Econometric Theory, **29**, 2013, 1–27.
doi:10.1017/S0266466612000205

NONPARAMETRIC COINTEGRATING REGRESSION WITH NNH ERRORS

QIYING WANG AND YING XIANG RACHEL WANG
The University of Sydney

This paper studies a nonlinear cointegrating regression model with nonlinear nonstationary heteroskedastic error processes. We establish uniform consistency for the conventional kernel estimate of the unknown regression function and develop a two-stage approach for the estimation of the heterogeneity generating function.

1. INTRODUCTION

It is now well known that the linear, time-invariant structure of the traditional cointegration model is often too restrictive and simplistic for modeling purposes. In most cases, even if a simple linear cointegrating relationship exists, it is unrealistic to assume this is not subject to any changes over the time span in which the variables are recorded. To address this issue, there is a large literature on parameter instability and time-varying coefficient models; see, e.g., Park and Hahn (1999), Cai, Li, and Park (2009), and Xiao (2009). Quintos and Phillips (1993) and Sun, Cai, and Li (2008) considered the problem of testing for parameter constancy. It is worth noting that despite the various modifications, these models are linear in spirit.

An alternative to time-varying coefficient models is nonlinear models. Granger and Teräsvirta (1993) offer many empirical examples where nonlinear models are desirable. Indeed, given the prevalence of nonlinear relationships in economics, it is expected that nonlinear cointegration captures the features of many long-run relationships in a more realistic manner. A typical nonlinear cointegrating regression model has the form

$$y_t = f(x_t) + \epsilon_t, \qquad t = 1, 2, \ldots, n, \tag{1.1}$$

where $\{\epsilon_t\}$ is a zero mean equilibrium error, x_t is a nonstationary regressor, and $f(\cdot)$ is an unknown function to be estimated with the observed data $\{y_t, x_t\}_{t=1}^n$.

With given observations (x_t, y_t) that may include nonstationary components, the issues on estimation and inference of the unknown $f(\cdot)$ have become

The authors thank the co-editor Professor Saikkonen and two referees for helpful comments on the original version. The main idea of this paper greatly benefited from discussions with Professor Phillips when the first author visited Singapore Management University in April 2009. Wang acknowledges partial research support from the Australian Research Council. Address correspondence to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia; e-mail: qiying.wang@sydney.edu.au.

© Cambridge University Press 2012

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

increasingly of interest in the past decade. In this regard, Phillips and Park (1998) studied nonparametric autoregression in the context of a random walk. Karlsen and Tjøstheim (2001) and Guerre (2004) studied nonparametric estimation for certain nonstationary processes in the framework of recurrent Markov chains. Karlsen, Myklebust, and Tjøstheim (2007) developed an asymptotic theory for nonparametric estimation of a time series regression equation involving stochastically nonstationary time series. Karlsen et al. (2007) address the function estimation problem for a possibly nonlinear cointegrating relation, providing an asymptotic theory of estimation and inference for nonparametric forms of cointegration. Under similar conditions and using related Markov chain methods, Schienle (2008) investigated additive nonlinear versions of (1.1) and obtained a limit theory for nonparametric regressions under smooth backfitting. More recently, Wang and Phillips (2009a, 2011) and Cai et al. (2009) considered an alternative treatment by making use of local time limit theory and, instead of recurrent Markov chains, worked with partial sum representations of the type $x_t = \sum_{j=1}^{t} \xi_j$ where ξ_j is a general linear process. In another paper, Wang and Phillips (2009b) considered the errors u_t to be serially dependent and cross correlated with the regressor x_t for small lags. For other related current work, we refer to Kasparis and Phillips (2009), Park and Phillips (1999, 2001), Bandi (2004), Gao, Maxwell, Lu, and Tjøstheim (2009a, 2009b), Choi and Saikkonen (2004, 2010), Marmer (2008), and Chen and Zhang (2010).

Although an extensive literature exists for estimation theory of $f(\cdot)$, less attention has been paid to studying the error structure, which belongs to a subclass of stochastic volatility processes. Just as it is unreasonable to assume the cointegrating relationship remains linear and unchanged throughout time, the long-run equilibrium error is unlikely to be homoskedastic over a long time span. It is well known that stock markets have variances that change substantially with time, rendering them much more volatile than many economic variables that can be used to explain their movements. In these situations, it makes sense to incorporate time heterogeneity into the error term, and this motivates us to investigate the model (1.1) with error structure

$$\epsilon_t = \sigma(x_t)\, u_t, \tag{1.2}$$

where $\sigma(\cdot)$ is a heterogeneity generating function (HGF) as defined in Park (2002) and for a filtration $\mathcal{F}_t$ to which x_{t+1} is adapted, $\{u_t,\ \mathcal{F}_t\}$ forms a martingale difference.

This incorporation idea is not new in stationary regression models. Various error structures that generate heteroskedasticity can be found in, e.g., Robinson (1987), Harvey and Robinson (1988), Hansen (1995), and Kristensen (2009). On the nonstationary front, Kim and Park (2010) investigated a linear cointegration model in which the conditional heteroskedasticity of the error is generated by a smooth deterministic function of time. The error process considered in this paper is known as a nonlinear nonstationary heteroskedastic (NNH) process and was first proposed in Park (2002), where the author examined the asymptotic behavior

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

of the sample statistics of NNH processes generated by integrable and asymptotically homogeneous functions and pointed out that NNH models offer an attractive alternative to autoregressive conditional heteroskedastic (ARCH) type volatility models because they not only demonstrate the desirable properties of volatility clustering and leptokurtosis, which are common traits in financial and economic series, but also allow for the possibility of explaining the source of volatility in terms of economic variables. In related works, the NNH error structure is incorporated into a stationary regression model and a linear cointegration model by Chung and Park (2007). Without involving the cointegrating relationship, Han and Park (2008) dealt with the NNH model with an ARCH component and Han and Zhang (2009) investigated the asymptotics for the kernel estimator of $\sigma(x)$.

The main purpose of the present paper is to estimate the HGF in the model given by (1.1) and (1.2). We consider a kernel-type estimator $\hat{\sigma}(x)$ of $\sigma(x)$. It is shown that $\hat{\sigma}(x)$ converges to $\sigma(x)$ in probability uniformly over a compact set and its limit distribution is normal under appropriate normalization. New technical mechanisms are developed to investigate the asymptotics for $\hat{\sigma}(x)$. In particular, we establish uniform consistency for the conventional kernel estimate of the unknown regression function and consider the lower bound for a class of functionals of nonstationary data over a compact set. These results are interesting in their own right and will be useful in other related research fields.

We finally remark that the two-step kernel-type estimator $\hat{\sigma}(x)$ considered in this paper is quite natural and simple. Our main result provides a theoretical framework that allows the extension from the stationary time series to a near $I(1)$ process. Uniform consistency is not only important in estimation theory but also is useful in other fields such as model specification testing. Existing studies mainly focus on the observations coming from a stationary data set. A recent work in this regard is Hansen (2008), where readers can also find a sequence of related references. Gao, Li, and Tjøstheim (2009) investigated uniform consistency of nonparametric kernel estimators for nonstationary time series. In their work, the authors considered the situation where the regressor is a recurrent Markov chain, and assumed that the error term is independent of the regressor. The setting on the model in this paper is different. We consider a near integrated process with innovations being a linear process, and we remove the restriction on the independence between the regressor and the error terms.

The remainder of the paper is organized as follows. Section 2 explains the model in more detail and defines some variables that are central to our development of the asymptotics. Section 3 outlines the asymptotic properties of the kernel estimator for both f and σ. Section 4 concludes. All the mathematical proofs are contained in Section 5.

2. THE MODEL AND ASSUMPTIONS

To estimate the HGF $\sigma(x)$, we reorganize the model driven by (1.1) and (1.2) as follows:

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

$$y_t = f(x_t) + \sigma(x_t)\, u_t, \tag{2.1}$$

$$\left[y_t - f(x_t)\right]^2 = \sigma^2(x_t) + \sigma^2(x_t)\left(u_t^2 - 1\right). \tag{2.2}$$

As $u_t^2 - 1$ may form a martingale difference, from the observations on x_t and $y_t - \hat{f}(x_t)$, where $\hat{f}(x)$ is the first-stage estimate of $f(x)$ in model (2.1), the conventional kernel estimate of $\sigma^2(x)$ in model (2.2) is given by

$$\hat{\sigma}^2(x) = \frac{\sum_{t=1}^n \left[y_t - \hat{f}(x_t)\right]^2 K\left[(x_t - x)/h\right]}{\sum_{t=1}^n K\left[(x_t - x)/h\right]}, \tag{2.3}$$

where $K(x)$ is a nonnegative real function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$.

The limit behavior of $\hat{\sigma}(x)$ depends on the choice of $\hat{f}(x)$. The present paper adopts the Nadaraya–Watson estimator defined by

$$\hat{f}(x) = \frac{\sum_{s=1}^n y_s\, K\left[\left(x_s - x\right)/h\right]}{\sum_{s=1}^n K\left[\left(x_s - x\right)/h\right]}, \tag{2.4}$$

where the kernel $K(x)$ and the bandwidth h are chosen to be the same as in (2.3) for technical convenience.

Throughout the paper, let $\{\xi_j, j \geq 1\}$ be a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k\, \epsilon_{j-k}, \tag{2.5}$$

where $\{\epsilon_j, -\infty < j < \infty\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}\epsilon_0 = 0$, $\mathbb{E}\epsilon_0^2 = 1$ and the characteristic function $\varphi(t)$ of ϵ_0 satisfies $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$. The coefficients $\phi_k, k \geq 0$, are assumed to satisfy $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$. The absolute summable condition is commonly used in the literature. We also use the following assumptions in the asymptotic development.

Assumption 1. $x_t = \rho\, x_{t-1} + \xi_t$, $(x_0 \equiv 0)$, where $\rho = 1 + \tau/n$ with τ being a constant.

Assumption 2.

(a) $\{u_t, \mathcal{F}_t, 1 \leq t \leq n\}$ is a martingale difference sequence satisfying $\mathbb{E}\left(u_t^2 | \mathcal{F}_{t-1}\right) = 1$, $\mathbb{E}\left(u_t^4 | \mathcal{F}_{t-1}\right) \to \Lambda^2$, almost surely (a.s.), where $\Lambda^2 > 1$ is a constant, and $\sup_{1 \leq t \leq n} \mathbb{E}\left(|u_t|^{4\nu} | \mathcal{F}_{t-1}\right) < \infty$ a.s. for some $\nu > 1$.
(b) x_t is adapted to $\mathcal{F}_{t-1}$, $t = 1, 2, \ldots, n$.

