\int \Big[\frac{\hat{p}_{j,k}(u, x^j, x^k)}{p_k(u, x^k) p_j(u, x^j)} - \frac{p_{j,k}(u, x^j, x^k)}{p_k(u, x^k) p_j(u, x^j)}\Big]^2 p_k(u, x^k) p_j(u, x^j) dx^j dx^k = o_p(1)$$

$$\int \Big[\frac{\hat{p}_{j,k}(u, x^j, x^k)}{p_k(u, x^k) \hat{p}_j(u, x^j)} - \frac{p_{j,k}(u, x^j, x^k)}{p_k(u, x^k) p_j(u, x^j)}\Big]^2 p_k(u, x^k) p_j(u, x^j) dx^j dx^k = o_p(1)$$

uniformly for $u \in I_h$. Furthermore, for each $u \in I_h$, $\hat{p}_j(u, \cdot)$ and $\hat{p}_{j,k}(u, \cdot)$ vanish outside the support of $p_j(u, \cdot)$ and $p_{j,k}(u, \cdot)$, respectively.

This condition as well as (A4) and (A8) can easily be proven by using the uniform convergence results for the kernel densities derived in Lemma C2.

(A3) There exists a finite constant C such that with probability tending to 1,

$$\int \hat{m}_j^2(u, x^j) p_j(u, x^j) dx^j < \infty$$

uniformly for $u \in I_h$.

Both this condition and (A5) directly follow from Lemmas C3 and C4, which describe the asymptotic behaviour of the variance part $\hat{m}_j^A$ and the bias part $\hat{m}_j^B$ of the Nadaraya-Watson estimate $\hat{m}_j$.

(A4) There exists a finite constant C such that with probability tending to 1,

$$\sup_{x^k \in I_h} \int \frac{\hat{p}_{j,k}^2(u, x^j, x^k)}{\hat{p}_k^2(u, x^k) p_j(u, x^j)} dx^j \leq C$$

for all $j \neq k$ uniformly for $u \in I_h$.

(A5) There exists a finite constant C such that with probability tending to 1,

$$\int \hat{m}_j^A(u, x^j)^2 p_j(u, x^j) dx^j \leq C$$
$$\int \hat{m}_j^B(u, x^j)^2 p_j(u, x^j) dx^j \leq C$$

uniformly for $u \in I_h$.

(A6) For $j \neq k$, it holds that

$$\sup_{x^j \in I_h} \Big| \int \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j)} \hat{m}_k^A(u, x^k) dx^k \Big| = o_p(h^2)$$
$$\Big\| \int \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j)} \hat{m}_k^A(u, x^k) dx^k \Big\|_2 = o_p(h^2)$$

uniformly for $u \in I_h$, where $\| \cdot \|_2$ denotes the norm in the space $L_2(p_j(u, \cdot))$.

To prove (A6), it suffices to show that

$$\sup_{u \in I_h,\, x^j \in [0,1]} \Big| \int \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j)} \hat{m}_k^A(u, x^k) dx^k \Big| = O_p\Big(\sqrt{\frac{\log T}{Th}} \Big). \tag{1.70}$$

For the proof of (1.70), we write

$$\begin{aligned} S_{k,j}(u, x^j) &= \int \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j)} \hat{m}_k^A(u, x^k) dx^k \\ &= \int \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j) \hat{p}_k(u, x^k)} \hat{\psi}_k(u, x^k) dx^k, \end{aligned}$$

where $\hat{m}_k^A(u, x^k) = \hat{\psi}_k(u, x^k) / \hat{p}_k(u, x^k)$ with

$$\hat{\psi}_k(u, x^k) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\Big(u, \frac{t}{T}\Big) K_h(x^k, X_{t,T}^k) \varepsilon_{t,T}.$$

In a first step, we replace $S_{k,j}(u,x^j)$ by the term

$$S^*_{k,j}(u,x^j) = \int \frac{p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)p_k(u,x^k)}\hat{\psi}_k(u,x^k)dx^k$$

and show that the resulting error is asymptotically negligible. This is done as follows:

$$\begin{aligned}
&\sup_{u\in I_h, x^j\in[0,1]} \left|S_{k,j}(u,x^j) - S^*_{k,j}(u,x^j)\right| \\
&= \sup_{u,x^j}\Big|\int \Big\{\frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)\hat{p}_k(u,x^k)} - \frac{\kappa_0(x^j)\kappa_0(x^k)p_{j,k}(u,x^j,x^k)}{\kappa_0(x^j)p_j(u,x^j)\kappa_0(x^k)p_k(u,x^k)}\Big\}\hat{\psi}_k(u,x^k)dx^k\Big| \\
&= O_p\Big(\sqrt{\frac{\log T}{Th^3}} + h\Big)O_p\Big(\sqrt{\frac{\log T}{Th^2}}\Big) \\
&= O_p\Big(\frac{\log T}{Th^{5/2}} + \sqrt{\frac{\log T}{T}}\Big),
\end{aligned}$$

as $\hat{\psi}_k(u,x^k) = O_p(\sqrt{\log T/Th^2})$ and the term in curly brackets is of the order $O_p(\sqrt{\log T/Th^3} + h)$ uniformly in u, x^j, and x^k. In a second step, we show that

$$\sup_{u\in I_h, x^j\in[0,1]} \left|S^*_{k,j}(u,x^j)\right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big).$$

To prove this, we write

$$S^*_{k,j}(u,x^j) = \frac{1}{T_{[0,1]^d}}\sum_{t=1}^{T} w_{k,j}(u,x^j,X^k_{t,T})\varepsilon_{t,T} \tag{1.71}$$

with

$$\begin{aligned}
w_{k,j}(u,x^j,X^k_{t,T}) &= I(X_{t,T}\in[0,1]^d)K_h\Big(u,\frac{t}{T}\Big) \\
&\quad\times\Big(\int \frac{p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)p_k(u,x^k)}K_h(x^k,X^k_{t,T})dx^k\Big).
\end{aligned}$$

Applying the techniques from the proof of Theorem 1.5 to (1.71) completes the proof of (1.70), which in turn yields (A6).

(A8) It holds that

$$\sup_{x^j\in I_h}\int\Big|\frac{p_{j,k}(u,x^j,x^k)}{p_j(u,x^j)p_k(u,x^k)} - \frac{\hat{p}_{j,k}(u,x^j,x^k)}{\hat{p}_j(u,x^j)\hat{p}_k(u,x^k)}\Big|p_k(u,x^k)dx^k = o_p(1)$$

uniformly for $u\in I_h$.

(A9) There exist deterministic functions

$$\alpha_{T,0}(u), \alpha_{T,1}(u, x^1), \ldots, \alpha_{T,d}(u, x^d)$$

$$\gamma_{T,1}(u), \ldots, \gamma_{T,d}(u)$$

and a function $\beta(u, x)$ (not depending on T) such that uniformly for $u \in I_h$

$$\int \alpha^2_{T,j}(u, x^j) p_j(u, x^j) dx^j < \infty \tag{1.72}$$

$$\int \beta^2(u, x) p(u, x) dx < \infty \tag{1.73}$$

$$\sup_{x^1 \in I_h, \ldots, x^d \in I_h} |\beta(u, x)| < \infty \tag{1.74}$$

$$\int \alpha_{T,j}(u, x^j) \hat{p}_j(u, x^j) dx^j = \gamma_{T,j}(u) + o_p(h^2) \tag{1.75}$$

with $\gamma_{T,j}(u) = O(h^2)$ and

$$\sup_{u, x^j \in I_h} \left| \hat{m}^B_j(u, x^j) - \hat{\mu}_{T,0}(u) - \hat{\mu}_{T,j}(u, x^j) \right| = o_p(h^2) \tag{1.76}$$

$$\sup_{u \in I_h} \int \left| \hat{m}^B_j(u, x^j) - \hat{\mu}_{T,0}(u) - \hat{\mu}_{T,j}(u, x^j) \right|^2 p_j(u, x^j) dx^j = o_p(h^4). \tag{1.77}$$

Here, $\hat{\mu}_{T,0}(u)$ is some random function and

$$\hat{\mu}_{T,j}(u, x^j) = \alpha_{T,0}(u) + \alpha_{T,j}(u, x^j) + \sum_{k \neq j} \int \alpha_{T,k}(u, x^k) \frac{\hat{p}_{j,k}(u, x^j, x^k)}{\hat{p}_j(u, x^j)} dx^k$$
$$+ h^2 \int \beta(u, x) \frac{p(u, x)}{p_j(u, x^j)} dx^{-j}.$$

We finally prove (A9). Equations (1.76) and (1.77) immediately follow from the uniform expansion of the bias part $\hat{m}^B_j$ proven in Lemma C4. Furthermore, it is trivial to see that (1.72)–(1.74) are fulfilled for $\alpha_{T,j}(u, x^j)$ and $\beta(u, x)$ as defined in Lemma C4. Finally, straightforward calculations yield a term $\gamma_{T,j}(u)$ in (1.75) which is of order h^2 uniformly for $u \in I_h$.

This completes the proof of Theorems 1.9 and 1.10. □

Chapter 2

Comparing Nonparametric Fits In Locally Stationary Regression Models

2.1 Introduction

Many economic and financial time series applications are marked by two main features. Firstly, the relationship between the variables of interest may be nonlinear. To model the relationship between a variable and its own lags, for example, a linear autoregressive process is often inappropriate. Nonlinear autoregressive structures such as threshold models are needed to get a satisfactory description of the data. Secondly, the relationship of the variables may change over time. In many cases, it is very plausible that two economic variables relate differently to each other in different economic situations.
A flexible framework which is able to capture both nonlinearities and structural change is given by the nonparametric regression model

$$Y_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} \quad \text{for } t = 1, \dots, T \tag{2.1}$$

with $\mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0$, where the function m may vary over time and the regressors $X_{t,T} = (X^1_{t,T}, \dots, X^d_{t,T})$ are locally stationary. The concept of local stationarity was introduced by Dahlhaus (cf. [4], [5], and [6]). Heuristically speaking, a process is locally stationary if over short time spans, i.e. locally in time, it behaves approximately stationary. A detailed description of model (2.1) including a rigorous definition of local stationarity can be found in Section 2.2.

In this chapter, we are interested in the question whether the regression function m in model (2.1) has the same shape at two different time points. Put differently, we want to know whether the function $m(u, \cdot)$ at some rescaled time point $u \in [0, 1]$ is identical to the function $m(v, \cdot)$ at another time point v. To decide upon this issue, we develop a kernel-based nonparametric testing procedure. The test statistic measures a weighted L_2-distance between kernel estimates of $m(u, \cdot)$ and $m(v, \cdot)$ and is introduced in Section 2.3.

The asymptotic properties of the statistic are analyzed in Section 2.5. To improve the finite sample behaviour of the test, we propose a wild bootstrap procedure in Section 2.6 and derive the asymptotic properties thereof. The limit behaviour of the test statistic will turn out to be mainly driven by a quadratic form. Not much is known about the asymptotic behaviour of quadratic forms in a locally stationary framework. To our knowledge, Lee & Subba Rao [25] are the only ones who have tackled this issue so far. However, they analyze a type of quadratic form which does not cover our case. The main theoretical challenge thus lies in the derivation of a limit theory for the quadratic form which shows up in our setting.

There is a large literature on testing structural change in nonparametric time series regression. One strand of the literature deals with structural breaks in the nonparametric regression function. There, the main issue is to localize and estimate the size of the structural breaks (see e.g. Delgado & Hidalgo [10]). Another strand of the literature is concerned with testing the hypothesis that the regression function is time-invariant. Different types of statistics have been proposed to deal with this testing problem: Hidalgo [19] for example has developed a conditional moment test, Su & Xiao [39] have suggested a CUSUM type test.

Our testing problem is closely related but not identical to testing whether a nonparametric regression function is time-invariant. Rather than testing whether the function $m(u, \cdot)$ is the same for *all* time points $u \in [0, 1]$, we test whether it is the same at *two* different time points u and v. From an applied point of view, both testing issues are interesting and complement each other. In many economic and financial applications, the question arises whether the regression function is fully stable over time. Equally interestingly, one may want to know whether the function is the same in two different situations, e.g. at a time point before a crisis and one during it.

2.2 The Model

Before we introduce the test statistic, we have a more detailed look at the underlying model (2.1),

$$Y_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} \quad \text{with} \quad \mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0.$$

The components of the model, namely the function m, the regressors $X_{t,T}$, and the residuals $\varepsilon_{t,T}$, are required to have the following main properties:

(i) The function m is not allowed to vary over time in whatever way. In particular, we do not allow for sudden structural changes. Instead, we assume that m varies smoothly over time. The exact smoothness conditions are listed in Section 2.4.

(ii) As already noted in the introduction, we do not restrict the regressors to be strictly stationary. Instead, we allow the triangular array $\{X_{t,T} : t = 1, \ldots, T\}$ to be locally stationary, which for our purpose is defined as follows:

Definition 2.1. *The process $\{X_{t,T}\}$ is locally stationary if for each time point $u \in [0,1]$ there exists an associated process $\{X_t(u)\}$ with the following two properties:*

(i) $\{X_t(u)\}$ is strictly stationary with density $f_{X_t(u)}$,

(ii) it holds that

$$\big\|X_{t,T} - X_t(u)\big\| \leq \Big(\Big|\frac{t}{T} - u\Big| + \frac{1}{T}\Big) U_{t,T}(u) \quad a.s.,$$

where $\{U_{t,T}(u)\}$ is a process of positive variables satisfying $\mathbb{E}[(U_{t,T}(u))^\rho] < C$ for some $\rho > 0$ and $C < \infty$ independent of u, t, and T. $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^d$.

(iii) We finally put some contraints on the residual process $\{\varepsilon_{t,T} : t = 1 \ldots, T\}$. To derive the asymptotic characteristics of the test statistic, we want to make use of a central limit theorem for martingale difference arrays. To be able to do so, we assume that

$$\mathbb{E}[\varepsilon_{t,T}|\mathcal{F}_{t-1,T}] = 0 \quad \text{with} \quad \mathcal{F}_{t-1,T} = \sigma(X_{t,T}, X_{t-1,T}, \varepsilon_{t-1,T}, \ldots, X_{1,T}, \varepsilon_{1,T}).$$

This rules out autocorrelation in the error terms $\varepsilon_{t,T}$. However, it allows for heteroskedasticity. In particular, it allows the residual process to be of the form

$$\varepsilon_{t,T} = \sigma\Big(\frac{t}{T}, X_{t,T}\Big)\varepsilon_t \tag{2.2}$$

with a time-varying volatility function σ and an i.i.d. process $\{\varepsilon_t\}$ having the property that ε_t is independent of $X_{s,T}$ for $s \le t$. To keep the notation in the proofs as simple as possible, we restrict attention to the residual process (2.2) in what follows.

An important class of processes that fit into the framework (2.1) is given by the nonlinear autoregressive model

$$X_{t,T} = m\Big(\frac{t}{T}, X_{t-1,T}^{t-d}\Big) + \sigma\Big(\frac{t}{T}, X_{t-1,T}^{t-d}\Big)\varepsilon_t \tag{2.3}$$

with $X_{t-1,T}^{t-d} = (X_{t-1,T}, \ldots, X_{t-d,T})$ and i.i.d. variables ε_t. One can show that under suitable low-level conditions on m, σ, and the residuals ε_t, the components of model (2.3) have the properties (i)–(iii). In particular, the autoregressive process $\{X_{t,T}\}$ can be shown to be locally stationary and strongly mixing with mixing coefficients that decay exponentially fast to zero. For a detailed analysis of model (2.3) and a proof of these results see Chapter 1.

2.3 The Test Statistic

We want to test whether the regression function $m(u, \cdot)$ at some time point $u \in [0, 1]$ partly (or even fully) coincides with the function $m(v, \cdot)$ at another time point v. The null hypothesis is thus given by

$$H_0: \ m(u, \cdot) = m(v, \cdot) \quad \pi\text{-}a.s.,$$

where π is some weight function and $(u, v) \in (0, 1)^2$ is some fixed pair of rescaled time points. The null hypothesis can equivalently be expressed as

$$H_0: \int \big[m(u, x) - m(v, x)\big]^2 \pi(x)dx = 0.$$

A natural way to come up with a test statistic for this problem is to replace the unknown functions $m(u, \cdot)$ and $m(v, \cdot)$ in the above L_2-distance by estimates $\hat{m}(u, \cdot)$ and $\hat{m}(v, \cdot)$ and to rescale appropriately. This yields the weighted L_2-test statistic

$$S_T = Th^{1+d/2} \int \big[\hat{m}(u, x) - \hat{m}(v, x)\big]^2 \pi(x)dx,$$

where $\hat{m}$ is a Nadaraya-Watson estimate given by

$$\hat{m}(u,x) = \frac{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) Y_{t,T}}{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)}.$$

In this definition, K denotes a one-dimensional kernel function and we use the notation $K_h(x) = h^{-1}K(x/h)$. For simplicity, we work with a product kernel and assume that the bandwidth is the same in each direction. In addition, we assume that the weight function π has bounded support.[1]

In what follows, we analyze the asymptotic behaviour of S_T under the null hypothesis as well as under fixed and local alternatives. The alternative hypothesis is given by

$$H_1 : \int \big[m(u,x) - m(v,x)\big]^2 \pi(x) dx > 0.$$

This treats the fixed alternative case, where $m(u,\cdot)$ and $m(v,\cdot)$ are some fixed pair of different functions. To get a rough impression of the power of the test, we additionally examine local alternatives, i.e. alternatives that converge to H_0 as the sample size grows. To formulate these alternatives, we define the sequence of functions

$$m_T(w,z) = m(w,z) + c_T \Delta(w,z),$$

where $c_T \to 0$, the function Δ is continuous and equals zero in a neighbourhood around u, and m satisfies the null hypothesis, i.e. $m(u,\cdot) = m(v,\cdot)$ π-a.s. The process $\{Y_{t,T}\}$ is thus given by

$$Y_{t,T} = m_T\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T} = m\Big(\frac{t}{T}, X_{t,T}\Big) + c_T \Delta\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T}. \tag{2.4}$$

If the process $\{Y_{t,T}\}$ is generated according to (2.4), we move along the sequence of local alternatives

$$H_{1,T} : \int \big[m(u,x) - m_T(v,x)\big]^2 \pi(x) dx = c_T^2 \int \Delta^2(v,x) \pi(x) dx.$$

In this case, the weighted L_2-distance between the regression function at time point u and that at time point v gets smaller as the sample size increases, i.e. the hypothesis $H_{1,T}$ comes closer and closer to H_0 as T tends to infinity.

[1] It is possible to allow for unbounded support by letting the limits of the integral in S_T diverge to infinity at an appropriate rate as the sample size increases.

2.4 Assumptions

To derive the asymptotic properties of the test statistic S_T, we make the following assumptions.

(C1) The process $\{X_{t,T}\}$ is locally stationary in the sense of Definition 2.1 with some $\rho \geq 1$. Thus, for each time point $u \in [0,1]$, there exists a strictly stationary process $\{X_t(u)\}$ with density $f(u,x) := f_{X_t(u)}(x)$ such that $\|X_{t,T} - X_t(u)\| \leq (|\frac{t}{T} - u| + \frac{1}{T})U_{t,T}(u)$ with $\mathbb{E}[(U_{t,T}(u))^\rho] \leq C$.

(C2) The array $\{X_{t,T}, \varepsilon_{t,T}\}$ is strongly mixing with mixing coefficients that converge exponentially fast to zero, i.e. $\alpha(k) \leq Ca^k$ for some $a < 1$.

(C3) The bandwidth h satisfies $Th^{2d+1} \to \infty$. Moreover, let r be a natural number with $r > d/2$ such that $Th^{4r+1} \to 0$ and $Th^{2r+1+d/2} \to 0$.

(C4) The kernel K is bounded and has compact support, i.e. $K(v) = 0$ for all $|v| > C_1$ with some $C_1 < \infty$. Moreover, K is Lipschitz, i.e. $|K(v) - K(v')| \leq L|v - v'|$ for some $L < \infty$ and all $v, v' \in \mathbb{R}$. Finally, K satisfies the conditions

$$\textstyle\int K(z)dz = 1, \quad \int z^j K(z)dz = 0 \text{ for } j = 1, \ldots, r-1.$$

(C5) For each $u \in [0,1]$, let $f(u,\cdot)$ be the density of $X_t(u)$. The functions f and m are r-times continuously differentiable. Moreover, $\inf_{u\in[0,1],x\in S} f(u,x) > 0$, where S is the closure of the set $\{x \in \mathbb{R}^d \mid \pi(x) \neq 0\}$.

(C6) The residuals are of the form $\varepsilon_{t,T} = \sigma(\frac{t}{T}, X_{t,T})\varepsilon_t$. Here, σ is a Lipschitz continuous function and $\{\varepsilon_t\}$ is an i.i.d. process having the property that ε_t is independent of $X_{s,T}$ for $s \leq t$. The variables ε_t satisfy $\mathbb{E}[\varepsilon_t^{6+\delta}] < \infty$ for some small $\delta > 0$ and are normalized such that $\mathbb{E}[\varepsilon_t^2] = 1$.

(C7) Let $f_{X_{t,T}}$ be the density of $X_{t,T}$ and $f_{X_{t,T},X_{t+l,T}}$ the joint density of $(X_{t,T}, X_{t+l,T})$. For any compact set $S \subseteq \mathbb{R}^d$, there exists a constant $B = B(S) < \infty$ such that $\sup_{t,T} \sup_{x\in S} f_{X_{t,T}}(x) \leq B$ and

$$\sup_{t,T} \sup_{l>0} \sup_{x,x'\in S} \mathbb{E}\big[|\varepsilon_t| \big| X_{t,T} = x, X_{t+l,T} = x'\big] f_{X_{t,T},X_{t+l,T}}(x,x') \leq B.$$

We quickly give some remarks on the above assumptions. First note that we do not necessarily require exponential mixing rates as assumed in (C2). These could be replaced by sufficiently large polynomial rates. We nevertheless make the stronger assumption (C2) to keep the notation in the proofs as simple as possible. Assumptions (C3)–(C5) allow us to use higher-order kernels ($r > 2$) in the analysis of the test statistic. Note however, that we only need them if the dimension of the regressors d is larger than 3.

2.5 The Asymptotic Distribution of S_T

In this section, we summarize the results on the asymptotic behaviour of the test statistic S_T. The first theorem states that under the null, S_T weakly converges to a Gaussian distribution if we subtract a bias term that diverges to infinity.

Theorem 2.1. *Assume that (C1)–(C7) are fulfilled. Then under H_0,*

$$S_T - B_T(u,v) \xrightarrow{d} N\big(0, V(u,v)\big).$$

Here, $B_T(u,v) = B_T(u) + B_T(v)$ and $V(u,v) = V(u) + V(v)$, where

$$B_T(u) = h^{-d/2} \iint K^2(w) \prod_{j=1}^{d} K^2(z^j)\sigma^2(u - hw, x - hz) \times f(u - hw, x - hz)\frac{\pi(x)}{f^2(u,x)} dwdzdx$$

$$V(u) = 2\kappa_2^2 \int \mathcal{K}^2(z)dz \int \frac{[\sigma^2(u,x)]^2\pi^2(x)}{f^2(u,x)}dx$$

with $\kappa_2 = \int K^2(w)dw$ and $\mathcal{K}(z) = \int \prod_{j=1}^{d} K(w^j) \prod_{j=1}^{d} K(w^j + z^j)dw$. The expressions $B_T(v)$ and $V(v)$ are defined analogously.

We now turn to the behaviour of S_T under fixed alternatives. The next theorem shows that S_T (corrected by the bias term $B_T(u,v)$) diverges in probability to infinity under H_1. The test based on the statistic $S_T - B_T(u,v)$ is thus consistent against fixed alternatives.

Theorem 2.2. *Assume that (C1)–(C7) are fulfilled. Then under H_1,*

$$(Th^{1+d/2})^{-1}\big(S_T - B_T(u,v)\big) \xrightarrow{P} \int \big[m(u,x) - m(v,x)\big]^2 \pi(x)dx > 0.$$

We finally examine the behaviour of S_T under local alternatives to get an idea of the quality of the test. According to the next theorem, the asymptotic power of the test against alternatives of the form $m + c_T\Delta$ with $c_T = (Th^{1+d/2})^{-1/2}$ and m satisfying the null hypothesis is constant for all functions Δ having the same weighted L_2-norm. This behaviour is well-known from other kernel-based L_2-test statistics (see e.g. Härdle & Mammen [18]).

Theorem 2.3. *Assume that (C1)–(C7) are fulfilled and let* $c_T = (Th^{1+d/2})^{-1/2}$*. Then under* $H_{1,T}$*,*

$$S_T - B_T(u,v) \xrightarrow{d} N\Big(\int \Delta^2(v,x)\pi(x)dx, V(u,v)\Big)$$

with $B_T(u,v)$ *and* $V(u,v)$ *as defined in Theorem 2.1.*

To prove Theorem 2.3, we need the process $\{X_{t,T}\}$ to be locally stationary and strongly mixing under local alternatives. This is guaranteed as long as the regressors $X_{t,T}$ do not contain lagged values of $Y_{t,T}$. In the autoregressive case (2.3), however, it is not clear at all whether the process $\{X_{t,T}\}$ has these two properties. In this short note, we do not explore this issue any further. Instead, we simply exclude the autoregressive case when examining local alternatives.

