Econometric Theory, **31**, 2015, 811–859.
doi:10.1017/S0266466614000565

TESTING FOR STRUCTURAL CHANGE IN TIME-VARYING NONPARAMETRIC REGRESSION MODELS

MICHAEL VOGT
University of Konstanz

In this paper, we consider a nonparametric model with a time-varying regression function and locally stationary regressors. We are interested in the question whether the regression function has the same shape over a given time span. To tackle this testing problem, we propose a kernel-based L_2-test statistic. We derive the asymptotic distribution of the statistic both under the null and under fixed and local alternatives. To improve the small sample behavior of the test, we set up a wild bootstrap procedure and derive the asymptotic properties thereof. The theoretical analysis of the paper is complemented by a simulation study and a real data example.

1. INTRODUCTION

In this paper, we are concerned with a testing issue in the nonparametric regression framework

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

where $\mathbb{E}[\varepsilon_{t,T}|X_{t,T}] = 0$ and the covariates $X_{t,T} = (X_{t,T}^1, \ldots, X_{t,T}^d)$ have dimension d. The regression function m in this setting is allowed to vary over time. Moreover, the regressors $X_{t,T}$ are not restricted to be stationary but are allowed to be locally stationary. Intuitively speaking, a process is locally stationary if over short periods of time (i.e., locally in time) it behaves approximately stationary. Throughout the introduction, we stick to this intuitive concept of local stationarity. A formal definition is given later on. Note that as usual in the literature on locally stationary processes, we have rescaled the time argument of the regression function to the unit interval.

We are interested in the question whether the function $m(u, \cdot)$ in model (1) has the same shape for all time points u within a given time span $I \subset [0, 1]$. Put differently, we want to test for structural stability of the regression function over the time period I. The main goal of the paper is to develop a formal procedure

I would like to thank Oliver Linton and Enno Mammen for numerous helpful discussions and comments. Financial support by the ERC is gratefully acknowledged. Address correspondence to Michael Vogt, Department of Mathematics and Statistics, University of Konstanz, 78457 Konstanz, Germany; e-mail: michael.vogt@uni-konstanz.de.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

to do so. Depending on the application, the time span I of interest may take different forms. In some applications, it will be an interval. In others, it will be the union of disjoint intervals. In our data example in Section 6, for instance, we investigate whether the regression function of a volatility model has the same shape within two different crisis periods, i.e., within two disjoint time intervals. To accommodate these different situations, we do not restrict the time period I to be an interval but rather allow it to be a finite union of intervals.

As a special case, we can set $I = [0, 1]$. This amounts to testing for global stability of the regression function m. There is a wide range of articles which analyze this testing issue in a parametric setting where the function m is known up to a finite number of time-varying coefficients; see e.g., Chen and Hong (2012), Kristensen (2012), or Zhang and Wu (2012) for some more recent references. If the model at hand is an autoregressive process with time-varying parameters, then the issue of testing for parameter stability is closely related to that of testing for stationarity in a locally stationary setup. Procedures to test for stationarity within a general class of locally stationary processes have been developed, for example, by Paparoditis (2009, 2010), Dette, Preuß, and Vetter (2011), and Preuß, Vetter, and Dette (2012). These articles approach the problem in the frequency domain and construct procedures to test whether the spectral density of the time series process under investigation is time-invariant.

Whereas the issue of testing for structural stability has received much attention in the parametric regression case, there is relatively little work in the nonparametric case. Hidalgo (1995) considers the model $Y_{t,T} = g(\frac{t}{T}, X_t) + \varepsilon_t$, where the regressors X_t are restricted to be strictly stationary. He develops a conditional moment test for the hypothesis that $g(u, \cdot) = g(\cdot)$ for a fixed function $g(\cdot)$ and all time points $u \in [0, 1]$. Su and Xiao (2008) suggest a CUSUM type test in a similar setting. They allow for nonstationarities in the covariates but analyze the behavior of their test statistic only on a rather specific type of (local) alternatives. Finally, Su and White (2010) set up a test for structural change in partially linear models.

A related strand of the literature deals with estimating the location and size of structural breaks in nonparametric regression. Müller (1992), for example, considers a nonparametric regression model with a fixed design and derives both the convergence rates and a central limit theorem for the estimates of the location and the size of the breaks in the regression function. Wu and Chu (1993) propose a test for the number of breaks in a similar setting. Delgado and Hidalgo (2000) consider the model $Y_{t,T} = g(\frac{t}{T}, X_t) + \varepsilon_t$ and propose estimates for both the location and the size of the breaks in the function g, which may occur in time direction or in the direction of the regressor X_t.

In the following sections, we develop a new statistical method to test for time-invariance of the regression function m in the general framework (1) over a given time span I and provide a complete asymptotic theory for the behavior of our test. A detailed description of model (1) and of the main assumptions on the model components including a rigorous definition of local stationarity is given in Section 2. Our test statistic measures an L_2-distance between kernel

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

estimates of the functions $m(u,\cdot)$ and $m(v,\cdot)$ at different time points $u, v \in I$ and is introduced in Section 3. Its asymptotic properties are analyzed in Section 4. The limit behavior of the test statistic will turn out to be mainly driven by a quadratic form. Not much is known about the asymptotic behavior of quadratic forms in a locally stationary setting. To our knowledge, Lee and Subba Rao (2011) are the only ones who have analyzed a general class of quadratic forms for locally stationary processes so far. However, the class they consider 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 setup.

To improve the finite sample behavior of our test, we propose a wild bootstrap procedure in Section 4.3 and derive the asymptotic properties thereof. In addition, we carry out a simulation study in Section 5 which confirms that our bootstrap procedure works well in small samples. The simulation study is complemented by a financial data example in Section 6.

2. THE MODEL

In this section, we give a detailed description of the regression framework (1) which underlies our testing problem. 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:

(a) The function m is assumed to be smooth in time direction. We thus allow for gradual changes but rule out sudden structural breaks in time direction. The exact smoothness conditions on m are listed in Section 4.1. Some discussion on how our test statistic behaves in the presence of structural breaks is provided in Section 7.2.

(b) As already noted in the introduction, we do not restrict the regressors to be strictly stationary. Instead, we allow the triangular array $\{X_{t,T} : t = 1,\dots,T\}$ 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 rescaled 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*

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

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 $\mathbb{R}^d$.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

This definition of local stationarity has been introduced in Vogt (2012). Intuitively speaking, it says that the process $\{X_{t,T}\}$ can be approximated locally around each time point u by a strictly stationary process, namely the process $\{X_t(u)\}$. Similar definitions can be found, e.g., in Dahlhaus and Subba Rao (2006) or Koo and Linton (2012).

(c) The error process $\{\varepsilon_{t,T} : t = 1, \ldots, T\}$ is assumed to have the martingale difference property that

$$\mathbb{E}\big[\varepsilon_{t,T} \big| \{X_{s,T} : s \leq t\}, \{\varepsilon_{s,T} : s < t\}\big] = 0 \tag{2}$$

for all t. This rules out autocorrelation in the error terms but allows for heteroskedasticity. (2) is, for example, satisfied by residuals of the form

$$\varepsilon_{t,T} = \sigma\left(\frac{t}{T}, X_{t,T}\right)\varepsilon_t, \tag{3}$$

where σ is a time-varying volatility function and $\{\varepsilon_t\}$ is an i.i.d. (independent and identically distributed) process having the property that ε_t is independent of $X_{s,T}$ for $s \leq t$. Imposing a martingale difference structure on the regression errors is technically convenient, as it allows us to apply a central limit theorem for martingale differences in the proofs. Similar martingale difference type conditions on the error terms are quite common in the literature on nonparametric kernel-based tests; see e.g., Fan and Li (1999), Li (1999), or Su and Xiao (2008). To simplify the calculations in the proofs later on, we will work with the specific error structure (3) instead of the more general martingale difference structure (2). The proofs can however be easily adjusted to work under the assumption (2).

An important class of processes that fit into the framework (1) is given by the nonlinear autoregressive model

$$Y_{t,T} = m\left(\frac{t}{T}, Y_{t-1,T}^{t-d}\right) + \sigma\left(\frac{t}{T}, Y_{t-1,T}^{t-d}\right)\varepsilon_t \tag{4}$$

with $Y_{t-1,T}^{t-d} = (Y_{t-1,T}, \ldots, Y_{t-d,T})$ 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 (4) have the properties (a)–(c). In particular, the autoregressive process $\{Y_{t,T}\}$ 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 (4) and a proof of these results, see Vogt (2012).

The time-varying regression function m of model (1) can be estimated by nonparametric kernel techniques. In what follows, we work with a Nadaraya–Watson type estimator. Alternatively, our test statistic could be based on local linear or more generally on local polynomial smoothers; see Fan and Gijbels (1996) for a detailed account of the local polynomial estimation method. The Nadaraya–Watson estimator of m is given by

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

$$\hat{m}_h(u,x) = \frac{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) Y_{t,T}}{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)}. \tag{5}$$

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 employ a product kernel and assume that the bandwidth is the same in each direction. Our results can however be easily modified to allow for nonproduct kernels and different bandwidths. The asymptotic properties of the estimator $\hat{m}_h$ are summarized in the first subsection of the Appendix.

3. THE TEST STATISTIC

Let $I \subset [0,1]$ be an interval or a finite union of intervals. We want to test whether the regression function $m(u,\cdot)$ has the same shape at all time points $u \in I$. The null hypothesis is thus given by

$$H_0 : \; m(u,\cdot) = m(v,\cdot) \quad \pi\text{-}a.s. \text{ for (almost) all } u, v \in I,$$

where π is some weight function. The null hypothesis can equivalently be expressed as

$$H_0 : \int_{I\times I} \left(\int \left[m(u,x) - m(v,x) \right]^2 \pi(x) dx \right) du dv = 0.$$

A natural way to come up with a test statistic for this problem is to replace the unknown regression function m in the above L_2-distance by an estimator and to rescale appropriately. This yields the weighted L_2-test statistic

$$S_T = T h^{\frac{d+1}{2}} \int_{I\times I} \left(\int \left[\hat{m}_h(u,x) - \hat{m}_h(v,x) \right]^2 \pi(x) dx \right) du dv, \tag{6}$$

where $\hat{m}_h$ is the Nadaraya–Watson smoother defined in (5). For simplicity, we assume that the weight function π has bounded support.[1] In the literature, related L_2-type test statistics have been used to test for parametric specification. In contrast to our setting, these statistics measure an L_2-distance between a nonparametric and a (semi)parametric fit; see e.g., Härdle and Mammen (1993), Sergides and Paparoditis (2009), or Preuß, Vetter, and Dette (2011).

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

$$H_1 : \int_{I\times I} \left(\int \left[m(u,x) - m(v,x) \right]^2 \pi(x) dx \right) du dv > 0.$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

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 m satisfies the null hypothesis (i.e., is time-invariant within I). The process $\{Y_{t,T}\}$ is thus given by

$$Y_{t,T} = m_T\left(\frac{t}{T}, X_{t,T}\right) + \varepsilon_{t,T} = m\left(\frac{t}{T}, X_{t,T}\right) + c_T \Delta\left(\frac{t}{T}, X_{t,T}\right) + \varepsilon_{t,T}. \tag{7}$$

If the process $\{Y_{t,T}\}$ is generated according to (7), we move along the sequence of local alternatives

$$H_{1,T} : \int_{I\times I} \left(\int \left[m_T(u,x) - m_T(v,x) \right]^2 \pi(x) dx \right) du dv = c_T^2 \Delta_I \tag{8}$$

with $\Delta_I := \int_{I\times I} (\int [\Delta(u,x) - \Delta(v,x)]^2 \pi(x) dx) du dv$. Note that the weighted L_2-distance in (8) 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.

4. ASYMPTOTICS

In what follows, we describe the asymptotic properties of the test statistic S_T defined in (6). Section 4.1 lists the conditions needed to do so. In Section 4.2, we derive the asymptotic distribution of S_T under the null as well as under fixed and local alternatives. Moreover, we explain how to bootstrap S_T in Section 4.3 and show that the proposed bootstrap is consistent.

4.1. Assumptions

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 $\|X_{t,T} - X_t(u)\| \leq (|\frac{t}{T} - u| + \frac{1}{T}) U_{t,T}(u)$ a.s. (almost surely) 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 $\alpha(k)$ that converge exponentially fast to zero, i.e., $\alpha(k) \leq C a^k$ for some $a < 1$.

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

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

(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

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

(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 $\mathbb{E}[|\varepsilon_t|^{8+\delta}] < \infty$ 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 density of $(X_{t,T}, X_{t+l,T})$. For any compact set $A \subset \mathbb{R}^d$, there exists a constant $C = C(A) < \infty$ such that $\sup_{t,T} \sup_{x\in A} f_{X_{t,T}}(x) \leq C$ and

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

We briefly give some remarks on the above assumptions. We start with a reminder of the definition of an α-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

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

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

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

where $\sigma(Z)$ is the σ-field generated by Z. The array $\{Z_{t,T}\}$ is said to be α-mixing (or strongly mixing) if $\alpha(k) \to 0$ as $k \to \infty$. Note that we do not necessarily require exponentially decaying 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 1. If $d = 1$, we can set $r = 2$ and choose the bandwidth h to satisfy $CT^{-(\frac{1}{4}-\delta)} \leq h \leq CT^{-(\frac{1}{5}+\delta)}$ for some small $\delta > 0$ and a positive constant C. Finally, Assumption (C7) is required to derive the uniform convergence rates of the estimator $\hat{m}_h$. It is a modification of the conditions (5)–(7) used in Hansen (2008).

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

4.2. The Asymptotic Distribution of S_T

We now summarize the results on the asymptotic behavior of the test statistic S_T. To formulate the theorems, we introduce the following shorthands: Let $\omega(u) = \mathbb{I}(u \in I)$ as well as $\lambda_I = \int \omega(u)du$ and define

$$V = 8\lambda_I^2\big(\int \mathscr{K}^2(z)dz\big)\int\bigg(\int \frac{[\sigma^2(u,x)]^2\pi^2(x)}{f^2(u,x)}dx\bigg)\omega(u)du$$

with $\mathscr{K}(z) = \int \prod_{j=0}^d K(w^j)\prod_{j=0}^d K(w^j + z^j)dw$. As we will see later on, V plays the role of the asymptotic variance of S_T. In addition, we define

$$\begin{aligned}
B_{T,1} &= \lambda_I h^{-\frac{d+1}{2}}\int\Big(\int\ldots\int K^2(w)\prod_{j=1}^d K^2(z^j)\sigma^2(u-hw,x-hz)\\
&\quad\times f(u-hw,x-hz)\frac{\pi(x)}{f^2(u,x)}dwdzdx\Big)\omega(u)du\\
B_{T,2} &= h^{-\frac{d-1}{2}}\iint\Big(\int\ldots\int K(w)K(\delta+w)\prod_{j=1}^d K^2(z^j)\sigma^2(u-hw,x-hz)\\
&\quad\times f(u-hw,x-hz)\frac{\pi(x)}{f(u,x)f(u+\delta h,x)}dwdzdx\Big)\omega(u)\omega(u+\delta h)d\delta du
\end{aligned}$$

together with

$$\begin{aligned}
B_{T,3} &= \lambda_I h^{-\frac{d+1}{2}}\int\Big(\int\ldots\int K^2(w)\prod_{j=1}^d K^2(z^j)\big[\overline{m}(x-hz)-\overline{m}(x)\big]^2\\
&\quad\times f(u-hw,x-hz)\frac{\pi(x)}{f^2(u,x)}dwdzdx\Big)\omega(u)du\\
B_{T,4} &= h^{-\frac{d-1}{2}}\iint\Big(\int\ldots\int K(w)K(\delta+w)\prod_{j=1}^d K^2(z^j)\big[\overline{m}(x-hz)-\overline{m}(x)\big]^2\\
&\quad\times f(u-hw,x-hz)\frac{\pi(x)}{f(u,x)f(u+\delta h,x)}dwdzdx\Big)\omega(u)\omega(u+\delta h)d\delta du,
\end{aligned}$$

where $\overline{m}(x) = \lambda_I^{-1}\int m(u,x)\omega(u)du$. Using these definitions, we can introduce the term

$$B_T = 2\big[B_{T,1} - B_{T,2} + B_{T,3} - B_{T,4}\big],$$

which turns out to be the bias by which S_T has to be corrected later on. The two bias components $B_{T,3}$ and $B_{T,4}$ are due to the smoothing bias of the Nadaraya–Watson estimator $\hat{m}_h$. Importantly, they converge to zero as long as $d \leq 2$ and thus only contribute to the bias of the test statistic in higher dimensions. As can

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

be seen, both the variance V and the bias B_T depend on the volatility function σ. This dependence captures how the test statistic is affected by the presence of heteroskedasticity in the model.

With the above definitions at hand, we can now formulate the main results. The first theorem states that under the null hypothesis, S_T weakly converges to a Gaussian distribution if we subtract the bias term B_T that diverges to infinity.

THEOREM 4.1. *Assume that (C1)–(C7) are fulfilled and let I be a closed subset of* $(0,1)$. *Then under* H_0,

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

Going back to the formula for V, the asymptotic variance of the test statistic is seen to depend on the length of the time span I, i.e., on $\lambda_I = \int \omega(u)du$. This intuitively makes sense as the test statistic measures an L_2-distance between estimates of the regression function at different time points $u, v \in I$. The larger the time span I, the larger this L_2-distance gets, which becomes reflected in the asymptotic variance V via its dependence on λ_I.