Assumption 3. The kernel K satisfies that $\int_{-\infty}^{\infty} K(s) ds = 1$, $K(\cdot)$ has a compact support, and for any $x, y \in R$,

$$|K(x) - K(y)| \leq C\, |x - y|,$$

where C is a positive constant.

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

Assumption 4. For a compact set $\Omega \subset R$, there exist real constants $\epsilon > 0$, $0 < \alpha, \beta \leq 1$, and $C_1, C_2 > 0$ such that $\sigma(x)$ is positive and

$$|f(y) - f(x)| \leq C_1 |y - x|^{\alpha}, \qquad |\sigma(y) - \sigma(x)| \leq C_2 |y - x|^{\beta},$$

for any $x, y \in \Omega_\epsilon$, where $\Omega_\epsilon = \{y : |y - x| \leq \epsilon, \text{where } x \in \Omega\}$.

Assumption 1 allows for both a nonstationary ($\tau = 0$) and a near nonstationary ($\tau \neq 0$) regressor and is standard in the near integration regression framework. See, e.g., Phillips (1987, 1988), Chan and Wei (1987), and Wang and Phillips (2009b). In the simplest case, it includes the example in which x_t is a random walk on i.i.d. random variables.

Assumption 2 is standard as in the stationary situation in which we impose a martingale structure so that $\text{cov}(u_{t+1}, x_t) = \mathbb{E}[x_t \mathbb{E}(u_{t+1} \mid \mathcal{F}_t)] = 0$. The restrictions that $\mathbb{E}\left(u_t^2 | \mathcal{F}_{t-1}\right) = 1$ and $\mathbb{E}\left(u_t^4 | \mathcal{F}_{t-1}\right) \to \Lambda^2$, a.s., are required because of the model structure (2.2). These restrictions are not necessary in the investigation of the asymptotics related to $\hat{f}(x)$, even for uniform convergence. In the latter situation, we use a less restricted Assumption 2*. See Theorem 3.2 in the next section.

Assumption 3 is a standard condition on $K(x)$ as in the stationary situation. The Lipschitz condition on $K(x)$ is not necessary if we only investigate the pointwise asymptotics. See Remark 3.3 for further details.

Assumption 4 requires a Lipschitz-type condition in a small neighborhood of the targeted compact set for the functionals to be estimated. This condition is quite weak and may host a wide set of functionals. Typical examples include that $f(x) = \theta_1 + \theta_2 x + \cdots + \theta_k x^{k-1}$; $f(x) = \alpha + \beta x^{\gamma}$; $f(x) = x(1 + \theta x)^{-1} I(x \geq 0)$; $f(x) = \left(\alpha + \beta e^x\right) / \left(1 + e^x\right)$. The function $\sigma(x)$ can have the same form as in $f(x)$, but we require that the x and the parameters are restricted on a space so that $\sigma(x)$ is positive. Also notice that Assumption 4 implies that

$$\sup_{x \in \Omega_\epsilon} |f(x)| < \infty \quad \text{and} \quad \sup_{x \in \Omega_\epsilon} \sigma(x) < \infty.$$

This fact will be repeatedly used in the proofs of the main results without further explanation.

3. MAIN RESULTS AND OUTLINE OF THE PROOFS

We have the following asymptotic results for the estimation of the HGF $\sigma(x)$ given in the model (2.1) and (2.2). Except when explicitly mentioned, we denote by $[a]$ the integer part of a positive constant a in what follows.

THEOREM 3.1. *Under Assumptions 1–4, for any h satisfying $h \to 0$ and $nh^{2+4/[\nu]} \to \infty$, we have*

$$\sup_{x \in \Omega} \left| \hat{\sigma}^2(x) - \sigma^2(x) \right| = O_P \left\{ \left(nh^{2+4/[\nu]}\right)^{-[\nu]/[2(2[\nu]+1)]} + h^{\min\{\alpha, \beta\}} \right\}, \tag{3.1}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

where Ω is defined as in Assumption 4. Furthermore, if in addition u_t are independent of x_t, then for any h satisfying $nh^2/\log^4 n \to \infty$ and $nh^{2+4\min\{2\alpha,\beta\}} \to 0$,

$$(T_n(z_1), T_n(z_2), \ldots, T_n(z_k)) \to_d N(0, \Sigma), \tag{3.2}$$

where $z_1, \ldots, z_k \in \Omega$ all are different,

$$T_n(z_j) = \left\{ \sum_{t=1}^{n} K\left[(x_t - z_j)/h\right] \right\}^{1/2} \left(\hat{\sigma}^2(z_j) - \sigma^2(z_j)\right),$$

and

$$\Sigma = \operatorname{diag}\left\{\sigma^4(z_1), \ldots, \sigma^4(z_k)\right\} \left(\Lambda^2 - 1\right) \int_{-\infty}^{\infty} K^2(s)ds.$$

Remark 3.1. The result (3.1) shows that $\hat{\sigma}^2(x)$ is a consistent estimator of $\sigma^2(x)$ uniformly over a compact set. If the error term u_t is assumed to be bounded by a constant, the restriction on the choice of h can be relaxed, and the convergence rate in (3.1) can be improved. Indeed, if in addition $|u_t| \le C$, then for any h satisfying $h \to 0$ and $nh^2/\log^4 n \to \infty$,

$$\sup_{x\in\Omega} \left|\hat{\sigma}^2(x) - \sigma^2(x)\right| = O_P\left[\left(nh^2\right)^{-1/4} \log n + h^{\min\{\alpha,\beta\}}\right]. \tag{3.3}$$

The result (3.2) provides a distributional property of the estimator $\hat{\sigma}^2(x)$ by imposing independence between u_t and x_t. Note that $\hat{\sigma}^2(x)$ involves the cross terms related to u_t and x_s for all $1 \le s \le n$ (see (3.4)) through $\hat{f}(x_t)$. It is not clear at the moment how to provide an accurate estimation for these cross terms, which are required in the investigation of distribution convergence, without the independence assumption.

Remark 3.2. To outline the essentials of the arguments in the proof of Theorem 3.1, we split $\hat{\sigma}^2(x) - \sigma^2(x)$ as

$$\begin{aligned}
\hat{\sigma}^2(x) - \sigma^2(x) =& \frac{\sum_{t=1}^{n} \sigma^2(x)\left(u_t^2 - 1\right) K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} \\
&+ \frac{\sum_{t=1}^{n} \left[\sigma^2(x_t) - \sigma^2(x)\right] u_t^2 K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} \\
&+ \frac{\sum_{t=1}^{n} \left(f(x_t) - \hat{f}(x_t)\right)^2 K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} \\
&+ \frac{2\sum_{t=1}^{n} \sigma(x_t) u_t \left(f(x_t) - \hat{f}(x_t)\right) K[(x_t - x)/h]}{\sum_{t=1}^{n} K[(x_t - x)/h]} \\
:=& I_{1n}(x) + I_{2n}(x) + I_{3n}(x) + I_{4n}(x).
\end{aligned} \tag{3.4}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

Although $I_{2n}(x)$ is easy to handle, the estimates of other terms over a compact set require new mechanisms. In particular, we need uniform asymptotics of $\hat{f}(x)$ on Ω_ϵ for a small ϵ. The result in this regard is presented in the following theorem, which is interesting in itself.

The following assumption is weaker than Assumption 2.

Assumption 2*.

(a) $\{u_t, \mathcal{F}_t, 1 \le t \le n\}$ is a martingale difference sequence satisfying $\sup_{1\le t\le n} \mathbb{E}(|u_t|^{2p}|\mathcal{F}_{t-1}) < \infty$ a.s. for some integer $p \ge 1$.
(b) x_t is adapted to $\mathcal{F}_{t-1}$, $t = 1, 2, ..., n$.

THEOREM 3.2. *Under Assumptions 1, 2*, 3, and 4, for any h satisfying* $h \to 0$ *and* $nh^{2+4/p} \to \infty$, *we have*

$$\sup_{x\in\Omega_\epsilon} \left|\hat{f}(x) - f(x)\right| = O_P\left\{\left(nh^{2+4/p}\right)^{-p/[2(2p+1)]} + h^{\min\{\alpha,\beta\}}\right\}, \tag{3.5}$$

where Ω_ϵ *is defined as in Assumption 4. If in addition* $|u_t| \le C$, *then, for any h satisfying* $h \to 0$ *and* $nh^2/\log^4 n \to \infty$,

$$\sup_{x\in\Omega_\epsilon} \left|\hat{f}(x) - f(x)\right| = O_P\left\{\left(nh^2\right)^{-1/4}\log n + h^{\min\{\alpha,\beta\}}\right\}. \tag{3.6}$$

Remark 3.3. A better result can be obtained if we are only interested in the pointwise asymptotics for $\hat{f}(x)$. Indeed, as in Wang and Phillips (2009a, 2009b) with minor modification, we may show that, for each fixed $x \in \Omega$,

$$\hat{f}(x) - f(x) = O_P\left\{\left(nh^2\right)^{-1/4} + h^{\min\{\alpha,\beta\}}\right\}. \tag{3.7}$$

Furthermore $\hat{f}(x)$ has an asymptotic distribution that is mixing normal, under minor additional conditions to Assumption 2*. More details are given in Wang and Phillips (2009a, 2009b).

It is interesting to note that, in a pointwise situation, the term $h^{\min\{\alpha,\beta\}}$ can be improved if we put a bias term on the left-hand side of (3.7). See Wang and Phillips (2011). It is not clear at the moment whether there are similar properties for the uniform consistency (3.5) and (3.6). The issue seems to be difficult because of its complexity. One of the referees argues that if the Nadaraya–Watson estimator $\hat{f}(x)$ is replaced by the local linear nonparametric estimator (e.g., Fan and Gijbels, 1996), defined by

$$\hat{f}^L(x) = \sum_{i=1}^n w_i y_i \Big/ \sum_{i=1}^n w_i, \qquad w_i = K[(x_i - x)/h]\left\{V_{n,2} - (X_i - x)V_{n,1}\right\},$$

where $V_{n,j} = \sum_1^n K[(x_{i-}x)/h](x_{i-}x)^j$, the order for the bias term might be improved. This conjecture raises an interesting topic for future research.

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

Remark 3.4. It is interesting to notice that, in the stationary situation, the sharp rate of convergence in (3.6) is $O_P[(\log n/nh)^{1/2}]$. See, e.g., Hansen (2008). There is an essential difference for the rate of convergence between stationary and nonstationary time series. The reason behind the difference is mainly because, in the nonstationary case, the amount of time spent by the process around any specific point is of order $\sqrt{n}$ rather than n. More explanation can be found in Remark 3.3 of Wang and Phillips (2009a).