2.6 Bootstrapping S_T

Theorem 2.1 allows us to approximate the distribution of the test statistic S_T by a Gaussian distribution. It is however well-known that in nonparametric hypothesis testing, the test statistic converges rather slowly to the asymptotic distribution (see e.g. Härdle & Mammen [18] or Li & Wang [26]). The approximation in finite samples is thus rather poor in many cases. Moreover, the bias and variance expressions $B_T(u,v)$ and $V(u,v)$ contain unknown functions. Replacing them by consistent estimates results in further approximation errors.

A common way to improve the finite sample behaviour of a test is to use bootstrap methods. In what follows, we set up a wild bootstrap procedure. As shown in the proof of Theorem 2.1, under the null hypothesis, it holds that

$$S_T = U_T(u) + U_T(v) + o_p(1)$$

with

$$U_T(u) = Th^{1+d/2} \int \Big(\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}K_{x,t,T}\varepsilon_{t,T}\Big)^2 \frac{\pi(x)}{\hat{f}^2(u,x)}dx$$
$$U_T(v) = Th^{1+d/2} \int \Big(\frac{1}{T}\sum_{t=1}^{T} K_{v,t,T}K_{x,t,T}\varepsilon_{t,T}\Big)^2 \frac{\pi(x)}{\hat{f}^2(v,x)}dx,$$

where $K_{u,t,T} = K_h(u - \frac{t}{T})$ and $K_{x,t,T} = \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)$ for short. Thus, under the null, S_T is asymptotically equivalent to the sum of two quadratic forms $U_T(u)$

and $U_T(v)$. This allows us to imitate the distribution of S_T by bootstrapping the two quadratic forms $U_T(u)$ and $U_T(v)$ rather than the whole statistic S_T (cp. Kreiss et al. [22]). Note that the distribution of $U_T(u)+U_T(v)$ does not depend on whether the null hypothesis is true or not. The bootstrap estimate of $U_T(u) + U_T(v)$ thus mimics the distribution of the statistic S_T under the null hypothesis no matter whether the null holds or not.

The bootstrap sample is generated as follows. To construct bootstrap residuals $\varepsilon^*_{t,T}$, denote the estimated residuals by

$$\hat{\varepsilon}_{t,T} = Y_{t,T} - \hat{m}\Big(\frac{t}{T}, X_{t,T}\Big).$$

Letting $\{\eta_t\}$ be some sequence of i.i.d. variables with zero mean and unit variance that is independent of $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$, we define

$$\varepsilon^*_{t,T} = \hat{\varepsilon}_{t,T} \cdot \eta_t.$$

The bootstrap residuals have the following properties: They are conditionally independent given the sample $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$. Moreover, they mimic the first two moments of the errors $\varepsilon_{t,T}$. In particular, $\mathbb{E}^*[\varepsilon^*_{t,T}] = 0$ and $\mathbb{E}^*[(\varepsilon^*_{t,T})^2] = \hat{\varepsilon}^2_{t,T}$, where $\mathbb{E}^*[\,\cdot\,] = \mathbb{E}[\,\cdot\,|\{Y_{t,T}, X_{t,T}\}_{t=1}^T]$. As we do not do any resampling for the regressors $X_{t,T}$, we arrive at the bootstrap sample $\{X_{t,T}, \varepsilon^*_{t,T}\}_{t=1}^T$.

Replacing the residuals $\varepsilon_{t,T}$ in the quadratic forms $U_T(u)$ and $U_T(v)$ by the bootstrap residuals $\varepsilon^*_{t,T}$, we obtain the bootstrap statistic

$$S^*_T = U^*_T(u) + U^*_T(v)$$

with

$$U^*_T(u) = Th^{1+d/2} \int \Big(\frac{1}{T}\sum_{t=1}^T K_{u,t,T} K_{x,t,T} \varepsilon^*_{t,T}\Big)^2 \frac{\pi(x)}{\hat{f}^2(u,x)} dx$$

$$U^*_T(v) = Th^{1+d/2} \int \Big(\frac{1}{T}\sum_{t=1}^T K_{v,t,T} K_{x,t,T} \varepsilon^*_{t,T}\Big)^2 \frac{\pi(x)}{\hat{f}^2(v,x)} dx.$$

The next theorem shows that the above defined wild bootstrap is consistent. To formulate the result, we let $\mathbb{P}^*(\,\cdot\,) := \mathbb{P}(\,\cdot\,|\{Y_{t,T}, X_{t,T}\}_{t=1}^T)$.

Theorem 2.4. *Assume that (C1)–(C7) are fulfilled. Then*

$$S^*_T - B_T(u,v) \stackrel{d}{\longrightarrow} N\big(0, V(u,v)\big)$$

conditional on the sample $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$ with probability tending to one. Put differently, $P^*\big(S^*_T - B_T(u,v) \le x\big) \stackrel{P}{\longrightarrow} \Phi(x)$, where Φ is a Gaussian distribution function with mean zero and variance $V(u,v)$.

2.7 Concluding Remarks

In this chapter, we have developed a nonparametric procedure to test whether the time-varying regression function in model (2.1) has the same form at two different time points. We have proposed a kernel-based L_2-statistic and have examined its asymptotic properties. In particular, we have shown that after subtracting a bias term that diverges to infinity, the statistic weakly converges to a normal distribution (both under the null and under local alternatives). To improve the small sample behaviour, we have additionally set up a wild bootstrap procedure and have shown that it is consistent.
There are a couple of other interesting testing issues in the framework (2.1) which may be approached quite similarly as the testing problem at hand:

- Rather than testing whether the time-varying regression function is the same at two different time points, one may ask the question whether it is the same over a whole time interval. Let $I \subset [0,1]$ be the (rescaled) time interval to be tested. A possible test statistic is given by

$$S'_T = n'_T \int_{I \times I} \Big(\int \big[\hat{m}(u,x) - \hat{m}(v,x)\big]^2 \pi(x) dx \Big) du dv,$$

  where n'_T is an appropriately chosen scaling factor diverging to infinity. Alternatively, one could use the statistic

$$S''_T = n''_T \int_I \Big(\int \big[\hat{m}(u,x) - \tilde{m}(x)\big]^2 \pi(x) dx \Big) du,$$

  which compares $\hat{m}$ with an estimate $\tilde{m}$ that does not localize in time but is based on all data points in the time interval I. Obviously, the statistics S'_T and S''_T are very similar to S_T. We thus conjecture that the proving techniques of this chapter can be used to derive the asymptotic distribution of these statistics and to set up a wild bootstrap procedure.

- Another interesting testing issue is whether the nonparametric function m can be replaced by a parametric specification with time-varying coefficients. One way to approach this problem is to measure the L_2-distance between a nonparametric and a parametric fit of the regression function. Similarly as in Härdle & Mammen [18], one may want to artificially smooth the parametric estimate to get rid of certain bias terms. The resulting test statistic will again be similar in structure to S_T.

Appendix

In what follows, we prove Theorems 2.1–2.3 and 2.4. Throughout the appendix, we use the symbol C to denote a universal real constant which may take a different value on each occurrence.

Auxiliary Results

To analyze the asymptotic behaviour of the test statistic S_T, we need some results on uniform convergence of the Nadaraya-Watson estimate $\hat{m}(u,x)$. To formulate these results, we split up the expression $\hat{m}(u,x) - m(u,x)$ into different components according to

$$\hat{m}(u,x) - m(u,x) = \frac{1}{\hat{f}(u,x)} \big(\hat{g}^V(u,x) + \hat{g}^B(u,x) \big)$$

with

$$\hat{f}(u,x) = \frac{1}{T} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)$$
$$\hat{g}^V(u,x) = \frac{1}{T} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)\varepsilon_{t,T}$$
$$\hat{g}^B(u,x) = \frac{1}{T} \sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)\Big[m\Big(\frac{t}{T}, X_{t,T}\Big) - m(u,x)\Big].$$

The following two lemmas summarize the convergence behaviour of these three components.

Lemma A1. *Let (C1)–(C7) be fulfilled. Then for any compact subset $S \subset \mathbb{R}^d$,*

$$\sup_{u\in[0,1],x\in S} \big|\hat{f}(u,x) - \mathbb{E}[\hat{f}(u,x)]\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big)$$
$$\sup_{u\in[0,1],x\in S} \big|\hat{g}^B(u,x) - \mathbb{E}\big[\hat{g}^B(u,x)\big]\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big)$$
$$\sup_{u\in[0,1],x\in S} \big|\hat{g}^V(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}}\Big).$$

Lemma A2. *Let (C1)–(C7) be fulfilled and let $I_h = [C_1 h, 1 - C_1 h]$. Then for any compact subset $S \subset \mathbb{R}^d$,*

$$\sup_{u \in I_h, x \in S} \big|\mathbb{E}[\hat{f}(u,x)] - f(u,x)\big| = O\Big(h^r + \frac{1}{Th^{d+1}}\Big)$$
$$\sup_{u \in I_h, x \in S} \big|\mathbb{E}[\hat{g}^B(u,x)]\big| = O\Big(h^r + \frac{1}{Th^{d}}\Big).$$

Combining these two lemmas immediately yields the following result.

Lemma A3. *Let (C1)–(C7) be fulfilled and let $I_h = [C_1 h, 1 - C_1 h]$. Then for any compact subset $S \subset \mathbb{R}^d$,*

$$\sup_{u \in I_h, x \in S} \big|\hat{f}(u,x) - f(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}} + h^r\Big)$$
$$\sup_{u \in I_h, x \in S} \big|\hat{m}(u,x) - m(u,x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^{d+1}}} + h^r\Big).$$

Lemmas A1–A3 directly follow from the results of Chapter 1. Note that Lemmas A1 and A3 can be shown to hold almost surely rather than only in probability. This is easily seen when inspecting the proofs in Chapter 1 and keeping in mind that the model variables are geometrically mixing.

Proof of Theorems 2.1–2.3

In what follows, we give the proof of Theorem 2.3. Theorem 2.1 is obtained by setting the function Δ equal to zero in the proof. Some straightforward additional considerations yield Theorem 2.2.
Using the shorthands $K_{u,t,T} = K_h(u - \frac{t}{T})$ and $K_{x,t,T} = \prod_{j=1}^d K_h(x^j - X_{t,T}^j)$, we can rewrite the statistic S_T as

$$S_T = Th^{1+d/2} \int \big[V_T(u,v,x) + B_T(u,v,x)\big]^2 \pi(x) dx$$

with

$$V_T(u,v,x) = \frac{1}{T}\sum_{t=1}^T K_{u,t,T} K_{x,t,T} \varepsilon_{t,T} \Big/ \hat{f}(u,x) - \frac{1}{T}\sum_{t=1}^T K_{v,t,T} K_{x,t,T} \varepsilon_{t,T} \Big/ \hat{f}(v,x)$$

$$B_T(u,v,x) = \frac{1}{T}\sum_{t=1}^T K_{u,t,T} K_{x,t,T} m_T\Big(\frac{t}{T}, X_{t,T}\Big) \Big/ \hat{f}(u,x)$$

$$-\frac{1}{T}\sum_{t=1}^{T}K_{v,t,T}K_{x,t,T}m_T\Big(\frac{t}{T},X_{t,T}\Big)\Big/\hat{f}(v,x).$$

Theorem 2.3 immediately follows from the following three lemmas.

Lemma A4. *Under (C1)–(C7), it holds that*

$$Th^{1+d/2}\int V_T^2(u,v,x)\pi(x)dx - B_T(u,v) \xrightarrow{d} N\big(0,V(u,v)\big).$$

Lemma A5. *Under (C1)–(C7), it holds that*

$$Th^{1+d/2}\int B_T(u,v,x)V_T(u,v,x)\pi(x)dx = o_p(1).$$

Lemma A6. *Under (C1)–(C7), it holds that*

$$Th^{1+d/2}\int B_T^2(u,v,x)\pi(x)dx = \int \Delta^2(v,x)\pi(x)dx + o_p(1).$$

We now give the proofs of the above lemmas.

Proof of Lemma A4. We write

$$Th^{1+d/2}\int V_T^2(u,v,x)\pi(x)dx = U_T(u,v)$$

with $U_T(u,v) = U_T(u) + U_T(v)$ and

$$U_T(u) = Th^{1+d/2}\int\Big(\frac{1}{T}\sum_{t=1}^{T}K_{u,t,T}K_{x,t,T}\varepsilon_{t,T}\Big)^2\frac{\pi(x)}{\hat{f}^2(u,x)}dx$$

$$U_T(v) = Th^{1+d/2}\int\Big(\frac{1}{T}\sum_{t=1}^{T}K_{v,t,T}K_{x,t,T}\varepsilon_{t,T}\Big)^2\frac{\pi(x)}{\hat{f}^2(v,x)}dx.$$

In what follows, we show that

$$U_T(u) - B_T(u) \xrightarrow{d} N(0,V(u)) \tag{2.5}$$

$$U_T(v) - B_T(v) \xrightarrow{d} N(0,V(v)). \tag{2.6}$$

Combining the arguments for (2.5) and (2.6) with the fact that $K_{u,t,T}K_{v,t,T} = 0$ for all $t = 1,\dots,T$ (provided T is large enough), it is straightforward to see that

$$U_T(u,v) - (B_T(u) + B_T(v)) \xrightarrow{d} N(0,V(u)+V(v)),$$

which yields the result.

In the remainder of the proof, we give the arguments for (2.5), the ones for (2.6) being exactly the same. To start with, we split up $U_T(u)$ into two parts according to

$$U_T(u) = U_{T,1}(u) + U_{T,2}(u) + o_p(1)$$

with

$$U_{T,1}(u) = \frac{h^{1+d/2}}{T} \sum_{t=1}^{T} K_{u,t,T}^2 \Big(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)} dx \Big) \varepsilon_{t,T}^2$$

$$U_{T,2}(u) = \frac{h^{1+d/2}}{T} \sum_{t=1}^{T} \sum_{s \neq t} K_{u,t,T} K_{u,s,T} \Big(\int K_{x,t,T} K_{x,s,T} \frac{\pi(x)}{f^2(u,x)} dx \Big) \varepsilon_{t,T} \varepsilon_{s,T},$$

where we have used the uniform convergence results of Lemmas A1–A3 to replace the kernel density $\hat{f}(u,x)$ by the true density $f(u,x)$. We now show that

$$U_{T,1}(u) = B_T(u) + o_p(1) \tag{2.7}$$

$$V(u)^{-1/2} U_{T,2}(u) \xrightarrow{d} N(0,1). \tag{2.8}$$

This completes the proof of (2.5). □

Proof of (2.7). It suffices to show that $\mathrm{Var}(U_{T,1}(u)) = o(1)$ and $\mathbb{E}[U_{T,1}(u)] = B_T(u) + o(1)$. The first claim easily follows from exploiting the mixing conditions on the model variables. To prove the second claim, we proceed as follows: To start with, we successively replace $X_{t,T}$ with the approximating variables $X_t(\frac{t}{T})$, using the fact that $\|X_{t,T} - X_t(u)\| \le (|\frac{t}{T} - u| + \frac{1}{T}) U_{t,T}(u)$. Similar arguments as in the proofs of Theorems 1.6 and 1.7 in Chapter 1 yield that

$$\begin{aligned}
\mathbb{E}[U_{T,1}(u)] &= \frac{h^{1+d/2}}{T} \sum_{t=1}^{T} K_{u,t,T}^2 \mathbb{E}\Big[\Big(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)} dx \Big) \sigma^2\Big(\frac{t}{T}, X_{t,T}\Big) \Big] \\
&= \frac{h^{1+d/2}}{T} \sum_{t=1}^{T} K_{u,t,T}^2 \mathbb{E}\Big[\Big(\int \prod_{j=1}^{d} K_h^2\Big(x^j - X_t^j\Big(\frac{t}{T}\Big)\Big) \\
&\qquad \times \frac{\pi(x)}{f^2(u,x)} dx \Big) \sigma^2\Big(\frac{t}{T}, X_t\Big(\frac{t}{T}\Big)\Big) \Big] + o(1) \\
&= \frac{h^{1+d/2}}{T} \sum_{t=1}^{T} K_{u,t,T}^2 \int \Big(\int \prod_{j=1}^{d} K_h^2(x^j - z^j) \\
&\qquad \times \sigma^2\Big(\frac{t}{T}, z\Big) f\Big(\frac{t}{T}, z\Big) dz \Big) \frac{\pi(x)}{f^2(u,x)} dx + o(1).
\end{aligned}$$

Since

$$\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}^{2}\sigma^{2}\Big(\frac{t}{T},z\Big)f\Big(\frac{t}{T},z\Big) = \int K_{h}^{2}(u-w)\sigma^{2}(w,z)f(w,z)dw + O\Big(\frac{1}{Th^{3}}\Big)$$

uniformly in u and z, we further get that

$$\begin{aligned}\mathbb{E}[U_{T,1}(u)] = h^{1+d/2}\int \Big(\iint K_{h}^{2}(u-w)\prod_{j=1}^{d} K_{h}^{2}(x^{j}-z^{j}) \\ \times \sigma^{2}(w,z)f(w,z)dwdz\Big)\frac{\pi(x)}{f^{2}(u,x)}dx + o(1) \\ = B_{T}(u) + o(1).\end{aligned}$$

□

Proof of (2.8). We rewrite $U_{T,2}(u)$ as

$$U_{T,2}(u) = \sum_{t=1}^{T} Z_{t,T}(u)$$

with

$$Z_{t,T}(u) = 2\frac{h^{1+d/2}}{T}\sum_{s<t} K_{u,t,T}K_{u,s,T}\Big(\int K_{x,t,T}K_{x,s,T}\frac{\pi(x)}{f^{2}(u,x)}dx\Big)\varepsilon_{t,T}\varepsilon_{s,T}.$$

Note that under (C6), $\{Z_{t,T}(u), \mathcal{F}_{t,T}\}$ with $\mathcal{F}_{t,T} = \sigma(X_{t+1,T}, X_{t,T}, \varepsilon_{t,T}, \ldots, X_{1,T}, \varepsilon_{1,T})$ is a martingale difference array. We can thus use a central limit theorem for martingale difference arrays (in particular Theorem 1 in Chapter 8 of Pollard [37]) to show that $\sum_{t=1}^{T} Z_{t,T}(u)$ is asymptotically normal. It suffices to verify the following conditions:

(CLT1) $\sum_{t=1}^{T} \mathbb{E}[Z_{t,T}^{4}(u)] \to 0$.

(CLT2) $\sum_{t=1}^{T} \mathbb{E}[Z_{t,T}^{2}(u)|\mathcal{F}_{t-1,T}] \xrightarrow{P} V(u)$.

This yields (2.8). □

Proof of (CLT1). We can write

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[Z_{t,T}^{4}(u)] = \frac{16h^{2d+4}}{T^{4}}\sum_{t=1}^{T}\sum_{s,s',s'',s'''\in\mathcal{S}_{t}}\int\ldots\int\mathbb{E}\Big[W_{t,T}(w,x,y,z) \\ \times W_{s,T}(w)W_{s',T}(x)W_{s'',T}(y)W_{s''',T}(z)\Big]\frac{\pi(w)\ldots\pi(z)}{f^{2}(u,w)\ldots f^{2}(u,z)}dw\ldots dz,\end{aligned}$$

where $\mathcal{S}_t$ denotes the set of index combinations (s, s', s'', s''') with $s, s', s'', s''' < t$,

$$W_{t,T}(w,x,y,z) = K^4_{u,t,T} K_{w,t,T} K_{x,t,T} K_{y,t,T} K_{z,t,T} \varepsilon^4_{t,T},$$
$$W_{s,T}(w) = K_{u,s,T} K_{w,s,T} \varepsilon_{s,T},$$

and $W_{s',T}(x)$, $W_{s'',T}(y)$, $W_{s''',T}(z)$ denote analogous expressions. We partition $\mathcal{S}_t$ into the subsets

$$\mathcal{S}^{(1)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}_t \mid \text{the indices } s,s',s'',s''' \text{ are all different}\big\}$$
$$\mathcal{S}^{(2)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}_t \mid \text{exactly two of the indices } s,s',s'',s''' \text{ are the same}\big\}$$
$$\mathcal{S}^{(3)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}_t \mid \text{exactly three of the indices } s,s',s'',s''' \text{ are the same}\big\}$$
$$\mathcal{S}^{(4)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}_t \mid \text{the indices } s,s',s'',s''' \text{ are all the same}\big\}$$
$$\mathcal{S}^{(5)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}_t \mid \text{the indices } s,s',s'',s''' \text{ form two different pairs}\big\}$$

and write

$$\sum_{t=1}^T \mathbb{E}[Z^4_{t,T}(u)] = Q^{(1)}_T + \ldots + Q^{(5)}_T$$

with

$$Q^{(i)}_T = \frac{16h^{2d+4}}{T^4} \sum_{t=1}^T \sum_{s,s',s'',s''' \in \mathcal{S}^{(i)}_t} \int \cdots \int \mathbb{E}\Big[W_{t,T}(w,x,y,z)$$
$$\times W_{s,T}(w) W_{s',T}(x) W_{s'',T}(y) W_{s''',T}(z)\Big] \frac{\pi(w)\ldots\pi(z)}{f^2(u,w)\ldots f^2(u,z)} dw \ldots dz$$

for $i = 1, \ldots, 5$. In the remainder of the proof, the terms $Q^{(1)}_T, \ldots, Q^{(5)}_T$ are considered one after the other.

We start with $Q^{(1)}_T$. An index k is said to be separated from another index k', if the two indices are further away from each other than $C_2 \log T$ for some large constant $C_2 < \infty$ to be chosen later on, i.e. $|k - k'| > C_2 \log T$. Using this definition, we split up the index set $\mathcal{S}^{(1)}_t$ into the two parts

$$\mathcal{S}^{(1,a)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}^{(1)}_t \;\big|\; \text{none of the indices } s,s',s'',s''' \text{ are separated from the index } t\big\}$$
$$\mathcal{S}^{(1,b)}_t = \big\{(s,s',s'',s''') \in \mathcal{S}^{(1)}_t \;\big|\; \text{at least one of the indices } s,s',s'',s''' \text{ is separated from the index } t\big\}$$

and write $Q^{(1)}_T = Q^{(1,a)}_T + Q^{(1,b)}_T$, the sums in $Q^{(1,a)}_T$ and $Q^{(1,b)}_T$ running over $\mathcal{S}^{(1,a)}_t$ and $\mathcal{S}^{(1,b)}_t$, respectively. First consider $Q^{(1,b)}_T$ and take a tuple $(s,s',s'',s''') \in \mathcal{S}^{(1,b)}_t$.