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

THEOREM 4.2. *Assume that (C1)–(C7) are fulfilled and let I be a closed subset of* $(0,1)$. *Then under* H_1,

$$(Th^{\frac{d+1}{2}})^{-1}(S_T - B_T) \xrightarrow{P} \int_{I\times I} \left(\int \left[m(u,x) - m(v,x) \right]^2 \pi(x)dx \right) dudv > 0.$$

We finally examine the behavior 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 $c_T = (Th^{(d+1)/2})^{-1/2}$ and m satisfying the null hypothesis is constant for all functions Δ with the same value Δ_I, where Δ_I is specified in Theorem 4.3. A similar behavior is well-known from other kernel-based L_2-test statistics (see e.g., Härdle and Mammen, 1993).

THEOREM 4.3. *Assume that (C1)–(C7) are fulfilled, let I be a closed subset of* $(0,1)$ *and let* $c_T = (Th^{(d+1)/2})^{-1/2}$. *Then under* $H_{1,T}$,

$$S_T - B_T \xrightarrow{d} N(\Delta_I, V)$$

with $\Delta_I = \int_{I\times I} (\int [\Delta(u,x) - \Delta(v,x)]^2 \pi(x)dx)dudv$.

To prove Theorem 4.3, we require the process $\{X_{t,T}\}$ to be locally stationary and strongly mixing under local alternatives. This is guaranteed by the conditions (C1) and (C2). Note, however, that it is not clear at all whether the autoregressive

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

process $\{Y_{t,T}\}$ defined in (4) fulfills (C1) and (C2) under local alternatives. In this paper, we do not explore this issue any further. Instead, we simply exclude the autoregressive case when examining local alternatives.

4.3. Bootstrapping S_T

Theorem 4.1 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., Li and Wang, 1998 or Härdle and Mammen, 1993). The approximation in finite samples is thus rather poor in many cases. Moreover, the bias and variance expressions B_T and V contain unknown functions. Replacing them by consistent estimates results in further approximation errors.

A common way to improve the finite sample behavior of a test is to use bootstrap methods. In what follows, we set up a wild bootstrap procedure. This type of bootstrap has been applied in a nonparametric time series context before; see e.g., Franke, Kreiss, and Mammen (2002) or Kreiss, Neumann, and Yao (2008). An alternative procedure that is often used for dependent data is the block bootstrap by Künsch (1989). This bootstrap technique is however problematic to apply in our setting as it does not appropriately mimic the regression structure of the model. This issue has been pointed out by Härdle and Mammen (1993) in an i.i.d. setting for a standard empirical bootstrap (i.e., for a block bootstrap with a block length of one). If we are working with the autoregressive model (4), another possibility is to use the autoregression bootstrap examined in Franke et al. (2002). This procedure has the advantage that it captures the dependence of the data by mimicking its autoregressive structure. However, its asymptotic properties are much harder to analyze than those of the wild bootstrap.

The wild bootstrap sample in our framework is given by $\{Y^*_{t,T}, X_{t,T}\}_{t=1}^T$, where we define

$$Y^*_{t,T} = \tilde{m}_g(X_{t,T}) + \varepsilon^*_{t,T}.$$

The bootstrap residuals are constructed as

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

where $\hat{\varepsilon}_{t,T} = Y_{t,T} - \hat{m}_h(\frac{t}{T}, X_{t,T})$ are the estimated residuals and $\{\eta_t\}$ is some sequence of i.i.d. variables with zero mean and unit variance that is independent of $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$. Moreover, the estimate $\tilde{m}_g$ is defined as

$$\tilde{m}_g(x) = \frac{\int \hat{m}_g(u,x)\omega(u)du}{\int \omega(u)du}$$

with $\omega(u) = \mathbb{I}(u \in I)$ and a bandwidth g. It approximates the average regression function $\overline{m}(x) = \int m(u,x)\omega(u)du / \int \omega(u)du$, which under the null is equal to

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

the true time-invariant regression function on the time span I. The bandwidth g is required to satisfy $g \gg h$. The exact conditions on the oversmoothed bandwidth g are summarized in condition (C8) below.

Given the bootstrap sample $\{Y^*_{t,T}, X_{t,T}\}_{t=1}^T$, we can define the Nadaraya–Watson estimator

$$\hat{m}^*_h(u,x) = \frac{\sum_{t=1}^T K_h(u-\frac{t}{T})\prod_{j=1}^d K_h(x^j - X^j_{t,T})Y^*_{t,T}}{\sum_{t=1}^T K_h(u-\frac{t}{T})\prod_{j=1}^d K_h(x^j - X^j_{t,T})}.$$

Replacing the estimate $\hat{m}_h$ by $\hat{m}^*_h$ in the statistic S_T, we obtain the bootstrap test statistic

$$S^*_T = Th^{\frac{d+1}{2}} \int_{I\times I} \Big(\int \big[\hat{m}^*_h(u,x) - \hat{m}^*_h(v,x)\big]^2 \pi(x)dx \Big) dudv.$$

In Theorem 4.4, we will show that the statistic S^*_T approximates the distribution of S_T under the null hypothesis no matter whether the null is true or not. This intuitively makes sense: Both under the null and under alternatives, the smoother $\tilde{m}_g$ estimates a function that is time-invariant over the time span I. The bootstrap sample $\{Y^*_{t,T}, X_{t,T}\}_{t=1}^T$ thus imitates the model under the null even if the alternative is true. For this reason, S^*_T should mimic the distribution of S_T under the null hypothesis no matter whether the null holds or not.

To formulate the asymptotic properties of the bootstrap statistic S^*_T, we let $\mathbb{P}^*(\,\cdot\,) := \mathbb{P}(\,\cdot\,|\{Y_{t,T}, X_{t,T}\}_{t=1}^T)$ and introduce the assumption

(C8) The bandwidth g is such that $h^{(d+1)/2}g^{-d} \to 0$ and $hg^{-d/2} \to 0$. Moreover, $Tg^{2r}h^2 \to 0$, $Tg^{2r}h^{(d+1)/2} \to 0$, and $Th^{d+1}g^{2r-d} \to 0$.

The first two conditions on g specify how strongly we have to oversmooth the estimate $\tilde{m}_g$. Note that the latter three conditions can always be fulfilled if r is large enough, i.e., if we work with higher-order kernels and the functions f and m are assumed to be sufficiently smooth. In the case with $d = 1$ and $r = 2$, the above conditions simplify to $h/g \to 0$ and $Tg^4h \to 0$. These are fulfilled, for example, by setting $g = (\log T)h$ and choosing h to satisfy $h \le CT^{-(\frac{1}{5}+\delta)}$ for some small $\delta > 0$ and a constant C.

The next theorem shows that the above defined wild bootstrap is consistent.

THEOREM 4.4. *Let (C1)–(C8) be fulfilled and let I be a closed subset of* $(0,1)$. *Then*

$$S^*_T - B_T \xrightarrow{d} N(0,V)$$

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

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

5. SIMULATIONS

In this section, we examine the small sample behavior of the proposed wild bootstrap procedure in a Monte Carlo experiment. To do so, we simulate the autoregressive model

$$Y_{t,T} = m\left(\frac{t}{T}, Y_{t-1,T}\right) + \varepsilon_t \tag{9}$$

with i.i.d. errors ε_t that follow a standard normal distribution. We consider a linear and a nonlinear specification of the regression function m:

(a) The linear specification is given by $m_i^L(u, y) = a_i(u)y$ with different coefficient functions a_i. In particular, for $i = 0, 1, 2$, we let

$$a_i(u) = -0.5 + c_i(u - 0.5)\mathbb{I}(u > 0.5)$$

with $c_0 = 0$, $c_1 = 1$, and $c_2 = 2$. Whereas the function m_0^L is time-invariant, m_1^L and m_2^L smoothly vary over time. In addition, we consider the parameter function

$$a_3(u) = -0.5\mathbb{I}(u \leq 0.5) + 0.5\mathbb{I}(u > 0.5)$$

to examine how our method performs in the presence of sudden structural breaks.

(b) The nonlinear specification of m is

$$m_i^{NL}(u, y) = \begin{cases} -b_i(u) & \text{for } y < -(4k+1) \\ b_i(u)\sin(0.5\pi y) & \text{for } |y| \leq (4k+1) \\ b_i(u) & \text{for } y > (4k+1), \end{cases}$$

where k is an extremely large natural number, e.g., $k = 10^{12}$. The parameter functions b_i determine the time-varying amplitude of the sine curve $b_i(u)\sin(0.5\pi y)$ and are defined by

$$b_i(u) = 2 + c_i \arctan(75(u - 0.5))$$

for $i = 0, 1, 2$, where $c_0 = 0$, $c_1 = 0.1$, and $c_2 = 0.2$. As before, we additionally consider a parameter function with a structural break. In particular, since $b_2(0) \approx 1.7$ and $b_2(1) \approx 2.3$, we work with the jump function

$$b_3(u) = 1.7\mathbb{I}(u \leq 0.5) + 2.3\mathbb{I}(u > 0.5).$$

By Theorem 3.2 in Vogt (2012), the autoregressive process (9) with the regression function m_i^{NL} is locally stationary for $i = 0, 1, 2$. Note that the function $m_i^{NL}(u, y)$ only differs from the sine curve $b_i(u)\sin(0.5\pi y)$ in the tail region where $|y| > (4k + 1)$. Clearly, if k is sufficiently large,

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

then the autoregressive process $\{Y_{t,T}\}$ will basically never wander into this region (at least not for any practically relevant sample size). We can thus regard the function $m_i^{NL}(u, y)$ in the simulated model as being equal to the sine curve $b_i(u)\sin(0.5\pi y)$.

In what follows, we denote the model with the regression function m_i^L by $\left(M_i^L\right)$ and analogously use the label $\left(M_i^{NL}\right)$ for the model with the function m_i^{NL}. In $\left(M_0^L\right)$ and $\left(M_0^{NL}\right)$, the regression function is time-invariant over the whole sample, whereas in the other models it varies over time, the time-variation in $\left(M_2^L\right)$ and $\left(M_2^{NL}\right)$ being stronger than in $\left(M_1^L\right)$ and $\left(M_1^{NL}\right)$, respectively. The models $\left(M_0^L\right)$ and $\left(M_0^{NL}\right)$ allow us to investigate the behavior of the bootstrap test under the null hypothesis, whereas the other models show how the bootstrap works on different alternatives.

For each model, we draw $N = 1,000$ samples of length $T = 1,000$ and produce $B = 1,000$ bootstrap replications for each sample. The time span I of interest is chosen as $[0.2, 0.3] \cup [0.7, 0.8]$. To produce the simulation results, we use Nadaraya–Watson smoothers with an Epanechnikov kernel. Both in the linear and nonlinear settings, we choose the bandwidth in time direction to equal 0.15 and the bandwidth in the direction of the regressor $Y_{t-1,T}$ to equal 0.75. To check the robustness of our results, we have varied the bandwidth in time direction between 0.1 and 0.2 and the bandwidth in the direction of $Y_{t-1,T}$ between 0.5 and 1.0. As this yields very similar results, we do not report them here. The bandwidth g must converge to zero more slowly than h from a theoretical perspective. However, as long as the order of h is between $O(T^{-(1/4-\delta)})$ and $O(T^{-(1/5+\delta)})$, we may pick any bandwidth g with $h/g \to 0$. Hence, g may differ from h only slightly and converge to zero almost as quickly. In the simulations, we ignore this technicality and simply set $g = h$. Finally, let $\underline{q}$ and $\overline{q}$ be the 5% and 95% quantiles of the process $\{Y_{t,T}\}$ in model $\left(M_0^L\right)$ or in model $\left(M_0^{NL}\right)$, depending on whether we consider the linear or the nonlinear simulation setup. The weight function π is chosen as $\pi(x) = \mathbb{I}(x \in [\underline{q}, \overline{q}])$. We thus ignore the extreme tails of the process $\{Y_{t,T}\}$ to avoid boundary effects.

We now turn to the simulation results. To start with, we report the results for the linear model setup. The left-hand table of Figure 1 gives the actual size of the bootstrap test in model $\left(M_0^L\right)$ for the most common nominal levels 0.05, 0.10, and 0.15. The plot on the right-hand side is a graphical presentation of the relationship between nominal and actual levels. It shows that the actual levels are very stable around the nominal ones.

The power results in $\left(M_1^L\right)$, $\left(M_2^L\right)$, and $\left(M_3^L\right)$ are summarized in Figure 2. The table on the left-hand side reports the estimated power of the bootstrap test for different nominal levels. A graphical illustration of the power is given by the size–power curves in the right-hand plot. Overall, the test appears to have good power against the chosen alternatives. One can nicely see that the power increases as we move from model $\left(M_1^L\right)$ to $\left(M_2^L\right)$, i.e., as we move further away from the null

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

nominal size	actual size
0.05	0.06
0.10	0.12
0.15	0.17

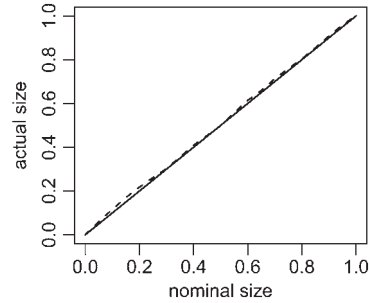


FIGURE 1. Comparison of nominal and actual size in model $\left(M_0^L\right)$. The dashed line in the right-hand figure gives the actual size of the bootstrap test in $\left(M_0^L\right)$ for each nominal level with target indicated as a solid line.

nominal size	power in $\left(M_1^L\right)$	power in $\left(M_2^L\right)$	power in $\left(M_3^L\right)$
0.05	0.69	0.94	1.0
0.10	0.78	0.99	1.0
0.15	0.83	0.99	1.0

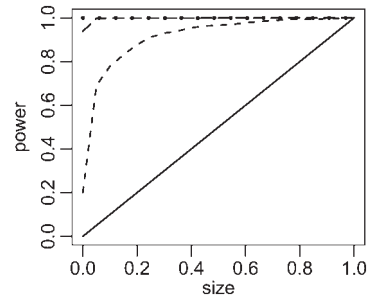


FIGURE 2. Estimated power in the models $\left(M_1^L\right)$, $\left(M_2^L\right)$, and $\left(M_3^L\right)$. The dashed line in the right-hand plot is the size–power curve for $\left(M_1^L\right)$, the long-dashed line is the one for $\left(M_2^L\right)$, and the dotted line corresponds to $\left(M_3^L\right)$.

hypothesis of time-invariance on the time span I. Moreover, the results for $\left(M_3^L\right)$ suggest that the test does not only have good power against smooth alternatives but also against sudden structural breaks.

The results in Figures 1 and 2 were generated as follows: Let $S_{T,n}$ be the value of the test statistic S_T in the n-th simulated sample. Moreover, let $S^*_{T,n,b}$ be the value of the bootstrap statistic S^*_T in the b-th bootstrap sample generated from the n-th simulation. Finally, denote by F^*_n the empirical distribution function calculated from the sample of bootstrap values $\{S^*_{T,n,b}\}_{b=1}^B$. With these definitions at hand, we proceed as follows: (1) Fix a nominal level α. (2) For each simulated sample n, calculate the $(1-\alpha)$-quantile $q^*_{\alpha,n}$ of F^*_n. (3) Compute the actual level and power corresponding to the nominal level α as $\frac{1}{N}\sum_{n=1}^N \mathbb{I}(S_{T,n} > q^*_{\alpha,n})$.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

We now turn to the simulation results for the nonlinear model specifications. Figure 3 summarizes the behavior of the bootstrap test in the setting $\left(M_0^{NL}\right)$. As can be seen, the actual levels approximate the nominal ones fairly accurately. The power properties in the settings $\left(M_1^{NL}\right)$, $\left(M_2^{NL}\right)$, and $\left(M_3^{NL}\right)$ are presented in Figure 4. As in the linear setup, our method appears to have good power both against smooth and nonsmooth alternatives. Even in model $\left(M_1^{NL}\right)$ where the parameter function b_1 varies over time only slightly, it detects the time-variation in the regression function reasonably well.

Overall, the simulation results suggest that the proposed bootstrap procedure has good size and power properties. The bootstrap thus provides an accurate way to calculate critical values of the test in practice.

nominal size	actual size
0.05	0.07
0.10	0.13
0.15	0.18

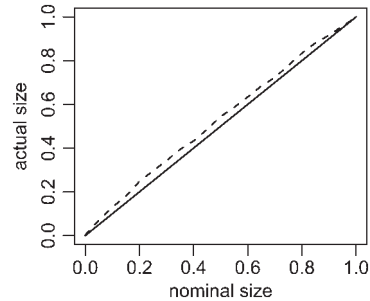


FIGURE 3. Comparison of nominal and actual size in model $\left(M_0^{NL}\right)$. The dashed line in the right-hand figure gives the actual size of the bootstrap test in $\left(M_0^{NL}\right)$ for each nominal level with target indicated as a solid line.

nominal size	power in $\left(M_1^{NL}\right)$	power in $\left(M_2^{NL}\right)$	power in $\left(M_3^{NL}\right)$
0.05	0.34	0.90	0.89
0.10	0.47	0.95	0.94
0.15	0.56	0.97	0.96

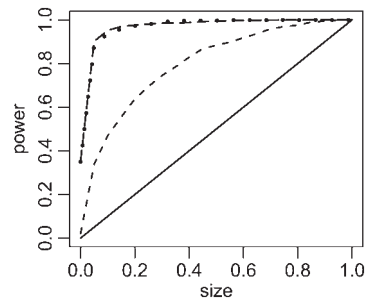


FIGURE 4. Estimated power in the models $\left(M_1^{NL}\right)$, $\left(M_2^{NL}\right)$, and $\left(M_3^{NL}\right)$. The dashed line in the right-hand plot is the size–power curve for $\left(M_1^{NL}\right)$, the long-dashed line is the one for $\left(M_2^{NL}\right)$, and the dotted line corresponds to $\left(M_3^{NL}\right)$.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

6. APPLICATION

We now apply our bootstrap procedure to a sample of financial volatility data. In particular, we consider the daily volatilities of the S&P 500 index. As a daily volatility measure, we use the so-called high–low range which is defined as the difference between the highest and lowest logarithmic price of a day. Alternatively, we could base our analysis on other volatility measures such as realized volatility. The high–low range has a long history in finance and has been employed in a variety of studies such as Schwert (1990), Alizadeh, Brandt, and Diebold (2002), and Martens and van Dijk (2007). As in many other papers, we work with the logarithmic transform of the high–low range. Our sample of logarithmic range data spans the period from 01/2000 to 12/2010 and is depicted in the left-hand panel of Figure 5. The right-hand panel provides some descriptive statistics. As can be seen, the data are slightly skewed and exhibit a moderate amount of kurtosis.