The proof of Theorem 3.2 is quite technical. For the sake of reading convenience, we separate the key steps into the following propositions, which are interesting in themselves.

PROPOSITION 3.1. *Under Assumptions 1, 2*, and 3, for any compact set* Ω^* *and* h *satisfying* $h \to 0$ *and* $nh^{2+4/p} \to \infty$,

$$\sup_{x\in\Omega^*} \frac{\sum_{t=1}^n u_t K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]} = O_P\left[\left(nh^{2+4/p}\right)^{-p/[2(2p+1)]}\right]. \tag{3.8}$$

If in addition $|u_j| \le C$ *and* $nh^2/\log^4 n \to \infty$, *the result (3.8) can be improved to*

$$\sup_{x\in\Omega^*} \frac{\sum_{t=1}^n u_t K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]} = o_P\left[\left(nh^2\right)^{-1/4}\log n\right]. \tag{3.9}$$

PROPOSITION 3.2. *Under Assumptions 1 and 3, for any compact set* Ω^* *and* $\eta > 0$, *there exist* $n_0 > 0$ *and* $M_0 > 0$ *such that, for all* $n \ge n_0$, $M \ge M_0$, *and* h *satisfying* $h \to 0$ *and* $nh^2 \log^{-4} n \to \infty$,

$$P\left(\inf_{x\in\Omega^*} \sum_{t=1}^n K[(x_t - x)/h] \ge \sqrt{n}h/M\right) \ge 1 - \eta. \tag{3.10}$$

Remark 3.5. Gao, Li, and Tjøstheim (2009) investigated a similar uniform consistency as in Theorem 3.2 in the situation where the regressor x_t is a recurrent Markov chain and the error term u_t is independent of x_t. Not only the setting on the model but also the techniques used in the current paper are different from those in Gao et al (2009). We develop our own techniques by establishing the key result (3.10). This key result may be useful in other fields. As in Gao et al (2009), it is possible to extend our result from the compact set Ω_ϵ to a range that $|x| \le T_n$, where $T_n \to \infty$. However, this kind of extension may require quite different techniques. Because Theorem 3.2 is enough for the purpose of this paper, we leave the extension for future research.

4. CONCLUSION

In this paper, we consider a nonlinear cointegrating regression model with nonlinear nonstationary heteroskedastic (NNH) error processes. A two-stage approach

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

is proposed to find an estimator $\hat{\sigma}(x)$ of the heterogeneity generating function (HGF) $\sigma(x)$. For a wide class of $\sigma(x)$, it is shown that $\hat{\sigma}(x)$ is consistent uniformly over a compact set and has an asymptotic distribution that is normal under appropriate normalization. New technical mechanisms are developed to investigate the asymptotics for $\hat{\sigma}(x)$. In particular, we establish uniform consistency for the conventional kernel estimate of the unknown regression function and consider the lower bound for a class of functionals of nonstationary data over a compact set. These results are interesting in their own right and will be useful in other related research fields.

5. PROOFS OF MAIN RESULTS

This section provides proofs of the main results. We start with Proposition 3.2. The techniques used in the proof of this proposition also provide some insights into other main results. Throughout this section, we denote constants by $C, C_1, C_2, ...$, which may be different at each appearance.

Proof of Proposition 3.2. Without loss of generality, assume $\phi_0 \neq 0$. Otherwise the proof follows from a routine modification. Let $a < b$ be integers such that $\Omega^* \subseteq [a, b]$ and

$$y_j = a + j\,(b-a)m_n^{-1}, \qquad j = 0, 1, 2, ..., m_n, \tag{5.1}$$

where $m_n = [(nh^2)^{1/4}h^{-2}]$. Write $\mathcal{F}_t^* = \sigma(\epsilon_s, s \le t)$ for $t \ge 1$. Note that

$$\begin{aligned}
&\inf_{x\in\Omega^*} \sum_{t=1}^{n} \mathbb{E}\left(K[(x_t - x)/h] \mid \mathcal{F}_{t-1}^*\right) \\
&\quad \le \inf_{x\in\Omega^*} \sum_{t=1}^{n} K[(x_t - x)/h] \\
&\qquad + \sup_{x\in[a,b]} \left| \sum_{t=1}^{n} \left\{ K[(x_t - x)/h] - \mathbb{E}\left(K[(x_t - x)/h] \mid \mathcal{F}_{t-1}^*\right)\right\}\right|
\end{aligned} \tag{5.2}$$

and the second term on the right-hand side of (5.2) is less than

$$\begin{aligned}
&\max_{0\le j\le m_n-1} \sup_{x\in[y_j, y_{j+1}]} \sum_{t=1}^{n} \Big\{ \left|K[(x_t - x)/h] - K[(x_t - y_j)/h]\right| \\
&\qquad\qquad + \mathbb{E}\left(\left|K[(x_t - x)/h] - K[(x_t - y_j)/h]\right| \mid \mathcal{F}_{t-1}^*\right)\Big\} \\
&\quad + \max_{0\le j\le m_n} \left| \sum_{t=1}^{n} \left\{ K[(x_t - y_j)/h] - \mathbb{E}\left(K[(x_t - y_j)/h] \mid \mathcal{F}_{t-1}^*\right)\right\}\right| \\
&:= \lambda_{1n} + \lambda_{2n}.
\end{aligned}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

It is readily seen from $nh^2\log^{-4}n\to\infty$ that (3.10) will follow if we prove

$$\lambda_{1n}=O_P\left[(nh^2)^{1/4}\log n\right], \tag{5.3}$$

$$\lambda_{2n}=O_P\left[(nh^2)^{1/4}\log n\right], \tag{5.4}$$

and for any $\eta>0$, there exist $n_0>0$ and $M_0>0$ such that, for all $n\ge n_0$ and $M\ge M_0$,

$$P\left(\inf_{x\in\Omega^*}\sum_{t=1}^{n}\mathbb{E}\left(K\left[(x_t-x)/h\right]\mid\mathcal{F}_{t-1}^*\right)\ge\sqrt{n}h/M\right)\ge 1-\eta. \tag{5.5}$$

To establish (5.3)–(5.5), we start with some preliminaries. First notice that

$$\begin{aligned}x_t&=\sum_{j=1}^{t}\rho^{t-j}\,\xi_j=\sum_{j=1}^{t}\rho^{t-j}\sum_{i=-\infty}^{j}\epsilon_i\phi_{j-i}\\&=x_t^*+x_t'+\phi_0\,\epsilon_0,\end{aligned} \tag{5.6}$$

where $x_t^*=\sum_{j=1}^{t}\rho^{t-j}\sum_{i=-\infty}^{0}\epsilon_i\phi_{j-i}$ depends only on $(\ldots,\epsilon_{-1},\epsilon_0)$ and

$$x_t'=\sum_{j=1}^{t}\rho^{t-j}\sum_{i=1}^{j}\epsilon_i\phi_{j-i}-\phi_0\,\epsilon_0=\sum_{i=1}^{t-1}\epsilon_i\sum_{j=0}^{t-i}\rho^{t-j-i}\,\phi_j.$$

Write $a_{t,i}=\rho^{t-i}\sum_{j=0}^{t-i}\rho^{-j}\,\phi_j$ and $d_t^2=\sum_{i=1}^{t-1}a_{t,i}^2=\mathbb{E}(x_t')^2$. The result (7.14) in Wang and Phillips (2009b) shows that $a_{t,i}\ge e^{-|\tau|}\,|\phi|/4$ whenever $1\le i\le t/2$ and $t\le n$ is sufficiently large. This implies that $d_t^2\asymp t$, where the notation $a_n\asymp b_n$ denotes that there exist constants $c_1,c_2>0$ such that $c_1\le a_n/b_n\le c_2$, and as in the proof of Corollary 2.2 of Wang and Phillips (2009a),

$$\int_{-\infty}^{\infty}\left|\mathbb{E}e^{i\lambda x_t/\sqrt{t}}\right|d\lambda\le C\int_{-\infty}^{\infty}\left|\mathbb{E}e^{i\lambda x_t'/\sqrt{t}}\right|d\lambda<\infty, \tag{5.7}$$

$$\int_{-\infty}^{\infty}\left|\mathbb{E}e^{i\lambda(x_t^*+x_t')/\sqrt{t}}\right|d\lambda\le C\int_{-\infty}^{\infty}\left|\mathbb{E}e^{i\lambda x_t'/\sqrt{t}}\right|d\lambda<\infty, \tag{5.8}$$

uniformly for $t\ge 1$. The results (5.7) and (5.8) imply that

F. $x_t/\sqrt{t}$ and $(x_t^*+x_t')/\sqrt{t}$ have densities $\nu_{1t}(x)$ and $\nu_{2t}(x)$, respectively, and both $\nu_{1t}(x)$ and $\nu_{2t}(x)$ are uniformly bounded on t and x by a constant C.

See, e.g., Lukács (1970, Thm. 3.2.2). Furthermore, by defining $\sum_k^j=0$ if $j<k$,

$$I_1(\lambda)=\mathbb{E}\exp\left\{i\lambda\sum_{q=s}^{t-1}\epsilon_q\,a_{t,q}\right\},\qquad I_2(\lambda,\lambda_1)=\mathbb{E}\exp\left\{i\sum_{q=1}^{s-1}\epsilon_q\left(\lambda\,a_{t,q}-\lambda_1\,a_{s,q}\right)\right\}, \tag{5.9}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

for $s \le t$, the same arguments as in the proof of Lemma 7.2 of Wang and Phillips (2009b) yield that there exist $\gamma_1, \gamma_2 > 0$ such that

$$I_1(\lambda) \le \begin{cases} e^{-\gamma_1 (t-s)} & \text{if } |\lambda| \ge 1, \\ e^{-\gamma_2 \lambda^2 (t-s)} & \text{if } |\lambda| \le 1, \end{cases} \tag{5.10}$$

$$I_2(\lambda, \lambda_1) \le \exp\left\{-\gamma_1 \#(\Omega_1) - \gamma_2 B_1 (\lambda_1 - \lambda B_2/B_1)^2\right\}, \tag{5.11}$$

where $\#(A)$ denotes the number of elements in A, Ω_1 (Ω_2, respectively) denotes the set of $1 \le q \le s/2$ such that $|\lambda a_{t,q} - \lambda_1 a_{s,q}| \ge 1$ ($|\lambda a_{t,q} - \lambda_1 a_{s,q}| < 1$, respectively), and

$$B_1 = \sum_{q \in \Omega_2} a_{s,q}^2 \quad \text{and} \quad B_2 = \sum_{q \in \Omega_2} a_{t,q} a_{s,q}.$$

Note that $\Omega_1 + \Omega_2 = s/2$ and recall that $a_{s,q} \ge e^{-|\tau|} |\phi|/4$ for all sufficiently large s. It is readily seen from (5.11) that

$$I_2(\lambda, \lambda_1) \le e^{-\gamma_1 \sqrt{s}} I_{(\#(\Omega_1) \ge \sqrt{s})} + e^{-\gamma_2 s (\lambda_1 - \lambda B_2/B_1)^2} I_{(\#(\Omega_1) \le \sqrt{s})}, \tag{5.12}$$

for some $\gamma_1, \gamma_2 > 0$.