W.l.o.g. we can restrict attention to tuples with $t > s > s' > s'' > s'''$ and $|s'' - s'''| > C_2 \log T$. (All other cases can be treated in exactly the same way.) As the model variables are mixing (with exponential decay), we can use Davydov's inequality to get

$$\begin{aligned}
&\Big|\mathbb{E}\Big[W_{t,T}(w,x,y,z)W_{s,T}(w)W_{s',T}(x)W_{s'',T}(y)W_{s''',T}(z)\Big]\Big| \\
&= \Big|\mathrm{Cov}\Big(W_{t,T}(w,x,y,z)W_{s,T}(w)W_{s',T}(x)W_{s'',T}(y), W_{s''',T}(z)\Big)\Big| \\
&\le C\alpha(C_2 \log T)^{\delta/(2+\delta)}\Big(\mathbb{E}\big|W_{s''',T}(z)\big|^{2+\delta}\Big)^{1/(2+\delta)} \\
&\quad \times \Big(\mathbb{E}\big|W_{t,T}(w,x,y,z)W_{s,T}(w)W_{s',T}(x)W_{s'',T}(y)\big|^{2+\delta}\Big)^{1/(2+\delta)} \\
&\le CT^{-C_3},
\end{aligned}$$

where C_3 is a large positive constant (which can be chosen as large as desired by picking C_2 large enough). This immediately yields that $Q_T^{(1,b)} \le CT^{-C_4}$ with some arbitrarily large constant C_4. As a result, the term $Q_T^{(1,b)}$ can be asymptotically neglected. We next turn to $Q_T^{(1,a)}$. As none of the indices s, s', s'', s''' are separated from t, the number of elements contained in $\mathcal{S}_t^{(1,a)}$ is smaller than $C(\log T)^4$ for each given t. As a consequence,

$$Q_T^{(1,a)} \le C\frac{h^{2d+4}}{T^4}\frac{(\log T)^4}{h^{4d+8}}\sum_{t=1}^{T} K_{u,t,T}^4 \le C\frac{(\log T)^4}{T^3 h^{2d+3}} \to 0.$$

Putting everything together, we arrive at $Q_T^{(1)} \to 0$.
By analogous arguments, we obtain that $Q_T^{(i)} \to 0$ for $i = 2, \dots, 5$. Consider for example $Q_T^{(2)}$. Because of symmetry considerations, we can assume w.l.o.g. that $s \ge s' \ge s'' \ge s'''$. Given this, the following cases are possible:

$$(a)\ t > s = s' > s'' > s''' \quad (b)\ t > s > s' = s'' > s''' \quad (c)\ t > s > s' > s'' = s'''.$$

For each of these three cases, we can distinguish between different scenarios in which some of the indices are separated from each other or not. Playing through all these possibilities and exploiting the mixing conditions similarly as in the analysis of $Q_T^{(1)}$, we get that $Q_T^{(2)} \to 0$. By similar case distinctions, we can show that $Q_T^{(i)} \to 0$ for $i = 3, 4, 5$ as well. □

Proof of (CLT2). To show (CLT2), it suffices to verify that

$$\sum\nolimits_{t=1}^{T} \big(\mathbb{E}[Z_{t,T}^2(u)|\mathcal{F}_{t-1,T}] - \mathbb{E}[Z_{t,T}^2(u)]\big) \xrightarrow{P} 0 \tag{2.9}$$

$$\sum\nolimits_{t=1}^{T} \mathbb{E}[Z_{t,T}^2(u)] \to V(u). \tag{2.10}$$

We first prove (2.9). Using the shorthands

$$\begin{aligned} W_{t,T}(x,y) &= K_{u,t,T}^2 K_{x,t,T} K_{y,t,T} \sigma^2\Big(\frac{t}{T}, X_{t,T}\Big) \\ W_{s,T}(x) &= K_{u,s,T} K_{x,s,T} \varepsilon_{s,T}, \end{aligned}$$

we have

$$\begin{aligned} &\mathbb{E}\Big(\sum_{t=1}^{T}\big(\mathbb{E}[Z_{t,T}^2(u)|\mathcal{F}_{t-1,T}] - \mathbb{E}[Z_{t,T}^2(u)]\big)\Big)^2 \\ &= \frac{16h^{2d+4}}{T^4}\sum_{t,t'=1}^{T}\sum_{s,s'<t}\sum_{s'',s'''<t'}\int\dots\int\Big(\mathbb{E}\big[W_{t,T}(w,x)W_{t',T}(y,z)W_{s,T}(w) \\ &\quad \times W_{s',T}(x)W_{s'',T}(y)W_{s''',T}(z)\big] - \mathbb{E}\big[W_{t,T}(w,x)W_{s,T}(w)W_{s',T}(x)\big] \\ &\quad \times \mathbb{E}\big[W_{t',T}(y,z)W_{s'',T}(y)W_{s''',T}(z)\big]\Big)\frac{\pi(w)\dots\pi(z)}{f^2(u,w)\dots f^2(u,z)}dw\dots dz. \end{aligned}$$

We now apply the same strategy as in the proof of (CLT1): By symmetry considerations, we can assume w.l.o.g. that $t \ge t'$, $s \ge s'$, and $s'' \ge s'''$. Thus, the following cases are possible:

$$(a)\ t \ge t' \ge s, s', s'', s''' \quad (b)\ t > s > t' \ge s', s'', s''' \quad (c)\ t > s, s' \ge t' > s'', s'''.$$

Each of these three cases can be further split up into subcases. For case (a), we can for example distinguish between the following possibilities:

- $(a1)$ the indices s, s', s'', s''' are all different
- $(a2)$ exactly two of the indices s, s', s'', s''' are the same
- $(a3)$ exactly three of the indices s, s', s'', s''' are the same
- $(a4)$ the indices s, s', s'', s''' are all the same
- $(a5)$ the indices s, s', s'', s''' form two different pairs.

Repeating the arguments from the proof of (CLT1), we can play through all these cases and proceed analogously for (b) and (c) to arrive at

$$\mathbb{E}\Big(\sum_{t=1}^{T}\big(\mathbb{E}[Z_{t,T}^2(u)|\mathcal{F}_{t-1,T}] - \mathbb{E}[Z_{t,T}^2(u)]\big)\Big)^2 \to 0,$$

which immediately implies (2.9).

Using the mixing conditions on the model variables, successively replacing $X_{t,T}$ by the approximating variables $X_t(\frac{t}{T})$ and then exploiting the smoothness conditions on m, σ and the densities f, we further obtain (2.10), thus completing the proof. □

Proof of Lemma A5. First recall that $m_T(w,z) = m(w,z) + c_T\Delta(w,z)$ with m satisfying the null hypothesis, i.e. $m(u,\cdot) = m(v,\cdot)$ π-a.s. We thus have

$$\frac{1}{T}\sum_{t=1}^{T} K_{v,t,T}K_{x,t,T}m_T\Big(\frac{t}{T}, X_{t,T}\Big) = m(v,x)\hat{f}(v,x) + \frac{1}{T}\sum_{t=1}^{T} K_{v,t,T}K_{x,t,T}\Delta_{t,T}(v,x) + \frac{c_T}{T}\sum_{t=1}^{T} K_{v,t,T}K_{x,t,T}\Delta\Big(\frac{t}{T}, X_{t,T}\Big) \qquad (2.11)$$

with $\Delta_{t,T}(v,x) = m(\frac{t}{T}, X_{t,T}) - m(v,x)$. This allows us to write

$$Th^{1+d/2}\int B_T(u,v,x)V_T(u,v,x)\pi(x)dx = Q_T(u) + Q_T(v) + R_T(v) + o_p(1)$$

with

$$Q_T(u) = \frac{h^{1+d/2}}{T}\int \sum_{t,s=1}^{T} K_{u,t,T}K_{u,s,T}K_{x,t,T}\varepsilon_{t,T}K_{x,s,T}\Delta_{s,T}(u,x)\frac{\pi(x)}{f^2(u,x)}dx$$

$$Q_T(v) = \frac{h^{1+d/2}}{T}\int \sum_{t,s=1}^{T} K_{v,t,T}K_{v,s,T}K_{x,t,T}\varepsilon_{t,T}K_{x,s,T}\Delta_{s,T}(v,x)\frac{\pi(x)}{f^2(v,x)}dx$$

$$R_T(v) = \frac{h^{1+d/2}c_T}{T}\int \sum_{t,s=1}^{T} K_{v,t,T}K_{v,s,T}K_{x,t,T}\varepsilon_{t,T}K_{x,s,T}\Delta\Big(\frac{s}{T}, X_{s,T}\Big)\frac{\pi(x)}{f^2(v,x)}dx,$$

where we have used the uniform convergence results from Lemmas A1–A3 to replace the kernel density estimates $\hat{f}(u,x)$ and $\hat{f}(v,x)$ by the true densities $f(u,x)$ and $f(v,x)$.

We start by analyzing $Q_T(u)$. As a first step, the term is split up into two components:

$$Q_T(u) = Q_{T,1}(u) + Q_{T,2}(u)$$

with

$$Q_{T,1}(u) = \frac{h^{1+d/2}}{T}\int \sum_{t=1}^{T} K^2_{u,t,T}K^2_{x,t,T}\varepsilon_{t,T}\Delta_{t,T}(u,x)\frac{\pi(x)}{f^2(u,x)}dx$$

$$Q_{T,2}(u) = \frac{h^{1+d/2}}{T}\int \sum_{t=1}^{T}\sum_{s\neq t} K_{u,t,T}K_{u,s,T}K_{x,t,T}\varepsilon_{t,T}K_{x,s,T}\Delta_{s,T}(u,x)\frac{\pi(x)}{f^2(u,x)}dx.$$

It is easy to see that $\mathbb{E}[Q_{T,1}^2(u)] \leq C\frac{h^2}{Th^{d+1}} \to 0$, which immediately implies that $Q_{T,1}(u) = o_p(1)$. To cope with the term $Q_{T,2}(u)$, we further decompose it into two parts:

$$Q_{T,2}(u) = Q_{T,2,V}(u) + Q_{T,2,B}(u)$$

with

$$\begin{aligned} Q_{T,2,V}(u) &= \frac{h^{1+d/2}}{T}\int\sum_{t=1}^{T}\sum_{s\neq t} K_{u,t,T}K_{u,s,T}K_{x,t,T}\varepsilon_{t,T} \\ &\quad\times\big(K_{x,s,T}\Delta_{x,s,T}(u) - \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\big)\frac{\pi(x)}{f^2(u,x)}dx \\ Q_{T,2,B}(u) &= \frac{h^{1+d/2}}{T}\int\sum_{t=1}^{T}\sum_{s\neq t} K_{u,t,T}K_{u,s,T}K_{x,t,T}\varepsilon_{t,T}\mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\frac{\pi(x)}{f^2(u,x)}dx. \end{aligned}$$

The second moment of $Q_{T,2,V}(u)$ is given by the expression

$$\begin{aligned} \mathbb{E}[Q_{T,2,V}^2(u)] &= \Big(\frac{h^{1+d/2}}{T}\Big)^2\iint\sum_{t,t'=1}^{T}\sum_{s\neq t, s'\neq t'} K_{u,t,T}K_{u,s,T}K_{u,t',T}K_{u,s',T} \\ &\quad\times\mathbb{E}\Big[K_{x,t,T}\varepsilon_{t,T}K_{y,t',T}\varepsilon_{t',T}\big(K_{x,s,T}\Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\big) \\ &\quad\big(K_{y,s',T}\Delta_{s',T}(u,y) - \mathbb{E}[K_{y,s',T}\Delta_{s',T}(u,y)]\big)\Big]\frac{\pi(x)\pi(y)}{f^2(u,x)f^2(u,y)}dxdy. \end{aligned}$$

Using similar techniques as in the proof of (CLT1), this expression can be shown to converge to zero, which yields that $Q_{T,2,V}(u) = o_p(1)$. Furthermore,

$$Q_{T,2,B}(u) = Th^{1+d/2}\int W_{T,1}(u,x)W_{T,2}(u,x)\frac{\pi(x)}{f^2(u,x)}dx + o_p(1) \tag{2.12}$$

with

$$\begin{aligned} W_{T,1}(u,x) &= \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}\mathbb{E}[K_{x,t,T}\Delta_{t,T}(u,x)] \\ W_{T,2}(u,x) &= \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}X_{x,t,T}\varepsilon_{t,T}. \end{aligned}$$

Replacing the occurrences of $X_{t,T}$ in $W_{T,1}(u,x)$ by the approximating variables $X_t(\frac{t}{T})$ analogously as in the proof of (2.7) yields that

$$\begin{aligned} W_{T,1}(u,x) &= \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}\int\prod_{j=1}^{d} K_h(x^j - z^j) \\ &\quad\times\Big(m\Big(\frac{t}{T},z\Big) - m(u,x)\Big)f\Big(\frac{t}{T},z\Big)dz + O\Big(\frac{1}{Th^d}\Big) \end{aligned}$$

uniformly in u and x. Since

$$\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}\Big(m\Big(\frac{t}{T},z\Big)-m(u,x)\Big)f\Big(\frac{t}{T},z\Big) \\ =\int K_h(u-w)\big(m(w,z)-m(u,x)\big)f(w,z)dw+O\Big(\frac{1}{Th^2}\Big)$$

uniformly in u, x and z, we further get that

$$W_{T,1}(u,x)=\iint K_h(u-w)\prod_{j=1}^{d}K_h(x^j-z^j) \\ \times\big(m(w,z)-m(u,x)\big)f(w,z)dwdz+O\Big(\frac{1}{Th^d}+\frac{1}{Th^2}\Big).$$

Finally, exploiting the smoothness conditions on m and f together with the properties of the higher-order kernels, standard arguments yield that

$$W_{T,1}(u,x)=O\Big(h^r+\frac{1}{Th^d}+\frac{1}{Th^2}\Big) \tag{2.13}$$

uniformly in u and x. We thus obtain that

$$\mathbb{E}\Big(Th^{1+d/2}\int W_{T,1}(u,x)W_{T,2}(u,x)\frac{\pi(x)}{f^2(u,x)}dx\Big)^2 \\ =T^2h^{d+2}\iint W_{T,1}(u,x)W_{T,1}(u,y) \\ \times\Big(\frac{1}{T^2}\sum_{t=1}^{T}K_{u,t,T}^2\mathbb{E}\Big[K_{x,t,T}K_{y,t,T}\varepsilon_{t,T}^2\Big]\Big)\frac{\pi(x)\pi(y)}{f^2(u,x)f^2(u,y)}dxdy \\ =O\Big(T^2h^{d+2}\Big(h^r+\frac{1}{Th^d}+\frac{1}{Th^2}\Big)^2\frac{1}{Th}\Big)=o(1).$$

Recalling (2.12), this implies that $Q_{T,2,B}(u)=o_p(1)$. As a result,

$$Q_T(u)=o_p(1)$$

and analogously $Q_T(v)=o_p(1)$. Similar arguments can be used to show that $R_T(v)=o_p(1)$. This completes the proof. □

Proof of Lemma A6. Using (2.11), recalling that $c_T=(Th^{1+d/2})^{-1/2}$ and applying the uniform convergence results of Lemmas A1–A3 to replace the kernel densities $\hat{f}(u,x)$ and $\hat{f}(v,x)$ by the true densities $f(u,x)$ and $f(v,x)$, we obtain

$$Th^{1+d/2}\int B_T^2(u,v,x)\pi(x)dx=Q_T(v)+R_T(u)+R_T(v)+W_T(v)+o_p(1)$$

with

$$Q_T(v) = \int \Big(\frac{1}{T}\sum_{t=1}^{T} K_{v,t,T}K_{x,t,T}\Delta\Big(\frac{t}{T}, X_{t,T}\Big)\Big)^2 \frac{\pi(x)}{f^2(v,x)}dx$$

$$R_T(u) = Th^{1+d/2}\int \Big(\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}K_{x,t,T}\Delta_{t,T}(u,x)\Big)^2 \frac{\pi(x)}{f^2(u,x)}dx$$

$$R_T(v) = Th^{1+d/2}\int \Big(\frac{1}{T}\sum_{t=1}^{T} K_{v,t,T}K_{x,t,T}\Delta_{t,T}(v,x)\Big)^2 \frac{\pi(x)}{f^2(v,x)}dx$$

$$W_T(v) = \frac{h^{1+d/2}c_T}{T}\int \sum_{t,s=1}^{T} K_{v,t,T}K_{v,s,T}K_{x,t,T}K_{x,s,T}\Delta_{t,T}(v,x)\Delta\Big(\frac{s}{T}, X_{s,T}\Big)\frac{\pi(x)}{f^2(v,x)}dx,$$

where as in Lemma A5, we use the abbrevation $\Delta_{t,T}(v,x) = m(\frac{t}{T}, X_{t,T}) - m(v,x)$. It is easy to see that

$$Q_T(v) = \int \Delta^2(v,x)\pi(x)dx + o_p(1).$$

To analyze the term $R_T(u)$, we decompose it according to

$$R_T(u) = R_{T,1}(u) + R_{T,2}(u) + R_{T,3}(u)$$

with

$$R_{T,1}(u) = \frac{h^{1+d/2}}{T}\int \sum_{t,s=1}^{T} K_{u,t,T}K_{u,s,t}\big(K_{x,t,T}\Delta_{t,T}(u,x) - \mathbb{E}[K_{x,t,T}\Delta_{t,T}(u,x)]\big)$$
$$\times \big(K_{x,s,T}\Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\big)\frac{\pi(x)}{f^2(u,x)}dx$$

$$R_{T,2}(u) = \frac{2h^{1+d/2}}{T}\int \sum_{t,s=1}^{T} K_{u,t,T}K_{u,s,t}\mathbb{E}[K_{x,t,T}\Delta_{t,T}(u,x)]$$
$$\times \big(K_{x,s,T}\Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\big)\frac{\pi(x)}{f^2(u,x)}dx$$

$$R_{T,3}(u) = \frac{h^{1+d/2}}{T}\int \sum_{t,s=1}^{T} K_{u,t,T}K_{u,s,t}\mathbb{E}[K_{x,t,T}\Delta_{t,T}(u,x)]$$
$$\times \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\frac{\pi(x)}{f^2(u,x)}dx.$$

In what follows, these three terms are considered separately. To handle the term $R_{T,1}(u)$, we further split it up into two parts,

$$R_{T,1}(u) = R_{T,1,a}(u) + R_{T,1,b}(u),$$

where

$$R_{T,1,a}(u) = \frac{h^{1+d/2}}{T} \int \sum_{t=1}^{T} K_{u,t,T}^2 \big(K_{x,t,T}\Delta_{t,T}(u,x) - \mathbb{E}[K_{x,t,T}\Delta_{t,T}(u,x)]\big)^2 \frac{\pi(x)}{f^2(u,x)} dx$$

$$R_{T,1,b}(u) = \frac{h^{1+d/2}}{T} \int \sum_{t=1}^{T} \sum_{s \neq t} K_{u,t,T} K_{u,s,t} \big(K_{x,t,T}\Delta_{t,T}(u,x) - \mathbb{E}[K_{x,t,T}\Delta_{t,T}(u,x)]\big)$$
$$\times \big(K_{x,s,T}\Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\big) \frac{\pi(x)}{f^2(u,x)} dx.$$

Arguing analogously as in the proof of (2.13) yields $\mathbb{E}[R_{T,1,a}(u)] = o(1)$, which in turn gives that $R_{T,1,a}(u) = o_p(1)$. By similar arguments as in the proof of (CLT1), we further obtain that $R_{T,1,b}(u) = o_p(1)$. As a result, $R_{T,1}(u) = o_p(1)$. Repeating the arguments used to analyze the term $Q_{T,2,B}(u)$ in Lemma A5, we obtain that $R_{T,2}(u) = o_p(1)$. Finally, to prove that $R_{T,3}(u) = o_p(1)$, we again use arguments similar to those for (2.13). These yield

$$R_{T,3}(u) = Th^{1+d/2} \int \Big\{ \frac{1}{T} \sum_{t=1}^{T} K_{u,t,T} \mathbb{E}\Big[\prod_{j=1}^{d} K_h\Big(x^j - X_t^j\Big(\frac{t}{T}\Big)\Big)$$
$$\times \Big(m\Big(\frac{t}{T}, X_t\Big(\frac{t}{T}\Big)\Big) - m(u,x)\Big)\Big]\Big\}^2 \frac{\pi(x)}{f^2(u,x)} dx + O(h^{d/2})$$
$$= O\Big(Th^{1+d/2}\Big(h^r + \frac{1}{Th^2}\Big)^2\Big) + O(h^{d/2}) = o(1).$$

Putting everything together, we arrive at

$$R_T(u) = o_p(1)$$

and analogously at $R_T(v) = o_p(1)$. Slightly modifying the above arguments, we get that $W_T(v) = o_p(1)$ as well. □

Proof of Theorem 2.4

The proof mimics the arguments of Lemma A4 in the bootstrap world. We write

$$S_T^* = U_T^*(u) + U_T^*(v)$$

with

$$U_T^*(u) = Th^{1+d/2} \int \Big(\frac{1}{T} \sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \varepsilon_{t,T}^* \Big)^2 \frac{\pi(x)}{\hat{f}^2(u,x)} dx$$

$$U_T^*(v) = Th^{1+d/2} \int \Big(\frac{1}{T} \sum_{t=1}^{T} K_{v,t,T} K_{x,t,T} \varepsilon_{t,T}^* \Big)^2 \frac{\pi(x)}{\hat{f}^2(v,x)} dx.$$

As in Lemma A4, the two terms $U_T^*(u)$ and $U_T^*(v)$ can be analyzed separately. We restrict attention to $U_T^*(u)$, the arguments for $U_T^*(v)$ being the same. Using the uniform convergence results from Lemmas A1–A3, one can show that $\frac{1}{T}\sum_{t=1}^T K_{u,t,T}K_{x,t,T}\varepsilon_{t,T}^* = O_p(\sqrt{\log T/Th^{d+1}} + h^r)$ uniformly in u and x. This allows us to write

$$U_T^*(u) = U_{T,1}^*(u) + U_{T,2}^*(u) + o_p(1)$$

with

$$\begin{aligned}
U_{T,1}^*(u) &= \frac{h^{1+d/2}}{T}\sum_{t=1}^T K_{u,t,T}^2\Big(\int K_{x,t,T}^2\frac{\pi(x)}{f^2(u,x)}dx\Big)(\varepsilon_{t,T}^*)^2 \\
U_{T,2}^*(u) &= \frac{h^{1+d/2}}{T}\sum_{t=1}^T\sum_{s\neq t} K_{u,t,T}K_{u,s,T}\Big(\int K_{x,t,T}K_{x,s,T}\frac{\pi(x)}{f^2(u,x)}dx\Big)\varepsilon_{t,T}^*\varepsilon_{s,T}^*.
\end{aligned}$$

In what follows, we show that

$$U_{T,1}^*(u) = B_T(u) + o_p(1) \tag{2.14}$$

$$\mathbb{P}^*(U_{T,2}^*(u) \le x) \xrightarrow{P} \Phi_{0,V(u)}(x), \tag{2.15}$$

where $\Phi_{0,V(u)}$ is a Gaussian distribution function with mean zero and variance $V(u)$. Combining (2.14) and (2.15) immediately yields that $\mathbb{P}^*(U_T^*(u) - B_T(u) \le x) \xrightarrow{P} \Phi_{0,V(u)}(x)$. This completes the proof.

Proof of (2.14). Noting that

$$\begin{aligned}
\hat{\varepsilon}_{t,T}^2 = \varepsilon_{t,T}^2 &+ 2\varepsilon_{t,T}\Big[m\Big(\frac{t}{T},X_{t,T}\Big) - \hat{m}\Big(\frac{t}{T},X_{t,T}\Big)\Big] \\
&+ \Big[m\Big(\frac{t}{T},X_{t,T}\Big) - \hat{m}\Big(\frac{t}{T},X_{t,T}\Big)\Big]^2,
\end{aligned} \tag{2.16}$$

we have that

$$\begin{aligned}
\mathbb{E}^*[U_{T,1}^*(u)] &= \frac{h^{1+d/2}}{T}\sum_{t=1}^T K_{u,t,T}^2\Big(\int K_{x,t,T}^2\frac{\pi(x)}{f^2(u,x)}dx\Big)\hat{\varepsilon}_{t,T}^2 \\
&= \frac{h^{1+d/2}}{T}\sum_{t=1}^T K_{u,t,T}^2\Big(\int K_{x,t,T}^2\frac{\pi(x)}{f^2(u,x)}dx\Big)\varepsilon_{t,T}^2 + o_p(1) \\
&= \frac{h^{1+d/2}}{T}\sum_{t=1}^T K_{u,t,T}^2\mathbb{E}\Big[\Big(\int K_{x,t,T}^2\frac{\pi(x)}{f^2(u,x)}dx\Big)\varepsilon_{t,T}^2\Big] + o_p(1) \\
&= \mathbb{E}[U_{T,1}(u)] + o_p(1).
\end{aligned}$$

From Lemma A4 we already know that $\mathbb{E}[U_{T,1}(u)] = B_T(u) + o(1)$, leaving us with

$$\mathbb{E}^*[U^*_{T,1}(u)] = B_T(u) + o_p(1).$$

Moreover, it is easy to see that $U^*_{T,1}(u) - \mathbb{E}^*[U^*_{T,1}(u)] = o_p(1)$. □

Proof of (2.15). We rewrite $U^*_{T,2}(u)$ as

$$U^*_{T,2}(u) = \sum_{t,s=1}^{T} w^*_{s,t,T}$$

with

$$w^*_{s,t,T} = \begin{cases} \frac{h^{1+d/2}}{T} K_{u,t,T} K_{u,s,T} \Big(\int K_{x,t,T} K_{x,s,T} \frac{\pi(x)}{f^2(u,x)} dx \Big) \varepsilon^*_{t,T} \varepsilon^*_{s,T} & \text{for } t \neq s \\ 0 & \text{otherwise.} \end{cases}$$

As the bootstrap residuals are independent conditional on the sample $\{Y_{t,T}, X_{t,T}\}$, we can directly use the results of de Jong [20] on quadratic forms to show (2.15). In particular, it suffices to show that the following three conditions are satisfied (see Theorem 2.1 in [20]):

(CLT1*) $\text{Var}^*(U^*_{T,2}(u)) \xrightarrow{P} V(u)$.

(CLT2*) $\text{Var}^*(U^*_{T,2}(u))^{-1} \max_{1 \le s \le T} \sum_{t=1}^{T} \text{Var}^*(w^*_{s,t,T}) \xrightarrow{P} 0$.

(CLT3*) $\text{Var}^*(U^*_{T,2}(u))^{-2}\ \mathbb{E}^*[U^*_{T,2}(u)^4] \xrightarrow{P} 3$.