In recent years, a wide range of AR-type models have been proposed to model the daily time series dynamics of volatility measures such as realized volatility or the high–low range. As most volatility measures are characterized by slowly decaying sample autocorrelations, long-memory models like ARFIMA have been employed quite frequently; see Andersen, Bollerslev, Diebold, and Labys (2003) and Pong, Shackleton, Taylor, and Xu (2004) among others. However, as pointed out, for example, in Mikosch and Stărică (2004), the long-memory behavior may be spuriously generated by structural change in the volatility process. This has led several authors to use AR-type models with time-varying parameters for modeling volatility; see e.g., Chen, Härdle, and Pigorsch (2010). Related time-varying volatility models have been proposed in Mercurio and Spokoiny (2004), Spokoiny (2009), Čížek, Härdle, and Spokoiny (2009), and Härdle, Hautsch, and Mihoci (2012) among others.

Following the above considerations, we set up an autoregressive model with a time-varying regression function to capture the dynamics of daily volatility. Letting $r_{t,T}$ denote the logarithm of the daily range, we consider the model

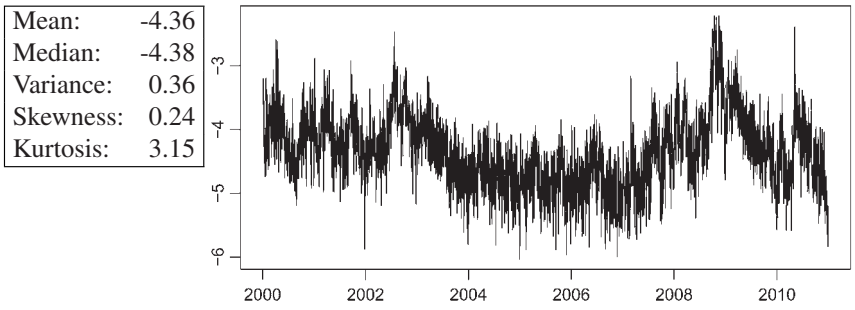


FIGURE 5. The left-hand panel shows the logarithmic daily high–low range of the S&P 500 index; the right-hand panel summarizes some descriptive statistics.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

$$r_{t,T} = m_0\left(\frac{t}{T}\right) + m_1\left(\frac{t}{T}, r_{t-1,T}\right) + \varepsilon_{t,T}, \tag{10}$$

where $\mathbb{E}[\varepsilon_{t,T}|r_{t-1,T}] = 0$ and $m_0(\frac{t}{T}) = \mathbb{E}[r_{t,T}]$ is the unconditional mean. The function m_1 in model (10) describes the volatility dynamics apart from level shifts, which are captured by the mean function m_0. We are interested in the question whether m_0 and m_1 vary over time. To test for time-invariance of m_0, a standard test for parameter stability may be used. For this reason, we restrict attention to the function m_1. To test for time-invariance of m_1, we modify our procedure as follows: Rewrite (10) as

$$r^c_{t,T} = m_1\left(\frac{t}{T}, r_{t-1,T}\right) + \varepsilon_{t,T}, \tag{11}$$

where $r^c_{t,T} = r_{t,T} - m_0(\frac{t}{T})$ is the demeaned version of $r_{t,T}$. If the demeaned variables $r^c_{t,T}$ were known, we could simply apply our procedure to the sample $\{r^c_{t,T}, r_{t-1,T}\}$. However, as the variables $r^c_{t,T}$ are in fact unknown, we replace them by estimates $\hat{r}^c_{t,T}$ and then apply our approach to the sample $\{\hat{r}^c_{t,T}, r_{t-1,T}\}$. Here, we let $\hat{r}^c_{t,T} = r_{t,T} - \hat{m}_{0,h}(\frac{t}{T})$ and use the kernel average $\hat{m}_{0,h}(u) = \frac{1}{T}\sum_{t=1}^{T} K_h(u - \frac{t}{T})r_{t,T}$ to estimate the mean function m_0 at the time point u. Moreover, the function m_1 is estimated by a Nadaraya–Watson smoother $\hat{m}_{1,h}$ calculated from the sample $\{\hat{r}^c_{t,T}, r_{t-1,T}\}$. This slightly modified approach has very similar asymptotic properties as those discussed in the previous sections. In particular, the wild bootstrap is still consistent. The details are given in Section 7.1.

We now apply model (11) to the sample of daily range data from Figure 5 and use our bootstrap procedure to test for time-invariance of the function m_1. As a first example, we check whether the function has the same shape within two different time periods in our sample, a period shortly after the technology bubble burst (08/2000–01/2001) and a period at the beginning of the recent financial crisis

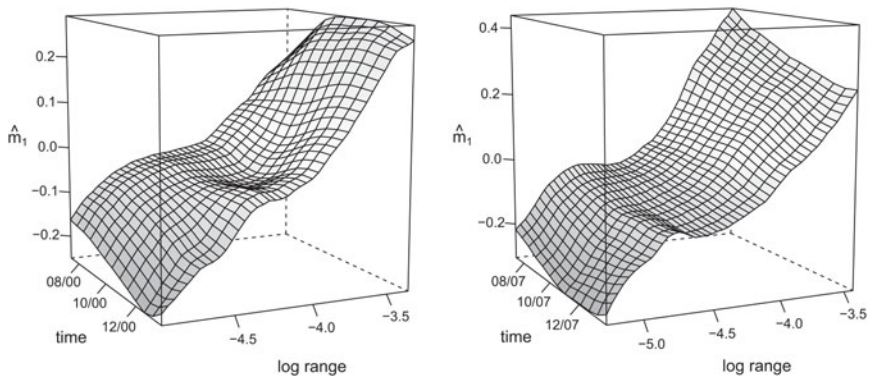


FIGURE 6. Estimates of the time-varying function m_1. The left panel shows the estimates for the time span 08/2000–01/2001 and the right one gives the estimates for the period 08/2007–01/2008.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

(08/2007–01/2008). Figure 6 shows the estimates of m_1 within these two periods. As can be seen, the fits are fairly stable within each of the two time intervals. Our bootstrap test confirms this impression. Applying it to each of the two intervals separately, we find a p-value roughly between 0.2 and 0.7 for a wide range of different bandwidths. This is strong evidence that the function m_1 is time-invariant within each of the two periods.

We now tackle the question whether the function m_1 has the same shape within both time periods. Let $I_A, I_B \subset [0,1]$ be the two rescaled time intervals that correspond to the periods 08/2000–01/2001 and 08/2007–01/2008, respectively. Given that m_1 is time-invariant within each of the intervals I_A and I_B, we can estimate it on I_k $(k = A, B)$ by the averaged smoother $\tilde{m}_{1,h}^{(k)}(x) := \int_{I_k} \hat{m}_{1,h}(u,x)du / \int_{I_k} du$. To get a first impression of how different the function m_1 is on the two intervals I_A and I_B, we plot $\tilde{m}_{1,h}^{(A)}$ and $\tilde{m}_{1,h}^{(B)}$ in Figure 7.

The two fits in Figure 7 have a similar shape and pointwise confidence bands which strongly overlap. However, there are also clear differences between them and regions where the confidence intervals are separated. To clarify whether the differences between the two fits are due to estimation error or to a structural change in the function m_1, we apply our bootstrap procedure to the union of the two intervals I_A and I_B. The estimated p-values for a variety of different bandwidths are summarized in Table 1.

The p-values in Table 1 lie in the region between 0.29 and 0.59. Thus, the bootstrap test does not reject the null hypothesis of stability at the usual significance levels. This is evidence that the function m_1 in our volatility model has the same shape on both intervals I_A and I_B, i.e., both when the technology bubble burst and when the recent financial crisis broke out.

As a second example, we test whether the function m_1 in model (11) is time-invariant over the last six months of 2008. This is a fairly turbulent time period,

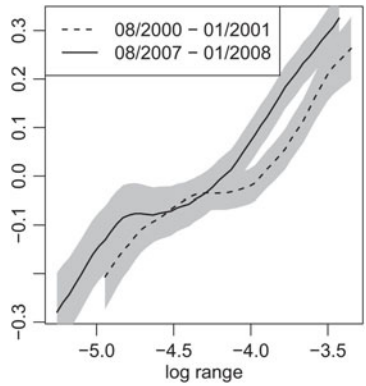


FIGURE 7. Averaged estimates $\tilde{m}_{1,h}^{(A)}$ (dashed line) and $\tilde{m}_{1,h}^{(B)}$ (solid line) for the two time periods 08/2000–01/2001 and 08/2007–01/2008. The gray shaded areas are the corresponding 95% pointwise confidence intervals.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

TABLE 1. Estimated p-values for testing the hypothesis that the function m_1 is time-invariant on $I_A \cup I_B$. h_u is the bandwidth in time direction and h_x the bandwidth in the direction of the covariate

h_u \ h_x	0.25	0.3	0.35	0.4	0.45	0.5
60 days	0.42	0.41	0.38	0.35	0.35	0.29
80 days	0.53	0.51	0.47	0.43	0.42	0.37
100 days	0.59	0.47	0.47	0.43	0.40	0.34
120 days	0.50	0.46	0.37	0.36	0.34	0.29

TABLE 2. Estimated p-values for testing the hypothesis that the function m_1 is time-invariant over the last six months of 2008. h_u is the bandwidth in time direction and h_x the bandwidth in the direction of the covariate

h_u \ h_x	0.25	0.3	0.35	0.4	0.45	0.5
60 days	0.000	0.000	0.006	0.026	0.067	0.160
80 days	0.000	0.000	0.005	0.006	0.009	0.038
100 days	0.002	0.005	0.009	0.007	0.013	0.022
120 days	0.004	0.010	0.016	0.016	0.015	0.008

which in particular includes the bankruptcy of Lehman Brothers. Table 2 presents the estimated p-values of our procedure for a number of different bandwidths.

As can be seen, the p-values all lie below 0.16, most of them taking values smaller than 0.05. Hence, our bootstrap test provides strong evidence against the null hypothesis, suggesting that the function m_1 is time-varying in the period around the bankruptcy of Lehman Brothers.

7. CONCLUDING REMARKS

In this paper, we have developed a nonparametric procedure to test whether the time-varying regression function in model (1) has the same shape over a time period $I \subset [0, 1]$. In our theoretical analysis, we have examined the asymptotic behavior of the proposed approach. In particular, we have shown that after subtracting a bias term that diverges to infinity, the test statistic weakly converges to a normal distribution (both under the null and under local alternatives). To improve the small sample behavior of the test, we have set up a wild bootstrap procedure and have shown that it is consistent. Our procedure may be extended in various directions. We close the paper by outlining some of them.

7.1. A Modified Version of the Test

When our test rejects the null, this is evidence for time-variation in the function m. The test being very general, it however remains silent about which features of the function are time-varying. In particular, the variation may stem from a simple

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

unconditional mean shift, or it may be due to changes in more subtle features of the function. In what follows, we explain how to modify our approach to make it a bit more informative about which aspects of the function are changing over time.

In a first step, we reformulate model (1). Letting $m_0(\frac{t}{T}) := \mathbb{E}[Y_{t,T}] = \mathbb{E}[m(\frac{t}{T}, X_{t,T})]$ be the unconditional mean function, we can rewrite (1) as

$$Y_{t,T} = m_0\Big(\frac{t}{T}\Big) + m_1\Big(\frac{t}{T}, X_{t,T}\Big) + \varepsilon_{t,T}, \tag{12}$$

where $m_1(u,x) := m(u,x) - m_0(u)$ and by construction $\mathbb{E}[m_1(\frac{t}{T}, X_{t,T})] = 0$ for all time points t. We thus split up the regression function m into two components: the unconditional mean m_0 and its "demeaned" version m_1. Given the formulation (12) of our model, we may now test separately whether the functions m_0 and m_1 are changing over time. This allows us to distinguish simple mean shifts from time-variations in more subtle aspects of the regression function.

To test whether m_0 is changing over time, a simplified version of our test statistic may be employed which measures an L_2-distance between kernel estimators of m_0. Alternatively, one of the many tests for parameter stability may be used. To test whether m_1 is time-varying, we adapt our test statistic to the situation at hand: Let the mean function m_0 be estimated by the kernel average

$$\hat{m}_{0,h}(u) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) Y_{t,T}$$

and let $\hat{Y}^c_{t,T} = Y_{t,T} - \hat{m}_{0,h}(\frac{t}{T})$ be the approximately demeaned version of $Y_{t,T}$. Based on the sample $\{\hat{Y}^c_{t,T}, X_{t,T}\}$, we then construct the Nadaraya–Watson smoother

$$\hat{m}_{1,h}(u,x) = \frac{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X^j_{t,T}) \hat{Y}^c_{t,T}}{\sum_{t=1}^{T} K_h(u - \frac{t}{T}) \prod_{j=1}^{d} K_h(x^j - X^j_{t,T})}$$

to estimate the function m_1. Replacing the estimator $\hat{m}_h$ in the test statistic S_T by $\hat{m}_{1,h}$, we obtain the modified statistic

$$S^c_T = T h^{\frac{d+1}{2}} \int_{I\times I} \Big(\int \big[\hat{m}_{1,h}(u,x) - \hat{m}_{1,h}(v,x)\big]^2 \pi(x) dx \Big) du dv.$$

Analogously as in Section 4.3, we can set up a wild bootstrap version of S^c_T. Slightly abusing notation, we let the bootstrap sample be given by $\{Y^*_{t,T}, X_{t,T}\}_{t=1}^T$ with

$$Y^*_{t,T} = \hat{m}_{0,g}\Big(\frac{t}{T}\Big) + \tilde{m}_{1,g}(X_{t,T}) + \varepsilon^*_{t,T},$$

where the bootstrap residuals are the same as in Section 4.3 and we define $\tilde{m}_{1,g}(x) = \int \hat{m}_{1,g}(u,x)\omega(u)du / \int \omega(u)du$. The bootstrap statistic $S^{c,*}_T$ is now constructed from the sample $\{Y^*_{t,T}, X_{t,T}\}_{t=1}^T$ in the same way as S^c_T from the original data.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

The statistic S_T^c and its bootstrap counterpart $S_T^{c,*}$ have very similar asymptotic properties as the statistics S_T and S_T^*. In particular, under the null that m_1 is time-invariant over the time span I, it holds that

$$S_T^c - B_T^c \xrightarrow{d} N(0, V), \tag{13}$$

where V is the variance expression defined in Section 4.2 and B_T^c is a slightly modified version of the bias term B_T. In particular, $B_T^c = B_{T,1}^c + B_{T,2}^c + B_{T,3}^c$, where $B_{T,1}^c$ has exactly the same form as B_T with m being replaced by m_1. Moreover,

$$B_{T,2}^c = Th^{\frac{d+1}{2}} \int\int \frac{1}{T^2} \sum_{s,t=1}^{T} w_{u,s,T} w_{u,t,T} \times \mathbb{E}\Big[\Big(\varepsilon_{s,T} + m_1\Big(\frac{s}{T}, X_{s,T}\Big)\Big)\Big(\varepsilon_{t,T} + m_1\Big(\frac{t}{T}, X_{t,T}\Big)\Big)\Big] \frac{\pi(x)\omega(u)}{f^2(u,x)} dxdu$$

$$B_{T,3}^c = Th^{\frac{d+1}{2}} \int\int \frac{1}{T^2} \sum_{s,t=1}^{T} K_{u,s,T} w_{u,t,T} \times \mathbb{E}\Big[K_{x,s,T}\varepsilon_{s,T}\Big(\varepsilon_{t,T} + m_1\Big(\frac{t}{T}, X_{t,T}\Big)\Big)\Big] \frac{\pi(x)\omega(u)}{f^2(u,x)} dxdu,$$

where $w_{u,t,T} = \frac{1}{T}\sum_{s=1}^{T} K_{u,s,T} K_{\frac{s}{T},t,T}$ and we use the shorthands $K_{u,t,T} = K_h(u - \frac{t}{T})$ as well as $K_{x,t,T} = \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)$. The additional bias components $B_{T,2}^c$ and $B_{T,3}^c$ result from the fact that we approximate the trend function m_0 by the smoother $\hat{m}_{0,h}$ in the test statistic. Importantly, they are at most of the order $O(h^{\frac{d-1}{2}} \log T)$ and thus of much smaller order than the leading bias $S_{T,1}^c$. In addition to the distribution result (13), the wild bootstrap can again be shown to be consistent,

$$S_T^{c,*} - B_T^c \xrightarrow{d} N(0, V) \tag{14}$$

conditional on the sample $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$ with probability tending to one both under the null and under alternatives. The results (13) and (14) essentially follow from the proofs of Theorems 4.1 and 4.4. Specifically, S_T^c and $S_T^{c,*}$ can be seen to consist of the same components as S_T and S_T^* together with some additional terms, which can be analyzed by slightly modifying the arguments from the proofs.