We are now ready to prove (5.3)–(5.5).

Start with (5.3). Recall that $K(x)$ has a compact support. Without loss of generality, suppose $K(x) = 0$ when $x \notin [c, d]$. Because $x_t/\sqrt{t}$ has a uniformly bounded density $\nu_{1t}(x)$ by the fact **F**, it is readily seen from Assumption 3 that

$$\begin{aligned} \mathbb{E}\lambda_{1n} &\le 2 \sum_{t=1}^{n} \mathbb{E} \max_{0 \le j \le m_n - 1} \sup_{x \in [y_j, y_{j+1}]} \left| K[(x_t - x)/h] - K[(x_t - y_j)/h] \right| \\ &\le 2 \sum_{t=1}^{n} \int_{-\infty}^{\infty} \max_{0 \le j \le m_n - 1} \sup_{x \in [y_j, y_{j+1}]} \left| K[(\sqrt{t}\, y - x)/h] - K[(\sqrt{t} y - y_j)/h] \right| \\ &\qquad \times \nu_{1t}(y) dy \\ &\le 2 \sum_{t=1}^{n} \frac{h}{\sqrt{t}} \int_{-|a|/h - |c|}^{|b|/h + |d|} \max_{0 \le j \le m_n - 1} \sup_{x \in [y_j, y_{j+1}]} \left| K(y - x/h) - K(y - y_j/h) \right| \\ &\qquad \times \nu_{1t}\left(yh/t\right) dy \\ &\le C \sum_{t=1}^{n} \frac{h^{-1}}{\sqrt{t}} \max_{0 \le j \le m_n - 1} \left| y_{j+1} - y_j \right| \\ &\le C \sqrt{n}\, m_n^{-1} h^{-1} \le C \left(nh^2\right)^{1/4}. \end{aligned} \tag{5.13}$$

Hence, $P(\lambda_{1n} \ge (nh^2)^{1/4} \log n) \le C/\log n$, which yields the required (5.3).

To prove (5.4), write $Z_{jt} = K[(x_t - y_j)/h] - \mathbb{E}\big(K[(x_t - y_j)/h] \mid \mathcal{F}_{t-1}^*\big)$. We first claim that

$$\begin{aligned} \max_{0 \le j \le m_n} \sum_{t=1}^{n} \mathbb{E}\big(Z_{jt}^2 \mid \mathcal{F}_{t-1}^*\big) &\le 2 \sum_{t=1}^{n} \max_{0 \le j \le m_n} \mathbb{E}\big(K^2[(x_t - y_j)/h] \mid \mathcal{F}_{t-1}^*\big) \\ &= O_P\big(\sqrt{n}h\big). \end{aligned} \tag{5.14}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

In fact, by recalling that $\int_{-\infty}^{\infty}|\mathbb{E}e^{it\epsilon_0}|dt<\infty$, the $\phi_0\epsilon_0$ has a density, say, $d(x)$. This, together with (5.6) and the independence between ϵ_0 and $x_{1t}:=x_t^*+x_t'$, yields that

$$\sup_{x\in\Omega^*}\mathbb{E}\big(K^2[(x_t-x)/h]\mid\mathcal{F}_{t-1}^*\big)\le C\int_{-\infty}^{\infty}K[(x_{1t}+y)/h]\sup_{x\in\Omega^*}d(y+x)dy. \tag{5.15}$$

Hence, by noting that

$$\mathbb{E}K[(x_{1t}+y)/h]=\int_{-\infty}^{\infty}K\big(s\sqrt{t}/h+y/h\big)v_{2t}(s)ds\le Ch/\sqrt{t}$$

uniformly on $y\in R$ because of the fact **F**, it follows that

$$\begin{aligned}\mathbb{E}\Big[\max_{0\le j\le m_n}\sum_{t=1}^{n}\mathbb{E}\big(Z_{jt}^2\mid\mathcal{F}_{t-1}^*\big)\Big]&\le 2\sum_{t=1}^{n}\int_{-\infty}^{\infty}\mathbb{E}K[(x_{1t}+y)/h]\sup_{x\in\Omega^*}d(y+x)dy\\&\le Ch\sum_{t=1}^{n}t^{-1/2}\le C_1\sqrt{n}h,\end{aligned} \tag{5.16}$$

where we have used the fact that $\int_{-\infty}^{\infty}\sup_{x\in\Omega^*}d(y+x)dy<\infty$, as Ω^* is a compact set. The claim (5.14) follows from (5.16) by Markov's inequality.

By virtue of (5.14), it follows from $|Z_{jt}|\le C$, $nh^2\log^{-4}n\to\infty$, and an exponential inequality for the martingale given by de la Peña (1999) that, for any $0<\delta<1$,

$$\begin{aligned}&P\big(\lambda_{2n}\ge(\sqrt{n}h)^{1/2}\log n\big)\\&\quad\le P\left[\lambda_{2n}\ge(\sqrt{n}h)^{1/2}\log n,\quad\max_{0\le j\le m_n}\sum_{t=1}^{n}\mathbb{E}\big(Z_{jt}^2\mid\mathcal{F}_{t-1}^*\big)\le\sqrt{n}h\log^{\delta}n\right]+o(1)\\&\quad\le\sum_{j=0}^{m_n}P\left[\sum_{t=1}^{n}Z_{jt}\ge(\sqrt{n}h)^{1/2}\log n,\quad\sum_{t=1}^{n}\mathbb{E}\big(Z_{jt}^2\mid\mathcal{F}_{t-1}^*\big)\le\sqrt{n}h\log^{\delta}n\right]+o(1)\\&\quad\le C(nh^2)^{1/4}h^{-2}\exp\left\{-\frac{\sqrt{n}h\log^2 n}{2\sqrt{n}h\log^{\delta}n+C(\sqrt{n}h)^{1/2}\log n}\right\}+o(1)\\&\quad\le C(nh^2)^{1/4}h^{-2}e^{-C_1\log^{2-\delta}n}+o(1)=o(1),\end{aligned} \tag{5.17}$$

which implies the required (5.4).

The proof of (5.5) is more laborious. First note that, as in the proof of (5.15),

$$\begin{aligned}\inf_{x\in\Omega^*}\mathbb{E}\big(K[(x_t-x)/h]\mid\mathcal{F}_{t-1}^*\big)&=\inf_{x\in\Omega^*}\int_{-\infty}^{\infty}K[(x_{1t}+y-x)/h]d(y)dy\\&\ge\int_{|y|\le A_0}K[(x_{1t}+y)/h]\inf_{x\in\Omega^*}d(y+x)dy,\end{aligned}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

where A_0 is a constant chosen such that $I := \int_{|y|\le A_0} \inf_{x\in\Omega^*} d(y+x)\,dy > 0$. Because

$$P\left(\inf_{x\in\Omega^*}\sum_{t=1}^{n}\mathbb{E}\left(K[(x_t-x)/h]\mid\mathcal{F}_{t-1}^*\right)\ge\sqrt{n}h/M\right)\ge P(\Gamma_n\ge\sqrt{n}h/M),$$

where

$$\Gamma_n=\int_{|y|\le A_0}\sum_{t=1}^{n}K[(x_{1t}+y)/h]\inf_{x\in\Omega^*}d(y+x)\,dy,$$

(5.5) will follow if we prove

$$\frac{\Gamma_n}{\sqrt{n}h}\to_D\Gamma, \tag{5.18}$$

where Γ is a random variable satisfying $P(\Gamma>0)=1$. To prove (5.18), we first assume that

Con: $K(x)$ is continuous and $\hat{K}(x)$ has a compact support,
where $\hat{K}(x)=\int_{-\infty}^{\infty}e^{ixt}K(t)dt$. (5.19)

This restriction will be removed later. Under condition (5.19), we have $K(x)=1/2\pi\int_{-\infty}^{\infty}e^{i\lambda x}\hat{K}(-\lambda)d\lambda$. This implies that, for each $A>0$,

$$\begin{aligned}\Gamma_n&=\int_{|y|\le A_0}\left[V_{1yA}+V_{2yA}\right]\inf_{x\in\Omega^*}d(y+x)\,dy\\&:=\Gamma_{1nA}+\Gamma_{2nA},\end{aligned} \tag{5.20}$$

where, with $c_n=\sqrt{n}/h$,

$$V_{1yA}=\frac{1}{2\pi\,c_n}\sum_{t=1}^{n}\int_{|\lambda|\ge A}e^{i\lambda(x_{1t}+y)/\sqrt{n}}\hat{K}(-\lambda/c_n)d\lambda,$$

$$V_{2yA}=\frac{1}{2\pi\,c_n}\sum_{t=1}^{n}\int_{|\lambda|<A}e^{i\lambda(x_{1t}+y)/\sqrt{n}}\hat{K}(-\lambda/c_n)d\lambda.$$

Recall that $0<I=\int_{|y|\le A_0}\inf_{x\in\Omega^*}d(y+x)\,dy\le\int\inf_{x\in\Omega^*}d(y+x)\,dy<\infty$, because Ω^* is a compact set. It is readily seen that, to prove (5.18), it suffices to show that

$$\sup_{|y|\le A_0}\mathbb{E}\left|V_{1yA}\right|=o(\sqrt{n}h), \tag{5.21}$$

and on a richer probability space that also holds a standard Brownian motion $W(t)$ such that

$$\sup_{|y|\le A_0}\left|\frac{V_{2yA}}{\sqrt{n}h}-\frac{1}{\phi}L_G(1,0)\right|=o_P(1), \tag{5.22}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

as $n \to \infty$ first and then $A \to \infty$, where $G(t) = W(t) + \kappa \int_0^t e^{\kappa(t-s)} W(s) ds$ and $L_G(r, x)$ is a local time of the Gaussian process $G(t)$ expressed as

$$L_G(t,x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \int_0^t e^{iuG(s)} ds\, du. \tag{5.23}$$

Here and in what follows the right-hand side of (5.23) is understood as a limitation in L_2 of $L_G^N(t,x)$, as $N \to \infty$, defined as

$$L_G^N(t,x) = \frac{1}{2\pi} \int_{-N}^{N} e^{-iux} \int_0^t e^{iuG(s)} ds\, du.$$

The process $\{L_\zeta(t,s), t \geq 0, s \in R\}$ is said to be the local time of a measurable process $\{\zeta(t), t \geq 0\}$ if, for any locally integrable function $T(x)$,

$$\int_0^t T[\zeta(s)] ds = \int_{-\infty}^{\infty} T(s) L_\zeta(t,s) ds, \quad \text{all } t \in R, \tag{5.24}$$

with probability one. For more details in these regards, we refer to Berman (1969).