To show (CLT1*), we proceed similarly to the proof of (2.14). The details are omitted. For the proof of (CLT2*), note that

$$\max_{1 \le s \le T} \sum_{t=1}^{T} \text{Var}^*(w^*_{s,t,T}) \le C \iint \max_{1 \le s \le T} \Big(h^{d+1} K^2_{u,s,T} |K_{x,s,t} K_{y,s,T}| \hat{\varepsilon}^2_{s,T} \Big) \times \Big(\frac{h}{T^2} \sum_{t=1}^{T} K^2_{u,t,T} |K_{x,t,T} K_{y,t,T}| \hat{\varepsilon}^2_{t,T} \Big) \frac{\pi(x)\pi(y)}{f^2(u,x) f^2(u,y)} dxdy.$$

Using (2.16) together with the fact that $\max_{1 \le s \le T} \varepsilon^2_{s,T} = O_p(T^{2/\nu})$ for $\nu = 6 + \delta$, we obtain that

$$\max_{1 \le s \le T} \Big(h^{d+1} K^2_{u,s,T} |K_{x,s,t} K_{y,s,T}| \hat{\varepsilon}^2_{s,T} \Big) = O_p\Big(\frac{T^{2/\nu}}{h^{d+1}} \Big). \tag{2.17}$$

Moreover, it is easily seen that

$$\frac{h}{T^2}\sum_{t=1}^{T} K_{u,t,T}^2\Big(\iint |K_{x,t,T}K_{y,t,T}|\frac{\pi(x)\pi(y)}{f^2(u,x)f^2(u,y)}dxdy\Big)\hat{\varepsilon}_{t,T}^2 = O_p\Big(\frac{1}{T}\Big). \tag{2.18}$$

Combining (2.17) and (2.18), we arrive at

$$\max_{1\le s\le T}\sum_{t=1}^{T}\mathrm{Var}^*(w_{s,t,T}^*) = O_p\Big(\frac{1}{T^{1-2/\nu}h^{d+1}}\Big) = o_p(1),$$

the last equality following from the conditions on the bandwidth h listed in (C3). This shows (CLT2*). For the proof of (CLT3*), we use that

$$\begin{aligned}
\mathbb{E}^*[U_{T,2}^*(u)^4] &= \sum_{t\neq s, t'\neq s', t''\neq s'', t'''\neq s'''} \mathbb{E}^*\big[w_{s,t,T}^* w_{s',t',T}^* w_{s'',t'',T}^* w_{s''',t''',T}^*\big] \\
&= 12\sum_{t_1\neq t_2\neq t_3\neq t_4}\mathbb{E}^*\big[(w_{t_1,t_2,T}^*)^2(w_{t_3,t_4,T}^*)^2\big] + 8\sum_{t_1\neq t_2}\mathbb{E}^*\big[(w_{t_1,t_2,T}^*)^4\big] \\
&\quad + 48\sum_{t_1\neq t_2\neq t_3\neq t_4}\mathbb{E}^*\big[w_{t_1,t_2,T}^* w_{t_2,t_3,T}^* w_{t_3,t_4,T}^* w_{t_4,t_1,T}^*\big] \\
&\quad + 192\sum_{t_1\neq t_2\neq t_3}\mathbb{E}^*\big[w_{t_1,t_2,T}^*(w_{t_1,t_3,T}^*)^2 w_{t_2,t_3,T}^*\big] \\
&\quad + 48\sum_{t_1\neq t_2\neq t_3}\mathbb{E}^*\big[(w_{t_1,t_2,T}^*)^2(w_{t_2,t_3,T}^*)^2\big] \\
&=: Q_{T,1} + Q_{T,2} + Q_{T,3} + Q_{T,4} + Q_{T,5}.
\end{aligned}$$

Exploiting the mixing conditions on the model variables yields that $Q_{T,i} = o_p(1)$ for $i = 2,\ldots,5$. Moreover, noting that $\mathrm{Var}^*(U_{T,2}^*(u)) = 2\sum_{t_1\neq t_2} E^*(w_{t_1,t_2,T}^*)^2$, it is easily seen that $Q_{T,1} = 3\mathrm{Var}^*(U_{T,2}^*(u))^2 + o_p(1)$. This completes the proof of (CLT3*). □

Chapter 3

Locally Stationary Multiplicative Volatility Modelling

3.1 Introduction

Given the ever-changing economic and financial environment, it is quite plausible that many financial time series behave in a nonstationary way. Especially over longer horizons, structural changes may occur. Thus, the technical assumption of stationarity is likely to be violated in many cases. This issue has been pointed out by numerous authors in recent years. In particular, it has been claimed that many interesting stylized facts of financial return and volatility series can be neatly explained by employing nonstationary models (see e.g. Mikosch & Stărică [33], [34], and [35]).

An attractive way to deal with nonstationarities in financial time series is the theory on locally stationary processes introduced by Dahlhaus (cf. [4], [5], and [6]). Intuitively speaking, a process is locally stationary if over short time spans (i.e. locally in time) it behaves approximately stationary. In recent years, many locally stationary models have been proposed in the financial time series context. Usually, these models are extensions of parametric time series models where the parameters are allowed to change smoothly over time. Within the family of ARCH models, for example, Dahlhaus & Subba Rao [8] have introduced a class of ARCH processes with time-varying parameters.

A closely related locally stationary model which has been explored in a number of studies is given by the equation

$$Y_{t,T} = \tau\left(\frac{t}{T}\right)\varepsilon_t \quad \text{for } t = 1, \ldots, T, \tag{3.1}$$

where $Y_{t,T}$ are log-returns, τ is a smooth deterministic function of time and $\{\varepsilon_t\}$ is a standard stationary GARCH process with $\mathbb{E}[\varepsilon_t^2] = 1$. As usual in the literature on locally stationary models, the time-varying parameter τ does not depend on real time t, but on rescaled time $\frac{t}{T}$. We comment on this feature in more detail in Section 3.2. Model (3.1) has been considered for example in Feng [13], where the τ-function is estimated nonparametrically. Engle & Rangel [12] work with a closely related model, where the τ-component is modelled parametrically as a flexible exponential spline function. A multivariate generalization of model (3.1) is studied in Linton & Hafner [28].

Model (3.1) can be considered as a GARCH process with time-varying parameters, with certain restrictions imposed on the parameter functions. In particular, the unconditional volatility level $\mathbb{E}[Y_{t,T}^2]$ is given by the time-dependent function $\tau^2(\frac{t}{T})$, which is allowed to vary smoothly over time. In reality, the volatility level is unlikely to change deterministically over time. Instead it reflects and varies with changes in the economic and financial environment. Therefore, the τ-function should depend on certain economic and financial variables. In model (3.1), these dependencies are not modelled explicitly. Instead, rescaled time serves as a catch-all for omitted explanatory variables.

These considerations show that in a more realistic version of model (3.1), the τ-function should depend on economic and financial influences. However, there is clearly no way to come up with a model that incorporates all relevant variables. One way to deal with this is to use rescaled time as a proxy for the omitted variables. To formalize these ideas, we propose the model

$$Y_{t,T} = \tau\Big(\frac{t}{T}, X_t\Big)\varepsilon_t, \tag{3.2}$$

where $Y_{t,T}$ are log-returns, X_t is an $\mathbb{R}^d$-valued random vector of economic or financial covariates and τ is a smooth function of time and the variables X_t. As before, $\{\varepsilon_t\}$ is a standard GARCH process. To countervail the curse of dimensionality, we split up the τ-function into multiplicative components thus yielding the model

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big)\prod_{j=1}^{d}\tau_j(X_t^j)\varepsilon_t, \tag{3.3}$$

where τ_0 and τ_j for $j = 1, \ldots, d$ are smooth functions of time and the regressors X_t^j, respectively. As will be seen in Section 3.2, the multiplicative specification of the τ-function in (3.3) not only avoids the curse of dimensionality but also allows a direct interpretation of the various components.

In the following sections, we give an in-depth theoretical treatment of model (3.3). The complete formulation of the model together with its assumptions is given in Section 3.2. In Section 3.3, we propose a two-step procedure to estimate both the nonparametric and the parametric components of the model. To estimate the nonparametric functions τ_j for $j = 0, \dots, d$, we extend the smooth backfitting procedure of Mammen et al. [29] to our locally stationary stetting. Having estimates $\tilde{\tau}_j$ of the functions τ_j, we can construct approximate expressions $\tilde{\varepsilon}_t$ of the GARCH variables ε_t. This allows us to estimate the GARCH parameters of the model via approximate quasi-maximum likelihood methods in a second step. Consistency and asymptotic normality of our estimators are shown in Section 3.4.

The contribution of this chapter is twofold. From a technical point of view, we extend the asymptotic results for model (3.1) to a more general framework in which the τ-function depends both on rescaled time and stochastic regressors. This vastly complicates both steps of the asymptotic analysis and as a result, we cannot extend existing proving techniques as provided in Linton & Hafner [28] in a straightforward manner. In terms of volatility modelling, we introduce a flexible framework which allows to capture both nonstationarities and influences from the economic and financial environment. As the component functions τ_j in our model are completely nonparametric, we are able to explore the form of the relationship between volatility and its potential sources. Therefore, our model allows us to extend existing parametric studies on the sources of volatility as conducted e.g. in Engle & Rangel [12] and Ghysels, Engle & Sohn [11].

To illustrate the usefulness of our model and to complement the technical analysis, we present an empirical example in Section 3.5. There, the model is applied to S&P 500 return data using various interest rate spreads as explanatory variables.

3.2 Model

In this section, we rigorously introduce our model. We observe a sample of log-returns $Y_{t,T}$ and covariates X_t for $t = 1, \dots, T$, where $X_t = (X_t^1, \dots, X_t^d)$ is an $\mathbb{R}^d$-valued random vector. The return series is assumed to follow the process

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X_t^j)\varepsilon_t \quad \text{for } t = 1, \dots, T \tag{3.4}$$

with

$$\varepsilon_t = \sigma_t \eta_t \quad \text{and} \quad \sigma_t^2 = w_0 + a_0 \varepsilon_{t-1}^2 + b_0 \sigma_{t-1}^2.$$

Here, τ_0 and τ_j $(j = 1, \dots, d)$ are smooth nonparametric functions of time and the stochastic regressors, respectively. Furthermore, $\{\varepsilon_t\}$ is a strictly stationary GARCH process with i.i.d. residuals η_t that satisfy $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = 1$. For simplicity, we restrict attention to the GARCH(1,1) specification.

In order to conduct meaningful asymptotics, we let the function τ_0 depend on rescaled time $\frac{t}{T}$ rather than on real time t. Thus, τ_0 is defined on $(0, 1]$ rather than on $\{1, \dots, T\}$. In what follows, we denote rescaled time by $x_0 \in (0, 1]$. It relates to observed time $t \in \{0, \dots, T\}$ through the mapping $t = [x_0 T]$, where $[x]$ denotes the smallest integer weakly larger than x. If we defined the function τ_0 in terms of observed time, we would not get additional information on the structure of τ_0 around a particular time point t as the sample size T increases. Within the framework of rescaled time, in contrast, the function τ_0 is observed on a finer and finer grid on the unit interval as T grows. Thus, we obtain more and more information on the local structure of τ_0 around each point x_0 in rescaled time. This is the reason why we can make meaningful asymptotic considerations within this framework. A detailed discussion of the concept of rescaled time can be found in Dahlhaus [5].

We make the following assumptions on the model components.

(C1) The process $\{X_t, \varepsilon_t, \sigma_t\}$ is strictly stationary.

(C2) The functions τ_0 and τ_j $(j = 1, \dots, d)$ are twice (continuously) differentiable, strictly positive, and bounded away from zero. The second derivatives are Lipschitz continuous.

(C3) The variables X_t and ε_t satisfy the condition that $\mathbb{E}[\varepsilon_t^2 | X_t] = \mathbb{E}[\varepsilon_t^2]$ almost surely.

Further technical assumptions needed for deriving asymptotic results are given in the relevant sections.

By assuming in (C1) that the covariates X_t and the errors ε_t are strictly stationary, we restrict the potential sources of nonstationarity in our model. Nonstationarities stem exclusively from the time-varying trend function τ_0. If we pinned this function down at a particular value, say $\bar{\tau}_0$, the resulting model would be strictly stationary. Thus, the function τ_0 is supposed to catch all nonstationary fluctuations in the model. Note that it is possible to weaken (C1) to allow for local stationarities in the covariates X_t. We conjecture that we would obtain almost identical asymptotic results in this case. We elaborate on this point in Section 3.6.

Conditions (C1) and (C2) ensure that the process $\{Y_{t,T}\}$ is locally stationary. Using the smoothness of τ_0, we have

$$\big|Y_{t,T} - Y_t(x_0)\big| \le C\Big|\frac{t}{T} - x_0\Big|U_t, \tag{3.5}$$

where C is a constant independent of x_0, t and T, $Y_t(x_0) = \tau_0(x_0)\prod_{j=1}^d \tau_j(X_t^j)\varepsilon_t$, and $U_t = \prod_{j=1}^d \tau_j(X_t^j)\varepsilon_t$. Note that both $\{Y_t(x_0)\}$ and $\{U_t\}$ are strictly stationary processes. As $U_t = O_p(1)$, we obtain from (3.5) that

$$\big|Y_{t,T} - Y_t(x_0)\big| = O_p\Big(\Big|\frac{t}{T} - x_0\Big|\Big). \tag{3.6}$$

Therefore, if $\frac{t}{T}$ is close to x_0, then $Y_{t,T}$ is close to $Y_t(x_0)$ at least in a stochastic sense. Put differently, locally in time, the process $\{Y_{t,T}\}$ is close to the stationary process $\{Y_t(x_0)\}$. In this sense, the process $\{Y_{t,T}\}$ is locally stationary.
We close this section with a remark on the interpretation of the nonparametric components of model (3.4). First, note that the functions $\tau_0, \dots, \tau_d$ and the GARCH residual ε_t are only identified up to a multiplicative constant in model (3.4). Thus we are free to rescale them in a suitable way. For instance, if we normalize the components such that $\mathbb{E}[\varepsilon_t^2] = 1$, then (C3) implies that

$$\mathbb{E}[Y_{t,T}^2|X_t] = \tau_0^2\Big(\frac{t}{T}\Big)\prod_{j=1}^d \tau_j^2(X_t^j). \tag{3.7}$$

Thus, the product of the τ-components gives the volatility at time t conditional on the covariates X_t. If we additionally scale the model to satisfy $\mathbb{E}[\prod_{j=1}^d \tau_j^2(X_t^j)] = 1$, we obtain that

$$\mathbb{E}[Y_{t,T}^2] = \tau_0^2\Big(\frac{t}{T}\Big),$$

i.e. the deterministic function of time $\tau_0^2(\frac{t}{T})$ gives the time-varying unconditional volatility. In (3.7), $\tau_0^2(\frac{t}{T})$ thus specifies the unconditional volatility level and the product of the remaining components $\prod_{j=1}^d \tau_j^2(X_t^j)$ is the multiplicative factor by which the volatility conditional on X_t deviates from the unconditional level.

3.3 Estimation Procedure

We now turn to the two-step estimation procedure alluded to in the introduction. In the first step, we provide estimates of the nonparametric functions $\tau_0, \dots, \tau_d$. In the second step, we use these nonparametric estimates to obtain estimators of the GARCH parameters. The following assumptions ensure that the various steps of our procedure are well-defined.

(C4) The conditional volatility σ_t^2 is bounded away from zero and the GARCH residuals η_t have a density with respect to Lebesgue measure which is bounded in a neighbourhood of zero.

(C5) The variables X_t and ε_t are such that $\mathbb{E}[\log \varepsilon_t^2 | X_t] = \mathbb{E}[\log \varepsilon_t^2] = 0$.

(C6) The variables X_t have compact support, say $[0,1]^d$.

Assumptions (C4) and (C5) are needed for the first estimation step, as will become clear in the next subsection. Note that (C4) ensures $\log \varepsilon_t^2$ to be finite almost surely and that it is thus required for (C5) to make sense. (C6) is only needed for the second estimation step. For the first step, we could allow the support of X_t to be unbounded and estimate the functions $\tau_0, \ldots, \tau_d$ uniformly over compact subsets of the support. However, for ease of notation, we assume (C6) throughout the chapter.

3.3.1 Estimation of the Nonparametric Model Components

In order to estimate the nonparametric functions $\tau_0, \ldots, \tau_d$, we do not consider the multiplicative model (3.4) directly. Instead we transform (3.4) to obtain an additive structure by squaring and taking the logarithm. This yields

$$Z_{t,T} = m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + u_t, \tag{3.8}$$

where $Z_{t,T} := \log Y_{t,T}^2$, $m_j := \log \tau_j^2$ for $j = 0, \ldots, d$, and $u_t := \log \varepsilon_t^2$. The above transformation is well-behaved under assumptions (C2) and (C4). The functions $m_0, \ldots, m_d$ in (3.8) are only identified up to an additive constant. To identify them, we assume that

$$\int_0^1 m_0(x_0)dx_0 = 0 \quad \text{and} \quad \int_{\mathbb{R}} m_j(x_j)p_j(x_j)dx_j = 0 \quad \text{for } j = 1, \ldots, d,$$

where p_j is the marginal density of X_t^j. With this normalization, we can rewrite (3.8) as

$$Z_{t,T} = m_c + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j) + u_t, \tag{3.9}$$

where m_c is a constant and by assumption (C5), $\mathbb{E}[u_t|X_t] = 0$. By a slight abuse of notation, the normalized functions are again labelled as $m_0, \ldots, m_d$. In what follows, we will write $x = (x_0, x_{-0})$ with x_0 denoting a point in rescaled time and $x_{-0} = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Equation (3.9) has the form of an additive regression model with component functions $m_0, \dots, m_d$ and error term u_t. It is nonstandard in the sense that it contains the deterministic trend function m_0. As a consequence, the model dynamics are not stationary any more. We obtain estimators $\tilde{m}_0, \dots, \tilde{m}_d$ of the functions $m_0, \dots, m_d$ by extending the smooth backfitting approach introduced by Mammen et al. [29] to allow for these nonstationarities. Estimators of the functions $\tau_0, \dots, \tau_d$ in (3.4) are then obtained by setting

$$\tilde{\tau}_j = \sqrt{\exp(\tilde{m}_j)}$$

for $j = 0, \dots, d$. In the remainder of this subsection we introduce the smooth backfitting estimators $\tilde{m}_0, \dots, \tilde{m}_d$. For simplicity, we restrict attention to smooth backfitting based on Nadaraya-Watson estimators. Alternatively, the approach could be based on local linear smoothers.

Before introducing the estimator in our setting, we reconsider the standard stationary case. To do this, we fix the time argument at some point $x_0 \in (0, 1]$, thus leaving us with the additive regression model

$$Z_t(x_0) = m_c + m_0(x_0) + \sum_{j=1}^{d} m_j(X_t^j) + u_t, \tag{3.10}$$

where the dependent variables $Z_t(x_0)$ are strictly stationary and $m_c + m_0(x_0)$ is the model constant. In order to define the smooth backfitting estimators for this standard model, we introduce the function spaces

$$\mathcal{F}(p) = \{g : \mathbb{R}^d \to \mathbb{R} \mid \int g^2(x_{-0}) p(x_{-0}) dx_{-0} < \infty\}$$
$$\mathcal{F}_{\text{add}}(p) = \{g \in \mathcal{F}(p) \mid g(x_{-0}) = g_1(x_1) + \dots + g_d(x_d) \ (p \text{ a.s.})\},$$

where p is the joint density of the regressors $X_t = (X_t^1, \dots, X_t^d)$, $\mathcal{F}(p)$ is the class of $L^2(p)$-functions and $\mathcal{F}_{\text{add}}(p)$ is the subclass of functions that allow an additive representation. Further, let $\Pi(Z_t(x_0)|X_t)$ be the orthogonal projection of $Z_t(x_0)$ onto the space of additive functions $\mathcal{F}_{\text{add}}(p)$ and define

$$m(x_0, \cdot) := \mathbb{E}[Z_t(x_0)|X_t] = \arg\min_{g \in \mathcal{F}(p)} \mathbb{E}\big(Z_t(x_0) - g(X_t)\big)^2$$
$$m_{\text{add}}(x_0, \cdot) := \Pi(Z_t(x_0)|X_t) = \arg\min_{g \in \mathcal{F}_{\text{add}}(p)} \mathbb{E}\big(Z_t(x_0) - g(X_t)\big)^2.$$

Note that the additive regression function in model (3.10) is given by $m_{\text{add}}(x_0, \cdot)$. Using basic properties of orthogonal projections we obtain that

$$\begin{aligned} m_{\text{add}}(x_0, \cdot) &= \arg \min_{g \in \mathcal{F}_{\text{add}}(p)} \mathbb{E}\big(Z_t(x_0) - g(X_t)\big)^2 \\ &= \arg \min_{g \in \mathcal{F}_{\text{add}}(p)} \mathbb{E}\big(m(x_0, X_t) - g(X_t)\big)^2 \\ &= \arg \min_{g \in \mathcal{F}_{\text{add}}(p)} \int_{\mathbb{R}^d} \big(m(x_0, x_{-0}) - g(x_{-0})\big)^2 p(x_{-0}) dx_{-0}. \end{aligned} \tag{3.11}$$

Thus, $m_{\text{add}}(x_0, \cdot)$ solves the projection problem as formulated in (3.11). The smooth backfitting estimator of $m_{\text{add}}(x_0, \cdot)$ is now defined as the solution to an empirical version of (3.11), where the functions $m(x_0, \cdot)$ and p are replaced by d-dimensional kernel estimators.

In order to extend the above approach to our framework, we proceed as follows: We regard rescaled time as an additional regressor and let the L_2-projection in (3.11) cover the time dimension. With $q(x) := I(x_0 \in (0, 1])p(x_{-0})$, this leads to the projection equation

$$\begin{aligned} m_{\text{add}} &= \arg \min_{g \in \mathcal{F}_{\text{add}}(q)} \int_0^1 \int_{\mathbb{R}^d} \big(m(x_0, x_{-0}) - g(x_0, x_{-0})\big)^2 p(x_{-0}) dx_{-0}\, dx_0 \\ &= \arg \min_{g \in \mathcal{F}_{\text{add}}(q)} \int_{\mathbb{R}^{d+1}} \big(m(x) - g(x)\big)^2 q(x) dx, \end{aligned} \tag{3.12}$$

where similar to the standard case

$$\begin{aligned} \mathcal{F}(q) &= \{g : \mathbb{R}^{d+1} \to \mathbb{R} \mid \int_{\mathbb{R}^{d+1}} g^2(x) q(x) dx < \infty\} \\ \mathcal{F}_{\text{add}}(q) &= \{g \in \mathcal{F}(q) \mid g(x) = g_0(x_0) + g_1(x_1) + \cdots + g_d(x_d) \ (q \text{ a.s.})\}. \end{aligned}$$

The L_2-projection in (3.12) is with respect to the density function $q(x) = I(x_0 \in (0, 1])p(x_{-0})$ on $\mathbb{R}^{d+1}$. Thus, rescaled time is treated in a similar way to an additional stochastic regressor which is uniformly distributed over $(0, 1]$ and independent of the variables X_t. The intuition for this is the following: Firstly, as the variables X_t are strictly stationary, their distribution is time-invariant. In this sense their stochastic behaviour is independent of rescaled time $\frac{t}{T}$. Thus rescaled time behaves similarly to an additional stochastic variable that is independent of X_t. Secondly, as the points $\frac{t}{T}$ are evenly spaced over the unit interval, a variable with a uniform distribution closely replicates the pattern of rescaled time.

Just as for the standard stationary model without time trend component we define the smooth backfitting estimator as the solution to an empirical counterpart of the projection problem (3.12) with m and q replaced by kernel estimators $\hat{m}$ and $\hat{q}$. Hence, the smooth backfitting estimator $\tilde{m}$ of m_{add} is given by

$$\tilde{m} = \arg\min_{g \in \mathcal{F}_{\text{add}}(\hat{q})} \int_{\mathbb{R}^{d+1}} \big(\hat{m}(x) - g(x)\big)^2 \hat{q}(x) dx, \tag{3.13}$$

where the minimization is done under the constraints

$$\int \tilde{m}_j(x_j)\hat{p}_j(x_j)dx_j = 0 \quad \text{for } j = 0, \ldots, d,$$

and the function class $\mathcal{F}_{\text{add}}(\hat{q})$ is defined as before with q replaced by the estimate $\hat{q}$. In the above formulas, $\hat{p}_j$ is a kernel estimator of p_j for $j = 0, \ldots, d$, where we define $p_0(x_0) = I(x_0 \in (0,1])$. Explicit expressions for these estimators are given below in (3.15) and (3.18). We further define $\hat{q}(x) = \frac{1}{T}\sum_{t=1}^T K_h(x_0, \frac{t}{T}) \prod_{k=1}^d K_h(x_k, X_t^k)$ and let $\hat{m}(x) = \sum_{t=1}^T K_h\big(x_0, \frac{t}{T}\big) \prod_{k=1}^d K_h(x_k, X_t^k) Y_{t,T} \big/ \hat{q}(x)$ be a $(d+1)$-dimensional Nadaraya-Watson smoother. In these definitions,

$$K_h(v, w) = \frac{K_h(v - w)}{\int_0^1 K_h(s - w) ds}$$

is a modified kernel weight, where $K_h(v) = \frac{1}{h}K(\frac{v}{h})$ and the kernel function $K(\cdot)$ integrates to one. These weights have the property that $\int_0^1 K_h(v, w) dv = 1$ for all v, which is needed to derive the asymptotic results of the smooth backfitting estimators.