7.2. Structural Breaks

The theoretical results of the paper have been derived under the assumption that the function m varies smoothly over time. An interesting issue is whether our procedure is not only able to detect smooth changes but also sudden breaks in m. To clarify this issue, let m have a break at the time point $u_0 \in (0, 1)$ and assume for simplicity that it is time-invariant otherwise. Moreover, let $I \subset (0, 1)$ be

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

a time period whose interior contains the time point u_0. We now use our procedure to test whether m is time-invariant within the period I. To do so, let $I_1 = I \setminus [u_0 - C_1 h, u_0 + C_1 h]$ and write

$$S_T(J) = Th^{\frac{d+1}{2}} \int_{J \times J} \Big(\int \big[\hat{m}_h(u,x) - \hat{m}_h(v,x)\big]^2 \pi(x) dx \Big) du dv$$

for any time period $J \subset [0,1]$. Noting that $S_T(I) \geq S_T(I_1)$, we get

$$\begin{aligned}(Th^{\frac{d+1}{2}})^{-1}\big(S_T(I) - B_T(I)\big) &= (Th^{\frac{d+1}{2}})^{-1} S_T(I) + o_p(1) \\ &\geq (Th^{\frac{d+1}{2}})^{-1} S_T(I_1) + o_p(1), \end{aligned} \tag{15}$$

where $B_T(I)$ is the bias term that corresponds to the statistic $S_T(I)$. Furthermore, using the arguments from the proof of Theorems 4.1–4.3, it is easy to see that

$$(Th^{\frac{d+1}{2}})^{-1} S_T(I_1) \xrightarrow{P} \int_{I \times I} \Big(\int \big[m(u,x) - m(v,x)\big]^2 \pi(x) dx \Big) du dv > 0. \tag{16}$$

Combining (15) and (16) immediately yields that our test is consistent against breaks of fixed size. To get a better idea of the power of the test in the presence of breaks, one should also analyze its behavior in situations where the function m has a break of shrinking size. As far as we can see, this can however not be done by a straightforward modification of our proofs but requires substantially new and different arguments.

7.3. Additive Models

Our procedure being fully nonparametric, it suffers from the curse of dimensionality: When the dimension d of the covariates is large, the convergence rates of the Nadaraya–Watson smoothers are rather slow resulting in a poor behavior of the test statistics.

One way to circumvent the curse of dimensionality is to put a bit of structure on the regression function m. In particular, one may assume that it splits up into time-varying additive components, thus yielding the model

$$Y_{t,T} = m_0\Big(\frac{t}{T}\Big) + \sum_{j=1}^{d} m_j\Big(\frac{t}{T}, X_{t,T}^j\Big) + \varepsilon_{t,T}. \tag{17}$$

The component functions m_j in (17) can be estimated by smooth backfitting methods as introduced in Mammen, Linton, and Nielsen (1999). As shown in Vogt (2012), the resulting estimators uniformly converge to the true component functions at the usual two-dimensional nonparametric rate, no matter how large the full dimension d. Thus, the convergence rate does not deteriorate as the dimension grows.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

To cope with the case of high dimensions d, it would be useful to extend our testing theory to the additive setting (17). In particular, we would like to test (i) whether the additive regression function $m(u,x) = m_0(u) + \sum_{j=1}^{d} m_j(u, x^j)$ as a whole varies over time and (ii) whether a specific component function $m_j(u, x^j)$ is time-varying. Even though it is by no means trivial to extend our theory to tackle these issues, it may be possible to do so along the following lines: Mammen and Park (2005) have derived stochastic higher-order expansions for smooth backfitting estimators in an i.i.d. setting. These expansions link the backfitting estimators to the underlying Nadaraya–Watson smoothers which are used as pilot estimates in the algorithm. We conjecture that similar higher-order expansions may be derived in the setup (17) and that it is possible to extend our proofs with the help of these expansions.

NOTE

1. The case of unbounded support can be handled as follows: The uniform convergence results on the Nadaraya–Watson estimator in Lemmas A.1–A.3 in the Appendix can be extended to hold over diverging compact sets. We can thus allow for unbounded support of π by letting the limits of the integrals in S_T diverge to infinity at an appropriate rate as the sample size increases.

REFERENCES

Alizadeh, S., M.W. Brandt, & F.X. Diebold (2002) Range-based estimation of stochastic volatility models. *Journal of Finance* 57, 1047–1091.

Andersen, T., T. Bollerslev, F. Diebold, & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer.

Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* 2, 107–144.

Chen, Y., W. Härdle, & U. Pigorsch (2010) Localized realized volatility modeling. *Journal of the American Statistical Association* 105, 1376–1393.

Chen, B. & Y. Hong (2012) Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80, 1157–1183.

Čížek, P., W. Härdle, & V. Spokoiny (2009) Adaptive pointwise estimation in time-inhomogeneous conditional heteroscedasticity models. *The Econometrics Journal* 12, 248–271.

Dahlhaus, R. & S. Subba Rao (2006) Statistical inference for time-varying ARCH processes. *Annals of Statistics* 34, 1075–1114.

de Jong, P. (1987) A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields* 75, 261–277.

Delgado, M.A. & J. Hidalgo (2000) Nonparametric inference on structural breaks. *Journal of Econometrics* 96, 113–144.

Dette, H., P. Preuß, & M. Vetter (2011) A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association* 106, 1113–1124.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman & Hall.

Fan, Y. & Q. Li (1999) Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.

Franke, J., J.-P. Kreiss, & E. Mammen (2002) Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* 8, 1–37.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Hansen, B.E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.
Härdle, W., N. Hautsch, & A. Mihoci (2012) Local adaptive multiplicative error models for high-frequency forecasts. *Journal of Applied Econometrics*, forthcoming.
Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.
Hidalgo, J. (1995) A nonparametric conditional moment test for structural stability. *Econometric Theory* 11, 671–698.
Koo, B. & O. Linton (2012) Estimation of semiparametric locally stationary diffusion models. *Journal of Econometrics* 170, 210–233.
Kreiss, J.-P., M.H. Neumann, & Q. Yao (2008) Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and Its Interface* 1, 367–380.
Kristensen, D. (2012) Nonparametric detection and estimation of structural change. *The Econometrics Journal* 15, 420–461.
Künsch, H.R. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* 17, 1217–1241.
Lee, J. & S. Subba Rao (2011) A note on quadratic forms of nonstationary stochastic processes. Preprint.
Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics* 92, 101–147.
Li, Q. & S. Wang (1998) A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.
Mammen, E., O. Linton, & J. Nielsen (1999) The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* 27, 1443–1490.
Mammen, E. & B.U. Park (2005) Bandwidth selection for smooth backfitting in additive models. *Annals of Statistics* 33, 1260–1294.
Martens, M. & D. van Dijk (2007) Measuring volatility with the realized range. *Journal of Econometrics* 138, 181–207.
Mercurio, D. & V. Spokoiny (2004) Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics* 32, 577–602.
Mikosch, T. & C. Stărică (2004) Non-stationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* 86, 378–390.
Müller, H.-G. (1992) Change points in nonparametric regression analysis. *Annals of Statistics* 20, 737–761.
Paparoditis, E. (2009) Testing temporal constancy of the spectral structure of a time series. *Bernoulli* 15, 1190–1221.
Paparoditis, E. (2010) Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association* 105, 839–851.
Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer.
Pong, S., M.B. Shackleton, S.J. Taylor, & X. Xu (2004) Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking & Finance* 28, 2541–2563.
Preuß, P., M. Vetter, & H. Dette (2011) Testing semiparametric hypotheses in locally stationary processes. *Scandinavian Journal of Statistics*, forthcoming.
Preuß, P., M. Vetter, & H. Dette (2012) A test for stationarity based on empirical processes. *Bernoulli*, forthcoming.
Schwert, G.W. (1990) Stock volatility and the crash of '87. *Review of Financial Studies* 3, 77–102.
Sergides, M. & E. Paparoditis (2009) Frequency domain tests of semiparametric hypotheses for locally stationary processes. *Scandinavian Journal of Statistics* 36, 800–821.
Spokoiny, V. (2009) Multiscale local change point detection with applications to value-at-risk. *Annals of Statistics* 37, 1405–1436.
Su, L. & H. White (2010) Testing structural change in partially linear models. *Econometric Theory* 26, 1761–1806.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Su, L. & Z. Xiao (2008) Testing structural change in time-series nonparametric regression models. *Statistics and Its Interface* 1, 347–366.

Vogt, M. (2012) Nonparametric regression for locally stationary time series. *Annals of Statistics* 40, 2601–2633.

Wu, J. S. & C.K. Chu (1993) Kernel-type estimators of jump points and values of a regression function. *Annals of Statistics* 21, 1545–1566.

Zhang, T. & W.B. Wu (2012) Inference of time-varying regression models. *Annals of Statistics* 40, 1376–1402.

APPENDIX

In what follows, we prove Theorems 4.1–4.3 and 4.4. 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 behavior of the test statistic S_T, we need some results on uniform convergence of the Nadaraya–Watson estimator $\hat{m}_h$. To formulate these results, we split up the expression $\hat{m}_h(u,x) - m(u,x)$ into different components according to

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

with

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

$$\hat{g}_h^V(u,x) = \frac{1}{T}\sum_{t=1}^{T} K_h\Big(u - \frac{t}{T}\Big) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)\varepsilon_{t,T}$$

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

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

LEMMA A.1. *Let (C1)–(C7) be fulfilled. Then*

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

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

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

where the compact set $S \subset \mathbb{R}^d$ has been defined in (C5).

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

LEMMA A.2. *Let (C1)–(C7) be fulfilled and let* $I_h = [C_1 h, 1 - C_1 h]$. *Then*

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

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

Combining these two lemmas immediately yields the following result.

LEMMA A.3. *Let (C1)–(C7) be fulfilled and let* $I_h = [C_1 h, 1 - C_1 h]$. *Then*

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

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

Lemmas A.1–A.3 directly follow from the results in Vogt (2012).

Proof of Theorems 4.1–4.3

In what follows, we give the proof of Theorem 4.3. Theorem 4.1 is obtained by setting the function Δ equal to zero in the proof. Some straightforward additional considerations yield Theorem 4.2.

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

$$S_T = Th^{\frac{d+1}{2}} \int_{I \times I} \Big(\int \big[V_T(u,v,x) + B_T(u,v,x) \big]^2 \pi(x) dx \Big) du dv$$

with

$$\begin{aligned} V_T(u,v,x) &= \frac{1}{T} \sum_{t=1}^T K_{u,t,T} K_{x,t,T} \varepsilon_{t,T} \Big/ \hat{f}_h(u,x) \\ &\quad - \frac{1}{T} \sum_{t=1}^T K_{v,t,T} K_{x,t,T} \varepsilon_{t,T} \Big/ \hat{f}_h(v,x) \\ B_T(u,v,x) &= \frac{1}{T} \sum_{t=1}^T K_{u,t,T} K_{x,t,T} m_T\Big(\frac{t}{T}, X_{t,T}\Big) \Big/ \hat{f}_h(u,x) \\ &\quad - \frac{1}{T} \sum_{t=1}^T K_{v,t,T} K_{x,t,T} m_T\Big(\frac{t}{T}, X_{t,T}\Big) \Big/ \hat{f}_h(v,x). \end{aligned}$$

Theorem 4.3 immediately follows from the following three lemmas.

LEMMA A.4. *Under (C1)–(C7), it holds that*

$$Th^{\frac{d+1}{2}} \int_{I \times I} \Big(\int V_T^2(u,v,x) \pi(x) dx \Big) du dv - 2(B_{T,1} - B_{T,2}) \xrightarrow{d} N(0, V).$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

LEMMA A.5. *Under (C1)–(C7), it holds that*

$$Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int B_T(u,v,x) V_T(u,v,x) \pi(x) dx \right) du dv = o_p(1).$$

LEMMA A.6. *Under (C1)–(C7), it holds that*

$$Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int B_T^2(u,v,x) \pi(x) dx \right) du dv = \Delta_I + 2(B_{T,3} - B_{T,4}) + o_p(1)$$

with $\Delta_I = \int_{I\times I} (\int [\Delta(u,x) - \Delta(v,x)]^2 \pi(x) dx) du dv$.

We now give the proofs of the above lemmas. Throughout, we use the notation

$$\Lambda_{T,1}(u,x) = \frac{1}{T} \sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \varepsilon_{t,T}$$

$$\Lambda_{T,2}(u,x) = \frac{1}{T} \sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \Delta_{t,T}(u,x)$$

$$\Lambda_{T,3}(u,x) = \frac{1}{T} \sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \Delta\left(\frac{t}{T}, X_{t,T}\right)$$

with $\Delta_{t,T}(u,x) = m(\frac{t}{T}, X_{t,T}) - m(u,x)$.

Proof of Lemma A.4. Let

$$U_T = Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int V_T^2(u,v,x) \pi(x) dx \right) du dv.$$

Using the shorthand $\lambda_I = \int \omega(\varphi) d\varphi$, we obtain that $U_T = 2(U_{T,1} - U_{T,2})$ with

$$U_{T,1} = \lambda_I Th^{\frac{d+1}{2}} \int_I \left(\int \Lambda_{T,1}^2(u,x) \frac{\pi(x)}{\hat{f}_h^2(u,x)} dx \right) du$$

$$U_{T,2} = Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda_{T,1}(u,x) \Lambda_{T,1}(v,x) \frac{\pi(x)}{\hat{f}_h(u,x) \hat{f}_h(v,x)} dx \right) du dv.$$

In what follows, we show that

$$U_{T,1} - B_{T,1} \xrightarrow{d} N(0, V_1) \tag{A.1}$$

$$U_{T,2} - B_{T,2} = o_p(1) \tag{A.2}$$

with $V_1 = V/4$. Combining (A.1) and (A.2) completes the proof. To show (A.1), we split up $U_{T,1}$ into two parts according to

$$U_{T,1} = U_{T,1}^B + U_{T,1}^V + o_p(1)$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

with

$$U_{T,1}^B = \frac{\lambda_I h^{\frac{d+1}{2}}}{T} \int_I \sum_{t=1}^T K_{u,t,T}^2 \left(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)} dx \right) \varepsilon_{t,T}^2 du$$

$$U_{T,1}^V = \frac{\lambda_I h^{\frac{d+1}{2}}}{T} \int_I \sum_{t=1}^T \sum_{s \neq t} K_{u,t,T} K_{u,s,T} \left(\int K_{x,t,T} K_{x,s,T} \frac{\pi(x)}{f^2(u,x)} dx \right) \varepsilon_{t,T} \varepsilon_{s,T} du,$$

where we have used the uniform convergence results of Lemmas A.1–A.3 to replace the kernel density $\hat{f}_h(u,x)$ by the true density $f(u,x)$. In the sequel, we show that

$$U_{T,1}^B = B_{T,1} + o_p(1) \tag{A.3}$$

$$(V_1)^{-1/2} U_{T,1}^V \xrightarrow{d} N(0,1). \tag{A.4}$$

This immediately yields (A.1). To prove (A.2), we analogously decompose $U_{T,2}$ into two parts, $U_{T,2} = U_{T,2}^B + U_{T,2}^V + o_p(1)$, and show that $U_{T,2}^B = B_{T,2} + o_p(1)$ as well as $U_{T,2}^V = o_p(1)$. This can be done by repeating part of the arguments used to show (A.3) and (A.4). ■

Proof of (A.3). It suffices to show that $\mathrm{Var}\big(U_{T,1}^B\big) = o(1)$ and $\mathbb{E}\big[U_{T,1}^B\big] = B_{T,1} + o(1)$. The first claim easily follows by 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 $\|X_{t,T} - X_t(u)\| \leq (|\frac{t}{T} - u| + \frac{1}{T}) U_{t,T}(u)$. This can be achieved by the same techniques as in the proof of Theorem 4.2 in Vogt (2012), which yield that

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

Next note that

$$\begin{aligned}
&\frac{1}{T} \sum_{t=1}^T K_{u,t,T}^2 \sigma^2\Big(\frac{t}{T}, z\Big) f\Big(\frac{t}{T}, z\Big) \\
&\quad = \int K_h^2(u-w) \sigma^2(w,z) f(w,z) dw + O\Big(\frac{1}{Th^3}\Big)
\end{aligned} \tag{A.5}$$

uniformly in u and z. This follows from the fact that the sum on the left-hand side can be regarded as a Riemann approximation of the integral on the right. Exploiting the

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

smoothness conditions on K, σ, and f, the approximation error can be calculated to be of the order $(Th^3)^{-1}$. With the help of (A.5), we finally arrive at

$$\mathbb{E}\left[U_{T,1}^B\right] = \lambda_I h^{\frac{d+1}{2}} \int_I \int \left(\iint K_h^2(u-w) \prod_{j=1}^{d} K_h^2(x^j - z^j) \times \sigma^2(w,z) f(w,z) dw dz \right) \frac{\pi(x)}{f^2(u,x)} dx du + o(1)$$
$$= B_{T,1} + o(1). \quad \blacksquare$$

Proof of (A.4). We rewrite $U_{T,1}^V$ as

$$U_{T,1}^V = \sum_{t=1}^{T} Z_{t,T}$$

with

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

Note that under (C6), $\{Z_{t,T}, \mathscr{F}_{t,T}\}$ with $\mathscr{F}_{t,T} = \sigma(X_{t+1,T}, X_{t,T}, \varepsilon_{t,T}, \ldots, X_{1,T}, \varepsilon_{1,T})$ 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, 1984) to show that $\sum_{t=1}^{T} Z_{t,T}$ is asymptotically normal. According to the theorem in Pollard (1984), it suffices to verify the following conditions:

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

(CLT2) $\sum_{t=1}^{T} \mathbb{E}[Z_{t,T}^2 | \mathscr{F}_{t-1,T}] \xrightarrow{P} V_1$.