By recalling that $x_{1t} = x_t^* + x_t'$ and the independence between x_t^* and x_t', we have that

$$\begin{aligned}
\mathbb{E}|V_{1yA}|^2 &\leq \frac{Ch^2}{n} \sum_{s,t=1}^{n} \int_{|\lambda| \geq A} \int_{|\lambda_1| \geq A} \left|\hat{K}(-\lambda/c_n)\right| \left|\hat{K}(-\lambda_1/c_n)\right| \\
&\quad \times \left|\mathbb{E} e^{i(\lambda x_{1t} - \lambda_1 x_{1s})/\sqrt{n}}\right| d\lambda\, d\lambda_1 \\
&\leq \frac{Ch^2}{n} \sum_{s,t=1}^{n} \int_{|\lambda| \geq A} \int_{|\lambda_1| \geq A} \left|\hat{K}(-\lambda/c_n)\right| \left|\hat{K}(-\lambda_1/c_n)\right| \\
&\quad \times \left|\mathbb{E} e^{i(\lambda x_t' - \lambda_1 x_s')/\sqrt{n}}\right| d\lambda\, d\lambda_1 \\
&\leq \frac{Ch^2}{n} \sum_{s \leq t}^{n} \int_{|\lambda| \geq A} I_1\left(\frac{\lambda}{\sqrt{n}}\right) \left|\hat{K}(-\lambda/c_n)\right| d\lambda \\
&\quad \times \int_{|\lambda_1| \geq A} I_2\left(\frac{\lambda}{\sqrt{n}}, \frac{\lambda_1}{\sqrt{n}}\right) \left|\hat{K}(-\lambda_1/c_n)\right| d\lambda_1,
\end{aligned} \tag{5.25}$$

where $I_1(s)$ and $I_2(s,t)$ are defined as in (5.9). It follows from (5.12) that

$$\begin{aligned}
&\int_{|\lambda_1| \geq A} I_2\left(\frac{\lambda}{\sqrt{n}}, \frac{\lambda_1}{\sqrt{n}}\right) \left|\hat{K}(-\lambda_1/c_n)\right| d\lambda_1 \\
&\quad \leq e^{-\gamma_1 \sqrt{s}} \int \left|\hat{K}(-\lambda_1/c_n)\right| d\lambda_1 + C \int e^{-\gamma_2' s \lambda_1^2/n} d\lambda_1 \\
&\quad \leq C\left[c_n e^{-\gamma_1 \sqrt{s}} + (n/s)^{1/2}\right].
\end{aligned}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

It follows from (5.10) that

$$\begin{aligned}
&\int_{|\lambda|\geq A} I_1\left(\frac{\lambda}{\sqrt{n}}\right)\left|\hat{K}(-\lambda/c_n)\right|d\lambda \\
&\quad\leq e^{-\gamma_1(t-s)}\int_{|\lambda|\geq\sqrt{n}}\left|\hat{K}(-\lambda/c_n)\right|d\lambda + C\int_{A\leq|\lambda|\leq\sqrt{n}} e^{-\gamma_2\lambda^2(t-s)/n}d\lambda \\
&\quad\leq C\left[c_n e^{-\gamma_1(t-s)} + \int_{A\leq|\lambda|\leq\sqrt{n}} e^{-\gamma_2\lambda^2(t-s)/n}d\lambda\right].
\end{aligned}$$

Taking these estimates into (5.25), simple calculations show that

$$\begin{aligned}
\sup_{x\in\omega^*}\mathbb{E}|V_{1yA}|^2 &\leq \frac{Ch^2}{n}\sum_{t-s=0}^{n}\left[c_n e^{-\gamma_1(t-s)} + \int_{A\leq|\lambda|\leq\sqrt{n}} e^{-\gamma_2\lambda^2(t-s)/n}d\lambda\right] \\
&\quad\times\sum_{s=1}^{n}\left[c_n e^{-\gamma_1\sqrt{s}} + (n/s)^{1/2}\right] \\
&\leq Ch^2\left[c_n + \int_{A\leq|\lambda|\leq\sqrt{n}}\sum_{t-s=0}^{n} e^{-\gamma_2\lambda^2(t-s)/n}d\lambda\right] \\
&\leq C\sqrt{n}h + nh^2\int_{|\lambda|\geq A}\int_0^1 e^{-\gamma_2\lambda^2 x}\,dx\,d\lambda \\
&= o(nh^2),
\end{aligned}$$

as $n\to\infty$ first and then $A\to\infty$, which yields the required (5.21).

To establish (5.22), first notice that $\{\epsilon_j, j\in Z\}$ can be redefined on a richer probability space that also holds a standard Brownian motion $W(t)$ such that

$$\sup_{0\leq t\leq 1}\left|x_{[nt],n} - G(t)\right| = o_P(1), \tag{5.26}$$

where $x_{k,n} = x_k/(\sqrt{n}\phi)$ and $G(t) = W(t) + \kappa\int_0^t e^{\kappa(t-s)}W(s)ds$. Indeed, on a richer probability space,

$$\sup_{0\leq t\leq 1}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_j - W(t)\right| = o_P(1),$$

See, e.g., Csörgö and Révész (1981). Using this result in place of the fact that $\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_j \Rightarrow W(t)$ on $D[0,1]$, the same technique as in the proof of Phillips (1987) (also see Chan and Wei, 1987) yields that

$$\sup_{0\leq t\leq 1}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\rho^{[nt]-j}\epsilon_j - G(t)\right| = o_P(1).$$

The result (5.26) can now be obtained by the same argument, with minor modifications, as in the proof of Proposition 7.1 in Wang and Phillips (2009b). On the

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

other hand, simple calculations show that

$$\sup_{|\lambda|\le A,|y|\le A_0}\left|e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)-\hat{K}(0)\right|=o(1)$$

as $c_n\to\infty$. This, together with (5.26) and the fact that Γ_{2nA} can be rewritten as

$$\begin{aligned}\frac{\Gamma_{2nA}}{\sqrt{n}h}&=\frac{1}{2\pi}\int_{|\lambda|<A}e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)\frac{1}{n}\sum_{t=1}^{n}e^{i\lambda\phi x_{k,n}}\,d\lambda\\&=\frac{1}{2\pi}\int_{|\lambda|<A}e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)\int_0^1 e^{i\lambda\phi x_{[tn],n}}dt\,d\lambda,\end{aligned}$$

yields that, for any $A>0$,

$$\begin{aligned}&\sup_{|y|\le A_0}\left|\frac{V_{2yA}}{\sqrt{n}h}-\frac{\hat{K}(0)}{2\pi}\int_{|\lambda|\le A}\int_0^1 e^{i\lambda\phi G(t)}dt\,d\lambda\right|\\&\quad\le\ C\sup_{|\lambda|\le A,|y|_0}\left|e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)-\hat{K}(0)\right|\\&\qquad+\frac{|\hat{K}(0)|}{2\pi}\int_{|\lambda|\le A}\int_0^1\left|e^{i\lambda\phi(x_{[nt],n}-G(t))}-1\right|dtd\lambda\\&\quad\le\ o(1)+O(1)\sup_{0\le t\le 1}\left|x_{[nt],n}-G(t)\right|=o_P(1).\end{aligned}$$

This proves (5.22) because $\hat{K}(0)=\int K(x)dx=1$ and

$$\frac{1}{2\pi}\int_{|\lambda|\le A}\int_0^1 e^{i\lambda\phi G(t)}dt\,d\lambda\to_P\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_0^1 e^{i\lambda\phi G(t)}dt\,d\lambda=\frac{1}{\phi}L_G(1,0),$$

by the expression (5.23).

Finally, we remove the restriction (5.19). For this we only need to notice that, for any $\epsilon>0$, there exist $K^+_{\delta_0}(x)$ and $K^-_{\delta_0}(x)$ such that $K^-_{\delta_0}(x)\le K(x)\le K^+_{\delta_0}(x)$, both $K^+_{\delta_0}(x)$ and $K^-_{\delta_0}(x)$ satisfy (5.19), $\int_{-\infty}^{\infty}(|K^+_{\delta_0}(x)|+|K^-_{\delta_0}(x)|)dx<\infty$, and

$$\int_{-\infty}^{\infty}\left[K^+_{\delta_0}(x)-K^-_{\delta_0}(x)\right]dx<\epsilon. \tag{5.27}$$

The constructions of $K^+_{\delta_0}(x)$ and $K^-_{\delta_0}(x)$ are similar to the proof of Theorem 4.2.1 in Borodin and Ibragimov (1995). We omit the details.