By differentiation, we can show that the solution to the projection problem (3.13) is characterized by the system of integral equations

$$\tilde{m}_j(x_j) = \hat{m}_j(x_j) - \sum_{k \neq j} \int \tilde{m}_k(x_k) \frac{\hat{p}_{k,j}(x_k, x_j)}{\hat{p}_j(x_j)}\, dx_k - \tilde{m}_c \tag{3.14}$$

$$\int \tilde{m}_j(x_j)\hat{p}_j(x_j)dx_j = 0$$

for $j = 0, \ldots, d$ with $\tilde{m}_c = \frac{1}{T}\sum_{t=1}^T Z_{t,T}$. The kernel estimators which show up in (3.14) are given by

$$\hat{p}_j(x_j) = \frac{1}{T}\sum_{t=1}^T K_h(x_j, X_t^j) \tag{3.15}$$

$$\hat{p}_{j,k}(x_j, x_k) = \frac{1}{T}\sum_{t=1}^T K_h(x_j, X_t^j) K_h(x_k, X_t^k) \tag{3.16}$$

$$\hat{m}_j(x_j) = \frac{1}{T}\sum_{t=1}^T K_h(x_j, X_t^j) Z_{t,T} / \hat{p}_j(x_j). \tag{3.17}$$

for $j, k = 1, \dots, d$. Here, $\hat{p}_j$ is the one-dimensional kernel density estimator of the marginal density p_j of X_t^j, $\hat{p}_{j,k}$ is the two-dimensional kernel density estimate of the joint density $p_{j,k}$ of (X_t^j, X_t^k), and $\hat{m}_j(x_j)$ is a one-dimensional local constant smoother. Further,

$$\hat{p}_0(x_0) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) \tag{3.18}$$

$$\hat{p}_{0,k}(x_0, x_k) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) K_h(x_k, X_t^k) \tag{3.19}$$

$$\hat{m}_0(x_0) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) Z_{t,T}/\hat{p}_0(x_0). \tag{3.20}$$

Note that it would be more natural to define $\hat{p}_0(x_0) = I(x_0 \in (0, 1])$, as we already know the "true density" of rescaled time. However, for technical reasons, we set $\hat{p}_0(x_0) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_0, \frac{t}{T})$. This creates a behaviour of $\hat{p}_0$ at the boundary of the support $(0, 1]$ analogous to that of $\hat{p}_j$ for $j = 1, \dots, d$.[1]
A solution to the set of equations (3.14) can be obtained by an iterative projection algorithm, which converges for arbitrary starting values, see Mammen et al. [29], who also establish the asymptotic properties of this solution under very general high order conditions. To prove consistency and asymptotic normality of our estimators, we show that these high order conditions are satisfied in our framework.

3.3.2 Estimation of the Parametric Model Components

To motivate the second step in our estimation procedure, we first consider an infeasible estimator of the model parameters. Suppose that the nonparametric components $\tau_0^2, ..., \tau_d^2$ were known. In this situation, the GARCH variables ε_t^2 would be observable, since

$$\varepsilon_t^2 = \frac{Y_{t,T}^2}{\tau_0^2(\frac{t}{T}) \prod_{k=1}^{d} \tau_k^2(X_k^t)}. \tag{3.21}$$

[1] Alternatively, we could define $\hat{p}_0(x_0) = \int_0^1 K_h(x_0, v)dv$. (Note that $\int_0^1 K_h(x_0, v)dv = 1$ for $x_0 \in [2C_1 h, 1 - 2C_1 h]$, where $[-C_1, C_1]$ is the support of the kernel function K.) Moreover, we could set $\hat{p}_{0,k}(x_0, x_k) = \hat{p}_0(x_0)\hat{p}_k(x_k)$, thereby exploiting the "independence" of rescaled time and the other regressors.

The GARCH parameters $\phi_0 := (w_0, a_0, b_0)$ could thus be estimated by standard quasi maximum likelihood methods, where the quasi log-likelihood is given by

$$l_T(\phi) = -\sum_{t=1}^{T} \Big(\log v_t^2(\phi) + \frac{\varepsilon_t^2}{v_t^2(\phi)} \Big). \tag{3.22}$$

Here, $\phi = (w, a, b)$ and

$$v_t^2(\phi) = \begin{cases} \frac{w}{1-b} & \text{for } t = 1 \\ w + a\varepsilon_{t-1}^2 + b v_{t-1}^2(\phi) & \text{for } t = 2, \dots, T \end{cases} \tag{3.23}$$

is the conditional volatility of the GARCH process with starting value $v_0^2(\phi) = w/(1-b)$.
As the functions $\tau_0^2, \dots, \tau_d^2$ are not observed, we cannot apply this standard approach. However, given the estimates $\tilde{\tau}_0^2, \dots, \tilde{\tau}_d^2$ from the first estimation step, we can replace ε_t^2 by the standardized residuals

$$\tilde{\varepsilon}_t^2 = \frac{Y_{t,T}^2}{\tilde{\tau}_0^2(\frac{t}{T}) \prod_{k=1}^{d} \tilde{\tau}_k^2(X_k^t)} \tag{3.24}$$

and use these as approximations to ε_t^2 in the quasi maximum likelihood estimation. The quasi log-likelihood then becomes

$$\tilde{l}_T(\phi) = -\sum_{t=1}^{T} \Big(\log \tilde{v}_t^2(\phi) + \frac{\tilde{\varepsilon}_t^2}{\tilde{v}_t^2(\phi)} \Big), \tag{3.25}$$

where analogously to (3.23),

$$\tilde{v}_t^2(\phi) = \begin{cases} \frac{w}{1-b} & \text{for } t = 1 \\ w + a\tilde{\varepsilon}_{t-1}^2 + b \tilde{v}_{t-1}^2(\phi) & \text{for } t = 2, \dots, T \end{cases} \tag{3.26}$$

is the approximate conditional volatility. Our estimator $\tilde{\phi}$ of the true parameter values ϕ_0 is now defined as

$$\tilde{\phi} = \arg\max_{\phi \in \Phi} \tilde{l}_T(\phi), \tag{3.27}$$

where the parameter space Φ is assumed to be compact. In comparison to this, the standard maximum likelihood estimator for the case in which the τ-components are known is defined as

$$\hat{\phi} = \arg\max_{\phi \in \Phi} l_T(\phi). \tag{3.28}$$

3.4 Asymptotics

In Subsection 3.4.1 we treat the nonparametric estimates $\tilde{\tau}_0, \ldots, \tilde{\tau}_d$ and in Subsection 3.4.2 we give results on the GARCH estimates $\tilde{\phi}$.
In order to derive the asymptotic properties of the nonparametric estimators, we need the following assumptions.

(C7) The kernel K is bounded, has compact support ($[-C_1, C_1]$, say) and is symmetric about zero. Moreover, it fulfills the Lipschitz condition that there exists a positive constant L such that $|K(u) - K(v)| \leq L|u - v|$.

(C8) The density p of X_t and the densities $p_{(0,l)}$ of (X_t, X_{t+l}), $l = 1, 2, \ldots$, are uniformly bounded. Furthermore, p is bounded away from zero on $[0, 1]^d$. The first partial derivatives of p exist and are continuous.

(C9) Let $Z_t = Z_{t,T} - m_0(\frac{t}{T})$. For some $\theta > \frac{8}{3}$, $\mathbb{E}[|Z_t|^\theta] < \infty$.

(C10) The conditional densities $f_{X_t|Z_t}$ of X_t given Z_t and $f_{X_t,X_{t+l}|Z_t,Z_{t+l}}$ of (X_t, X_{t+l}) given (Z_t, Z_{t+l}), $l = 1, 2, \ldots$, exist and are bounded.

(C11) The process $\{X_t, \varepsilon_t, \sigma_t\}$ is strongly mixing with mixing coefficients α satisfying $\alpha(k) \leq a^k$ for some $0 < a < 1$.

(C12) The bandwidth h satisfies either of the following:
(a) $T^{\frac{1}{5}} h \to c_h$ for some constant c_h.
(b) $T^{\frac{1}{4}+\delta} h \to c_h$ for some constant c_h and some small $\delta > 0$.

Note that the above assumptions are very similar to the conditions that can be found in Mammen et al. [29] for the strictly stationary case. It should also be mentioned that we do not necessarily require exponentially decaying mixing rates as assumed in (C11). These could alternatively be replaced by sufficiently high polynomial rates. We nevertheless make the stronger assumption (C11) to keep the notation and structure of the proofs as clear as possible.
Additionally to the above assumptions, we require the following conditions to hold for the GARCH estimates to be consistent and asymptotically normal.

(C13) The parameter space Φ is a compact subset of $\{\phi \in \mathbb{R}^3 \,|\, \phi = (w, a, b)$ with $0 < \underline{\kappa} \leq w, a \leq \overline{\kappa} < \infty$ and $0 \leq b < 1\}$ with constants $\underline{\kappa}$ and $\overline{\kappa}$. The true parameter $\phi_0 = (w_0, a_0, b_0)$ is an interior point of Φ and $a_0 + b_0 < 1$.

(C14) $\mathbb{E}[\varepsilon_t^{8+\delta}] < \infty$, for some $\delta > 0$.

(C15) $\mathbb{E}[\varepsilon_{t-i}^2 \varepsilon_{t-k}^2 | X_{t-k}^j] \leq C$ uniformly in i, j, and k.

(C13) is a standard assumption in the theory on GARCH models. (C14) and (C15) are rather technical assumptions that are only needed to show asymptotic normality of the GARCH estimates.

3.4.1 Asymptotics for the Nonparametric Model Components

We now give asymptotic results for the estimators $\tilde{\tau}_0, \ldots, \tilde{\tau}_d$ in our multiplicative model. First, we derive the asymptotic properties of the estimates $\tilde{m}_0, \ldots, \tilde{m}_d$ in the additively transformed model. From these results, we can directly infer the asymptotic behaviour of their multiplicative counterparts.
In view of the second estimation step, we require uniform as opposed to pointwise convergence as well as a uniform expansion of the estimates. The latter is provided in Appendix A. The former is given in the following theorem, which shows that $\tilde{m}_0, \ldots, \tilde{m}_d$ converge uniformly to the true functions at the usual one-dimensional nonparametric rates. The theorem also characterizes the asymptotic distribution of the nonparametric estimates.

Theorem 3.1. *Suppose that conditions (C1)–(C11) hold.*

(a) Assume that the bandwidth h satisfies (C12a) or (C12b). Then, for $I_h = [2C_1 h, 1 - 2C_1 h]$ and $I_h^c = [0, 2C_1 h) \cup (1 - 2C_1 h, 1]$,

$$\sup_{x_j \in I_h} \left| \tilde{m}_j(x_j) - m_j(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \tag{3.29}$$

$$\sup_{x_j \in I_h^c} \left| \tilde{m}_j(x_j) - m_j(x_j) \right| = O_p(h) \tag{3.30}$$

for all $j = 0, \ldots, d$.

(b) Assume that the bandwidth h satisfies (C12a). Then, for any $x_0, \ldots, x_d \in (0,1)$,

$$T^{\frac{2}{5}} \begin{bmatrix} \tilde{m}_0(x_0) - m_0(x_0) \\ \vdots \\ \tilde{m}_d(x_d) - m_d(x_d) \end{bmatrix} \xrightarrow{d} N(B_m(x), V_m(x))$$

with the bias term $B_m(x) = [c_h^2\beta_0(x_0), \dots, c_h^2\beta_d(x_d)]'$ *and the covariance matrix* $V_m(x) = \text{diag}(v_0(x_0), \dots, v_d(x_d))$. *Here,* $v_0(x_0) = c_h^{-1} c_K \sum_{l=-\infty}^{\infty} \gamma_u(l)$ *and* $v_j(x_j) = c_h^{-1} c_K \sigma_j^2(x_j)/p_j(x_j)$ *for* $j = 1, \dots, d$ *with* $c_K = \int K^2(u)du$, $\gamma_u(l) = Cov(u_t, u_{t+l})$ *and* $\sigma_j^2(x_j) = Var(u_t|X_t^j = x_j)$. *Furthermore, the functions* $\beta_j(x_j)$ *are the components of the* $L_2(p)$*-projection of the function* β *defined in Lemma C3 of Appendix C onto the space of additive functions.*

The rates of convergence given in Theorem 3.1(a) differ for the interior and boundary regions of the support of the covariates. In particular, the rate near the boundary in (3.30) is slower than in the interior (3.29). However, the slow convergence at the boundary does not pose a problem for the second estimation step as the size of the boundary region shrinks sufficiently fast as $T \to \infty$.
The asymptotic results for $\tilde{m}_0, \dots, \tilde{m}_d$ carry over to $\tilde{\tau}_0, \dots, \tilde{\tau}_d$ and their squared version. This is clear from the fact that $\tilde{\tau}_j = \sqrt{\exp(\tilde{m}_j)}$ for $j = 0, \dots, d$. As we are mainly interested in the squared version of the estimates, we report the asymptotic results for these in the following corollary.

Corollary 3.1. *Suppose that conditions (C1)–(C11) hold.*

(a) Under the conditions of Theorem 3.1(a), it holds that

$$\sup_{x_j \in I_h} \left| \tilde{\tau}_j^2(x_j) - \tau_j^2(x_j) \right| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \tag{3.31}$$

$$\sup_{x_j \in I_h^c} \left| \tilde{\tau}_j^2(x_j) - \tau_j^2(x_j) \right| = O_p(h) \tag{3.32}$$

for all $j = 0, \dots, d$.

(b) Under the conditions of Theorem 3.1(b), it holds that

$$T^{\frac{2}{5}} \begin{bmatrix} \tilde{\tau}_0^2(x_0) - \tau_0^2(x_0) \\ \vdots \\ \tilde{\tau}_d^2(x_d) - \tau_d^2(x_d) \end{bmatrix} \xrightarrow{d} N(B_\tau(x), V_\tau(x)),$$

where the bias $B_\tau(x)$ *and the variance* $V_\tau(x)$ *are as in Theorem 3.1(b) with* $c_h^2\beta_j(x_j)$ *replaced by* $\tau_j^2(x_j)c_h^2\beta_j(x_j)$ *and* $v_j(x_j)$ *replaced by* $\tau_j^4(x_j)v_j(x_j)$.

The main idea of the proofs is to exploit the fact that rescaled time behaves similarly to a random variable which has a uniform distribution on $(0, 1]$ and is independent of the other covariates. The details are given in Appendix A.

3.4.2 Asymptotics for the Parametric Model Components

Given the estimators for $\tau_0^2, \dots, \tau_d^2$ from the first step, the GARCH parameters ϕ_0 are estimated by $\tilde{\phi}$ as outlined in Subsection 3.3.2. In this subsection, we look at consistency and asymptotic normality of $\tilde{\phi}$. The following theorem establishes consistency.

Theorem 3.2. *Suppose that the bandwidth h satisfies (C12a) or (C12b). In addition, let assumptions (C1)–(C11) and (C13) be fulfilled. Then $\tilde{\phi}$ is a consistent estimator of ϕ_0, i.e.*

$$\tilde{\phi} \xrightarrow{P} \phi_0.$$

We next give a result on the limiting distribution of the GARCH estimates which shows that these are asymptotically normal.

Theorem 3.3. *Suppose that the bandwidth h satifies (C12b) and let assumptions (C1)–(C11) together with (C13)–(C15) be fulfilled. Then it holds that*

$$\sqrt{T}(\tilde{\phi} - \phi_0) \xrightarrow{d} N(0, \Sigma).$$

Details on the covariance matrix Σ can be found in Appendix B.

The proof of asymptotic normality is the theoretically most challenging part of this chapter. The details are postponed to the appendices. For now we will be content with providing an outline. By the usual Taylor expansion argument, we arrive at

$$\sqrt{T}(\tilde{\phi} - \phi_0) = -\Big(\frac{1}{T}\frac{\partial^2 \tilde{l}_T(\bar{\phi})}{\partial\phi\partial\phi^T}\Big)^{-1}\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial\phi},$$

where $\bar{\phi}$ is an intermediate point between $\tilde{\phi}$ and ϕ_0. As in the standard case, we can show that the second derivative on the right-hand side converges in probability to a deterministic matrix. The asymptotic distribution is thus determined by the term $\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial\phi}$, which we rewrite as

$$\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial\phi_i} = \underbrace{\frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i}}_{=:A_1} + \underbrace{\Big(\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial\phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i}\Big)}_{=:A_2}.$$

We will show that this term is asymptotically normal. The main challenge to do so is to derive a stochastic expansion of the term A_2. This requires rather involved and nonstandard arguments which are presented in detail in Appendix B.

In particular, we cannot just extend the arguments presented in Linton & Hafner [28] to fit our setting. Once we have provided the expansion of A_2, we are in a position to apply a central limit theorem to the sum $A_1 + A_2$, which completes the proof. We will see that the term A_2 is itself asymptotically normal and thus contributes to the limit distribution of $\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi}$. As a consequence, we obtain a larger asymptotic variance than in the standard case (where only the term A_1 occurs). This reflects the additional uncertainty that results from not knowing the functions $\tau_0, \dots, \tau_d$.

3.5 Application

To illustrate our model, we apply it to a sample of daily financial data spanning the period from the beginning of 2001 until the middle of 2010. The estimated model is given by

$$Y_{t,T} = \tau_0^2\left(\frac{t}{T}\right) \prod_{j=1}^{3} \tau_j^2(X_t^j)\varepsilon_t^2, \tag{3.33}$$

where $Y_{t,T}$ are S&P 500 log-returns and the covariates are various interest rate spreads. Specifically, we include the default spread defined as the difference between the yield of Aaa and Baa bonds. This is commonly used as a measure for credit default risk. We also include the term spread between 10-year and 3-month treasuries in order to capture the slope of the yield curve. Finally we include the LOIS, the difference between the LIBOR and the return on the overnight interest swap, which is used as a measure for liquidity in the interbank market.[2]

The estimation results for the nonparametric model components are presented in Figure 3.1. The solid lines represent the estimators $\tilde{\tau}_j^2$, the dashed lines are the pointwise 5% confidence intervals. The bandwidths for the function fits are chosen by a rule of thumb following the application in Yu, Mammen & Park [42]. For $j = 1, 2, 3$, the estimates $\tilde{\tau}_j^2$ are normalized such that $\tilde{\tau}_j^2(x_j^m) = 1$, where x_j^m is the median of the j-th covariate X_t^j. This means that the effect of the j-th covariate on volatility is normalized to 1 if it takes a "normal" (i.e. its median) value. Note that due to this normalization, the estimate $\tilde{\tau}_0^2$ gives the time-varying unconditional volatility only up to a multiplicative constant.

[2]All data are taken from the Federal Reserve Bank.

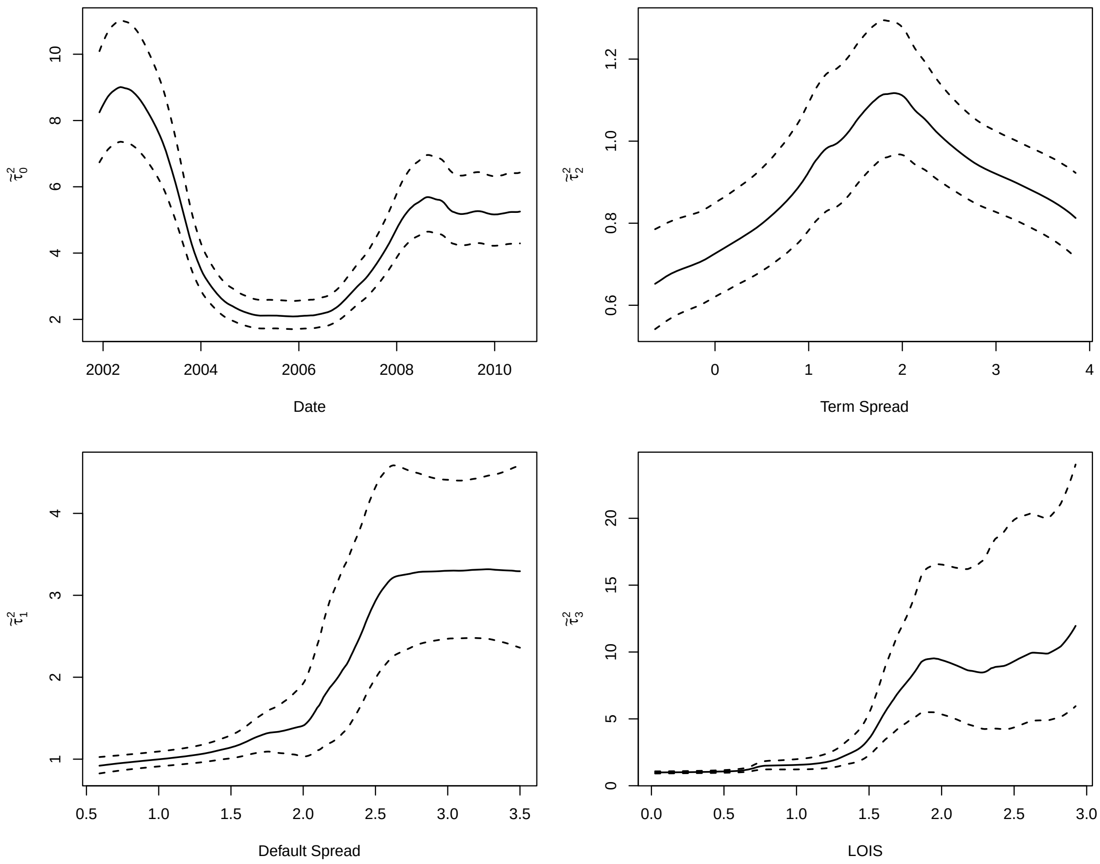

Figure 3.1: Multiplicative function fits

The curves in Figure 3.1 are to be interpreted as follows. As already mentioned, the estimate $\tilde{\tau}_0^2$ specifies the unconditional volatility level up to a multiplicative factor. Furthermore, as

$$\mathbb{E}[Y_{t,T}^2 | X_t] = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{3} \tau_j(X_t^j), \tag{3.34}$$

the estimates $\tilde{\tau}_j^2$ for $j = 1, 2, 3$ can be interpreted as the multiplicative effect of the covariates X_t^j on S&P 500 volatility. To illustrate this, let us compare volatility between two different settings. Hold the covariates X_t^{-j} fixed at some value x_{-j}. Change X_t^j from its median x_j^m to some value x_j. From (3.34), one can then see that volatility is changed by the factor $\tau_j^2(x_j)/\tau_j^2(x_j^m) = \tau_j^2(x_j)$. Consequently, the fits $\tilde{\tau}_j^2(x_j)$ estimate the factor by which volatility gets increased or dampened, when the j-th covariate changes from a normal value (i.e. its median) to some other more extreme value.

We now look at the estimated component functions in Figure 3.1 one after the other, starting with the functions that depend on the three spreads.

- The bottom left hand panel shows that the effect of the default spread on volatility is an increasing monotonic function. Interestingly, it is highly nonlinear. In particular, for relatively low as well as for moderate values around the median, the function is close to one. This means that there is next to no effect on volatility in these areas. However, for large values of the default spread there is a sharp increase of the function. Thus, a large value of the spread, i.e. high risk of credit default, dramatically increases S&P 500 volatility.

- The bottom right hand panel depicts the effect of the LOIS. The overall shape is similar to the one for the default spread. The only real difference is the range of the function and thus the size of the effect on volatility.

- In the top right hand panel, one can see the estimated effect of the term spread. The function has an inverted U-shape. This means that both for very large and very small values of the term spread volatility is dampened. It is hard to give a clear intuition for this result. A possible explanation is that the term spread is not only a risk premium but also an indicator for the future state of the economy. The shape of the curve may reflect the interaction of these two roles of the spread.

We next turn to the discussion of $\tilde{\tau}_0^2$. A rescaled version of $\tilde{\tau}_0^2$, which estimates the unconditional volatility level, is given by the solid line in Figure 3.2. The dashed line is the estimated unconditional volatility obtained from the simpler model (3.1) without covariates.
Both curves in Figure 3.2 clearly show the volatility increase in the two recent crises. The hump at the beginning of the sample corresponds to the aftermath of the technology crisis and the events after 9/11, whereas the one at the end depicts the recent financial crisis. Interestingly, the unconditional volatility level in our model is much lower in the recent financial crisis than the level in the simpler model without covariates. This suggests that our regressors explain a considerable part of unconditional volatility in the recent crisis. During the earlier crisis, however, the difference between the two curves is not so striking. Thus, the explanatory power of our covariates in this period seems to be moderate (if it exists at all). This is

quite plausible as these variables are from the financial sector and the turbulence following 2001 was not primarily driven by events in this sector.

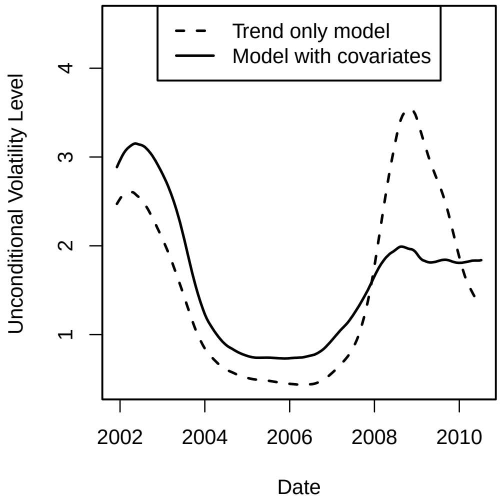

Figure 3.2: Time-varying unconditional volatilities

We finally come to the estimation results for the parametric model components. In Table 3.1, we compare the GARCH estimates of our model with the ones obtained from the simpler model (3.1) and from a standard GARCH(1,1) model.