This yields (A.4). $\blacksquare$

Proof of (CLT1). We can write

$$\sum_{t=1}^{T} \mathbb{E}\left[Z_{t,T}^4\right] = \frac{16\lambda_I^4 \mathbb{E}[\varepsilon_t^4] h^{2d+2}}{T^4} \times \iint \sum_{t=1}^{T} \sum_{(s,s',s'',s''') \in \mathscr{S}_t} \mathbb{E}\Big[W_{t,T}(\underline{u}, \underline{x}) W_{s,T}(u,x) W_{s',T}(u',x') \times W_{s'',T}(u'',x'') W_{s''',T}(u''',x''') \Big] \frac{\pi(x) \ldots \pi(x''') \omega(u) \ldots \omega(u''')}{f^2(u,x) \ldots f^2(u''',x''')} d\underline{x} d\underline{u},$$

where $\mathscr{S}_t$ denotes the set of index tuples (s,s',s'',s''') with $s,s',s'',s''' < t$, $\omega(u) = \mathbb{I}(u \in I)$, $\underline{u} = (u,u',u'',u''')$, and $\underline{x} = (x,x',x'',x''')$. Moreover,

$$W_{t,T}(\underline{u}, \underline{x}) = K_{u,t,T} K_{u',t,T} K_{u'',t,T} K_{u''',t,T} K_{x,t,T} K_{x',t,T} K_{x'',t,T} K_{x''',t,T} \left(\sigma^2\left(\frac{t}{T}, X_{t,T} \right) \right)^2,$$
$$W_{s,T}(u,x) = K_{u,s,T} K_{x,s,T} \varepsilon_{s,T},$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

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

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

and write

$$\sum_{t=1}^{T} \mathbb{E}\Big[Z_{t,T}^4\Big] = Q_T^{(1)} + \cdots + Q_T^{(5)}$$

with

$$Q_T^{(i)} = \frac{16\lambda_I^4 \mathbb{E}[\varepsilon_t^4] h^{2d+2}}{T^4} \iint \sum_{t=1}^{T} \sum_{(s,s',s'',s''') \in \mathscr{S}_t^{(i)}} \mathbb{E}\Big[W_{t,T}(\underline{u},\underline{x}) W_{s,T}(u,x) W_{s',T}(u',x') \times W_{s'',T}(u'',x'') W_{s''',T}(u''',x''')\Big] \frac{\pi(x)\ldots\pi(x''')\omega(u)\ldots\omega(u''')}{f^2(u,x)\ldots f^2(u''',x''')} d\underline{x}\,d\underline{u}$$

for $i = 1,\ldots,5$. In what follows, we show that $Q_T^{(i)} \to 0$ for $i = 1,\ldots,5$. To lay out the proving strategy, we give a detailed account of the arguments for the term $Q_T^{(3)}$. The other terms can be handled in an analogous way.

To analyze the term $Q_T^{(3)}$, we first have a closer look at the index set $\mathscr{S}_t^{(3)}$. Because of symmetry considerations, we can assume w.l.o.g. (without loss of generality) that $s \geq s' \geq s'' \geq s'''$. Given this, the following two cases are possible:

$$(A)\ s = s' = s'' > s''' \qquad (B)\ s > s' = s'' = s'''.$$

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 specified later on, i.e., $|k - k'| > C_2 \log T$. Using this definition, we can further split up case (A) into the two subcases

(A_1) $s = s' = s'' > s'''$ and s''' is separated from $s = s' = s''$

(A_2) $s = s' = s'' > s'''$ and s''' is not separated from $s = s' = s''$.

Analogously, we can distinguish between the two subcases

(B_1) $s > s' = s'' = s'''$ and s is separated from t

(B_2) $s > s' = s'' = s'''$ and s is not separated from t.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Introducing the index sets

$$\mathscr{S}_t^{(3,A_1)} = \{(s,s',s'',s''') \in \mathscr{S}_t \mid \text{the indices satisfy } (A_1)\}$$
$$\mathscr{S}_t^{(3,A_2)} = \{(s,s',s'',s''') \in \mathscr{S}_t \mid \text{the indices satisfy } (A_2)\}$$
$$\mathscr{S}_t^{(3,B_1)} = \{(s,s',s'',s''') \in \mathscr{S}_t \mid \text{the indices satisfy } (B_1)\}$$
$$\mathscr{S}_t^{(3,B_2)} = \{(s,s',s'',s''') \in \mathscr{S}_t \mid \text{the indices satisfy } (B_2)\},$$

we obtain that

$$Q_T^{(3)} = 4\big(Q_T^{(3,A_1)} + Q_T^{(3,A_2)} + Q_T^{(3,B_1)} + Q_T^{(3,B_2)}\big),$$

where the terms $Q_T^{(3,k)}$ are defined analogously as $Q_T^{(3)}$ for $k = A_1, A_2, B_1, B_2$. We now consider the terms $Q_T^{(3,k)}$ one after the other.

(A_1) As the model variables are mixing (with exponential decay) and s''' is separated from $s = s' = s''$, we can use Davydov's inequality (see e.g., Bosq, 1998, Cor. 1.1) to get

$$\begin{aligned}
&\Big|\mathbb{E}\Big[W_{t,T}(\underline{u},\underline{x})W_{s,T}(u,x)W_{s',T}(u',x')W_{s'',T}(u'',x'')W_{s''',T}(u''',x''')\Big]\Big| \\
&= \Big|\mathrm{Cov}\Big(W_{t,T}(\underline{u},\underline{x})W_{s,T}(u,x)W_{s',T}(u',x')W_{s'',T}(u'',x''), W_{s''',T}(u''',x''')\Big)\Big| \\
&\le C\alpha(C_2\log T)^{1-\frac{1}{4/3+\delta}-\frac{1}{4+\delta}}\Big(\mathbb{E}\big|W_{s''',T}(u''',x''')\big|^{4+\delta}\Big)^{\frac{1}{4+\delta}} \\
&\quad\times\Big(\mathbb{E}\big|W_{t,T}(\underline{u},\underline{x})W_{s,T}(u,x)W_{s',T}(u',x')W_{s'',T}(u'',x'')\big|^{4/3+\delta}\Big)^{\frac{1}{4/3+\delta}} \\
&\le CT^{-C_3},
\end{aligned}$$

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 $Q_T^{(3,A_1)} \le CT^{-C_4}$ with some arbitrarily large constant C_4. As a result, $Q_T^{(3,A_1)} \to 0$.

(A_2) As s''' is not separated from $s = s' = s''$, the number of elements contained in the index set $\mathscr{S}_t^{(3,A_2)}$ is smaller than $C\log T$ for each given t, where C is a large positive constant independent of t. From this, it is easy to infer that $Q_T^{(3,A_2)} \le C\frac{\log T}{T^2h^{2d+1}} \to 0$.

Using analogous arguments as for (A_1) and (A_2), we further obtain that $Q_T^{(3,B_1)} \to 0$ and $Q_T^{(3,B_2)} \to 0$. As a result, $Q_T^{(3)} \to 0$. ■

Proof of (CLT2). To show (CLT2), it suffices to verify that

$$\sum_{t=1}^{T}\big(\mathbb{E}[Z_{t,T}^2|\mathscr{F}_{t-1,T}] - \mathbb{E}[Z_{t,T}^2]\big) \xrightarrow{P} 0 \qquad \textbf{(A.6)}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

$$\sum_{t=1}^{T} \mathbb{E}[Z_{t,T}^2] \to V_1. \tag{A.7}$$

We first prove (A.6). Letting $\mathscr{S}$ be the set of index tuples

$$\mathscr{S} = \big\{(t,t',s,s',s'',s''') \;\big|\; 1 \le t,t',s,s',s'',s''' \le T \text{ with } s,s' < t \text{ and } s'',s''' < t'\big\},$$

we can write

$$\mathbb{E}\left(\sum_{t=1}^{T} \big(\mathbb{E}[Z_{t,T}^2|\mathscr{F}_{t-1,T}] - \mathbb{E}[Z_{t,T}^2]\big)\right)^2 = \frac{16\lambda_1^4 h^{2d+2}}{T^4} \sum_{(t,t',s,s',s'',s''')\in\mathscr{S}} \iint \Gamma_{t,t',s,s',s'',s'''}(\underline{u},\underline{x}) q(\underline{u},\underline{x}) d\underline{u} d\underline{x},$$

where we use the shorthands

$$q(\underline{u},\underline{x}) = \frac{\pi(x)\dots\pi(x''')\omega(u)\dots\omega(u''')}{f^2(u,x)\dots f^2(u''',x''')}$$

and

$$\begin{aligned}\Gamma_{t,t',s,s',s'',s'''}(\underline{u},\underline{x}) = {} & \mathbb{E}\big[W_{t,T}(u,u',x,x') W_{s,T}(u,x) W_{s',T}(u',x') \\ & \times W_{t',T}(u'',u''',x'',x''') W_{s'',T}(u'',x'') W_{s''',T}(u''',x''')\big] \\ & - \mathbb{E}\big[W_{t,T}(u,u',x,x') W_{s,T}(u,x) W_{s',T}(u',x')\big] \\ & \times \mathbb{E}\big[W_{t',T}(u'',u''',x'',x''') W_{s'',T}(u'',x'') W_{s''',T}(u''',x''')\big]\end{aligned}$$

with

$$\begin{aligned} W_{t,T}(u,u',x,x') &= K_{u,t,T} K_{u',t,T} K_{x,t,T} K_{x',t,T} \sigma^2\Big(\frac{t}{T}, X_{t,T}\Big) \\ W_{s,T}(u,x) &= K_{u,s,T} K_{x,s,T} \varepsilon_{s,T}. \end{aligned}$$

Similarly to the proof of (CLT1), we partition the index set $\mathscr{S}$ into different subsets. By symmetry considerations, we can restrict attention to subsets with $t \ge t'$, $s \ge s'$, and $s'' \ge s'''$. Thus, the following cases are possible:

$$(A)\ t \ge t' \ge s,s',s'',s''' \quad (B)\ t > s > t' \ge s',s'',s''' \quad (C)\ t > s,s' \ge t' > s'',s'''.$$

Let $\mathscr{S}^{(i)} \subseteq \mathscr{S}$ be the subset of index combinations which fulfill the requirements of case $i \in \{A,B,C\}$ and write

$$Q_T^{(i)} = \frac{16\lambda_1^4 h^{2d+2}}{T^4} \sum_{(t,t',s,s',s'',s''')\in\mathscr{S}^{(i)}} \iint \Gamma_{t,t',s,s',s'',s'''}(\underline{u},\underline{x}) q(\underline{u},\underline{x}) d\underline{u} d\underline{x}.$$

To complete the proof of (A.6), we establish that $Q_T^{(i)} \to 0$ for $i = A,B,C$. In what follows, we give the details of the proof for case (A). The other cases can be handled by analogous arguments.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

To show that $Q_T^{(A)} \to 0$, we first split up the case (A) into the following subcases:

(A_1) $t \geq t' \geq s, s', s'', s'''$ and the indices s, s', s'', s''' are all different

(A_2) $t \geq t' \geq s, s', s'', s'''$ and exactly two of the indices s, s', s'', s''' are the same

(A_3) $t \geq t' \geq s, s', s'', s'''$ and exactly three of the indices s, s', s'', s''' are the same

(A_4) $t \geq t' \geq s, s', s'', s'''$ and the indices s, s', s'', s''' are all the same

(A_5) $t \geq t' \geq s, s', s'', s'''$ and the indices s, s', s'', s''' form two different pairs.

Letting $\mathscr{S}^{(A_k)}$ and $Q_T^{(A_k)}$ for $k = 1, \ldots, 5$ be defined analogously as the expressions $\mathscr{S}^{(A)}$ and $Q_T^{(A)}$, it suffices to show that $Q_T^{(A_k)} \to 0$ for $k = 1, \ldots, 5$. We start with (A_1):

(A_1) By definition of case (A_1), the term $Q_T^{(A_1)}$ only contains elements whose indices s, s', s'', s''' are all different. W.l.o.g., we can assume that $t \geq t' \geq s > s' > s'' > s'''$. As in the proof of (CLT1), we say that an index k is separated from another index k' if $|k - k'| > C_2 \log T$ for some sufficiently large constant C_2. With this notation at hand, we can distinguish between the following situations:

$(A_{1\text{-}1})$ the indices fulfill (A_1) and s, s', s'', s''' are all separated from t'

$(A_{1\text{-}2})$ the indices fulfill (A_1) and only s', s'', s''' are separated from t'

$(A_{1\text{-}3})$ the indices fulfill (A_1) and only s'', s''' are separated from t'

$(A_{1\text{-}4})$ the indices fulfill (A_1) and only s''' is separated from t'

$(A_{1\text{-}5})$ the indices fulfill (A_1) and none of s, s', s'', s''' is separated from t'

In the case $(A_{1\text{-}1})$, we can apply Davydov's inequality to get

$$\begin{aligned}
|\Gamma_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x})| = \big|&\mathrm{Cov}\big(W_{t,T}(u, u', x, x') W_{t',T}(u'', u''', x'', x'''), \\
&\qquad W_{s,T}(u, x) W_{s',T}(u', x') W_{s'',T}(u'', x'') W_{s''',T}(u''', x''')\big) \\
&- \mathrm{Cov}\big(W_{t,T}(u, u', x, x'), W_{s,T}(u, x) W_{s',T}(u', x')\big) \\
&\quad \times \mathbb{E}\big[W_{t',T}(u'', u''', x'', x''') W_{s'',T}(u'', x'') W_{s''',T}(u''', x''')\big]\big| \\
\leq\ & C T^{-C_3},
\end{aligned}$$

where C_3 is a large positive constant (which can be chosen as large as desired by picking C_2 large enough). This immediately shows that the sum of terms $\Gamma_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x})$ whose indices s, s', s'', s''' are separated from t' can be asymptotically ignored, or put differently, $Q_T^{(A_{1\text{-}1})} \to 0$. By the same token, we obtain that $Q_T^{(A_{1\text{-}k})} \to 0$ for $k = 2, 3, 4$. It thus remains to show that $Q_T^{(A_{1\text{-}5})} \to 0$. As none of the indices s, s', s'', s''' is separated from t' in the case $(A_{1\text{-}5})$, the index set $\mathscr{S}^{(A_{1\text{-}5})}$ contains at most $CT^2(\log T)^4$ terms. Using this fact and bounding the elements of $Q_T^{(A_{1\text{-}5})}$ in an obvious way give that $Q_T^{(A_{1\text{-}5})} \to 0$. As a result, we arrive at $Q_T^{(A_1)} \to 0$.

Making use of the same techniques as above, we further obtain that $Q_T^{(A_k)} \to 0$ for $k = 2, 3, 4$. To cope with the term $Q_T^{(A_5)}$, some additional arguments are needed:

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

(A_5) By definition of (A_5), the indices s, s', s'', s''' form two pairs. More precisely, two different situations are possible: (i) $s = s'$ and $s'' = s'''$ or (ii) $s = s''$ and $s' = s'''$. We partition (A_5) into these two subcases which are denoted by $(A_{5\text{-}1})$ and $(A_{5\text{-}2})$, respectively. First consider the subcase $(A_{5\text{-}1})$. To keep the notation tractable, we introduce the shorthands

$$\begin{aligned}
\psi_t &= W_{t,T}(u, u', x, x') \\
\psi_{t'} &= W_{t',T}(u'', u''', x'', x''') \\
\psi_{s,s'} &= W_{s,T}(u, x) W_{s',T}(u', x') \\
\psi_{s'',s'''} &= W_{s'',T}(u'', x'') W_{s''',T}(u''', x''').
\end{aligned}$$

With this, we can write

$$\begin{aligned}
\Gamma_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x}) &= \mathbb{E}[\psi_t \psi_{t'} \psi_{s,s'} \psi_{s'',s'''}] - \mathbb{E}[\psi_t \psi_{s,s'}] \mathbb{E}[\psi_{t'} \psi_{s'',s'''}] \\
&= \mathrm{Cov}(\psi_t, \psi_{t'} \psi_{s,s'} \psi_{s'',s'''}) - \mathrm{Cov}(\psi_t, \psi_{s,s'}) \mathbb{E}[\psi_{t'} \psi_{s'',s'''}] \\
&\quad + \mathbb{E}[\psi_t] \mathrm{Cov}(\psi_{t'}, \psi_{s,s'} \psi_{s'',s'''}) - \mathbb{E}[\psi_t] \mathbb{E}[\psi_{s,s'}] \mathrm{Cov}(\psi_{t'}, \psi_{s'',s'''}) \\
&\quad + \mathbb{E}[\psi_t] \mathbb{E}[\psi_{t'}] \mathrm{Cov}(\psi_{s,s'}, \psi_{s'',s'''}),
\end{aligned}$$

i.e., we can reformulate $\Gamma_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x})$ in terms of covariance expressions. This allows us to employ the techniques from (A_1) again. Specifically, we first consider the case in which some of the indices are separated from each other and apply Davydov's inequality to show that this case is asymptotically negligible. In a second step, we can then take for granted that these indices are not separated, which enables us to bound the number of elements in the index set. Setting up the proof along these lines, we arrive at $Q_T^{(A_{5\text{-}1})} \to 0$.
We next turn to $(A_{5\text{-}2})$. Similarly as above, we can write

$$\begin{aligned}
&\Gamma_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x}) - \mathbb{E}[\psi_t] \mathbb{E}[\psi_{t'}] \mathbb{E}[\psi_{s,s'} \psi_{s'',s'''}] \\
&= \mathrm{Cov}(\psi_t, \psi_{t'} \psi_{s,s'} \psi_{s'',s'''}) - \mathrm{Cov}(\psi_t, \psi_{s,s'}) \mathbb{E}[\psi_{t'} \psi_{s'',s'''}] \\
&\quad + \mathbb{E}[\psi_t] \mathrm{Cov}(\psi_{t'}, \psi_{s,s'} \psi_{s'',s'''})
\end{aligned}$$

as well as

$$\begin{aligned}
&\mathbb{E}[\psi_t] \mathbb{E}[\psi_{t'}] \mathbb{E}[\psi_{s,s'} \psi_{s'',s'''}] \\
&= \mathbb{E}\big[W_{t,T}(u, u', x, x')\big] \mathbb{E}\big[W_{t',T}(u'', u''', x'', x''')\big] \\
&\quad \times \mathbb{E}\big[W_{s,T}(u, x) W_{s'',T}(u'', x'')\big] \mathbb{E}\big[W_{s',T}(u', x') W_{s''',T}(u''', x''')\big] \\
&=: \Gamma^*_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x}).
\end{aligned}$$