This completes the proof of (5.5) and hence also the proof of Proposition 3.2. ■

Proof of Proposition 3.1. By virtue of (3.10), it is only necessary to prove that

$$\sup_{x\in\Omega^*}\sum_{t=1}^{n}u_t\,K[(x_t-x)/h]=O_P\left[(nh^2)^{(p+1)/(2(2p+1))}\,h^{-(2)/(2p+1)}\right], \tag{5.28}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

and if in addition $|u_t| \le C$, then, for some $\delta > 0$,

$$\sup_{x\in\Omega^*} \sum_{t=1}^{n} u_t K[(x_t - x)/h] = o_P\left[(nh^2)^{1/4} \log^{1-\delta} n\right]. \tag{5.29}$$

To prove (5.28), we adopt the same partition as in (5.1) but with

$$m_n = \left[(nh^2)^{(p)/(2(2p+1))} h^{-(4p)/(2p+1)}\right].$$

It follows that

$$\begin{aligned}
&\sup_{x\in\Omega^*} \left| \sum_{t=1}^{n} u_t K[(x_t - x)/h] \right| \\
&\quad \le \max_{0\le j\le m_n-1} \sup_{x\in[y_j, y_{j+1}]} \sum_{t=1}^{n} |u_t| \left| K[(x_t - x)/h] - K[(x_t - y_j)/h] \right| \\
&\qquad + \max_{0\le j\le m_n} \left| \sum_{t=1}^{n} u_t K[(x_t - y_j)/h] \right| \\
&\quad := \lambda_{3n} + \lambda_{4n}.
\end{aligned} \tag{5.30}$$

Recall that x_t is adapted to $\mathcal{F}_{t-1}$ and $\sup_{1\le t\le n} \mathbb{E}(|u_t| \mid \mathcal{F}_{t-1}) < \infty$, as indicated in Assumption 2. As in the proof of (5.13), we have

$$\begin{aligned}
\mathbb{E}\lambda_{3n} &\le \sum_{t=1}^{n} \sup_{1\le t\le n} \mathbb{E}(|u_t| \mid \mathcal{F}_{t-1}) \\
&\quad \times \mathbb{E} \max_{0\le j\le m_n-1} \sup_{x\in[y_j, y_{j+1}]} \left| K[(x_t - x)/h] - K[(x_t - y_j)/h] \right| \\
&\le C \sqrt{n} m_n^{-1} h^{-1}.
\end{aligned} \tag{5.31}$$

Under $\mathbb{E}|u_t|^{2p} < \infty$, the well-known martingale moment inequality yields that

$$\begin{aligned}
\mathbb{E}\lambda_{4n}^{2p} &\le \sum_{j=0}^{m_n} \mathbb{E} \left| \sum_{t=1}^{n} u_t K[(x_t - y_j)/h] \right|^{2p} \\
&\le C \sum_{j=0}^{m_n} \mathbb{E} \left\{ \sum_{t=1}^{n} K^2[(x_t - y_j)/h] \right\}^{p} \\
&\le C m_n \max_{0\le j\le m_n} \left\{ \sum_{t=1}^{n} \mathbb{E} K[(x_t - y_j)/h] \right. \\
&\qquad + \sum_{1\le s<t\le n} \mathbb{E}\left(K[(x_s - y_j)/h] K[(x_t - y_j)/h] \right) \\
&\qquad + \sum_{1\le t_1<t_2<\cdots<t_p\le n} \mathbb{E}\left(K[(x_{t_1} - y_j)/h] \right. \\
&\qquad\qquad \left. \left. \times K[(x_{t_2} - y_j)/h] \cdots K[(x_{t_p} - y_j)/h] \right) \right\},
\end{aligned} \tag{5.32}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

where we have used $K(x) \le C$. Note that, given on $\mathcal{F}_s$, $(x_t - x_s)/\sqrt{t-s}$ has a density $h_{s,t}(x)$ that is uniformly bounded by a constant C. See the proof of Proposition 7.2 of Wang and Phillips (2009b, p. 1934). Simple calculations show that, for $1 \le j \le p$,

$$\begin{aligned}\mathbb{E}\left\{K[(x_{t_j} - y_j)/h] \mid \mathcal{F}_{t_{j-1}}\right\} &= \int_{-\infty}^{\infty} K\left[\sqrt{t_j - t_{j-1}}\, y/h + (x_{t_{j-1}} - y_j)/h\right] h_{t_{j-1},t_j}(y)dy \\ &\le \frac{Ch}{\sqrt{t_j - t_{j-1}}} \int_{-\infty}^{\infty} K\left[y + (x_{t_{j-1}} - y_j)/h\right] dy \\ &\le C_1 h/\sqrt{t_j - t_{j-1}}.\end{aligned}$$

Taking this estimate into (5.32) and using the conditional arguments repeatedly, we obtain

$$\begin{aligned}\mathbb{E}\lambda_{4n}^{2p} &\le C\, m_n \left\{\sum_{t=1}^{n} \frac{h}{\sqrt{t}} + \sum_{1\le s<t\le n} \frac{h^2}{\sqrt{s}} \frac{1}{\sqrt{t-s}} \right. \\ &\qquad \left. + \sum_{1\le t_1<t_2<\cdots<t_p\le n} \frac{h^p}{\sqrt{t_1}} \frac{1}{\sqrt{t_2 - t_1}} \cdots \frac{1}{\sqrt{t_p - t_{p-1}}}\right\} \\ &\le C\, m_n\, (nh^2)^{p/2}, \end{aligned} \tag{5.33}$$

as $nh^2 \to \infty$. It follows from (5.31) and (5.33) that

$$\begin{aligned}\lambda_{3n} + \lambda_{4n} &= O_P\left[\sqrt{n} m_n^{-1} h^{-1} + m_n^{1/(2p)}\, (nh^2)^{1/4}\right] \\ &= O_P\left[(nh^2)^{(p+1)/(2(2p+1))}\, h^{-(2)/(2p+1)}\right],\end{aligned}$$

which yields (5.28).

The proof of (5.29) is similar to (5.28) but with

$$m_n = \left[\frac{(nh^2)^{1/4} h^{-2}}{\log^{1-\delta/2} n}\right].$$

In this case, we still have (5.30), and by (5.31)

$$\lambda_{3n} = O_P\left(\sqrt{n} m_n^{-1} h^{-1}\right) = o_P\left[\left(nh^2\right)^{1/4} \log^{1-\delta} n\right].$$

So it is only necessary to prove, for any $0 < \delta < 1$,

$$P\left[\lambda_{4n} \ge \left(nh^2\right)^{1/4} \log^{1-\delta/2} n\right] \to 0. \tag{5.34}$$

This follows from the same arguments as in the proof of (5.17). In fact, as in (5.16), we have

$$\begin{aligned}\mathbb{E}\left\{\max_{0\le j\le m_n} \sum_{t=1}^{n} K^2[(x_t - y_j)/h]\right\} &\le \mathbb{E}\left[\max_{0\le j\le m_n} \sum_{t=1}^{n} \mathbb{E}\left(K^2[(x_t - y_j)/h] \mid \mathcal{F}_{t-1}^*\right)\right] \\ &\le C\sqrt{n}\, h,\end{aligned}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

which implies that

$$\max_{0\le j\le m_n}\sum_{t=1}^{n}K^2\left[(x_t-y_j)/h\right]=O_P\left(\sqrt{n}h\right). \tag{5.35}$$

Hence, by recalling $nh^2/\log^4 n\to\infty$,

$$\begin{aligned}
&P\left(\lambda_{4n}\ge\left(\sqrt{n}h\right)^{1/2}\log^{1-\delta/2}n\right)\\
&\quad\le P\left[\lambda_{4n}\ge\left(\sqrt{n}h\right)^{1/2}\log^{1-\delta/2}n,\right.\\
&\qquad\left.\times\max_{0\le j\le m_n}\sum_{t=1}^{n}K^2\left[(x_t-y_j)/h\right]\le\sqrt{n}h\log^{\delta/2}n\right]+o(1)\\
&\quad\le\sum_{j=0}^{m_n}P\left[\sum_{t=1}^{n}u_jK\left[(x_t-y_j)/h\right]\ge\left(\sqrt{n}h\right)^{1/2}\log^{1-\delta/2}n,\right.\\
&\qquad\left.\times\sum_{t=1}^{n}K^2\left[(x_t-y_j)/h\right]\le\sqrt{n}h\log^{\delta/2}n\right]+o(1)\\
&\quad\le m_n\exp\left\{-\frac{\sqrt{n}h\log^{2-\delta}n}{2\sqrt{n}h\log^{\delta/2}n+C\left(\sqrt{n}h\right)^{1/2}\log^{1-\delta/2}n}\right\}+o(1)=o(1),
\end{aligned}$$

which yields (5.34). ■

Proof of Theorem 3.2. We may write $\hat{f}(x)-f(x)$ as

$$\begin{aligned}
\hat{f}(x)-f(x)&=\frac{\sum_{t=1}^{n}\sigma(x)u_tK\left[(x_t-x)/h\right]}{\sum_{t=1}^{n}K\left[(x_t-x)/h\right]}\\
&\quad+\frac{\sum_{t=1}^{n}\left[\sigma(x_t)-\sigma(x)\right]u_tK\left[(x_t-x)/h\right]}{\sum_{t=1}^{n}K\left[(x_t-x)/h\right]}\\
&\quad+\frac{\sum_{t=1}^{n}\left[f(x_t)-f(x)\right]K\left[(x_t-x)/h\right]}{\sum_{t=1}^{n}K\left[(x_t-x)/h\right]}\\
&:=\Theta_{1n}(x)+\Theta_{2n}(x)+\Theta_{3n}(x).
\end{aligned}\tag{5.36}$$

It follows from (3.8) that

$$\sup_{x\in\Omega_\epsilon}|\Theta_{1n}(x)|=O_P\left[\left(nh^{2+4/p}\right)^{-p/[2(2p+1)]}\right],$$

because $\sup_{x\in\Omega_\epsilon}|\sigma(x)|\le C$, by Assumption 4. Note that, for any $x\in\Omega_\epsilon$, there exists a $C_0>0$ such that $K[(x_t-x)/h]=0$ if $|x_t-x|\ge hC_0$. It is readily seen from Assumption 4 that, whenever n is sufficiently large,

$$\sup_{x\in\Omega_\epsilon}|\Theta_{3n}(x)|\le\frac{C_1\sum_{t=1}^{n}|x_t-x|^\alpha K\left[(x_t-x)/h\right]}{\sum_{t=1}^{n}K\left[(x_t-x)/h\right]}\le Ch^\alpha.$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

Similarly, by noting that $|u_t| - \mathbb{E}(|u_t| \mid \mathcal{F}_{t-1})$ forms a martingale difference satisfying Assumption 2*, we have

$$\begin{aligned}\sup_{x\in\Omega_\epsilon}|\Theta_{2n}(x)| &\le Ch^\beta \sup_{x\in\Omega_\epsilon}\frac{\sum_{t=1}^n |u_t|K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]}\\ &\le Ch^\beta\left\{\sup_{t\ge 1}\mathbb{E}(|u_t|\mid\mathcal{F}_{t-1})\right.\\ &\qquad\left.+\sup_{x\in\Omega_\epsilon}\frac{\sum_{t=1}^n\left[|u_t|-\mathbb{E}(|u_t|\mid\mathcal{F}_{t-1})\right]K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]}\right\}\\ &\le C_1h^\beta. \end{aligned}\tag{5.37}$$

Taking these estimates into (5.36), we derive (3.5). The proof of (3.6) is similar except using (3.9) instead of (3.8). We omit the details. ■