	$\tilde{w}$	$\tilde{a}$	$\tilde{b}$	$\tilde{a}+\tilde{b}$
Standard GARCH(1,1)	0.012	0.082	0.911	0.992
Model with trend	0.034	0.075	0.891	0.966
Model with trend and covariates	0.057	0.064	0.878	0.942

Table 3.1: GARCH parameter estimates

The sum of the two estimated parameters $\tilde{a}+\tilde{b}$ reported in the last column of Table 3.1 measures the persistence of shocks to volatility. One can see that this persistence measure decreases from 0.992 to 0.966 when accounting for time-varying unconditional volatility. This is in line with previous findings in the literature (compare e.g. Feng [13]). Including our covariates in the model further decreases the estimated persistence to 0.942. Note that the reported decrease in persistence is quite dramatic even though it may seem rather small at first sight (compare the discussion in Mikosch & Stărică [33] on this issue).

To sum up, our results suggest that we can explain a good deal of S&P 500 volatility by our model. However, for an in-depth analysis one would also need to validate the model. Specifically, such a step would aid in the choice of covariates. One possible model validation procedure is described in Nielsen & Sperlich [36]. Finally, from a practitioner's perspective it would also be interesting to look at the forecasting performance of the model.

3.6 Extensions

We use this section to discuss possible extensions and amendments to the model.

3.6.1 Estimation of the Covariance Matrix Σ

It is not at all trivial to construct a consistent estimate of the covariance matrix Σ introduced in Theorem 3.3. This results from the very complicated structure of Σ. In particular, the exact expression for Σ involves functions obtained from a higher order expansion of the stochastic part of the backfitting estimates (see Theorem A1 in Appendix A). It is very complicated to calculate the exact form of these functions and even more challenging to give consistent estimates for them. The construction of a consistent estimate of Σ is thus a difficult theoretical problem.

3.6.2 Efficiency Gains

We next discuss how to gain efficiency in the estimation of both the nonparametric and parametric components of the model. For this purpose, we adapt the procedure in Linton & Hafner [28].
First consider the nonparametric model components. If we knew the variables σ_t, we could divide the multiplicative model (3.4) by them to obtain

$$\frac{Y_{t,T}}{\sigma_t} = \tau_0\Big(\frac{t}{T}\Big)\prod_{j=1}^{d}\tau_j(X_t^j)\eta_t. \tag{3.35}$$

Squaring and taking the logarithm would then yield an additive regression model with error terms $v_t := \log\eta_t^2 - \mathbb{E}[\log\eta_t^2]$. These terms have a smaller variance than the errors $u_t = \log\varepsilon_t^2$ in the additive regression (3.9). In particular, $\mathrm{Var}(v_t) = \mathrm{Var}(\log\eta_t^2) \leq \mathrm{Var}(\log\sigma_t^2) + \mathrm{Var}(\log\eta_t^2) = \mathrm{Var}(u_t)$. This suggests that at least for $j = 1,\ldots,d$, the infeasible smooth backfitting estimates based on equation

(3.35) are more efficient in terms of asymptotic variance than our estimates.[3] Not knowing the variables σ_t, we could use our estimation procedure to get initial estimates of them. Plugging these estimates into (3.35), it should be possible to obtain feasible smooth backfitting estimates with smaller asymptotic variance.
We now come to the parametric model components. Again, it should be possible to adapt the procedure described in Linton & Hafner [28] to our setting in order to gain efficiency in the estimation of the parametric model parts. In the case of normally distributed GARCH residuals η_t, we may even be able to obtain estimates that reach the semiparametric efficiency bound. We omit the details and refer the interested reader to the description of the procedure in Linton & Hafner [28].

3.6.3 Locally Stationary Covariates

Finally, one may want to allow for locally stationary regressors in model (3.4). In this case,

$$Y_{t,T} = \tau_0\Big(\frac{t}{T}\Big) \prod_{j=1}^{d} \tau_j(X_{t,T}^j)\varepsilon_t \quad \text{for } t = 1, \dots, T,$$

where ε_t is a strictly stationary GARCH residual as before, but where the covariates $X_{t,T}$ now form a locally stationary process.
In this extended model, we face the following problem: If the regressors are locally stationary, their stochastic behaviour may change over time. As a consequence, rescaled time will not behave like an additional regressor any more that is *independent* of the other covariates. This drastically complicates the asymptotic analysis. In particular, it is not clear whether it is still possible to obtain one-dimensional convergence rates for the backfitting estimates.

3.7 Conclusion

We have proposed a new semiparametric volatility model, which generalizes the class of models $Y_{t,T} = \tau(\frac{t}{T})\varepsilon_t$, as for example considered in Feng [13] and Engle & Rangel [12]. These models are able to account for nonstationarities in the volatility process. In addition, we are able to include covariates in a nonparametric way, hence allowing us to flexibly capture the effects of the financial and economic environment.

[3] Whether the infeasible estimate for $j = 0$ is more efficient depends on the autocorrelations of the errors u_t. Specifically, there are efficiency gains if and only if $\sum_{k=-\infty}^{\infty} \mathrm{Cov}(u_0, u_k) > \mathrm{Var}(v_t)$.

We have derived the asymptotic theory both for the nonparametric and the parametric part of the model. To estimate the nonparametric model components, we have extended the smooth backfitting approach of Mammen et al. [29] to our nonstationary setting. Given the backfitting estimators, we were able to construct GARCH parameter estimates and to show that they are asymptotically normal. In particular, they converge at the fast parametric rate even though the nonparametric smoothers from the first step have slower nonparametric convergence rates. We have finally illustrated the strengths of our model by applying it to financial data. In particular, our semiparametric approach allows us to estimate the form of the relationship between volatility and its potential sources. Therefore, we manage to go beyond existing parametric approaches such as in Engle & Rangel [12] and Ghysels, Engle & Sohn [11].

Appendix A

In this appendix, we prove Theorem 3.1, which describes the asymptotic behaviour of our smooth backfitting estimates. For the proof, we split up the estimates into a "stochastic" part and a "bias" part. In Theorem A1, we provide a uniform expansion of the stochastic part. This result is an extension of a related expansion given in Mammen & Park [30] in the context of bandwidth selection in additive models. The bias part is treated in Theorem A2. The proof of both theorems requires uniform convergence results for the kernel smoothers that enter the backfitting procedure as pilot estimates. These results are summarized in Appendix C. As will turn out, both theorems are not only needed for the first estimation step but also for the derivation of the asymptotics of the GARCH estimates in the second step. In what follows, we use the symbol C to denote a finite real constant which may take a different value on each occurrence.

Proof of Theorem 3.1

We decompose the backfitting estimates $\tilde{m}_j$ into a stochastic part $\tilde{m}_j^A$ and a bias part $\tilde{m}_j^B$ according to

$$\tilde{m}_j(x_j) = \tilde{m}_j^A(x_j) + \tilde{m}_j^B(x_j).$$

The two components are defined by

$$\tilde{m}_j^S(x_j) = \hat{m}_j^S(x_j) - \sum_{k \neq j} \int \tilde{m}_k^S(x_k) \frac{\hat{p}_{k,j}(x_k, x_j)}{\hat{p}_j(x_j)} \, dx_k - \tilde{m}_c^S \tag{3.36}$$

for $S = A, B$. Here, $\hat{m}_k^A$ and $\hat{m}_k^B$ denote the stochastic part and the bias part of the Nadaraya-Watson pilote estimates defined as

$$\hat{m}_j^A(x_j) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)u_t/\hat{p}_j(x_j) \tag{3.37}$$

$$\hat{m}_j^B(x_j) = \frac{1}{T}\sum_{t=1}^{T} K_h(x_j, X_t^j)\big[m_c + m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j(X_t^j)\big]/\hat{p}_j(x_j) \tag{3.38}$$

for $j = 0, \ldots, d$, where we set $X_t^0 = \frac{t}{T}$ to shorten the notation. Furthermore, $\tilde{m}_c^A = \frac{1}{T}\sum_{t=1}^{T} u_t$ and $\tilde{m}_c^B = \frac{1}{T}\sum_{t=1}^{T}\{m_c + m_0(\frac{t}{T}) + \sum_{j=1}^{d} m_j(X_t^j)\}$. We now analyse the convergence behaviour of $\tilde{m}_j^A$ and $\tilde{m}_j^B$ separately.
We first provide a higher order expansion of the stochastic part $\tilde{m}_j^A$. The following result extends Theorem 6.1. in Mammen & Park [30] (in particular equation (6.3) of this theorem) to our locally stationary setting.

Theorem A1. *Suppose that assumptions (C1)–(C11) apply and that the bandwidth h satisfies (C12a) or (C12b). Then uniformly for $0 \leq x_j \leq 1$,*

$$\tilde{m}_j^A(x_j) = \hat{m}_j^A(x_j) + \frac{1}{T}\sum_{t=1}^{T} r_{j,t}(x_j)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big),$$

where $r_{j,t}(\cdot) := r_j(\frac{t}{T}, X_t, \cdot)$ are absolutely uniformly bounded functions with

$$|r_{j,t}(x_j') - r_{j,t}(x_j)| \leq C|x_j' - x_j|$$

for a constant $C > 0$.

Proof. We cannot apply Theorem 6.1. in Mammen & Park [30] directly, which treats the i.i.d. case. In what follows, we outline the arguments needed to deal with our locally stationary setting. For an additive function $g(x) = g_0(x_0) + \ldots + g_d(x_d)$, let

$$\hat{\psi}_j g(x) = g_0(x_0) + \ldots + g_{j-1}(x_{j-1}) + g_j^*(x_j) + g_{j+1}(x_{j+1}) + \ldots + g_d(x_d)$$

with

$$g_j^*(x_j) = -\sum_{k\neq j}\int g_k(x_k)\frac{\hat{p}_{j,k}(x_j, x_k)}{\hat{p}_j(x_j)}dx_k + \sum_{k=0}^{d}\int g_k(x_k)\hat{p}_k(x_k)dx_k.$$

Using the uniform convergence results from Appendix C and exploiting our model assumptions, we can show that Lemma 3 in Mammen et al. [29] applies in our case. For $\tilde{m}^A(x) = \tilde{m}_0^A(x_0) + \ldots + \tilde{m}_d^A(x_d)$, we therefore have the expansion

$$\tilde{m}^A(x) = \sum_{r=0}^{\infty} \hat{S}^r \hat{\tau}(x),$$

where $\hat{S} = \hat{\psi}_d \cdots \hat{\psi}_0$ and $\hat{\tau}(x) = \hat{\psi}_d \cdots \hat{\psi}_1[\hat{m}_0^A(x_0) - \hat{m}_{c,0}^A] + \ldots + \hat{\psi}_d[\hat{m}_{d-1}^A(x_{d-1}) - \hat{m}_{c,d-1}^A] + [\hat{m}_d^A(x_d) - \hat{m}_{c,d}^A]$ with $\hat{m}_{c,j}^A = \int \hat{m}_j^A(x_j)\hat{p}_j(x_j)dx_j$. Now decompose $\tilde{m}^A(x)$ according to

$$\tilde{m}^A(x) = \hat{m}^A(x) - \hat{m}_c^A + \sum_{r=0}^{\infty} \hat{S}^r(\hat{\tau}(x) - (\hat{m}^A(x) - \hat{m}_c^A)) + \sum_{r=1}^{\infty} \hat{S}^r(\hat{m}^A(x) - \hat{m}_c^A)$$

with $\hat{m}^A(x) = \hat{m}_0^A(x_0) + ... + \hat{m}_d^A(x_d)$ and $\hat{m}_c^A = \hat{m}_{c,0}^A + \ldots + \hat{m}_{c,d}^A$. We show that there exist absolutely bounded functions $a_t(x)$ with $|a_t(x) - a_t(y)| \leq C\|x - y\|$ for a constant C s.t.

$$\sum_{r=1}^{\infty} \hat{S}^r(\hat{m}^A(x) - \hat{m}_c^A) = \frac{1}{T}\sum_{t=1}^{T} a_t(x)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{3.39}$$

uniformly in x. A similar claim holds for the term $\sum_{r=0}^{\infty} \hat{S}^r(\hat{\tau}(x) - (\hat{m}^A(x) - \hat{m}_c^A))$. As $\hat{m}_c^A = (d+1)\frac{1}{T}\sum_{t=1}^{T} u_t$, this implies the result.
The idea behind the proof of (3.39) is as follows: From the definition of the operators $\hat{\psi}_j$, it can be seen that

$$\hat{S}(\hat{m}^A(x) - \hat{m}_c^A) = \sum_{j=0}^{d-1} \hat{\psi}_d \cdots \hat{\psi}_{j+1}\Big(\sum_{k=j+1}^{d} S_{j,k}(x_j)\Big) \tag{3.40}$$

with

$$S_{j,k}(x_j) = -\int \frac{\hat{p}_{j,k}(x_j, x_k)}{\hat{p}_j(x_j)}(\hat{m}_k^A(x_k) - \hat{m}_{c,k}^A)dx_k.$$

In what follows, we show that the terms $S_{j,k}(x_j)$ have the representation

$$S_{j,k}(x_j) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{p_{j,k}(x_j, X_t^k)}{p_j(x_j)p_k(X_t^k)} - 1\Big)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{3.41}$$

uniformly in x_j. Thus, they essentially have the desired form $\frac{1}{T}\sum_t w_{t,k}(x_j)u_t$ with some weights $w_{t,k}$. This allows us to infer that

$$\hat{S}(\hat{m}^A(x) - \hat{m}_c^A) = \frac{1}{T}\sum_{t=1}^{T} b_t(x)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{3.42}$$

uniformly in x with some absolutely bounded functions b_t satisfying $|b_t(x)-b_t(y)| \leq C\|x-y\|$ for some $C > 0$.
To show (3.41), we exploit the mixing behaviour of the variables X_t. Plugging the definition of $\hat{m}_k^A$ into the term $S_{j,k}$, we can write

$$S_{j,k}(x_j) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\int \frac{\hat{p}_{j,k}(x_j,x_k)}{\hat{p}_j(x_j)\hat{p}_k(x_k)}K_h(x_k,X_t^k)dx_k - 1\Big)u_t.$$

Then applying the uniform convergence results from Appendix C, we can replace the density estimates in the above expression by the true densities. This yields

$$\begin{aligned} S_{j,k}(x_j) &= -\frac{1}{T}\sum_{t=1}^{T}\Big(\int \frac{p_{j,k}(x_j,x_k)}{p_j(x_j)p_k(x_k)}K_h(x_k,X_t^k)dx_k - 1\Big)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big) \\ &=: S_{j,k}^*(x_j) + o_p\Big(\frac{1}{\sqrt{T}}\Big) \end{aligned}$$

uniformly for $x_j \in [0,1]$. In the final step, we show that

$$S_{j,k}^*(x_j) = -\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{p_{j,k}(x_j,X_t^k)}{p_j(x_j)p_k(X_t^k)} - 1\Big)u_t + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

again uniformly in x_j. This is done by applying a covering argument together with an exponential inequality for mixing variables. The employed techniques are similar to those used to establish the uniform convergence results of Appendix C. Finally, again using the results from Appendix C, it can be shown that

$$\sum_{r=0}^{\infty}\hat{S}^r(\hat{m}^A(x) - \hat{m}_c^A) = \sum_{r=0}^{\infty}S^{r-1}\hat{S}(\hat{m}^A(x) - \hat{m}_c^A) + o_p\Big(\frac{1}{\sqrt{T}}\Big) \tag{3.43}$$

uniformly in x, where S is defined analogously to $\hat{S}$ with the density estimates replaced by the true densities. Combining (3.42) and (3.43) yields the result. □

We now turn to the bias part $\tilde{m}_j^B$.

Theorem A2. *Suppose that (C1)–(C11) hold. If the bandwidth h satisfies (C12a), then*

$$\sup_{x_j \in I_h}|\tilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h^2) \tag{3.44}$$

$$\sup_{x_j \in I_h^c}|\tilde{m}_j^B(x_j) - m_j(x_j)| = O_p(h) \tag{3.45}$$

for $j = 0, \dots, d$. If the bandwidth satisfies (C12b), we have

$$\sup_{x_j \in I_h} \left| \tilde{m}_j^B(x_j) + \frac{1}{T} \sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j) \right| = O_p(h^2) \tag{3.46}$$

$$\sup_{x_j \in I_h^c} \left| \tilde{m}_j^B(x_j) + \frac{1}{T} \sum_{t=1}^{T} m_j(X_t^j) - m_j(x_j) \right| = O_p(h) \tag{3.47}$$

for $j = 0, \dots, d$.

Proof. The result follows from Theorem 3 in Mammen et al. [29]. To make sure that the latter theorem applies in our case, we have to show that the high-order conditions (A1)–(A5), (A8), and (A9) from [29] are fulfilled in our setting.[4] This can be achieved by using the results from Appendix C, in particular the expansion of $\hat{m}_j^B$ given in Lemma C3, and by following the arguments for the proof of Theorem 4 in [29]. To see that (3.44) has to be replaced by (3.46) in the undersmoothing case with $h = O(T^{-(\frac{1}{4}+\delta)})$, note that

$$\int \alpha_{T,k}(x_k)\hat{p}_k(x_k)dx_k = \frac{1}{T}\sum_{t=1}^{T} m_k(X_t^k) + O_p(h^2)$$

with $\frac{1}{T}\sum_{t=1}^{T} m_k(X_t^k) = O_p(\frac{1}{\sqrt{T}})$. Using this in the proof of Theorem 3 instead of $\int \alpha_{T,k}(x_k)\hat{p}_k(x_k)dx_k = \gamma_{T,j} + o_p(h^2)$ with $\gamma_{T,j} = O(h^2)$ gives (3.46). □

By combining Theorems A1 and A2, it is now straightforward to complete the proof of Theorem 3.1. □

Appendix B

This appendix contains the proofs of Theorems 3.2 and 3.3, which show consistency and asymptotic normality of the GARCH estimates. By far the most difficult part is the proof of asymptotic normality, which is split up into different bits. We first give the main steps of the argument, postponing the major technical issues to a series of lemmas. As already pointed out in Subsection 3.4.2, the main challenge of the proof is to derive a stochastic expansion of $\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi}$. This expansion is given in Lemmas B1–B3. Note that as in Appendix A, C denotes a finite real constant which may take a different value on each occurrence.

[4] Note that (A6) is not needed for the proof of Theorem 3 as opposed to the statement in [29].

Preliminary Remarks

To start with, we list some facts that are useful for the proof of Theorems 3.2 and 3.3. These facts concern the behaviour of the approximate GARCH variables $\tilde{\varepsilon}_t$ and of the conditional volatilities $\tilde{v}_t^2(\phi)$, which were defined in Subsection 3.3.2. For ease of notation, we use the shorthand $\tau(x) = \prod_{j=0}^d \tau_j(x_j)$ in what follows.

(G1) We can express $\tilde{\varepsilon}_t^2 - \varepsilon_t^2$ as

$$\tilde{\varepsilon}_t^2 - \varepsilon_t^2 = \varepsilon_t^2 \Big[\frac{\tau^2(\frac{t}{T}, X_t) - \tilde{\tau}^2(\frac{t}{T}, X_t)}{\tau^2(\frac{t}{T}, X_t)} + R_\varepsilon\Big(\frac{t}{T}, X_t\Big)\Big]$$

with $\sup_{x \in [0,1]^{d+1}} |R_\varepsilon(x)| = O_p(h^2)$.

(G2) The conditional volatility $v_t^2(\phi)$ has the expansion

$$v_t^2(\phi) = w \sum_{k=1}^{t-1} b^{k-1} + a \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 + b^{t-1} \frac{w}{1-b},$$

which yields that

$$\tilde{v}_t^2(\phi) - v_t^2(\phi) = \sum_{k=1}^{t-1} a b^{k-1} (\tilde{\varepsilon}_{t-k}^2 - \varepsilon_{t-k}^2).$$

(G3) It holds that

$$\max_{1 \le t \le T} \sup_{\phi \in \Phi} \big| \tilde{v}_t^2(\phi) - v_t^2(\phi) \big| = O_p(h).$$

(G4) It holds that

$$\frac{1}{\tilde{v}_t^2(\phi)} - \frac{1}{v_t^2(\phi)} = \frac{v_t^2(\phi) - \tilde{v}_t^2(\phi)}{v_t^2(\phi) v_t^2(\phi)} + R_t(\phi)$$

with $\max_{1 \le t \le T} \sup_{\phi \in \Phi} |R_t(\phi)| = O_p(h^2)$.

(G5) The derivatives of $v_t^2(\phi)$ with respect to the parameters w, a, and b are given by

$$\frac{\partial v_t^2(\phi)}{\partial w} = \sum_{k=1}^{t-1} b^{k-1} + \frac{b^{t-1}}{1-b}$$

$$\frac{\partial v_t^2(\phi)}{\partial a} = \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2$$

$$\frac{\partial v_t^2(\phi)}{\partial b} = w \Big(\sum_{k=1}^{t-1} (k-1) b^{k-2} + \frac{(t-1) b^{t-2}}{1-b} + \frac{b^{t-1}}{(1-b)^2} \Big) + a \sum_{k=1}^{t-1} (k-1) b^{k-2} \varepsilon_{t-k}^2.$$

The above facts are straightforward to verify. We thus omit the details.

Proof of Theorem 3.2

Let $l_T(\phi)$ and $\tilde{l}_T(\phi)$ be the likelihood functions introduced in (3.22) and (3.25) and define

$$l(\phi) = \mathbb{E}\Big[\frac{1}{T} l_T(\phi)\Big].$$

By the triangle inequality,

$$\sup_{\phi\in\Phi} \Big|\frac{1}{T}\tilde{l}_T(\phi) - l(\phi)\Big| \leq \sup_{\phi\in\Phi} \Big|\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T} l_T(\phi)\Big| + \sup_{\phi\in\Phi} \Big|\frac{1}{T} l_T(\phi) - l(\phi)\Big|.$$

From standard theory we know that

$$\sup_{\phi\in\Phi} \Big|\frac{1}{T} l_T(\phi) - l(\phi)\Big| = o_p(1)$$

and that $l(\phi)$ is a continuous function of ϕ with a unique maximum at ϕ_0. If we can further show that

$$\sup_{\phi\in\Phi} \Big|\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T} l_T(\phi)\Big| = o_p(1), \tag{3.48}$$

then standard theory on M-estimation implies $\tilde{\phi} \xrightarrow{P} \phi_0$.
In order to prove (3.48), we decompose the difference $\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T} l_T(\phi)$ into three parts (A), (B), and (C) and show that each of these is uniformly $o_p(1)$. We write

$$\begin{aligned}
&\frac{1}{T}\tilde{l}_T(\phi) - \frac{1}{T} l_T(\phi) \\
&= -\frac{1}{T}\sum_{t=1}^{T}\Big(\log \tilde{v}_t^2(\phi) + \frac{\tilde{\varepsilon}_t^2}{\tilde{v}_t^2(\phi)}\Big) + \frac{1}{T}\sum_{t=1}^{T}\Big(\log v_t^2(\phi) + \frac{\varepsilon_t^2}{v_t^2(\phi)}\Big) \\
&= \frac{1}{T}\sum_{t=1}^{T}\big(\log v_t^2(\phi) - \log \tilde{v}_t^2(\phi)\big) + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\Big(\frac{\tilde{v}_t^2(\phi) - v_t^2(\phi)}{\tilde{v}_t^2(\phi) v_t^2(\phi)}\Big) + \frac{1}{T}\sum_{t=1}^{T}\frac{\varepsilon_t^2 - \tilde{\varepsilon}_t^2}{\tilde{v}_t^2(\phi)} \\
&=: (A) + (B) + (C).
\end{aligned}$$

In order to prove that (A), (B), and (C) are uniformly $o_p(1)$, it suffices to show that

$$\max_{1\leq t\leq T}\sup_{\phi\in\Phi}\big|\tilde{v}_t^2(\phi) - v_t^2(\phi)\big| = o_p(1) \tag{3.49}$$

$$\frac{1}{T}\sum_{t=1}^{T}\big|\tilde{\varepsilon}_t^2 - \varepsilon_t^2\big| = o_p(1) \tag{3.50}$$

$$v_t^2(\phi) \geq v_{\min} > 0 \quad \text{and} \quad \tilde{v}_t^2(\phi) \geq v_{\min} > 0 \quad \text{for some constant } v_{\min}. \tag{3.51}$$

(3.49) is implied by (G3). Moreover, (3.51) is automatically satisfied, as by (C13)

$$v_t^2(\phi) = w \sum_{k=1}^{t-1} b^{k-1} + a \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 + b^{t-1} \frac{w}{1-b} \geq w \geq \underline{\kappa} > 0$$

and the same holds true for $\tilde{v}_t^2(\phi)$. For the proof of (3.50), we use (G1) together with Corollary 3.1 to obtain

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} \left| \tilde{\varepsilon}_t^2 - \varepsilon_t^2 \right| &\leq \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \left| \frac{\tau^2(\frac{t}{T}, X_t) - \tilde{\tau}^2(\frac{t}{T}, X_t)}{\tau^2(\frac{t}{T}, X_t)} + R_\varepsilon\Big(\frac{t}{T}, X_t\Big) \right| \\ &= O_p(h) \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 = O_p(h). \end{aligned}$$

□

Proof of Theorem 3.3

By the usual Taylor expansion argument, we obtain

$$0 = \frac{1}{T} \frac{\partial \tilde{l}_T(\tilde{\phi})}{\partial \phi} = \frac{1}{T} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} + \frac{1}{T} \frac{\partial^2 \tilde{l}_T(\bar{\phi})}{\partial \phi \partial \phi^T} (\tilde{\phi} - \phi_0)$$

with some intermediate point $\bar{\phi}$ between ϕ_0 and $\tilde{\phi}$. Rearranging and premultiplying by $\sqrt{T}$ yields

$$\sqrt{T}(\tilde{\phi} - \phi_0) = -\Big(\frac{1}{T} \frac{\partial^2 \tilde{l}_T(\bar{\phi})}{\partial \phi \partial \phi^T}\Big)^{-1} \frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi}.$$

In what follows, we show that

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} \xrightarrow{d} N(0, Q) \tag{3.52}$$

$$\frac{1}{T} \frac{\partial^2 \tilde{l}_T(\bar{\phi})}{\partial \phi \partial \phi^T} \xrightarrow{P} J, \tag{3.53}$$

where Q is some covariance matrix to be specified later on and J is an invertible deterministic matrix. This completes the proof.