The same reasoning as for $(A_{5\text{-}1})$ yields that

$$\begin{aligned}
Q_T^{(A_{5\text{-}2})} &= \frac{16 \lambda_I^4 h^{2d+2}}{T^4} \\
&\quad \times \sum_{(t,t',s,s',s'',s''') \in \mathscr{I}^{(A_{5\text{-}2})}} \iint \Gamma^*_{t,t',s,s',s'',s'''}(\underline{u}, \underline{x}) q(\underline{u}, \underline{x}) d\underline{u} d\underline{x} + o(1).
\end{aligned}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Recalling that $s = s''$ and $s' = s'''$ in case $(A_{5\text{-}2})$, introducing the shorthand

$$\kappa_{t,t',s,s'} = \int \cdots \int \left| K_{u,t,T} K_{u',t,T} K_{u'',t',T} K_{u''',t',T} \right.$$
$$\left. \times K_{u,s,T} K_{u'',s,T} K_{u',s',T} K_{u''',s',T} \right| \omega(u) \ldots \omega(u''') d\underline{u},$$

and w.l.o.g. assuming $d = 1$, we further obtain that

$$\begin{aligned}
&\left| \iint \Gamma^*_{t,t',s,s'}(\underline{u}, \underline{x}) q(\underline{u}, \underline{x}) d\underline{u} d\underline{x} \right| \\
&\leq C\kappa_{t,t',s,s'} \int \cdots \int \left(\int h^{-2} \left| K\left(\frac{x - x_t}{h}\right) K\left(\frac{x' - x_t}{h}\right) \right| f_{X_{t,T}}(x_t) dx_t \right) \\
&\qquad \times \left(\int h^{-2} \left| K\left(\frac{x - x_s}{h}\right) K\left(\frac{x'' - x_s}{h}\right) \right| f_{X_{s,T}}(x_s) dx_s \right) \\
&\qquad \times \left(\int h^{-2} \left| K\left(\frac{x'' - x_{t'}}{h}\right) K\left(\frac{x''' - x_{t'}}{h}\right) \right| f_{X_{t',T}}(x_{t'}) dx_{t'} \right) \\
&\qquad \times \left(\int h^{-2} \left| K\left(\frac{x' - x_{s'}}{h}\right) K\left(\frac{x''' - x_{s'}}{h}\right) \right| f_{X_{s',T}}(x_{s'}) dx_{s'} \right) \\
&\qquad \times \pi(x) \ldots \pi(x''') dx \ldots dx''' \\
&\leq C\kappa_{t,t',s,s'} \int \cdots \int \left(\int h^{-1} \left| K\left(\frac{x - x'}{h} + \varphi\right) K(\varphi) \right| d\varphi \right) \\
&\qquad \times \left(\int h^{-1} \left| K\left(\frac{x - x''}{h} + \varphi'\right) K(\varphi') \right| d\varphi' \right) \\
&\qquad \times \left(\int h^{-1} \left| K\left(\frac{x'' - x'''}{h} + \varphi''\right) K(\varphi'') \right| d\varphi'' \right) \\
&\qquad \times \left(\int h^{-1} \left| K\left(\frac{x' - x'''}{h} + \varphi'''\right) K(\varphi''') \right| d\varphi''' \right) \\
&\qquad \times \pi(x) \ldots \pi(x''') dx \ldots dx''' \\
&\leq C\kappa_{t,t',s,s'} \int \cdots \int \left(\int \left| K(\psi + \varphi) K(\varphi) \right| d\varphi \right) \\
&\qquad \times \left(\int h^{-1} \left| K\left(\frac{x' - x''}{h} + \psi + \varphi'\right) K(\varphi') \right| d\varphi' \right) \\
&\qquad \times \left(\int \left| K(\psi' + \varphi'') K(\varphi'') \right| d\varphi'' \right) \\
&\qquad \times \left(\int h^{-1} \left| K\left(\frac{x' - x''}{h} + \psi' + \varphi'''\right) K(\varphi''') \right| d\varphi''' \right) \\
&\qquad \times d\psi d\psi' \pi(x') \pi(x'') dx' dx'' \\
&\leq C\kappa_{t,t',s,s'} h^{-1}.
\end{aligned}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

In the general d-dimensional case, analogous calculations lead to the bound $C\kappa_{t,t',s,s'}h^{-d}$. With the help of this bound, it is straightforward to show that

$$\frac{16\lambda_I^4 h^{2d+2}}{T^4} \sum_{(t,t',s,s',s'',s''')\in\mathscr{S}^{(A_{5\text{-}2})}} \iint \Gamma^*_{t,t',s,s',s'',s'''}(\underline{u},\underline{x})q(\underline{u},\underline{x})d\underline{u}d\underline{x} = o(1).$$

As a result, $Q_T^{(A_5)} \to 0$.

Putting everything together, we finally arrive at

$$\mathbb{E}\left(\sum_{t=1}^{T}\left(\mathbb{E}[Z_{t,T}^2|\mathscr{F}_{t-1,T}] - \mathbb{E}[Z_{t,T}^2]\right)\right)^2 \to 0,$$

which implies (A.6).

Turning to the proof of (A.7), it holds that

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}[Z_{t,T}^2] = \frac{4\lambda_I^2 h^{d+1}}{T^2}\iint \sum_{t=1}^{T}\sum_{s,s'<t} K_{u,t,T}K_{u',t,T}K_{u,s,T}K_{u',s',T}\\ \times\left(\iint \mathbb{E}\Big[K_{x,t,T}K_{x',t,T}\varepsilon_{t,T}^2 K_{x,s,T}K_{x',s',T}\varepsilon_{s,T}\varepsilon_{s',T}\Big]\right.\\ \left.\times\frac{\pi(x)\pi(x')}{f^2(u,x)f^2(u',x')}dxdx'\right)\omega(u)\omega(u')dudu'.\end{aligned}$$

Once more taking advantage of the mixing conditions and applying Davydov's inequality, it is seen that

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\Big[Z_{t,T}^2\Big] = \frac{4\lambda_I^2 h^{d+1}}{T^2}\iint \sum_{t=1}^{T}\sum_{s<t} K_{u,t,T}K_{u',t,T}K_{u,s,T}K_{u',s,T}\\ \times\left(\iint \mathbb{E}\Big[K_{x,t,T}K_{x',t,T}\varepsilon_{t,T}^2\Big]\mathbb{E}\Big[K_{x,s,T}K_{x',s,T}\varepsilon_{s,T}^2\Big]\right.\\ \left.\times\frac{\pi(x)\pi(x')}{f^2(u,x)f^2(u',x')}dxdx'\right)\omega(u)\omega(u')dudu' + o(1).\end{aligned}$$

In a next step, we successively replace the nonstationary variables $X_{t,T}$ by the approximations $X_t(\frac{t}{T})$. To do so, we apply the techniques from the proof of Theorem 4.2 in Vogt (2012), which yield

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\Big[Z_{t,T}^2\Big] = \frac{4\lambda_I^2 h^{d+1}}{T^2}\iint \sum_{t=1}^{T}\sum_{s<t} K_{u,t,T}K_{u',t,T}K_{u,s,T}K_{u',s,T}\\ \times\left(\iint \mathbb{E}\Big[\prod_{j=1}^{d}K_h\Big(x^j - X_t^j\Big(\frac{t}{T}\Big)\Big)\prod_{j=1}^{d}K_h\Big((x')^j - X_t^j\Big(\frac{t}{T}\Big)\Big)\sigma^2\Big(\frac{t}{T},X_t\Big(\frac{t}{T}\Big)\Big)\Big]\right.\\ \times\mathbb{E}\Big[\prod_{j=1}^{d}K_h\Big(x^j - X_s^j\Big(\frac{s}{T}\Big)\Big)\prod_{j=1}^{d}K_h\Big((x')^j - X_s^j\Big(\frac{s}{T}\Big)\Big)\sigma^2\Big(\frac{s}{T},X_s\Big(\frac{s}{T}\Big)\Big)\Big]\\ \left.\times\frac{\pi(x)\pi(x')}{f^2(u,x)f^2(u',x')}dxdx'\right)\omega(u)\omega(u')dudu' + o(1).\end{aligned}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Exploiting the smoothness conditions on m, σ, and the densities f in a standard way, we can now infer that $\sum_{t=1}^{T} \mathbb{E}[Z_{t,T}^2] = V_1 + o(1)$, thus completing the proof. ■

Proof of Lemma A.5. First note that

$$\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} m_T\left(\frac{t}{T}, X_{t,T}\right) = m(u,x)\hat{f}_h(u,x) + \Lambda_{T,2}(u,x) + \Lambda_{T,3}(u,x), \tag{A.8}$$

since $m_T(w,z) = m(w,z) + c_T \Delta(w,z)$ with m satisfying the null hypothesis. This allows us to write

$$Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int B_T(u,v,x) V_T(u,v,x)\pi(x)dx \right) dudv$$
$$= 2\left(\lambda_I W_{T,1} - W_{T,2} + \lambda_I W_{T,3} - W_{T,4}\right) + o_p(1)$$

with

$$W_{T,1} = Th^{\frac{d+1}{2}} \int_I \left(\int \Lambda_{T,1}(u,x)\Lambda_{T,2}(u,x) \frac{\pi(x)}{f^2(u,x)} dx \right) du$$

$$W_{T,2} = Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda_{T,1}(u,x)\Lambda_{T,2}(v,x) \frac{\pi(x)}{f(u,x)f(v,x)} dx \right) dudv$$

$$W_{T,3} = c_T Th^{\frac{d+1}{2}} \int_I \left(\int \Lambda_{T,1}(u,x)\Lambda_{T,3}(u,x) \frac{\pi(x)}{f^2(u,x)} dx \right) du$$

$$W_{T,4} = c_T Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda_{T,1}(u,x)\Lambda_{T,3}(v,x) \frac{\pi(x)}{f(u,x)f(v,x)} dx \right) dudv,$$

where we have used the uniform convergence results from Lemmas A.1–A.3 to replace the kernel density estimates $\hat{f}_h(u,x)$ and $\hat{\tilde{f}}_h(v,x)$ by the true densities $f(u,x)$ and $f(v,x)$.

We start by analyzing $W_{T,1}$. As a first step, the term is split up into two components:

$$W_{T,1} = W_{T,1}^B + W_{T,1}^V$$

with

$$W_{T,1}^B = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t=1}^{T} K_{u,t,T}^2 K_{x,t,T}^2 \varepsilon_{t,T} \Delta_{t,T}(u,x) \frac{\pi(x)}{f^2(u,x)} dxdu$$

$$W_{T,1}^V = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s\neq t} K_{u,t,T} K_{u,s,T} K_{x,t,T} \varepsilon_{t,T} K_{x,s,T} \Delta_{s,T}(u,x) \frac{\pi(x)}{f^2(u,x)} dxdu.$$

It is easy to see that $\mathbb{E}(W_{T,1}^B)^2 \leq C\frac{h^2}{Th^{d+1}} \to 0$, which immediately implies that $W_{T,1}^B = o_p(1)$. To cope with the term $W_{T,1}^V$, we further decompose it into two parts:

$$W_{T,1}^V = W_{T,1,a}^V + W_{T,1,b}^V$$

with

$$W_{T,1,a}^V = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s\neq t} K_{u,t,T} K_{u,s,T} K_{x,t,T} \varepsilon_{t,T}$$
$$\times \left(K_{x,s,T}\Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T}\Delta_{s,T}(u,x)]\right) \frac{\pi(x)}{f^2(u,x)} dxdu$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

$$W_{T,1,b}^V = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t=1}^T \sum_{s \neq t} K_{u,t,T} K_{u,s,T} K_{x,t,T} \varepsilon_{t,T} \times \mathbb{E}[K_{x,s,T} \Delta_{s,T}(u,x)] \frac{\pi(x)}{f^2(u,x)} dx du.$$

The second moment of $W_{T,1,a}^V$ is given by the expression

$$\begin{aligned}\mathbb{E}\left(W_{T,1,a}^V\right)^2 = T^2 h^{d+1} \int_{I \times I} \iint \frac{1}{T^4} \sum_{t,t'=1}^T \sum_{s \neq t, s' \neq t'} K_{u,t,T} K_{u,s,T} K_{u',t',T} K_{u',s',T} \\ \times \mathbb{E}\Big[K_{x,t,T} \varepsilon_{t,T} \big(K_{x,s,T} \Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T} \Delta_{s,T}(u,x)] \big) \\ K_{x',t',T} \varepsilon_{t',T} \big(K_{x',s',T} \Delta_{s',T}(u',x') - \mathbb{E}[K_{x',s',T} \Delta_{s',T}(u',x')] \big) \Big] \\ \times \frac{\pi(x)\pi(x')\omega(u)\omega(u')}{f^2(u,x) f^2(u',x')} dx dx' du du'.\end{aligned}$$

Using similar techniques as in the proofs of (CLT1) and (CLT2), this expression can be shown to converge to zero, which yields that $W_{T,1,a}^V = o_p(1)$. We next turn to the term $W_{T,1,b}^V$. It holds that

$$W_{T,1,b}^V = Th^{\frac{d+1}{2}} \int_I \Big(\int \Lambda_{T,1}(u,x) \mathbb{E}\big[\Lambda_{T,2}(u,x)\big] \frac{\pi(x)}{f^2(u,x)} dx \Big) du + o_p(1). \tag{A.9}$$

Replacing the occurrences of $X_{t,T}$ in $\mathbb{E}\big[\Lambda_{T,2}(u,x)\big]$ by the approximating variables $X_t(\frac{t}{T})$ analogously as in the proof of (A.3) yields that

$$\begin{aligned}\mathbb{E}\big[\Lambda_{T,2}(u,x)\big] = \frac{1}{T} \sum_{t=1}^T K_{u,t,T} \int \prod_{j=1}^d K_h(x^j - z^j) \\ \times \Big(m\Big(\frac{t}{T}, z\Big) - m(u,x) \Big) f\Big(\frac{t}{T}, z\Big) dz + O\Big(\frac{1}{Th^d}\Big)\end{aligned}$$

uniformly in u and x. Next note that

$$\begin{aligned}&\frac{1}{T} \sum_{t=1}^T K_{u,t,T} \Big(m\Big(\frac{t}{T}, z\Big) - m(u,x) \Big) f\Big(\frac{t}{T}, z\Big) \\ &= \int K_h(u-w) \big(m(w,z) - m(u,x) \big) f(w,z) dw + O\Big(\frac{1}{Th^2}\Big)\end{aligned} \tag{A.10}$$

uniformly in u, x, and z, which follows upon regarding the term on the left-hand side as a Riemann sum. Using (A.10), we get

$$\begin{aligned}\mathbb{E}\big[\Lambda_{T,2}(u,x)\big] = \iint K_h(u-w) \prod_{j=1}^d K_h(x^j - z^j) \\ \times \big(m(w,z) - m(u,x) \big) f(w,z) dw dz + O\Big(\frac{1}{Th^d} + \frac{1}{Th^2}\Big).\end{aligned}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Finally, exploiting the smoothness conditions on m and f together with the properties of the higher-order kernels, standard arguments yield that

$$\mathbb{E}\big[\Lambda_{T,2}(u,x)\big] = O\left(h^r + \frac{1}{Th^d} + \frac{1}{Th^2}\right) \tag{A.11}$$

uniformly in u and x. We thus arrive at

$$\begin{aligned}
&\mathbb{E}\Big(Th^{\frac{d+1}{2}} \int_I \int \Lambda_{T,1}(u,x)\mathbb{E}\big[\Lambda_{T,2}(u,x)\big]\frac{\pi(x)}{f^2(u,x)}dxdu\Big)^2 \\
&\quad = T^2h^{d+1} \int_{I\times I} \iint \Big(\frac{1}{T^2}\sum_{t=1}^{T} K_{u,t,T}K_{u',t,T}\mathbb{E}\Big[K_{x,t,T}K_{x',t,T}\varepsilon_{t,T}^2\Big]\Big) \\
&\qquad \times \mathbb{E}\big[\Lambda_{T,2}(u,x)\big]\mathbb{E}\big[\Lambda_{T,2}(u',x')\big]\frac{\pi(x)\pi(x')}{f^2(u,x)f^2(u',x')}dxdx'dudu' \\
&\quad = O\Big(Th^{d+1}\Big(h^r + \frac{1}{Th^d} + \frac{1}{Th^2}\Big)^2\Big) = o(1).
\end{aligned}$$

Recalling (A.9), this implies that $W_{T,1,b}^V = o_p(1)$. As a result,

$$W_{T,1} = o_p(1)$$

and analogously $W_{T,2} = o_p(1)$. Similar arguments can be used to show that $W_{T,3} = o_p(1)$ as well as $W_{T,4} = o_p(1)$. This completes the proof. ■