Proof of Theorem 3.1. First we give the proof of (3.1), which is similar to that of (3.5). We still use the decomposition in (3.4). Recall that $\sup_{1\le t\le n}\mathbb{E}(|u_t^2|^{2[\nu]}|\mathcal{F}_{t-1}) < \infty$ a.s., where $[\nu]\ge 1$, by Assumption 2. It follows from (3.8) that

$$\sup_{x\in\Omega}|I_{1n}(x)| = O_P\left[\left(nh^{2+4/[\nu]}\right)^{-[\nu]/[2(2[\nu]+1)]}\right],$$

because $\sup_{x\in\Omega_\epsilon}|\sigma(x)|\le C$, by Assumption 4. Because Assumption 4 also implies that, for any $x, y\in\Omega_\epsilon$,

$$|\sigma^2(y)-\sigma^2(x)| \le 2|\sigma(y)-\sigma(x)|\sup_{x\in\Omega_\epsilon}|\sigma(x)| \le C|y-x|^\beta,$$

similarly to the estimate of (5.37), we have

$$\sup_{x\in\Omega}|I_{2n}(x)| \le Ch^\beta\frac{\sum_{t=1}^n u_t^2K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} = O_P(h^\beta).\tag{5.38}$$

Note that $h\to 0$ and, for any $x\in\Omega$, there exists a $C_0>0$ such that $K[(x_t-x)/h]=0$ if $|x_t-x|\ge hC_0$. This, together with (3.5), implies that, whenever n is sufficiently large,

$$\begin{aligned}\sup_{x\in\Omega}|I_{3n}(x)| &\le C\sup_{x\in\Omega_\epsilon}\left|\hat f(x)-f(x)\right|^2\\ &= O_P\left\{\left(nh^{2+4/[\nu]}\right)^{-[\nu]/(2[\nu]+1)}+h^{2\min\{\alpha,\beta\}}\right\}.\end{aligned}\tag{5.39}$$

Similarly, we have

$$\begin{aligned}\sup_{x\in\Omega}|I_{4n}(x)| &\le C\sup_{x\in\Omega_\epsilon}\left|\hat f(x)-f(x)\right|\sup_{x\in\Omega_\epsilon}|\sigma(x)|\frac{\sum_{t=1}^n|u_t|K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]}\\ &= O_P\left\{\left(nh^{2+4/[\nu]}\right)^{-[\nu]/[2(2[\nu]+1)]}+h^{\min\{\alpha,\beta\}}\right\}.\end{aligned}\tag{5.40}$$

Taking these estimates into (3.4), we obtain the required (3.1).

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

We next prove (3.2), and to do so we show

$$\sum_{j=1}^{k} b_j T_n(z_j) \to_d N\left(0, \sigma_1^2\right), \tag{5.41}$$

for any $(b_1, \ldots, b_k) \in \mathbb{R}^k$, where $\sigma_1^2 = (\Lambda^2 - 1) \int_{-\infty}^{\infty} K^2(u) du \sum_{j=1}^{k} b_j^2 \sigma^4(z_j)$. Define

$$\Delta_n(x) = \sum_{t=1}^{n} K[(x_t - x)/h], \qquad Z_{nt}(x) = \sigma^2(x) K[(x_t - x)/h] \Delta_n^{-1/2}(x),$$

and $Y_{nt} = \sum_{j=1}^{k} b_j Z_{nt}(z_j)$. Recall (3.4). We may write

$$\sum_{j=1}^{k} b_j T_n(z_j) = \sum_{t=1}^{n} \left(u_t^2 - 1\right) Y_{nt} + \sum_{j=1}^{k} b_j \, \Delta_n^{1/2}(z_j) \left[I_{2n}(z_j) + I_{3n}(z_j) + I_{4n}(z_j)\right].$$

Recall (5.35), i.e., $\Delta_n(x) = O_P(\sqrt{n}h)$ for each fixed $x \in \Omega$. The result (5.41) will follow if we prove

$$\sum_{t=1}^{n} \left(u_t^2 - 1\right) Y_{nt} \to_d N\left(0, \sigma_1^2\right) \tag{5.42}$$

and for each fixed $x \in \Omega$,

$$I_{jn}(x) = o_P[(nh^2)^{-1/4}], \qquad j = 2,\ 3,\ 4. \tag{5.43}$$

First prove (5.42). Write $\mathcal{F}_{nj} = \sigma(u_1, ..., u_j, x_1, ..., x_n), 1 \le j \le n$. Recalling that u_t is independent of x_t, it is readily seen that $\mathcal{F}_{nj} \subseteq \mathcal{F}_{n,j+1}$ for $1 \le j \le n, n \ge 1$, and $\{(u_t^2 - 1) Y_{nt}, \mathcal{F}_{nt}\}_{t=1}^{n}$ forms a martingale difference sequence. By the classical martingale limit theorem (see, e.g., Hall and Heyde, 1980 Cor. 3.1), it suffices to show that, for any $\nu > 1$,

$$\sum_{t=1}^{n} |Y_{nt}|^{2\nu} \to_P 0 \tag{5.44}$$

and

$$\sum_{t=1}^{n} Y_{nt}^2 \, \mathbb{E}\left[(u_t^2 - 1)^2 \mid \mathcal{F}_{t-1}\right] \to_P \sigma_1^2. \tag{5.45}$$

Note that $\inf_{x \in \Omega} \Delta_n^{-1}(x) = o_P(1)$ by (3.10). The proof of (5.44) is simple. Indeed it is readily seen that

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

$$\sum_{t=1}^{n} |Y_{nt}|^{2\nu} \le C \sum_{j=1}^{k} \sum_{t=1}^{n} \left|Z_{nt}(z_j)\right|^{2\nu}$$
$$\le C \sum_{j=1}^{k} \left|\sigma(z_j)\right|^{2\nu} \Delta_n^{-\nu}(z_j) \sum_{t=1}^{n} K[(x_t - z_j)/h]$$
$$\le C_1 \sum_{j=1}^{k} \Delta_n^{1-\nu}(z_j) = o_P(1),$$

which yields (5.44).

To prove (5.45), first notice that, for any a, b and $1 \le j \le k$,

$$\frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} \left\{ a\, K\left[(x_t - z_j)/h\right] + b\, K^2\left[(x_t - z_j)/h\right] \right\}$$
$$\to_D \int_{-\infty}^{\infty} \left[aK(x) + bK^2(x) \right] dx\, L_G(1,0),$$

by Proposition 7.2 of Wang and Phillips (2009b).[1] This implies that (recalling $\int K(x)dx = 1$)

$$\left\{ \frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} K\left[(x_t - z_j)/h\right],\ \frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} K^2\left[(x_t - z_j)/h\right] \right\}$$
$$\to_D \left\{ L_G(1,0), \int_{-\infty}^{\infty} K^2(x)dx\, L_G(1,0) \right\},$$

and hence by the continuous mapping theorem,

$$\frac{\sum_{t=1}^{n} K^2\left[(x_t - z_j)/h\right]}{\sum_{t=1}^{n} K\left[(x_t - z_j)/h\right]} \to_P \int_{-\infty}^{\infty} K^2(x)dx.$$

On the other hand, for fixed $z_i \ne z_j$, we have

$$K[(x_t - z_i)/h]\, K[(x_t - z_j)/h] = 0,$$

whenever n is sufficiently large, as $K(x)$ has a compact support. Simple calculations show that

$$\sum_{t=1}^{n} Y_{nt}^2 = \sum_{j=1}^{k} b_j^2 \sigma^4(z_j)\, \Delta_n^{-1}(z_j) \sum_{t=1}^{n} K^2[(x_t - z_j)/h]$$
$$+ 2 \sum_{1 \le i < j \le k} b_i b_j \sigma^2(z_i) \sigma^2(z_j) \Delta_n^{-1/2}(z_i) \Delta_n^{-1/2}(z_j)$$
$$\times \sum_{t=1}^{n} K[(x_t - z_i)/h] K[(x_t - z_j)/h]$$
$$\to_P \int_{-\infty}^{\infty} K^2(x)dx \sum_{j=1}^{k} b_j^2 \sigma^4(z_j). \qquad \textbf{(5.46)}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

This, together with the fact that

$$\mathbb{E}\big[(u_t^2-1)^2 \mid \mathcal{F}_{t-1}\big] = \mathbb{E}\big[u_t^4 \mid \mathcal{F}_{t-1}\big] - 1 \to_{\text{a.s.}} \Delta^2 - 1,$$

yields (5.45). The proof of (5.42) is now complete.

We next prove (5.43). The result (5.43) with $j = 2$ follows from (5.38) and $nh^{2+4\beta} \to 0$. To prove (5.43) with $j = 3$ and 4, for $x \in \Omega$, write

$$\Delta_{1n}(x) = \sum_{t=1}^{n} (f(x_t) - \hat{f}(x_t))^2 K[(x_t - x)/h], \tag{5.47}$$

$$\Delta_{2n}(x) = \sum_{t=1}^{n} \sigma(x_t) u_t (f(x_t) - \hat{f}(x_t)) K[(x_t - x)/h]. \tag{5.48}$$

We may rewrite $\Delta_{2n}(x)$ as

$$\begin{aligned}\Delta_{2n}(x) = &\sum_{\substack{s,t=1\\ s\neq t}}^{n} u_s u_t \sigma(x_s)\sigma(x_t)\, \Delta_n^{-1}(x_t) K[(x_s - x_t)/h] K[(x_t - x)/h] \\ &+ K(0) \sum_{s=1}^{n} u_s^2 \sigma^2(x_s)\, \Delta_n^{-1}(x_s) K[(x_s - x)/h] \\ &+ \sum_{\substack{s,t=1\\ s\neq t}}^{n} u_t \sigma(x_t) [f(x_s) - f(x_t)] \Delta_n^{-1}(x_t) K[(x_s - x_t)/h] K[(x_t - x)/h] \\ = &\Delta_{2n1}(x) + \Delta_{2n2}(x) + \Delta_{2n3}(x), \quad \text{say.}\end{aligned} \tag{5.49}$$