Proof of (3.52). We write

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} = \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} + \Big(\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i}\Big)$$

with

$$\frac{1}{\sqrt{T}}\frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial \phi_i} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\Big)\frac{1}{v_t^2} \qquad \text{(A)}$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{\partial \tilde{v}_t^2}{\partial \phi_i}\Big(\frac{1}{v_t^2}-\frac{1}{\tilde{v}_t^2}\Big) \qquad \text{(B)}$$

$$-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\frac{\varepsilon_t^2}{v_t^2}-\frac{\tilde{\varepsilon}_t^2}{v_t^2}\Big)\frac{\partial \tilde{v}_t^2}{\partial \phi_i}\frac{1}{\tilde{v}_t^2} \qquad \text{(C)}$$

$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2\Big(\frac{v_t^2-\tilde{v}_t^2}{v_t^2\tilde{v}_t^2}\Big)\frac{\partial \tilde{v}_t^2}{\partial \phi_i}\frac{1}{\tilde{v}_t^2}, \qquad \text{(D)}$$

where we use the abbreviations $v_t^2 = v_t^2(\phi_0)$ and $\tilde{v}_t^2 = \tilde{v}_t^2(\phi_0)$. In what follows, we show that (A) and (B) are asymptotically negligible, whereas (C) and (D) contribute to the limiting distribution.

We start with (A) and (B):

$$\begin{aligned}(A) &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i}-\frac{\partial v_t^2}{\partial \phi_i}\Big)\frac{1}{v_t^2} \\ &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{\sigma_t^2}\Big)\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i}-\frac{\partial v_t^2}{\partial \phi_i}\Big)\frac{1}{\sigma_t^2} \\ &\quad -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i}-\frac{\partial v_t^2}{\partial \phi_i}\Big)\Big(\frac{1}{v_t^2}-\frac{1}{\sigma_t^2}\Big)-\varepsilon_t^2\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i}-\frac{\partial v_t^2}{\partial \phi_i}\Big)\Big(\frac{1}{(v_t^2)^2}-\frac{1}{(\sigma_t^2)^2}\Big)\Big].\end{aligned}$$

Using (G2), we can show that $|\sigma_t^2 - v_t^2| = b^{t-1}|\sigma_1^2 - \frac{w}{1-b}|$. With this, it is easy to see that

$$(A) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\underbrace{\Big(1-\frac{\varepsilon_t^2}{\sigma_t^2}\Big)}_{=(1-\eta_t^2)}\Big(\frac{\partial \tilde{v}_t^2}{\partial \phi_i}-\frac{\partial v_t^2}{\partial \phi_i}\Big)\frac{1}{\sigma_t^2} + o_p(1).$$

As $(1-\eta_t^2)$ is a martingale difference, we can use results from empirical process theory to show that $(A) = o_p(1)$. Analogously, we obtain that $(B) = o_p(1)$.

Next we consider the terms (C) and (D). We restrict attention to (D), as this is the more complicated term. (C) can be treated analogously. Successively replacing the approximate expressions $\tilde{\varepsilon}_t^2$ and $\tilde{v}_t^2$ in (D) by the exact terms and using (G1) and (G3) to eliminate the resulting errors yields

$$(D) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^2\Big(\frac{v_t^2-\tilde{v}_t^2}{v_t^2 v_t^2}\Big)\frac{\partial v_t^2}{\partial \phi_i}\frac{1}{v_t^2} + o_p(1).$$

By analogous arguments as for (A) and (B), we can further replace some of the occurrences of v_t^2 by σ_t^2 to get

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\varepsilon_t^2}{\sigma_t^2} \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1).$$

Exploiting again the martingale difference structure of $(\frac{\varepsilon_t^2}{\sigma_t^2} - 1) = (\eta_t^2 - 1)$ gives

$$\begin{aligned}
(D) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\eta_t^2 - 1) \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big(\frac{v_t^2 - \tilde{v}_t^2}{\sigma_t^2 \sigma_t^2} \Big) \frac{\partial v_t^2}{\partial \phi_i} + o_p(1).
\end{aligned}$$

Once more using (G1)–(G3) and writing $m(x) = m_c + m_0(x_0) + \ldots + m_d(x_d)$ for short, we can infer that

$$\begin{aligned}
(D) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \underbrace{\frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2}}_{=:G_t} \sum_{k=1}^{t-1} a b^{k-1} (\varepsilon_{t-k}^2 - \tilde{\varepsilon}_{t-k}^2) + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \Big[\frac{\tau^2(\frac{t-k}{T}, X_{t-k}) - \tilde{\tau}^2(\frac{t-k}{T}, X_{t-k})}{\tau^2(\frac{t-k}{T}, X_{t-k})} + O_p(h^2) \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \Big[\frac{\exp(\xi_{t-k}) [m(\frac{t-k}{T}, X_{t-k}) - \tilde{m}(\frac{t-k}{T}, X_{t-k})]}{\exp(m(\frac{t-k}{T}, X_{t-k}))} \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \Big[m\Big(\frac{t-k}{T}, X_{t-k}\Big) - \tilde{m}\Big(\frac{t-k}{T}, X_{t-k}\Big) \Big] + o_p(1),
\end{aligned}$$

where the third equality is by a first order Taylor expansion with an intermediate point ξ_{t-k} between $m(\frac{t-k}{T}, X_{t-k})$ and $\tilde{m}(\frac{t-k}{T}, X_{t-k})$. We finally split up the difference $m(\frac{t-k}{T}, X_{t-k}) - \tilde{m}(\frac{t-k}{T}, X_{t-k})$ into its additive components and decompose the various components into their bias and stochastic parts. This yields

$$(D) = (D_c) - \sum_{j=0}^{d} (D_{V,j}) + \sum_{j=0}^{d} (D_{B,j}) + o_p(1)$$

with

$$(D_c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \Big[(m_c - \tilde{m}_c) + \sum_{j=0}^{d} \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) \Big]$$

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \tilde{m}_j^A(X_{t-k}^j)$$

$$(D_{B,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[m_j(X_{t-k}^j) - \tilde{m}_j^B(X_{t-k}^j) - \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) \Big]$$

for $j = 0, \ldots, d$, where for ease of notation we have used the shorthand $X_{t-k}^0 = \frac{t-k}{T}$. As in Appendix A, $\tilde{m}_j^A$ denotes the stochastic part of the backfitting estimate $\tilde{m}_j$ and $\tilde{m}_j^B$ denotes the bias part.

In Lemmas B1–B3, we will show that

$$(D_c) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{c,D} u_t + o_p(1) \tag{3.54}$$

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_{j,D}\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \tag{3.55}$$

$$(D_{B,j}) = o_p(1) \tag{3.56}$$

for all $j = 0, \ldots, d$ with $u_t = \log(\varepsilon_t^2)$. Here, $g_{c,D}$ is a constant which is specified in Lemma B2 and $g_{j,D}$ for $j = 0, \ldots, d$ are functions whose exact forms are given in Lemma B1. Using (C15), these functions are easily seen to be absolutely bounded by a constant independent of T. To summarize, we obtain that

$$(D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Big[g_{c,D} + \sum_{j=0}^{d} g_{j,D}\Big(\frac{t}{T}, X_t\Big) \Big] u_t + o_p(1).$$

Repeating the arguments from above, we can derive an analogous expression for (C). We thus get that

$$(C) + (D) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1)$$

with a function $g(\frac{t}{T}, X_t) = g_c + \sum_{j=0}^{d} g_j(\frac{t}{T}, X_t)$ whose additive components are absolutely bounded. Recalling that $(A) = o_p(1)$ and $(B) = o_p(1)$, we finally obtain that

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} - \frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g\Big(\frac{t}{T}, X_t\Big) u_t + o_p(1) \tag{3.57}$$

with an absolutely bounded function g.

We next consider the term $\frac{1}{\sqrt{T}} \frac{\partial l_T(\phi_0)}{\partial \phi_i}$ more closely. W.l.o.g. we can take $\phi_i = a$. (The case $\phi_i = b$ runs analogously and the case $\phi_i = w$ is much easier to handle.)

By similar arguments to before,

$$\begin{aligned}\frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i} &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(1-\frac{\varepsilon_t^2}{v_t^2}\Big)\frac{\partial v_t^2}{\partial\phi_i}\frac{1}{v_t^2}\\ &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\sum_{k=1}^{t-1}b^{k-1}\varepsilon_{t-k}^2+o_p(1).\end{aligned}$$

Furthermore,

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\sum_{k=1}^{t-1}b^{k-1}\varepsilon_{t-k}^2 &= \sum_{k=1}^{T-1}b^{k-1}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\varepsilon_{t-k}^2\\ &= \sum_{k=1}^{C_2\log T}b^{k-1}\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\varepsilon_{t-k}^2+o_p(1)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\sum_{k=1}^{\min_{t,T}}b^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)+o_p(1),\end{aligned}$$

where $C_2 > 0$ is a sufficiently large constant and $\min_{t,T} := \min\{t-1, C_2\log T\}$. For the second equality, we have used the fact that the weights b^k and b^i converge exponentially fast to zero as $i, k \to \infty$. This implies that only the sums up to $C_2 \log T$ with some constant C_2 are asymptotically relevant. Summing up, we have that

$$\frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i} = -\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big(\sum_{k=1}^{\min_{t,T}}b^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)+o_p(1). \tag{3.58}$$

Combining (3.57) and (3.58) yields

$$\begin{aligned}\frac{1}{\sqrt{T}}\frac{\partial\tilde{l}_T(\phi_0)}{\partial\phi_i} &= \frac{1}{\sqrt{T}}\frac{\partial l_T(\phi_0)}{\partial\phi_i}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g\Big(\frac{t}{T},X_t\Big)u_t+o_p(1)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big\{g\Big(\frac{t}{T},X_t\Big)u_t-\Big(\sum_{k=1}^{\min_{t,T}}b^{k-1}\varepsilon_{t-k}^2\Big)\Big(\frac{1-\eta_t^2}{\sigma_t^2}\Big)\Big\}+o_p(1)\\ &=: \frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_{t,T}+o_p(1),\end{aligned}$$

i.e. the term of interest can be written as a normalized sum of random variables $Z_{t,T}$ plus a term which is asymptotically negligible.

We now apply a central limit theorem for mixing arrays to the term $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{t,T}$. In particular, we employ the theorem of Francq & Zakoïan (2005), which allows the mixing coefficients of the array $\{Z_{t,T}\}$ to depend on the sample size T. Verifying the conditions of this theorem, we can conclude that

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi_i} \to N(0, \sigma^2)$$

with

$$\begin{aligned}
\sigma^2 &= \mathbb{E}\Big[\lambda_2(X_0)u_0\Big] - 2\mathbb{E}\Big[\lambda_1(X_0)u_0\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)\Big] \\
&\quad + \mathbb{E}\Big[\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)^2\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)^2\Big] + 2\mathbb{E}\big[\lambda_{1,1}(X_0, X_l)u_0u_l\big] \\
&\quad - 2\mathbb{E}\Big[\lambda_1(X_0)u_0\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{l-k}^2\Big)\Big(\frac{1-\eta_l^2}{\sigma_l^2}\Big)\Big] \\
&\quad - 2\mathbb{E}\Big[\lambda_1(X_l)u_l\Big(\sum_{k=1}^{\infty} b^{k-1}\varepsilon_{-k}^2\Big)\Big(\frac{1-\eta_0^2}{\sigma_0^2}\Big)\Big],
\end{aligned}$$

where we use the shorthand $\lambda_1(x) = \int_0^1 g(w,x)dw$, $\lambda_2(x) = \int_0^1 g^2(w,x)dw$, and $\lambda_{1,1}(x,x') = \int_0^1 g(w,x)g(w,x')dw$. Using the Cramer-Wold device, it is now easy to show that

$$\frac{1}{\sqrt{T}} \frac{\partial \tilde{l}_T(\phi_0)}{\partial \phi} \to N(0, Q).$$

The entries of the matrix Q can be calculated similarly to the expression σ^2. We omit the details as the formulas are rather lengthy and complicated. □

Proof of (3.53). By straightforward but tedious calculations it can be seen that

$$\sup_{\phi\in\Phi}\Big|\frac{1}{T}\frac{\partial^2 \tilde{l}_T(\phi)}{\partial\phi\partial\phi^T} - \frac{1}{T}\frac{\partial^2 l_T(\phi)}{\partial\phi\partial\phi^T}\Big| = o_p(1).$$

From standard theory for GARCH models, we further know that

$$\frac{1}{T}\frac{\partial^2 l_T(\bar{\phi})}{\partial\phi\partial\phi^T} \xrightarrow{P} J$$

with some invertible deterministic matrix J. This yields (3.53). □

In order to complete the proof of asymptotic normality of the GARCH estimates we still need to show that equations (3.54)–(3.56) are fulfilled for the terms (D_c), $(D_{V,j})$, and $(D_{B,j})$. We begin with the expansion of the variance components $(D_{V,j})$, as this is the technically most interesting part.

Lemma B1. *It holds that*

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{j,D}\Big(\frac{s}{T}, X_s\Big) u_s + o_p(1)$$

with

$$g_{j,D}\Big(\frac{s}{T}, X_s\Big) = g_{j,D}^{NW}(X_s^j) + g_{j,D}^{SBF}\Big(\frac{s}{T}, X_s\Big)$$

for $j = 0, \dots, d$. *The functions* $g_{j,D}^{NW}$ *and* $g_{j,D}^{SBF}$ *are absolutely bounded. Their exact form is given in the proof (see (3.63) and (3.66)–(3.68)).*

Proof. We start by giving a detailed exposition of the proof for $j \neq 0$. By Theorem A1, the stochastic part $\tilde{m}_j^A$ of the smooth backfitting estimate $\tilde{m}_j$ has the expansion

$$\tilde{m}_j^A(x_j) = \hat{m}_j^A(x_j) + \frac{1}{T} \sum_{s=1}^{T} r_{j,s}(x_j) u_s + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly in x_j, where $\hat{m}_j^A$ is the stochastic part of the Nadaraya-Watson pilot estimate and $r_{j,s}(\cdot) = r_j(\frac{s}{T}, X_s, \cdot)$ is Lipschitz continuous and absolutely bounded. With this result, we can decompose $(D_{V,j})$ as follows:

$$\begin{aligned}
(D_{V,j}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \tilde{m}_j^A(X_{t-k}^j) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \hat{m}_j^A(X_{t-k}^j) \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \Big[\frac{1}{T} \sum_{s=1}^{T} r_{j,s}(X_{t-k}^j) u_s\Big] + o_p(1) \\
&=: (D_{V,j}^{NW}) + (D_{V,j}^{SBF}) + o_p(1).
\end{aligned}$$

In the following, we will give the exact arguments needed to treat $(D_{V,j}^{NW})$. The line of argument for $(D_{V,j}^{SBF})$ is essentially identical although some of the steps are easier due to the properties of the $r_{j,s}$ functions.

W.l.o.g set $\phi_i = a$ and let $m_{i,k} = \max\{k+1, i+1\}$. Using $\partial v_t^2 / \partial a = \sum_{i=1}^{t-1} b^{i-1} \varepsilon_{t-i}^2$ and $\hat{m}_j^A(x_j) = \frac{1}{T} \sum_{s=1}^{T} K_h(x_j, X_s^j) u_s / \frac{1}{T} \sum_{v=1}^{T} K_h(x_j, X_v^j)$, we get

$$\begin{aligned}
(D_{V,j}^{NW}) &= \sum_{k=1}^{T-1} a b^{k-1} \sum_{i=1}^{T-1} b^{i-1} \\
&\quad \times \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{\frac{1}{T} \sum_{v=1}^{T} K_h(X_{t-k}^j, X_v^j)} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2 u_s\Big]. \qquad (3.59)
\end{aligned}$$

In a first step, we replace the sum $\frac{1}{T}\sum_{v=1}^{T} K_h(X_{t-k}^j, X_v^j)$ in (3.59) by a term which only depends on X_{t-k}^j and show that the resulting error is asymptotically negligible. Let $q_j(x_j) = \int_0^1 K_h(x_j, w)dw\ p_j(x_j)$. Furthermore define

$$
\begin{aligned}
B_j(x_j) &= \frac{1}{T}\sum_{v=1}^{T} \mathbb{E}[K_h(x_j, X_v^j)] - q_j(x_j) \\
V_j(x_j) &= \frac{1}{T}\sum_{v=1}^{T} \big(K_h(x_j, X_v^j) - \mathbb{E}[K_h(x_j, X_v^j)]\big).
\end{aligned}
$$

Notice that $\sup_{x_j\in[0,1]} |B_j(x_j)| = O_p(h)$ and $\sup_{x_j\in[0,1]} |V_j(x_j)| = O_p(\sqrt{\log T/Th})$. From the identity $\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j) = q_j(x_j) + B_j(x_j) + V_j(x_j)$ and a second order Taylor expansion of $(1+x)^{-1}$ we arrive at

$$
\begin{aligned}
\frac{1}{\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j)} &= \frac{1}{q_j(x_j)}\Big(1 + \frac{B_j(x_j)+V_j(x_j)}{q_j(x_j)}\Big)^{-1} \\
&= \frac{1}{q_j(x_j)}\Big(1 - \frac{B_j(x_j)+V_j(x_j)}{q_j(x_j)} + O_p(h^2)\Big) \qquad (3.60)
\end{aligned}
$$

uniformly in x_j. Plugging this decomposition into (3.59), we obtain

$$
\begin{aligned}
(D_{V,j}^{NW}) = &\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{K_h(X_{t-k}^j, X_s^j)}{q_j(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big] \\
&- (D_{V,j}^{NW,B}) - (D_{V,j}^{NW,V}) + o_p(1)
\end{aligned}
$$

with

$$
\begin{aligned}
(D_{V,j}^{NW,B}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j)\frac{B_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big] \\
(D_{V,j}^{NW,V}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j)\frac{V_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big].
\end{aligned}
$$

As $\sup_{x_j\in I_h} |B_j(x_j)| = O_p(h^2)$ and $\sup_{x_j\in I_h^c} |B_j(x_j)| = O_p(h)$, we can proceed similarly to the proof of Lemma $B3$ later on to show that $(D_{V,j}^{NW,B}) = o_p(1)$. Next we will show that $(D_{V,j}^{NW,V}) = o_p(1)$. Let $\mathbb{E}_v[\cdot]$ denote the expectation with respect to the variables indexed by v, then

$$\begin{aligned}
\big|(D_{V,j}^{NW,V})\big| &= \Big|\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{K_h(X_{t-k}^j, X_s^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 \\
&\qquad\qquad \times\Big(\frac{1}{T}\sum_{v=1}^{T}(K_h(X_{t-k}^j, X_v^j) - \mathbb{E}_v[K_h(X_{t-k}^j, X_v^j)])\Big)u_s\Big]\Big| \\
&\leq \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=m_{i,k}}^{T}\Big|\frac{1}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2\Big| \\
&\qquad\qquad \times \sup_{x_j\in[0,1]}\Big|\frac{1}{T}\sum_{v=1}^{T}(K_h(x_j, X_v^j) - \mathbb{E}_v[K_h(x_j, X_v^j)])\Big| \\
&\qquad\qquad \times \sup_{x_j\in[0,1]}\Big|\frac{1}{T}\sum_{s=1}^{T}K_h(x_j, X_s^j)u_s\Big|\Big) \\
&= O_p\Big(\frac{\log T}{Th}\Big)\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\underbrace{\Big(\frac{1}{\sqrt{T}}\sum_{t=m_{i,k}}^{T}\Big|\frac{1}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2\Big|\Big)}_{=O_p(\sqrt{T}) \text{ by Markov's inequality}} \\
&= O_p\Big(\frac{\log T}{Th}\sqrt{T}\Big) = o_p(1).
\end{aligned}$$

Together with the fact that $(D_{V,j}^{NW,B}) = o_p(1)$, this yields

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}u_s\Big] + o_p(1), \quad (3.61)$$

where we use the shorthand $\mu_t^{i,k} = (q_j(X_{t-k}^j)\sigma_t^2\sigma_t^2)^{-1}\varepsilon_{t-k}^2\varepsilon_{t-i}^2$.
In the next step, we replace the inner sum over t in (3.61) by a term that only depends on X_s^j and show that the resulting error can be asymptotically neglected. Define

$$\xi(X_{t-k}^j, X_s^j) := \xi_t^{i,k}(X_{t-k}^j, X_s^j) := K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k} - \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}],$$

where $\mathbb{E}_{-s}[\cdot]$ is the expectation with respect to all variables except for those depending on the index s. With the above notation at hand, we can write

$$\begin{aligned}
(D_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j)\mu_t^{i,k}]u_s\Big] \\
&\quad + (R_{V,j}^{NW}) + o_p(1),
\end{aligned}$$

where

$$
\begin{aligned}
(R_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big] \\
&= \sum_{k=1}^{C_2 \log T} ab^{k-1} \sum_{i=1}^{C_2 \log T} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big] + o_p(1) \quad (3.62)
\end{aligned}
$$

for some sufficiently large constant $C_2 > 0$. Once we show that $(R_{V,j}^{NW}) = o_p(1)$, we are left with

$$
\begin{aligned}
(D_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[K_h(X_{t-k}^j, X_s^j) \mu_t^{i,k}] u_s \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \frac{T - m_{i,k}}{T} \mathbb{E}_{-s}[K_h(X_{-k}^j, X_s^j) \mu_0^{i,k}] \Big) u_s + o_p(1).
\end{aligned}
$$

As the terms with $i, k \geq C_2 \log T$ are asymptotically negligible, we can expand the i and k sums to infinity, which yields

$$
\begin{aligned}
(D_{V,j}^{NW}) &= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}_{-s}[K_h(X_{-k}^j, X_s^j) \mu_0^{i,k}] \Big) u_s + o_p(1) \\
&=: \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{j,D}^{NW}(X_s^j) u_s + o_p(1) \quad (3.63)
\end{aligned}
$$

with

$$
\begin{aligned}
\mu_0^{i,k} &= \frac{1}{q_j(X_{-k}^j)} \frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-k}^2 \varepsilon_{-i}^2 \\
q_j(X_{-k}^j) &= \int_0^1 K_h(X_{-k}^j, w) dw \; p_j(X_{-k}^j).
\end{aligned}
$$

Thus it remains to show that $(R_{V,j}^{NW}) = o_p(1)$, which requires a lot of care. We will prove that the term in square brackets in (3.62) is $o_p(1)$ uniformly over $i, k \leq C_2 \log T$, which yields the desired result. It is easily seen that

$$
\begin{aligned}
P &:= \mathbb{P}\Big(\max_{i,k \leq C_2 \log T} \Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big| > \delta \Big) \\
&\leq \sum_{k=1}^{C_2 \log T} \sum_{i=1}^{C_2 \log T} \underbrace{\mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X_{t-k}^j, X_s^j) u_s \Big| > \delta \Big)}_{=: P_{i,k}}
\end{aligned}
$$

for a fixed $\delta > 0$. Then by Chebychev's inequality

$$
\begin{aligned}
P_{i,k} &\le \frac{1}{T^3\delta^2} \sum_{s,s'=1}^{T} \sum_{t,t'=m_{i,k}}^{T} \mathbb{E}\Big[\xi(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\Big] \\
&= \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\notin\Gamma_{i,k}} \mathbb{E}\Big[\xi(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\Big] \\
&\quad + \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \mathbb{E}\Big[\xi(X_{t-k}^j, X_s^j) u_s \xi(X_{t'-k}^j, X_{s'}^j) u_{s'}\Big] \\
&=: P_{i,k}^1 + P_{i,k}^2,
\end{aligned}
$$

where $\Gamma_{i,k}$ is the set of tuples (s, s', t, t') with $1 \le s, s' \le T$ and $m_{i,k} \le t, t' \le T$ such that one index is separated from the others. We say that an index, for instance t, is separated from the others if $\min\{|t-t'|, |t-s|, |t-s'|\} > C_3 \log T$, i.e. if it is further away from the other indices than $C_3 \log T$ for a constant C_3 to be chosen later on. We now analyse $P_{i,k}^1$ and $P_{i,k}^2$ separately.