Proof of Lemma A.6. Using (A.8), recalling that $c_T = (Th^{(d+1)/2})^{-1/2}$ and applying the uniform convergence results of Lemmas A.1–A.3 to replace the kernel densities $\hat{f}_h(u,x)$ and $\hat{f}_h(v,x)$ by the true densities $f(u,x)$ and $f(v,x)$, we obtain

$$\begin{aligned}
&Th^{\frac{d+1}{2}} \int_I \Big(\int B_T^2(u,v,x)\pi(x)dx\Big)du \\
&\qquad = 2\big(\lambda_I P_{T,1} - P_{T,2}\big) + 4\big(\lambda_I Q_{T,1} - Q_{T,2}\big) + 2\big(\lambda_I R_{T,1} - R_{T,2}\big) + o_p(1)
\end{aligned}$$

with

$$\begin{aligned}
P_{T,1} &= Th^{\frac{d+1}{2}} \int_I \left(\int \Lambda_{T,2}^2(u,x)\frac{\pi(x)}{f^2(u,x)}dx\right)du \\
P_{T,2} &= Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda_{T,2}(u,x)\Lambda_{T,2}(v,x)\frac{\pi(x)}{f(u,x)f(v,x)}dx\right)dudv \\
Q_{T,1} &= c_T Th^{\frac{d+1}{2}} \int_I \left(\int \Lambda_{T,2}(u,x)\Lambda_{T,3}(u,x)\frac{\pi(x)}{f^2(u,x)}dx\right)du \\
Q_{T,2} &= c_T Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda_{T,2}(u,x)\Lambda_{T,3}(v,x)\frac{\pi(x)}{f(u,x)f(v,x)}dx\right)dudv \\
R_{T,1} &= \int_I \left(\int \Lambda_{T,3}^2(u,x)\frac{\pi(x)}{f^2(u,x)}dx\right)du \\
R_{T,2} &= \int_{I\times I} \left(\int \Lambda_{T,3}(u,x)\Lambda_{T,3}(v,x)\frac{\pi(x)}{f(u,x)f(v,x)}dx\right)dudv.
\end{aligned}$$

Straightforward calculations yield that

$$2\big(\lambda_I R_{T,1} - R_{T,2}\big) = \int_{I\times I} \left(\int [\Delta(u,x) - \Delta(v,x)]^2\pi(x)dx\right)dudv + o_p(1).$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

To analyze the term $P_{T,1}$, we decompose it according to

$$P_{T,1} = P_{T,1,a} + 2P_{T,1,b} + P_{T,1,c}$$

with

$$P_{T,1,a} = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t,s=1}^{T} K_{u,t,T} K_{u,s,t} \big(K_{x,t,T} \Delta_{t,T}(u,x) - \mathbb{E}[K_{x,t,T} \Delta_{t,T}(u,x)]\big) \times \big(K_{x,s,T} \Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T} \Delta_{s,T}(u,x)]\big) \frac{\pi(x)}{f^2(u,x)} dxdu$$

$$P_{T,1,b} = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t,s=1}^{T} K_{u,t,T} K_{u,s,t} \mathbb{E}[K_{x,t,T} \Delta_{t,T}(u,x)] \times \big(K_{x,s,T} \Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T} \Delta_{s,T}(u,x)]\big) \frac{\pi(x)}{f^2(u,x)} dxdu$$

$$P_{T,1,c} = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t,s=1}^{T} K_{u,t,T} K_{u,s,t} \mathbb{E}[K_{x,t,T} \Delta_{t,T}(u,x)] \times \mathbb{E}[K_{x,s,T} \Delta_{s,T}(u,x)] \frac{\pi(x)}{f^2(u,x)} dxdu.$$

In what follows, these three terms are considered separately. To start with, the same arguments as used to analyze the term $W^V_{T,1,b}$ in Lemma A.5 yield that $P_{T,1,b} = o_p(1)$. Moreover, it holds that

$$P_{T,1,c} = Th^{\frac{d+1}{2}} \int_I \int \big(\mathbb{E}\big[\Lambda_{T,2}(u,x)\big]\big)^2 \frac{\pi(x)}{f^2(u,x)} dxdu.$$

Together with (A.11), this immediately implies that $P_{T,1,c} = o_p(1)$ as well. To handle the term $P_{T,1,a}$, we split it up into two parts,

$$P_{T,1,a} = P^B_{T,1,a} + P^V_{T,1,a},$$

where

$$P^B_{T,1,a} = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t=1}^{T} K^2_{u,t,T} \big(K_{x,t,T} \Delta_{t,T}(u,x) - \mathbb{E}[K_{x,t,T} \Delta_{t,T}(u,x)]\big)^2 \frac{\pi(x)}{f^2(u,x)} dxdu$$

$$P^V_{T,1,a} = Th^{\frac{d+1}{2}} \int_I \int \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s \neq t} K_{u,t,T} K_{u,s,t} \big(K_{x,t,T} \Delta_{t,T}(u,x) - \mathbb{E}[K_{x,t,T} \Delta_{t,T}(u,x)]\big) \times \big(K_{x,s,T} \Delta_{s,T}(u,x) - \mathbb{E}[K_{x,s,T} \Delta_{s,T}(u,x)]\big) \frac{\pi(x)}{f^2(u,x)} dxdu.$$

Applying the techniques from the proofs of (CLT1) and (CLT2), we can show that $\mathbb{E}(P^B_{T,1,a} - \mathbb{E}[P^B_{T,1,a}])^2 \to 0$ and thus

$$P^B_{T,1,a} - \mathbb{E}[P^B_{T,1,a}] = o_p(1).$$

Slightly varying the arguments for the proof of (A.11), we can further establish that

$$\mathbb{E}[P^B_{T,1,a}] = \lambda_I^{-1} B_{T,3},$$

which in turn yields that $P^B_{T,1,a} = \lambda_I^{-1} B_{T,3} + o_p(1)$. Finally, once more applying the techniques used to derive (CLT1) and (CLT2), we obtain that $P^V_{T,1,a} = o_p(1)$. As a result, $P_{T,1,a} = \lambda_I^{-1} B_{T,3} + o_p(1)$.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Putting everything together, we arrive at

$$P_{T,1} = \lambda_I^{-1} B_{T,3} + o_p(1).$$

Slightly modifying the above arguments, we further get that $P_{T,2} = B_{T,4} + o_p(1)$ as well as $Q_{T,1} = o_p(1)$ and $Q_{T,2} = o_p(1)$. ■

Proof of Theorem 4.4

The proof mimics the arguments for Theorems 4.1–4.3 in the bootstrap world. We write

$$S_T^* = Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \left[V_T^*(u,v,x) + B_T^*(u,v,x) \right]^2 \pi(x) dx \right) du dv$$

with

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

$$B_T^*(u,v,x) = \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \tilde{m}_g(X_{t,T}) \Big/ \hat{f}_h(u,x) - \frac{1}{T}\sum_{t=1}^{T} K_{v,t,T} K_{x,t,T} \tilde{m}_g(X_{t,T}) \Big/ \hat{f}_h(v,x)$$

and show that under the conditions of Theorem 4.4,

$$\textbf{(B1)} \qquad Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int (V_T^*(u,v,x))^2 \pi(x) dx \right) du dv - 2(B_{T,1} - B_{T,2}) \xrightarrow{d} N(0, V)$$

conditional on the sample $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$ with probability tending to one as well as

$$\textbf{(B2)} \qquad Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int B_T^*(u,v,x) V_T^*(u,v,x) \pi(x) dx \right) du dv = o_p(1)$$

$$\textbf{(B3)} \qquad Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int (B_T^*(u,v,x))^2 \pi(x) dx \right) du dv = 2(B_{T,3} - B_{T,4}) + o_p(1).$$

Combining these three statements completes the proof. ■

For the proof of (B1)–(B3), we use the notation

$$\Lambda_{T,1}^*(u,x) = \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \varepsilon_{t,T}^*$$

$$\Lambda_{T,2}^*(u,x) = \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T} K_{x,t,T} \Delta_{t,T}^*(u,x),$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

where $\Delta^*_{t,T}(u,x) = \tilde{m}_g(X_{t,T}) - \overline{m}(x)$ and $\overline{m}(x) = \int_I m(u,x)du / \int_I du$ as defined in Section 4.3. Moreover, we let $\mathbb{P}^*(\,\cdot\,) := \mathbb{P}(\,\cdot\,|\{Y_{t,T}, X_{t,T}\}_{t=1}^T)$. Analogously, $\mathbb{E}^*[\,\cdot\,]$ and $\mathrm{Var}^*(\,\cdot\,)$ are used to denote the expectation and variance conditional on the sample $\{Y_{t,T}, X_{t,T}\}_{t=1}^T$.

Proof of (B1). Letting

$$U_T^* = Th^{\frac{d+1}{2}} \int_{I\times I} \Big(\int (V_T^*(u,v,x))^2 \pi(x) dx \Big) du dv,$$

we can write $U_T^* = 2(U_{T,1}^* - U_{T,2}^*)$ with

$$U_{T,1}^* = \lambda_I Th^{\frac{d+1}{2}} \int_I \left(\int (\Lambda_{T,1}^*(u,x))^2 \frac{\pi(x)}{\hat{f}_h^2(u,x)} dx \right) du$$

$$U_{T,2}^* = Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda_{T,1}^*(u,x) \Lambda_{T,1}^*(v,x) \frac{\pi(x)}{\hat{f}_h(u,x)\hat{f}_h(v,x)} dx \right) du dv.$$

Similarly as in Lemma A.4, it suffices to show that

$$U_{T,1}^* - B_{T,1} \xrightarrow{d} N(0, V_1) \quad \text{cond. on the data with prob. tending to one} \tag{A.12}$$

$$U_{T,2}^* - B_{T,2} = o_p(1). \tag{A.13}$$

Note that (A.12) is equivalent to $\mathbb{P}^*(U_{T,1}^* - B_{T,1} \le x) \xrightarrow{P} \Phi(x)$ pointwise for each $x \in \mathbb{R}$, where Φ is a Gaussian distribution function with mean zero and variance V_1. In what follows, we restrict attention to the proof of (A.12). (A.13) follows by simpler but similar arguments. To prove (A.12), we decompose $U_{T,1}^*$ according to

$$U_{T,1}^* = U_{T,1}^{B,*} + U_{T,1}^{V,*} + o_p(1)$$

with

$$U_{T,1}^{B,*} = \frac{\lambda_I h^{\frac{d+1}{2}}}{T} \int_I \sum_{t=1}^T K_{u,t,T}^2 \left(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)} dx \right) (\varepsilon_{t,T}^*)^2 du$$

$$U_{T,1}^{V,*} = \frac{\lambda_I h^{\frac{d+1}{2}}}{T} \int_I \sum_{t=1}^T \sum_{s\neq t} K_{u,t,T} K_{u,s,T} \left(\int K_{x,t,T} K_{x,s,T} \frac{\pi(x)}{f^2(u,x)} dx \right) \varepsilon_{t,T}^* \varepsilon_{s,T}^* du,$$

where we have replaced the kernel density $\hat{f}_h(u,x)$ by the true density $f(u,x)$. To do so, we have exploited the fact 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 can be shown by using the uniform convergence results from Lemmas A.1–A.3 and noting that the array $\{\varepsilon_{t,T} \cdot \eta_t\}$ has the same mixing properties as $\{\varepsilon_{t,T}\}$ (cp. Theorem 5.2 in Bradley, 2005). In the sequel, we prove that

$$U_{T,1}^{B,*} = B_{T,1} + o_p(1) \tag{A.14}$$

$$\mathbb{P}^*(U_{T,1}^{V,*} \le x) \xrightarrow{P} \Phi(x). \tag{A.15}$$

Combining (A.14) and (A.15) immediately yields that $\mathbb{P}^*(U_{T,1}^* - B_{T,1} \le x) \xrightarrow{P} \Phi(x)$, thus completing the proof of (A.12). ■

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

Proof of (A.14). Noting that

$$\hat{\varepsilon}_{t,T}^2 = \varepsilon_{t,T}^2 + 2\varepsilon_{t,T}\left[m\left(\frac{t}{T}, X_{t,T}\right) - \hat{m}_h\left(\frac{t}{T}, X_{t,T}\right)\right] + \left[m\left(\frac{t}{T}, X_{t,T}\right) - \hat{m}_h\left(\frac{t}{T}, X_{t,T}\right)\right]^2, \tag{A.16}$$

we have

$$\begin{aligned}
\mathbb{E}^*\left[U_{T,1}^{B,*}\right] &= \frac{\lambda_I h^{\frac{d+1}{2}}}{T}\int_I \sum_{t=1}^T K_{u,t,T}^2\left(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)}dx\right)\hat{\varepsilon}_{t,T}^2 du \\
&= \frac{\lambda_I h^{\frac{d+1}{2}}}{T}\int_I \sum_{t=1}^T K_{u,t,T}^2\left(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)}dx\right)\varepsilon_{t,T}^2 du + o_p(1) \\
&= \frac{\lambda_I h^{\frac{d+1}{2}}}{T}\int_I \sum_{t=1}^T K_{u,t,T}^2 \mathbb{E}\left[\left(\int K_{x,t,T}^2 \frac{\pi(x)}{f^2(u,x)}dx\right)\varepsilon_{t,T}^2\right] du + o_p(1) \\
&= \mathbb{E}[U_{T,1}^B] + o_p(1),
\end{aligned}$$

where the second equality uses the fact that $m(u,x) - \hat{m}_h(u,x) = O_p(\sqrt{\log T/Th^{d+1}} + h^r)$ uniformly in u and x and the third one follows by an application of Davydov's inequality. From Lemma A.4 we already know that $\mathbb{E}[U_{T,1}^B] = B_{T,1} + o(1)$, leaving us with

$$\mathbb{E}^*\left[U_{T,1}^{B,*}\right] = B_{T,1} + o_p(1).$$

Moreover, it is not difficult to see that $U_{T,1}^{B,*} - \mathbb{E}^*[U_{T,1}^{B,*}] = o_p(1)$. ■

Proof of (A.15). We rewrite $U_{T,1}^{V,*}$ as

$$U_{T,1}^{V,*} = \sum_{t,s=1}^T w_{s,t,T}^*$$

with

$$w_{s,t,T}^* = \begin{cases} \frac{\lambda_I h^{\frac{d+1}{2}}}{T}\int_I K_{u,t,T}K_{u,s,T}\left(\int K_{x,t,T}K_{x,s,T}\frac{\pi(x)}{f^2(u,x)}dx\right)\varepsilon_{t,T}^*\varepsilon_{s,T}^* du & \text{for } t \neq s \\ 0 & \text{otherwise.} \end{cases}$$

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 (1987) on quadratic forms to show (A.15). In particular, it suffices to show that the following three conditions are satisfied (see Theorem 2.1 in de Jong, 1987):

(CLT1*) $\mathrm{Var}^*\left(U_{T,1}^{V,*}\right) \xrightarrow{P} V_1$.

(CLT2*) $\mathrm{Var}^*\left(U_{T,1}^{V,*}\right)^{-1} \max_{1\le s\le T}\sum_{t=1}^T \mathrm{Var}^*(w_{s,t,T}^*) \xrightarrow{P} 0$.

(CLT3*) $\mathrm{Var}^*\left(U_{T,1}^{V,*}\right)^{-2} \mathbb{E}^*(U_{T,1}^{V,*})^4 \xrightarrow{P} 3$.

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

To show (CLT1*), we proceed similarly to the proof of (A.14). The details are omitted. For the proof of (CLT2*), note that

$$\max_{1\le s\le T}\sum_{t=1}^{T}\mathrm{Var}^*(w^*_{s,t,T}) \le C\int_I\iint \max_{1\le s\le T}\Big(h^d|K_{u,s,T}|\big(\textstyle\int_I|K_{u',s,T}|du'\big)|K_{x,s,t}K_{x',s,T}|\hat\varepsilon^2_{s,T}\Big)$$
$$\times\Big(\frac{1}{T^2}\sum_{t=1}^{T}|K_{u,t,T}K_{x,t,T}K_{x',t,T}|\hat\varepsilon^2_{t,T}\Big)\pi(x)\pi(x')dxdx'du$$

with a sufficient large positive constant C. Using (A.16) together with the fact that $\max_{1\le s\le T}\varepsilon^2_{s,T} = O_p(T^{2/\nu})$ for $\nu = 8+\delta$, we obtain that

$$\max_{1\le s\le T}\Big(h^d|K_{u,s,T}|\big(\textstyle\int_I|K_{u',s,T}|du'\big)|K_{x,s,t}K_{x',s,T}|\hat\varepsilon^2_{s,T}\Big) = O_p\Big(\frac{T^{2/\nu}}{h^{d+1}}\Big). \tag{A.17}$$

Moreover, it is easily seen that

$$\frac{1}{T^2}\int_I\sum_{t=1}^{T}|K_{u,t,T}|\Big(\iint|K_{x,t,T}K_{x',t,T}|\pi(x)\pi(x')dxdx'\Big)\hat\varepsilon^2_{t,T}du$$
$$= O_p\Big(\frac{1}{T}\Big). \tag{A.18}$$

Combining (A.17) and (A.18), we arrive at

$$\max_{1\le s\le T}\sum_{t=1}^{T}\mathrm{Var}^*(w^*_{s,t,T}) = O_p\left(\frac{1}{T^{1-2/\nu}h^{d+1}}\right) = o_p(1),$$

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

$$\mathbb{E}^*\Big(U^{V,*}_{T,1}\Big)^4 = \sum_{t\ne s,t'\ne s',t''\ne s'',t'''\ne s'''}\mathbb{E}^*\Big[w^*_{s,t,T}w^*_{s',t',T}w^*_{s'',t'',T}w^*_{s''',t''',T}\Big]$$
$$= 12\sum_{t_1\ne t_2\ne t_3\ne t_4}\mathbb{E}^*\Big[\Big(w^*_{t_1,t_2,T}\Big)^2(w^*_{t_3,t_4,T})^2\Big] + 8\sum_{t_1\ne t_2}\mathbb{E}^*\Big[(w^*_{t_1,t_2,T})^4\Big]$$
$$+48\sum_{t_1\ne t_2\ne t_3\ne t_4}\mathbb{E}^*\Big[w^*_{t_1,t_2,T}w^*_{t_2,t_3,T}w^*_{t_3,t_4,T}w^*_{t_4,t_1,T}\Big]$$
$$+192\sum_{t_1\ne t_2\ne t_3}\mathbb{E}^*\Big[w^*_{t_1,t_2,T}(w^*_{t_1,t_3,T})^2w^*_{t_2,t_3,T}\Big]$$
$$+48\sum_{t_1\ne t_2\ne t_3}\mathbb{E}^*\Big[(w^*_{t_1,t_2,T})^2(w^*_{t_2,t_3,T})^2\Big]$$
$$=: Q^*_{T,1}+Q^*_{T,2}+Q^*_{T,3}+Q^*_{T,4}+Q^*_{T,5}.$$

Exploiting the mixing conditions on the model variables yields that $Q^*_{T,i} = o_p(1)$ for $i = 2,\dots,5$. Below, we give the proof for $Q^*_{T,3}$ which is the most difficult term to handle. As a result, we obtain that

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

$$\mathbb{E}^*\left(U_{T,1}^{V,*}\right)^4 = 12 \sum_{t_1\neq t_2\neq t_3\neq t_4} \mathbb{E}^*\left[(w_{t_1,t_2,T}^*)^2 (w_{t_3,t_4,T}^*)^2\right] + o_p(1).$$

Noting that $\mathrm{Var}^*(U_{T,1}^{V,*}) = 2\sum_{t_1\neq t_2} E^*(w_{t_1,t_2,T}^*)^2$, it is now easy to see that $\mathbb{E}^*(U_{T,1}^{V,*})^4 = 3\mathrm{Var}^*(U_{T,1}^{V,*})^2 + o_p(1)$. This completes the proof of (CLT3*).