Let $\Xi_n = I\big\{\inf_{x\in\Omega^*} \sum_{t=1}^{n} K[(x_t - x)/h] \geq \sqrt{n}h/M\big\}$, where $M \geq M_0$ is chosen as in (3.10). Note that, for $x \in \Omega$,

$$K[(x_s - x_t)/h] K[(x_t - x)/h] = 0, \tag{5.50}$$

if $|x_t - x| \geq C_0 h$ or $|x_s - x_t| \geq C_0 h$ for some $C_0 > 0$. By virtue of the independence between u_t and x_t, it is readily seen that there exists $\epsilon > 0$ such that

$$\begin{aligned}\mathbb{E}\Big\{\Delta_{2n1}^2(x)\Xi_n\Big\} &= \mathbb{E}\left\{\Xi_n \sum_{\substack{s,t=1\\ s\neq t}}^{n} u_s^2 u_t^2 \sigma^2(x_s)\sigma^2(x_t)\, \Delta_n^{-2}(x_t) K^2[(x_s - x_t)/h] \right.\\ &\qquad \left. \times\, K^2[(x_t - x)/h]\right\} \\ &\leq C \sup_{x\in\Omega_\epsilon} \sigma^4(x) \mathbb{E}\left\{\Xi_n \sum_{t=1}^{n} \Delta_n^{-1}(x_t) K^2[(x_t - x)/h]\right\} \\ &\leq \frac{CM}{\sqrt{n}h} \mathbb{E} \sum_{t=1}^{n} K^2[(x_t - x)/h] \leq C_1 M,\end{aligned} \tag{5.51}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

whenever n is sufficiently large. By virtue of this fact and (3.10), for any $\eta > 0$, there exists an $M_0 > 0$ such that for $M \geq M_0$,

$$\begin{aligned} P\left[\frac{\Delta_{2n1}(x)}{\Delta_n(x)} \geq (nh^2)^{-1/4}/M\right] &\leq P\left\{\inf_{x\in\Omega^*}\sum_{t=1}^{n} K\left[(x_t-x)/h\right] \leq \sqrt{n}h/M\right\} + \frac{M}{(nh^2)^{1/2}}\mathbb{E}\left\{\Delta_{2n1}^2(x)\Xi_n\right\} \\ &\leq \eta + \frac{C_1M^2}{(nh^2)^{1/2}}, \end{aligned} \tag{5.52}$$

which yields $(\Delta_{2n1}(x))/(\Delta_n(x)) = o_P[(nh^2)^{-1/4}]$. Similarly, $(\Delta_{2n2}(x))/(\Delta_n(x)) = o_P[(nh^2)^{-1/4}]$. As for $\Delta_{2n3}(x)$, it follows from (5.50) and Assumption 4 that

$$\begin{aligned} \mathbb{E}\Delta_{2n3}^2(x) &= \mathbb{E}\left\{\sum_{t=1}^{n} u_t^2\,\sigma^2(x_t)\,\Delta_n^{-2}(x_t)K^2[(x_t-x)/h] \right. \\ &\qquad \left. \times\left(\sum_{\substack{s=1\\ s\neq t}}^{n}[f(x_s)-f(x_t)]K[(x_s-x_t)/h]\right)^2\right\} \\ &\leq C\sup_{x\in\Omega_\epsilon}\sigma^2(x)\,h^{2\alpha}\,\mathbb{E}\sum_{t=1}^{n}K^2[(x_t-x)/h] \\ &\leq C_1\sqrt{n}h^{1+2\alpha}. \end{aligned} \tag{5.53}$$

Similar to the proof of (5.52) we have $(\Delta_{2n3}(x))/(\Delta_n(x)) = o_P[(nh^2)^{-1/4}]$, because $h \to 0$. Taking these estimates into (5.49), we prove

$$I_{4n} = \frac{2\Delta_{2n}(x)}{\Delta_n(x)} = o_P\left[(nh^2)^{-1/4}\right].$$

It is quite similar to show that

$$I_{3n} = \frac{\Delta_{1n}(x)}{\Delta_n(x)} = o_P\left[(nh^2)^{-1/4}\right].$$

Indeed, we may rewrite $\Delta_{1n}(x)$ as

$$\begin{aligned} \Delta_{1n}(x) \leq{}& \sum_{t=1}^{n}\Delta_n^{-2}(x_t)\,K[(x_t-x)/h]\left(\sum_{\substack{s=1\\ s\neq t}}^{n}[f(x_s)-f(x_t)]K[(x_s-x_t)/h]\right)^2 \\ &+\sum_{t=1}^{n}\Delta_n^{-2}(x_t)\,K[(x_t-x)/h]\left(\sum_{s=1}^{n}\sigma(x_s)u_s\,K[(x_s-x_t)/h]\right)^2 \\ :={}& \Delta_{1n1}(x)+\Delta_{1n2}(x). \end{aligned} \tag{5.54}$$

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

It follows from similar arguments as in (5.53) that

$$\Delta_{1n1}(x) \le Ch^{2\alpha} \sum_{t=1}^{n} K[(x_t - x)/h] = O_P\left(\sqrt{n}h^{1+2\alpha}\right) = o_P\left[\left(nh^2\right)^{1/4}\right],$$

whenever $nh^{2+8\alpha} \to 0$, which yields $(\Delta_{1n1}(x))/(\Delta_n(x)) = o_P[(nh^2)^{-1/4}]$. Note that, as in (5.51),

$$\begin{aligned}\mathbb{E}\{\Delta_{1n2}(x)\Xi_n\} &= \mathbb{E}\left\{\Xi_n \sum_{\substack{s,t=1 \\ s\neq t}}^{n} \sigma^2(x_s)\, u_s^2\, \Delta_n^{-2}(x_t)\, K^2[(x_s - x_t)/h] K[(x_t - x)/h]\right\} \\ &\le C \sup_{x\in\Omega_\epsilon} \sigma^2(x) \mathbb{E}\left\{\Xi_n \sum_{t=1}^{n} \Delta_n^{-1}(x_t) K[(x_t - x)/h]\right\} \\ &\le CM.\end{aligned}$$

As in the proof of (5.52), we have $(\Delta_{1n2}(x))/(\Delta_n(x)) = o_P[(nh^2)^{-1/4}]$. Taking these estimates into (5.54), we obtain the required $I_{3n}(x) = (\Delta_{2n1}(x))/(\Delta_n(x)) = o_P[(nh^2)^{-1/4}]$. The proof of (5.43) is now complete and hence that of Theorem 3.1. ■

NOTE

1. Proposition 7.2 of Wang and Phillips (2009b) requires $\int_{-\infty}^{\infty} |\hat{K}(t)|dt < \infty$, where $\hat{K}(t) = \int_{-\infty}^{\infty} e^{itx} K(x)dx$. But this condition is only used in the proof of tightness, not for the finite-dimensional distribution convergence.

REFERENCES

Bandi, F. (2004) On Persistence and Nonparametric Estimation (with an Application to Stock Return Predictability). Manuscript, University of Chicago.

Berman, S.M. (1969) Local times and sample function properties of stationary Gaussian processes. *Transactions of the American Mathematical Society* 137, 277–299.

Borodin, A.N. & I.A. Ibragimov (1995) *Limit Theorems for Functionals of Random Walks*. Proceedings of the Steklov Institute of Mathematics, vol. 195. *American Mathematical Society*.

Cai, Z., Q. Li, & J.Y. Park (2009) Functional-coefficient models for nonstationary time series data. *Journal of Econometrics* 148, 101–113.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary $AR(1)$ process. *Annals of Statistics* 15, 1050–1063.

Chen, J., D. Li, & L. Zhang (2010) Robust estimation in nonlinear cointegrating model. *Journal of Multivariate Analysis* 101, 707–717.

Choi, I. & P. Saikkonen (2004) Testing linearity in cointegrating smooth transition regressions. *Econometrics Journal* 7, 341–365.

Choi, I. & P. Saikkonen (2010) Tests for nonlinear cointegration. *Econometric Theory*, 26, 682–709.

Chung, H. & J.Y. Park (2007) Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics* 137, 230–259.

Csörgö, M. & P. Révész (1981) *Strong Approximations in Probability and Statistics*. Probability and Mathematical Statistics. Academic Press.

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

de la Pena, V.H. (1999) A general class of exponential inequalities for martingales and ratios. *Annals of Probability* 27, 537–564.

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

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009a) Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* 25, 1869–1892.

Gao, J., M. King, Z. Lu, & D. Tjøstheim (2009b) Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics* 37, 3893–3928.

Gao, J., D. Li, & D. Tjøstheim (2009) Uniform Consistency for Nonparametric Estimates in Null Recurrent Time Series. Working Paper series 0085, University of Adelaide, School of Economics.

Granger, C.W.J. & T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships*. Oxford University Press.

Guerre, E. (2004) Design-Adaptive Pointwise Nonparametric Regression Estimation for Recurrent Markov Time Series. CREST Working paper 2004-22.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application,* Probability and Mathematical Statistics. Academic Press.

Han, H. & J.Y. Park (2008) Time series properties of ARCH processes with persistent covariates. *Journal of Econometrics* 146, 275–292.

Han, H. & S. Zhang (2009) Nonstationary Semiparametric ARCH Model and Nonparametric Cointegrating Volatility Model. Working paper, National University of Singapore.

Hansen, B.E. (1995) Regression with nonstationary volatility. *Econometrica* 63, 1113–1132.

Hansen, B.E. (2008) Uniform convergence rate for kernel estimation with dependent data. *Econometric Theory* 24, 726–748.

Harvey, A.C. & P.M. Robinson (1988) Efficient estimation of nonstationary time series regression. *Journal of Time Series Analysis* 9, 201–214.

Karlsen, H.A., T. Myklebust, & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration model. *Annals of Statistics* 35, 252–259.

Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.

Kasparis, I. & P.C.B. Phillips (2009) Dynamic Misspecification in Nonparametric Cointegrating Regression. Cowles Foundation Discussion paper 1700.

Kim, C.S. & J.Y. Park (2010) Cointegrating regressions with time heterogeneity. *Econometric Review* 29, 397–438.

Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory* 25, 1433–1445.

Lukács, E. (1970) *Characteristic Functions*. Hafner Publishing.

Marmer, V. (2008) Nonlinearity, nonstationarity, and spurious forecasts. *Journal of Econometrics* 142, 1–27.

Park, J.Y. (2002) Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics* 110, 383–415.

Park, J.Y. & S.B. Hahn (1999) Cointegrating regressions with time varying coefficients. *Econometric Theory* 15, 664–703.

Park, J.Y. & P.C.B. Phillips (1999) Asymptotics for nonlinear transformation of integrated time series. *Econometric Theory* 15, 269–298.

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56, 1021–1044.

Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion paper 1181.

Quintos, C.E. & P.C.B. Phillips (1993) Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675–706.

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press

Robinson, P.M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891.
Schienle, M. (2008) Nonparametric nonstationary regression. Ph.D. dissertation, University of Mannheim.
Sun, Y., Z. Cai, & Q. Li (2008) Consistent Nonparametric Test on Parametric Smooth Coefficient Model with Nonstationary Data. Manuscript, University of Toronto.
Wang, Q. & P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.
Wang, Q. & P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
Wang, Q. & P.C.B. Phillips (2011) Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory* 27, 235–259.
Xiao, Z. (2009) Functional-coefficient cointegration models. *Journal of Econometrics* 152, 81–92.

https://doi.org/10.1017/S0266466612000205 Published online by Cambridge University Press