(a) First consider $P_{i,k}^1$. If a tuple (s, s', t, t') is not an element of $\Gamma_{i,k}$, then no index can be separated from the others. Since the index t cannot be separated, there exists an index, say t', such that $|t-t'| \le C_3 \log T$. Now take an index different from t and t', for instance s. Then by the same argument, there exists an index, say s', such that $|s-s'| \le C_3 \log T$. As a consequence, the number of tuples $(s, s', t, t') \notin \Gamma_{i,k}$ is smaller than $CT^2(\log T)^2$ for some constant C. Using (C14), this suffices to infer that

$$
\big|P_{i,k}^1\big| \le \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\notin\Gamma_{i,k}} \frac{C}{h^2} \le \frac{C}{\delta^2} \frac{(\log T)^2}{Th^2}.
$$

Hence, $|P_{i,k}^1| \le C\delta^{-2}(\log T)^{-3}$ uniformly in i and k.

(b) The term $P_{i,k}^2$ is more difficult to handle. We start by taking a cover $\{I_m\}_{m=1}^{M_T}$ of the compact support $[0,1]$ of X_{t-k}^j. The elements I_m are intervals of length $1/M_T$ given by $I_m = [\frac{m-1}{M_T}, \frac{m}{M_T})$ for $m = 1, \ldots, M_T - 1$ and $I_{M_T} = [1 - \frac{1}{M_T}, 1]$. The midpoint of the interval I_m is denoted by x_m. With this, we can write

$$
\begin{aligned}
K_h(X_{t-k}^j, X_s^j) = \sum_{m=1}^{M_T} & I(X_{t-k}^j \in I_m) \\
& \times \big[K_h(x_m, X_s^j) + (K_h(X_{t-k}^j, X_s^j) - K_h(x_m, X_s^j))\big]. \quad (3.64)
\end{aligned}
$$

Using (3.64), we can further write

$$
\begin{aligned}
\xi(X^j_{t-k}, X^j_s) &= \sum_{m=1}^{M_T} \Big\{ I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s) \mu_t^{i,k} \\
&\qquad - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s) \mu_t^{i,k}] \Big\} \\
&\quad + \sum_{m=1}^{M_T} \Big\{ I(X^j_{t-k} \in I_m)(K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s)) \mu_t^{i,k} \\
&\qquad - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m)(K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s)) \mu_t^{i,k}] \Big\} \\
&=: \xi_1(X^j_{t-k}, X^j_s) + \xi_2(X^j_{t-k}, X^j_s)
\end{aligned}
$$

and

$$
\begin{aligned}
P^2_{i,k} &= \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\big[\xi_1(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\big] \\
&\quad + \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\big[\xi_2(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\big] =: P^{2,1}_{i,k} + P^{2,2}_{i,k}.
\end{aligned}
$$

We first consider $P^{2,2}_{i,k}$. Set $M_T = CT(\log T)^3 h^{-3}$ and exploit the Lipschitz continuity of the kernel K to get that $|K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s)| \leq \frac{C}{h^2} |X^j_{t-k} - x_m|$. This gives us

$$
\begin{aligned}
\big|\xi_2(X^j_{t-k}, X^j_s)\big| &\leq \frac{C}{h^2} \sum_{m=1}^{M_T} \Big(\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\leq I(X^j_{t-k} \in I_m) M_T^{-1}} \mu_t^{i,k} \\
&\qquad + \mathbb{E}\big[\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\leq I(X^j_{t-k} \in I_m) M_T^{-1}} \mu_t^{i,k} \big] \Big) \\
&\leq \frac{C}{M_T h^2} \big(\mu_t^{i,k} + \mathbb{E}[\mu_t^{i,k}] \big). \qquad (3.65)
\end{aligned}
$$

Plugging (3.65) into the expression for $P^{2,2}_{i,k}$, we arrive at

$$
\begin{aligned}
\big|P^{2,2}_{i,k}\big| &\leq \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\Big[\big|\xi_2(X^j_{t-k}, X^j_s)\big| \big|u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\big| \Big] \\
&\leq \frac{1}{T^3 \delta^2} \frac{C}{M_T h^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \underbrace{\mathbb{E}\big[(\mu_t^{i,k} + \mathbb{E}[\mu_t^{i,k}]) |u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}| \big]}_{\leq C h^{-1}} \\
&\leq \frac{C}{\delta^2} \frac{1}{(\log T)^3}.
\end{aligned}
$$

We next turn to $P_{i,k}^{2,1}$. Write

$$P_{i,k}^{2,1} = \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \Big(\sum_{m=1}^{M_T} S_m\Big)$$

with

$$S_m = \mathbb{E}\Big[\big\{I(X_{t-k}^j \in I_m)K_h(x_m, X_s^j)\mu_t^{i,k} - \mathbb{E}_{-s}[I(X_{t-k}^j \in I_m)K_h(x_m, X_s^j)\mu_t^{i,k}]\big\} \\ \times u_s \xi(X_{t'-k}^j, X_{s'}^j)u_{s'}\Big]$$

and assume that an index, w.l.o.g. t, can be separated from the others. Choosing $C_3 \gg C_2$, we get

$$\begin{aligned} S_m &= \mathrm{Cov}\Big(I(X_{t-k}^j \in I_m)\mu_t^{i,k} - \mathbb{E}[I(X_{t-k}^j \in I_m)\mu_t^{i,k}], \\ &\qquad K_h(x_m, X_s^j)u_s\xi(X_{t'-k}^j, X_{s'}^j)u_{s'}\Big) \\ &\leq \frac{C}{h^2}(\alpha([C_3 - C_2]\log T))^{1-\frac{2}{p}} \leq \frac{C}{h^2}(a^{(C_3-C_2)\log T})^{1-\frac{2}{p}} \leq \frac{C}{h^2}T^{-C_4} \end{aligned}$$

with some $C_4 > 0$ by Davydov's inequality, where p is chosen slightly larger than 2. Note that the above bound is independent of i and k and that we can make C_4 arbitrarily large by choosing C_3 large enough. This shows that $|P_{i,k}^{2,1}| \leq C\delta^{-2}(\log T)^{-3}$ uniformly in i and k with some constant C.

Combining (a) and (b) yields that $P \to 0$ for each fixed $\delta > 0$. This implies that

$$(R_{V,j}^{NW,V}) = o_p(1),$$

which completes the proof for the term $(D_{V,j}^{NW})$.

As stated at the beginning of the proof, the term $(D_{V,j}^{SBF})$ can be treated in exactly the same way. Following analogous arguments as above, one obtains

$$\begin{aligned} (D_{V,j}^{SBF}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[r_{j,s}(X_{t-k}^j)\zeta_t^{i,k}]\ u_s\Big] + o_p(1) \\ &= \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}_{-s}[r_{j,s}(X_{-k}^j)\zeta_0^{i,k}]\Big)u_s + o_p(1) \\ &=: \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g_{j,D}^{SBF}\Big(\frac{s}{T}, X_s\Big)u_s + o_p(1) \end{aligned} \tag{3.66}$$

with $\zeta_t^{i,k} = (\sigma_t^2 \sigma_t^2)^{-1} \varepsilon_{t-k}^2 \varepsilon_{t-i}^2$.

Finally, the proofs for $j = 0$ are very similar but somewhat simpler and are thus omitted here. For completeness we provide the functions $g_{0,D}^{NW}$ and $g_{0,D}^{SBF}$:

$$g_{0,D}^{NW}\Big(\frac{s}{T}\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}\big[\frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-k}^2 \varepsilon_{-i}^2\big]\Big) \int_0^1 \frac{K_h(\frac{s}{T}, v)}{\int_0^1 K_h(v,w)dw} dv \quad (3.67)$$

$$g_{0,D}^{SBF}\Big(\frac{s}{T}, X_s\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}\big[\frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-k}^2 \varepsilon_{-i}^2\big]\Big) \int_0^1 r_{0,s}(w) dw. \quad (3.68)$$

□

Lemma B2. *It holds that*

$$(D_c) = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{c,D} u_s$$

with

$$g_{c,D} = \sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}\Big[\frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-i}^2 \varepsilon_{-k}^2\Big].$$

Proof. Using the fact that

$$\tilde{m}_c = \frac{1}{T} \sum_{s=1}^{T} Z_{s,T} = m_c + \frac{1}{T} \sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big) + \sum_{j=1}^{d} \frac{1}{T} \sum_{s=1}^{T} m_j(X_s^j) + \frac{1}{T} \sum_{s=1}^{T} u_s,$$

we arrive at

$$(D_c) = -\Big(\frac{1}{T} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2\Big)\Big(\frac{1}{\sqrt{T}} \sum_{s=1}^{T} u_s\Big)$$

with $G_t = \frac{\partial v_t^2}{\partial \phi_i} (\sigma_t^2 \sigma_t^2)^{-1}$. Now let $m_{i,k} = \max\{k+1, i+1\}$ and assume w.l.o.g. that $\phi_i = a$. Then

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 &= \frac{1}{T} \sum_{t=1}^{T} \Big(\sum_{i=1}^{t-1} b^{i-1} \varepsilon_{t-i}^2\Big) \frac{1}{\sigma_t^2 \sigma_t^2} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \\ &= \sum_{k=1}^{C_2 \log T} ab^{k-1} \sum_{i=1}^{C_2 \log T} b^{i-1} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 + o_p(1) \end{aligned}$$

with some sufficiently large constant C_2. Using Chebychev's inequality and exploiting the mixing properties of the variables involved, one can show that

$$\max_{i,k \le C_2 \log T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \Big(\frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 - \mathbb{E}\Big[\frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2\Big]\Big) = o_p(1).$$

This allows us to infer that

$$\frac{1}{T}\sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2 = \sum_{k=1}^{C_2\log T} ab^{k-1} \sum_{i=1}^{C_2\log T} b^{i-1}\frac{1}{T}\sum_{t=m_{i,k}}^{T} \mathbb{E}\Big[\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-i}^2\varepsilon_{t-k}^2\Big] + o_p(1)$$
$$= \sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\Big[\frac{1}{\sigma_0^2\sigma_0^2}\varepsilon_{-i}^2\varepsilon_{-k}^2\Big] + o_p(1),$$

which completes the proof. □

Lemma B3. *It holds that*

$$(D_{B,j}) = o_p(1)$$

for $j = 0,\dots,d$.

Proof. We start by considering the case $j = 0$: Define

$$J_h = \{t \in \{1,\dots,T\} : C_1 h \le \frac{t}{T} \le 1 - C_1 h\}$$
$$J_{h,c}^u = \{t \in \{1,\dots,T\} : 1 - C_1 h < \frac{t}{T}\}$$
$$J_{h,c}^l = \{t \in \{1,\dots,T\} : \frac{t}{T} < C_1 h\},$$

where $[-C_1, C_1]$ is the support of K. Using the uniform convergence rates from Theorem A2 and assuming w.l.o.g. that $\phi_i = a$, we get

$$|(D_{B,0})| = \Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial v_t^2}{\partial a}\frac{1}{\sigma_t^2\sigma_t^2}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-k}^2$$
$$\times\Big[m_0\Big(\frac{t-k}{T}\Big) - \tilde{m}_0^B\Big(\frac{t-k}{T}\Big) - \frac{1}{T}\sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big)\Big]\Big|$$
$$\le O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(t-k \in J_{h,c}^l)$$
$$+ O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(t-k \in J_{h,c}^u)$$
$$+ O_p(h^2)\frac{C}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(t-k \in J_h)$$
$$=: (D_{B,0}^{J_{h,c}^l}) + (D_{B,0}^{J_{h,c}^u}) + (D_{B,0}^{J_h}).$$

By Markov's inequality, $(D_{B,0}^{J_h}) = O_p(h^2\sqrt{T}) = o_p(1)$. Recognizing that

(i) $I(t-k \in J^u_{h,c}) \leq I(t \in J^u_{h,c})$ for all $k \in \{0,\dots,t-1\}$

(ii) $\sum_{t=1}^{T} I(t \in J^u_{h,c}) \leq C_1 T h$,

we get $(D_{B,0}^{J^u_{h,c}}) = O_p(h^2\sqrt{T}) = o_p(1)$ by another appeal to Markov's inequality. This just leaves $(D_{B,0}^{J^l_{h,c}})$, which is a bit more tedious. By a change of variable $j = t-k$,

$$\begin{aligned}
(D_{B,0}^{J^l_{h,c}}) &\leq O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon^2_{t-i}\sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon^2_j I(j \in J^l_{h,c}) \\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon^2_{t-i} I\Big(\Big[\frac{t}{2}\Big] \in J^l_{h,c}\Big)\sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon^2_j I(j \in J^l_{h,c}) \\
&\quad + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon^2_{t-i} I\Big(\Big[\frac{t}{2}\Big] \notin J^l_{h,c}\Big)\sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon^2_j I(j \in J^l_{h,c}) \\
&=: (A) + (B),
\end{aligned}$$

where $[x]$ denotes the smallest integer larger than x. Realizing that $[t/2] \in J^l_{h,c}$ only if $t < 2C_1 hT$, we get $(A) = O_p(h^2\sqrt{T}) = o_p(1)$ once again by Markov's inequality. In (B) we can truncate the summation over j at $[t/2]-1$, as $I(j \in J^l_{h,c}) = 0$ for $j \geq [t/2]$ if $[t/2] \notin J^l_{h,c}$. We thus obtain

$$\begin{aligned}
(B) &\leq O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon^2_{t-i}\sum_{j=1}^{[t/2]-1} ab^{t-j-1}\varepsilon^2_j \\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{[t/2]}\sum_{i=1}^{t-1} b^{i-1}\sum_{j=1}^{[t/2]-1} ab^{t-j-1-[t/2]}\varepsilon^2_{t-i}\varepsilon^2_j.
\end{aligned}$$

By a final appeal to Markov's inequality we arrive at

$$(B) = O_p(h)O_p\Big(\frac{1}{\sqrt{T}}\Big) = o_p(1),$$

thus completing the proof for $j = 0$.

Next consider the case $j \neq 0$. Similarly to before, we have

$$\begin{aligned}
|(D_{B,j})| &\leq O_p(h^2)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon^2_{t-i}\varepsilon^2_{t-k} I(X^j_{t-k} \in I_h) \\
&\quad + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon^2_{t-i}\varepsilon^2_{t-k} I(X^j_{t-k} \notin I_h)
\end{aligned}$$

$$= O_p(h^2\sqrt{T}) + O_p\Big(\frac{h}{\sqrt{T}}\Big) \underbrace{\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h)}_{=:R_T}$$

with $I_h = [2C_1h, 1-2C_1h]$ as defined in Theorem 3.1. Using (C15), it is easy to see that $R_T = O_p(h)$, which yields the result for $j \neq 0$. □

Appendix C

For completeness, we collect some standard type uniform convergence results in this appendix which are needed to prove Theorem 3.1. These can be shown by small modifications of standard arguments as given for example in Masry [32] or Hansen [17]. We start with the kernel density estimates $\hat{p}_j$ and $\hat{p}_{j,k}$. Using the notation $p_0(x_0) = I(x_0 \in (0,1])$, we have the following result.

Lemma C1. *Suppose that (C1)–(C11) hold and that the bandwidth h satisfies (C12a) or (C12b). Then*

$$\sup_{x_j \in I_h} \big|\hat{p}_j(x_j) - p_j(x_j)\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) + o(h) \quad (3.69)$$

$$\sup_{0 \le x_j \le 1} \big|\hat{p}_j(x_j) - \kappa_0(x_j)p_j(x_j)\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) + O(h) \quad (3.70)$$

$$\sup_{x_j, x_k \in I_h} \big|\hat{p}_{j,k}(x_j,x_k) - p_{j,k}(x_j,x_k)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^2}}\Big) + o(h) \quad (3.71)$$

$$\sup_{0 \le x_j, x_k \le 1} \big|\hat{p}_{j,k}(x_j,x_k) - \kappa_0(x_j)\kappa_0(x_k)p_{j,k}(x_j,x_k)\big| = O_p\Big(\sqrt{\frac{\log T}{Th^2}}\Big) + O(h) \quad (3.72)$$

for $j,k = 0,\dots,d$ with $j \neq k$, where $\kappa_0(v) = \int K_h(v,w)dw$ and $I_h = [2C_1h, 1-2C_1h]$.

Proof. We restrict attention to (3.72), the other results following by analogous arguments. For $j,k \neq 0$, we are in the standard strictly stationary setting and can immediately apply results from Masry [32] or Hansen [17]. The case $j=0$ and $k \neq 0$, where we simultaneously smooth in the direction of time and the regressor X_t^k, can be handled by using similar arguments. In particular, we apply a covering argument together with an exponential inequality for mixing variables. To do so, we have to show that $Th^2\mathrm{Var}(\hat{p}_{0,k}(x_0,x_k)) \le C$ uniformly in x_0 and x_k, which is achieved by exploiting that

$$\sup_{x_0 \in (0,1]} \frac{1}{T} \sum_{t=1}^{T} K_h^r\Big(x_0, \frac{t}{T}\Big) \leq \frac{C}{h^{r-1}}$$

$$\max_{l=1,\dots,T-1} \sup_{x_0 \in (0,1]} \frac{1}{T} \sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big) K_h\Big(x_0, \frac{t+l}{T}\Big) \leq \frac{C}{h}.$$

□

We now examine the convergence behaviour of the one-dimensional Nadaraya-Watson smoothers $\hat{m}_j$ defined in (3.17) and (3.20). For the stochastic part $\hat{m}_j^A$, we have

Lemma C2. *Under (C1)–(C11) together with (C12a) or (C12b),*

$$\sup_{x_j \in [0,1]} \big|\hat{m}_j^A(x_j)\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big) \tag{3.73}$$

for all $j = 0, \dots, d$.

Proof. The case $j \neq 0$ is standard. For the case $j = 0$, we have to modify the arguments in a similar vein to Lemma C1. □

For the bias part $\hat{m}_j^B$, we have the following expansion:

Lemma C3. *Under (C1)–(C11) together with (C12a) or (C12b),*

$$\sup_{x_j \in I_h} \big|\hat{m}_j^B(x_j) - \hat{\mu}_{T,j}(x_j)\big| = o_p(h^2) \tag{3.74}$$

$$\sup_{x_j \in I_h^c} \big|\hat{m}_j^B(x_j) - \hat{\mu}_{T,j}(x_j)\big| = O_p(h^2) \tag{3.75}$$

for all $j = 0, \dots, d$, *where*

$$\hat{\mu}_{T,j}(x_j) = \alpha_{T,0} + \alpha_{T,j}(x_j) + \sum_{k \neq j} \int \alpha_{T,k}(x_k) \frac{\hat{p}_{j,k}(x_j, x_k)}{\hat{p}_j(x_j)} dx_k + h^2 \int \beta(x) \frac{p(x)}{p_j(x_j)} dx_{-j}.$$

Here, $\alpha_{T,0} = m_c$ *and*

$$\alpha_{T,k}(x_k) = m_k(x_k) + m_k'(x_k) \frac{h\kappa_1(x_k)}{\kappa_0(x_k)}$$

$$\beta(x) = \sum_{k=0}^{d} \int u^2 K(u) du \Big(\frac{\partial \log p(x)}{\partial x_k} m_k'(x_k) + \frac{1}{2} m_k''(x_k) \Big)$$

with $\kappa_0(x_k) = \int K_h(x_k, w) dw$ *and* $\kappa_1(x_k) = \int K_h(x_k, w)(\frac{w - x_k}{h}) dw$.

Proof. The result can be proven by exploiting the smoothness conditions imposed on the densities p_j and $p_{j,k}$ as well as the functions m_j and by using the fact that for $l = 0, 1, 2$,

$$\frac{1}{T}\sum_{t=1}^{T} K_h\Big(x_0, \frac{t}{T}\Big)\Big(\frac{\frac{t}{T} - x_0}{h}\Big)^l = \int K_h(x_0, w)\Big(\frac{w - x_0}{h}\Big)^l dw + O\Big(\frac{1}{Th^r}\Big)$$

uniformly for $x_0 \in I_h$ with $r = 1$ and uniformly for $x_0 \in I_h^c$ with $r = 2$. We omit the details. Compare also the relevant parts in the proof of Theorem 4 in Mammen et al. [29]. □

Bibliography

[1] Alizadeh, S., Brandt, M.W. & Diebold, F.X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance* **57** 1047-1091.

[2] Chandler, G. & Polonik, W. (2006). Discrimination of locally stationary time series based on the excess mass functional. *Journal of the American Statistical Association* **101** 240-253.

[3] Chen, M. & Chen, G. (2000). Geometric ergodicity of nonlinear autoregressive models with changing conditional variances. *The Canadian Journal of Statistics* **28** 605-613.

[4] Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications* **62** 139-168.

[5] Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series Analysis*, (Eds. P. M. Robinson and M. Rosenblatt) **2**. Springer, New York.

[6] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25** 1-37.

[7] Dahlhaus, R., Neumann, M. H. & von Sachs, R. (1999). Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli* **5** 873-906.

[8] Dahlhaus, R. & Subba-Rao, S. (2006). Statistical inference for time-varying ARCH processes. *The Annals of Statistics* **34** 1075-1114.

[9] Dahlhaus, R. & Polonik, W. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1-39.

[10] Delgado, M. A. & Hidalgo, J. (2000). Nonparametric inference on structural breaks. *Journal of Econometrics* **96** 113-144.

[11] Engle, R. F., Ghysels, E. & Sohn, B. (2008). On the economic sources of stock market volatility. Working Paper.

[12] Engle, R. F. & Rangel, J. G. (2008). The Spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies* **21** 1187-1222.

[13] Feng, Y. (2004). Simultaneously modelling conditional heteroskedasticity and scale change. *Econometric Theory* **20** 563-596.

[14] Francq, C. & Zakoian, J.-M. (2005). A central limit theorem for mixing triangular arrays of variables whose dependence is allowed to grow with the sample size. *Econometric Theory* **21** 1165-1171.

[15] Fryzlewicz, P., Sapatinas, T. & Subba Rao, S. (2008). Normalised least squares estimation in time-varying ARCH models. *Annals of Statistics* **36** 742-786.

[16] Fryzlewicz, P. & Subba Rao, S. (2011). Mixing properties of ARCH and time-varying ARCH processes. *Bernoulli* **17** 320-346.

[17] Hansen, B.E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726-748.

[18] Härdle, W. & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* **21** 1926-1947.

[19] Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability. *Econometric Theory* **11** 671-698.

[20] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* **75** 261-277.

[21] Koo, B. & Linton, O. (2010). Semiparametric estimation of locally stationary diffusion models. Working Paper.

[22] Kreiss, J.-P., Neumann, M.H. & Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and its Interface* **1** 367-380.

[23] Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogeneous, dependent data. *Econometric Theory* **25** 1433-1445.

[24] Kristensen, D. (2011). Stationary approximations of time-inhomogenous Markov chains with applications. Mimeo.

[25] Lee, J. & Subba Rao, S. (2010). A note on quadratic forms of nonstationary stochastic processes. Preprint.

[26] Li, Q. & Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* **87** 145-165.

[27] Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and their Applications* **65** 69-80.

[28] Linton, O. & Hafner, C. M. (2010). Efficient estimation of a multivariate multiplicative volatility model. *The Journal of Econometrics* **159** 55-73.

[29] Mammen, E., Linton, O. & Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics* **27** 1443-1490.

[30] Mammen, E. & Park, B. U. (2005). Bandwidth selection for smooth backfitting in additive models. *The Annals of Statistics* **33** 1260-1294.

[31] Martens, M. & van Dijk, D. (2007). Measuring volatility with the realized range. *Journal of Econometrics* **138** 181-207.

[32] Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* **17** 571-599.

[33] Mikosch, T. & Stărică, C. (2000). Is it really long memory we see in financial returns? In *Extremes and Integrated Risk Management*, (Ed. P. Embrechts), pp. 149-168, London: Risk Books.

[34] Mikosch, T. & Stărică, C. (2003). Long-range dependence effects and ARCH modelling. In *Theory and Applications of Long Range Dependence*, (Eds. P. Doukhan, G. Oppenheim and M.S. Taqqu), pp. 439-459, Boston: Birkhäuser.

[35] Mikosch, T. & Stărică, C. (2004). Non-stationarities in financial time series, the long-range dependence, and IGARCH effects. *The Review of Economics and Statistics* **86** 378-390.

[36] Nielsen, J. P. & Sperlich, S. (2003). Prediction of stock returns: A new way to look at it. *Astin Bulletin* **33** 399-417.

[37] Pollard, D. (1984). *Convergence of stochastic processes*. Springer, New York.

[38] Rogers, L.C.G. & Satchell, S.E. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability* **1** 504-512.

[39] Su, L. & Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. *Statistics and its Interface* **1** 347-366.

[40] Wu, G. & Xiao, Z. (2002). A generalized partially linear model of asymmetric volatility. *Journal of Empirical Finance* **9** 287-319.

[41] Yang, D. & Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *Journal of Business* **73** 477-491.

[42] Yu, K., Mammen, E. & Park, B. U. (2009). Semiparametric additive regression: gains from the additive structure of the infinite-dimensional parameter. Preprint.

Lebenslauf

09/2007 – heute	Promotion in Volkswirtschaftslehre an der Universität Mannheim, Mitglied des Graduiertenkollegs CDSE
07/2006	Magisterabschluß an der Universität Heidelberg
10/2003 – 08/2004	Visiting Student in Philosophie und Mathematik an der Universität Oxford
10/2000 – 07/2006	Magisterstudium der Philosophie (Hauptfach), Mathematik (Nebenfach) und VWL (Nebenfach) an der Universität Heidelberg
07/1999	Abitur
09/1990 – 07/1999	Kopernikus Gymnasium Aalen-Wasseralfingen