We now provide the details of the proof that $Q_{T,3}^* = o_p(1)$. Using the shorthands

$$\hat\phi_{t_1} = |K_{x,t_1,T} K_{x''',t_1,T}| \hat\varepsilon_{t_1,T}^2$$
$$\hat\phi_{t_2} = |K_{x,t_2,T} K_{x',t_2,T}| \hat\varepsilon_{t_2,T}^2$$
$$\hat\phi_{t_3} = |K_{x',t_3,T} K_{x'',t_3,T}| \hat\varepsilon_{t_3,T}^2$$
$$\hat\phi_{t_4} = |K_{x'',t_4,T} K_{x''',t_4,T}| \hat\varepsilon_{t_4,T}^2$$

as well as

$$\kappa_{t_1,t_2,t_3,t_4} = |K_{u,t_1,T} K_{u,t_2,T} K_{u',t_2,T} K_{u',t_3,T} \times K_{u'',t_3,T} K_{u'',t_4,T} K_{u''',t_4,T} K_{u''',t_1,T}|,$$

it holds that

$$|Q_{T,3}^*| \le \frac{C\lambda_I^4 h^{2d+2}}{T^4} \sum_{t_1\neq t_2\neq t_3\neq t_4} \int\cdots\int \kappa_{t_1,t_2,t_3,t_4} \hat\phi_{t_1}\hat\phi_{t_2}\hat\phi_{t_3}\hat\phi_{t_4} \times \frac{\pi(x)\ldots\pi(x''')}{f^2(u,x)\ldots f^2(u''',x''')} d\underline{x}\,\omega(u)\ldots\omega(u''')d\underline{u}$$

with $\underline{u} = (u,u',u'',u''')$ and $\underline{x} = (x,x',x'',x''')$. Next, by (A.16), we have

$$\begin{aligned}\hat\phi_{t_1} &= |K_{x,t_1,T} K_{x''',t_1,T}| \varepsilon_{t_1,T}^2 + 2|K_{x,t_1,T} K_{x''',t_1,T}| \\ &\quad \times \left[m\left(\frac{t_1}{T}, X_{t_1,T}\right) - \hat m_h\left(\frac{t_1}{T}, X_{t_1,T}\right)\right]\varepsilon_{t_1,T} \\ &\quad + |K_{x,t_1,T} K_{x''',t_1,T}| \left[m\left(\frac{t_1}{T}, X_{t_1,T}\right) - \hat m_h\left(\frac{t_1}{T}, X_{t_1,T}\right)\right]^2 \\ &=: \phi_{t_1} + \overline{\phi}_{t_1} + \overline{\overline{\phi}}_{t_1}.\end{aligned}$$

We thus obtain

$$\begin{aligned}|Q_{T,3}^*| \le{}& \frac{C\lambda_I^4 h^{2d+2}}{T^4} \sum_{t_1\neq t_2\neq t_3\neq t_4} \int\cdots\int \kappa_{t_1,t_2,t_3,t_4} \phi_{t_1}\phi_{t_2}\phi_{t_3}\phi_{t_4} \\ &\times \frac{\pi(x)\ldots\pi(x''')}{f^2(u,x)\ldots f^2(u''',x''')} d\underline{x}\,\omega(u)\ldots\omega(u''')d\underline{u} \\ &+ \frac{C\lambda_I^4 h^{2d+2}}{T^4} \sum_{t_1\neq t_2\neq t_3\neq t_4} \int\cdots\int \kappa_{t_1,t_2,t_3,t_4} \overline{\phi}_{t_1}\phi_{t_2}\phi_{t_3}\phi_{t_4} \\ &\times \frac{\pi(x)\ldots\pi(x''')}{f^2(u,x)\ldots f^2(u''',x''')} d\underline{x}\,\omega(u)\ldots\omega(u''')d\underline{u} \\ &+ \cdots \\ =:{}& Q_{T,3,a}^* + Q_{T,3,b}^* + \cdots\end{aligned}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

In what follows, we show that $Q^*_{T,3,a} = o_p(1)$, the other terms being $o_p(1)$ by similar arguments. It holds that

$$\begin{aligned}
Q^*_{T,3,a} &= \frac{C\lambda_I^4 h^{2d+2}}{T^4} \sum_{t_1 \neq t_2 \neq t_3 \neq t_4} \int \cdots \int \kappa_{t_1,t_2,t_3,t_4} \\
&\quad \times \big(\phi_{t_1} - \mathbb{E}[\phi_{t_1}] + \mathbb{E}[\phi_{t_1}]\big)\big(\phi_{t_2} - \mathbb{E}[\phi_{t_2}] + \mathbb{E}[\phi_{t_2}]\big) \\
&\quad \times \big(\phi_{t_3} - \mathbb{E}[\phi_{t_3}] + \mathbb{E}[\phi_{t_3}]\big)\big(\phi_{t_4} - \mathbb{E}[\phi_{t_4}] + \mathbb{E}[\phi_{t_4}]\big) \\
&\quad \times \frac{\pi(x)\ldots\pi(x''')}{f^2(u,x)\ldots f^2(u''',x''')} d\underline{x}\omega(u)\ldots\omega(u''')d\underline{u} \\
&= \frac{C\lambda_I^4 h^{2d+2}}{T^4} \sum_{t_1 \neq t_2 \neq t_3 \neq t_4} \int \cdots \int \kappa_{t_1,t_2,t_3,t_4} \\
&\quad \times \Big\{ \mathbb{E}[\phi_{t_1}]\mathbb{E}[\phi_{t_2}]\mathbb{E}[\phi_{t_3}]\mathbb{E}[\phi_{t_4}] \\
&\qquad + (\phi_{t_1} - \mathbb{E}[\phi_{t_1}])\mathbb{E}[\phi_{t_2}]\mathbb{E}[\phi_{t_3}]\mathbb{E}[\phi_{t_4}] \\
&\qquad + (\phi_{t_1} - \mathbb{E}[\phi_{t_1}])(\phi_{t_2} - \mathbb{E}[\phi_{t_2}])\mathbb{E}[\phi_{t_3}]\mathbb{E}[\phi_{t_4}] \\
&\qquad + (\phi_{t_1} - \mathbb{E}[\phi_{t_1}])(\phi_{t_2} - \mathbb{E}[\phi_{t_2}])(\phi_{t_3} - \mathbb{E}[\phi_{t_3}])\mathbb{E}[\phi_{t_4}] \\
&\qquad + (\phi_{t_1} - \mathbb{E}[\phi_{t_1}])(\phi_{t_2} - \mathbb{E}[\phi_{t_2}])(\phi_{t_3} - \mathbb{E}[\phi_{t_3}])(\phi_{t_4} - \mathbb{E}[\phi_{t_4}]) + \cdots \Big\} \\
&\quad \times \frac{\pi(x)\ldots\pi(x''')}{f^2(u,x)\ldots f^2(u''',x''')} d\underline{x}\omega(u)\ldots\omega(u''')d\underline{u} \\
&=: Q_1 + \cdots + Q_5 + \cdots
\end{aligned}$$

We have that $Q_1 \leq Ch^d$. Moreover, exploiting the mixing conditions by means of Davydov's inequality similarly as in the proofs of (CLT1) and (CLT2), we can show that $\mathbb{E}[Q_i^2] = o(1)$ for $i = 2, \ldots, 5$. The other terms contained in $Q^*_{T,3,a}$ can be handled in the same way. Note that to apply Davydov's inequality in this context, we require the moment condition $\mathbb{E}[\varepsilon_t^{8+\delta}] < \infty$ to hold for some small $\delta > 0$. Proceeding along these lines, we finally arrive at $Q^*_{T,3,a} = o_p(1)$. ■

Proof of (B2). It holds that

$$Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int B^*_T(u,v,x)V^*_T(u,v,x)\pi(x)dx \right) dudv = 2\big(\lambda_I W^*_{T,1} - W^*_{T,2}\big)$$

with

$$W^*_{T,1} = Th^{\frac{d+1}{2}} \int_I \left(\int \Lambda^*_{T,1}(u,x)\Lambda^*_{T,2}(u,x) \frac{\pi(x)}{\hat{f}_h^2(u,x)} dx \right) du$$

$$W^*_{T,2} = Th^{\frac{d+1}{2}} \int_{I\times I} \left(\int \Lambda^*_{T,1}(u,x)\Lambda^*_{T,2}(v,x) \frac{\pi(x)}{\hat{f}_h(u,x)\hat{f}_h(v,x)} dx \right) dudv.$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

We first consider the term $W^*_{T,1}$ in more detail. Writing

$$\begin{aligned}\Lambda^*_{T,1}(u,x) &= \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}K_{x,t,T}\left(m\left(\frac{t}{T},X_{t,T}\right)-\hat{m}_h\left(\frac{t}{T},X_{t,T}\right)\right)\eta_t \\ &\quad+\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}K_{x,t,T}\varepsilon_{t,T}\eta_t \\ &=: \Lambda^*_{T,1,a}(u,x)+\Lambda^*_{T,1,b}(u,x)\end{aligned}$$

and

$$\begin{aligned}\Lambda^*_{T,2}(u,x) &= \frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}K_{x,t,T}\left[\tilde{m}_g(X_{t,T})-\overline{m}(X_{t,T})\right] \\ &\quad+\frac{1}{T}\sum_{t=1}^{T} K_{u,t,T}K_{x,t,T}\left[\overline{m}(X_{t,T})-\overline{m}(x)\right] \\ &=: \Lambda^*_{T,2,a}(u,x)+\Lambda^*_{T,2,b}(u,x),\end{aligned}$$

we can split up $W^*_{T,1}$ into additive components according to

$$\begin{aligned}W^*_{T,1} &= Th^{\frac{d+1}{2}}\int_I\left(\int \Lambda^*_{T,1,a}(u,x)\Lambda^*_{T,2,a}(u,x)\frac{\pi(x)}{\hat{f}^2_h(u,x)}dx\right)du \\ &\quad+Th^{\frac{d+1}{2}}\int_I\left(\int \Lambda^*_{T,1,a}(u,x)\Lambda^*_{T,2,b}(u,x)\frac{\pi(x)}{\hat{f}^2_h(u,x)}dx\right)du \\ &\quad+Th^{\frac{d+1}{2}}\int_I\left(\int \Lambda^*_{T,1,b}(u,x)\Lambda^*_{T,2,a}(u,x)\frac{\pi(x)}{\hat{f}^2_h(u,x)}dx\right)du \\ &\quad+Th^{\frac{d+1}{2}}\int_I\left(\int \Lambda^*_{T,1,b}(u,x)\Lambda^*_{T,2,b}(u,x)\frac{\pi(x)}{\hat{f}^2_h(u,x)}dx\right)du \\ &=: W^*_{T,1,aa}+W^*_{T,1,ab}+W^*_{T,1,ba}+W^*_{T,1,bb}.\end{aligned}$$

Using the uniform convergence results from Lemmas A.1–A.3 together with some straightforward additional arguments, it is not difficult to verify that

$$\Lambda^*_{T,1,a}(u,x) = O_p\left(\sqrt{\frac{\log T}{Th^{d+1}}}+h^r\right) \tag{A.19}$$

$$\Lambda^*_{T,1,b}(u,x) = O_p\left(\sqrt{\frac{\log T}{Th^{d+1}}}\right) \tag{A.20}$$

$$\Lambda^*_{T,2,a}(u,x) = O_p\left(\sqrt{\frac{\log T}{Tg^{d}}}+g^r\right) \tag{A.21}$$

$$\Lambda^*_{T,2,b}(u,x) = O_p\left(\sqrt{\frac{\log T}{Th^{d-1}}}+h^r\right) \tag{A.22}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

uniformly in u and x. Now define $\mathscr{C}_T$ to be the event that

$$\inf_{u,x} \big|\hat{f}(u,x)\big| \geq c > 0,$$

$$\sup_{u,x} \big|\hat{m}_h(u,x) - m(u,x)\big| \leq C\left(\sqrt{\frac{\log T}{Th^{d+1}}} + h^r\right),$$

$$\sup_{u,x} \big|\Lambda^*_{T,2,a}(u,x)\big| \leq C\left(\sqrt{\frac{\log T}{Tg^{d}}} + g^r\right), \quad \text{and}$$

$$\sup_{u,x} \big|\Lambda^*_{T,2,b}(u,x)\big| \leq C\left(\sqrt{\frac{\log T}{Th^{d-1}}} + h^r\right)$$

for some fixed constants $0 < c < C < \infty$. If the constant c is chosen sufficiently small and C is large enough, then the probability of $\mathscr{C}_T$ not occurring converges to zero. Put differently, $\mathbb{P}(\mathscr{C}_T^c) = o(1)$, where $\mathscr{C}_T^c$ denotes the complement of the event $\mathscr{C}_T$. This follows upon inspecting the proofs of the uniform convergence results A.1–A.3.

Keeping in mind the above remarks, we obtain that

$$\begin{aligned}\mathbb{P}\big(|W^*_{T,1,aa}| > \delta\big) &\leq \mathbb{P}\big(|W^*_{T,1,aa}| > \delta, \mathscr{C}_T\big) + \mathbb{P}(\mathscr{C}_T^c) \\ &= \mathbb{P}\big(\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}| > \delta\big) + o(1)\end{aligned}$$

for each fixed $\delta > 0$. Moreover,

$$\begin{aligned}\mathbb{P}\big(\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}| > \delta\big) &\leq \mathbb{E}\big[\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}|^2\big]/\delta^2 \\ &= \mathbb{E}\big[\mathbb{E}^*[\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}|^2]\big]/\delta^2.\end{aligned}$$

Exploiting the independence of the bootstrap residuals η_t, we obtain

$$\begin{aligned}&\mathbb{E}^*[\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}|^2] \\ &= \mathbb{I}(\mathscr{C}_T)T^2h^{d+1}\int_{I\times I}\iint \Lambda^*_{T,2,a}(u,x)\Lambda^*_{T,2,a}(u',x') \\ &\quad\times\left\{\frac{1}{T^2}\sum_{t=1}^{T} K_{u,t,T}K_{u',t,T}K_{x,t,T}K_{x',t,T}\Big(m\Big(\frac{t}{T},X_{t,T}\Big) - \hat{m}_h\Big(\frac{t}{T},X_{t,T}\Big)\Big)^2\right\} \\ &\quad\times\frac{\pi(x)\pi(x')}{\hat{f}_h^2(u,x)\hat{f}_h^2(u',x')}dxdx'\Big)dudu'\end{aligned}$$

and thus

$$\begin{aligned}&\mathbb{E}\Big[\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}|^2\Big] \\ &\quad\leq CT^2h^{d+1}\left(\sqrt{\frac{\log T}{Th^{d+1}}} + h^r\right)^2\left(\sqrt{\frac{\log T}{Tg^{d}}} + g^r\right)^2 \\ &\qquad\times\mathbb{E}\left[\int_{I\times I}\iint \frac{1}{T^2}\sum_{t=1}^{T}|K_{u,t,T}K_{u',t,T}K_{x,t,T}K_{x',t,T}|\pi(x)\pi(x')dxdx'dudu'\right] \\ &\quad= o(1),\end{aligned}$$

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press

where the inequality is by the definition of the event $\mathscr{C}_T$. As a result, $\mathbb{P}\big(\mathbb{I}(\mathscr{C}_T)|W^*_{T,1,aa}| > \delta\big) = o(1)$, which immediately yields that $W^*_{T,1,aa} = o_p(1)$. Repeating the above arguments, the terms $W^*_{T,1,ab}$, $W^*_{T,1,ba}$, and $W^*_{T,1,bb}$ can be shown to be $o_p(1)$ as well. We thus arrive at

$$W^*_{T,1} = o_p(1).$$

Analogously, we obtain that $W^*_{T,2} = o_p(1)$. This completes the proof. ■

Proof of (B3). The result easily follows by using the uniform convergence rates (A.21) and (A.22). ■

https://doi.org/10.1017/S0266466614000565 Published online by Cambridge University Press