Title
A Fixed-bandwidth View of the Pre-asymptotic Inference for Kernel Smoothing with Time Series Data

Authors
Kim, Min Seong
Sun, Yixiao
Yang, Jingjing

Publication Date
2016-01-04

A Fixed-bandwidth View of the Pre-asymptotic Inference for Kernel Smoothing with Time Series Data*

Min Seong Kim
Department of Economics
Ryerson University

Yixiao Sun
Department of Economics
UC San Diego

Jingjing Yang
Department of Economics
University of Nevada, Reno

September 16, 2015

Abstract

This paper develops robust testing procedures for nonparametric kernel methods in the presence of temporal dependence of unknown forms. Based on the fixed-bandwidth asymptotic variance and the pre-asymptotic variance, we propose a heteroskedasticity and autocorrelation robust (HAR) variance estimator that achieves double robustness — it is asymptotically valid regardless of whether the temporal dependence is present or not, and whether the kernel smoothing bandwidth is held constant or allowed to decay with the sample size. Using the HAR variance estimator, we construct the studentized test statistic and examine its asymptotic properties under both the fixed-smoothing and increasing-smoothing asymptotics. The fixed-smoothing approximation and the associated convenient t-approximation achieve extra robustness — it is asymptotically valid regardless of whether the truncation lag parameter governing the covariance weighting grows at the same rate as or a slower rate than the sample size. Finally, we suggest a simulation-based calibration approach to choose smoothing parameters that optimize testing oriented criteria. Simulation shows that the proposed procedures work very well in finite samples.

Keywords: heteroskedasticity and autocorrelation robust variance, calibration, fixed-smoothing asymptotics, fixed-bandwidth asymptotics, kernel density estimator, local polynomial estimator, t-approximation, testing-optimal smoothing-parameters choice, temporal dependence

JEL Classification Number: C12, C14, C22

*Email: minseong.kim@ryerson.ca, yisun@ucsd.edu, and jingjingy@unr.edu. For helpful comments, we thank seminar and conference participants at OSU, U of Guelph, CESG, KAEA-KEA, and 2015 ESWC. Kim gratefully acknowledges research support from SSHRC (430-2015-00527). Sun gratefully acknowledges partial research support from NSF under Grant No. SES-1530592. Address correspondence to Yixiao Sun, Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA.

1 Introduction

This paper proposes new robust testing procedures for nonparametric kernel methods in the presence of temporal dependence of unknown forms. An important issue in hypothesis testing with time series data is how to take nonparametric dependence into account in calculating the standard error. Dependence is typical in time series, and an estimator with positively dependent data tends to have larger variation than that with *iid* data. Therefore, if temporal dependence is not properly considered, we may have an over-rejection problem. For parametric models this has been a well-researched problem since Newey and West (1987) and Andrews (1991). It is now standard practice to use the heteroskedasticity and autocorrelation robust (HAR) standard error in empirical studies.

No such procedure for nonparametric kernel methods has been proposed in the literature. The fact that the distribution of a kernel estimator with dependent data is asymptotically equivalent to the distribution with *iid* data may have masked the need for more robust nonparametric kernel methods in finite samples. See Robinson (1983) for detail on the asymptotic equivalence. This is in sharp contrast to the parametric case. The asymptotic equivalence implies that the usual standard error formula with *iid* data is still valid for time series data in an asymptotic sense. However, in finite samples, temporal dependence does affect the sampling distribution of a kernel estimator. In particular, when a process is highly persistent and/or the sample size is not large enough, the asymptotic variance tends to understate the true finite sample variation of a kernel estimator, and this understated variation causes the usual asymptotic test to over-reject in finite samples. See, for example, Conley, Hansen and Liu (1997) and Pritsker (1998), who discuss this problem in kernel density estimation based on short term interest rates.

In developing new and more accurate testing procedures, the paper makes several contributions. Firstly, based on the fixed-bandwidth asymptotic variance and the "pre-asymptotic" variance of a kernel estimator, we construct a kernel based HAR variance estimator that captures temporal dependence. The "pre-asymptotic" approach has also been used in Chen, Liao and Sun (2014) in sieve inference on time series models. Here the "pre-asymptotic" variance is further justified using the new fixed-bandwidth asymptotics where the kernel-smoothing bandwidth h is held fixed. The proposed HAR variance estimator achieves double robustness — it is asymptotically valid regardless of whether the temporal dependence is present or not, and whether the kernel smoothing bandwidth is held constant or allowed to decay with the sample size.

Secondly, we consider the asymptotic properties of the HAR variance estimator and the associated test statistics under various specifications of the smoothing parameters. There are two smoothing parameters in our testing procedures. The first is the kernel smoothing bandwidth parameter h, and the second is the truncation lag parameter b for covariance weighting, which is parametrized as the ratio of the truncation lag S_T to the sample size T. Both h and b can be fixed or small in our asymptotic specifications. Under the small-h specification, $h \to 0$ but $hT \to \infty$. Similarly, under the small-b specification, $b \to 0$ but $bT \to \infty$.

Under the fixed-h and small-b specification, the asymptotic properties of the HAR variance estimator and the associated test statistic resemble what one would obtain in a parametric setting. Under the small-h and small-b specification, the asymptotic bias and variance of the HAR variance estimator are determined jointly by h and b. This is in contrast with the parametric setting where the bias and variance trade-off is dictated by b only.

Regardless of whether h is fixed or small, the fixed-b asymptotics delivers new limiting distributions of the HAR estimator and the associated t-statistic. Under the fixed-b asymptotics, the degree of (periodogram) smoothing is fixed, and the HAR variance estimator converges in distribution to a random variable which is proportional to the true variance. As a result, the t-statistic is asymptotically equivalent to a ratio of a standard normal variable to the square root of an independent random weighting variable. The randomness of the HAR variance estimator is embedded in this random weighting variable. The asymptotically equivalent distribution is nonstandard but easy to simulate because it is a function of T *iid* standard normal variables. We also extend Sun (2014a) to approximate the fixed-b asymptotic distribution by a Student's t-distribution.

Since Kiefer, Vogelsang and Bunzel (2000) and Kiefer and Vogelsang (2002a, 2002b, 2005), the fixed-smoothing asymptotics, which includes the fixed-b asymptotics as an example, has gained quite some attention in the econometrics literature. Jansson (2004) and Sun, Phillips and Jin (2008) show the higher order accuracy of the fixed-b asymptotic approximation as compared to the normal approximation. The fixed-smoothing asymptotics has also been employed in different settings, for example, Bester, Conley, Hansen and Vogelsang (2015) and Sun and Kim (2015) in the spatial setting, and Gonçalves (2012), Kim and Sun (2013) and Vogelsang (2012) in the panel data setting. Sun (2014b) proposes a fixed-smoothing asymptotic test in the two-step GMM setting. In a nonparametric setting, Chen, Liao and Sun (2014) propose the fixed-smoothing asymptotic tests for the sieve method with time series data. No such procedure is considered for the kernel method, though it is one of most popular estimation methods for nonparametric and semiparametric models. From this point of view, our paper makes an important contribution, providing autocorrelation robust tests based on nonparametric kernel estimators.

Finally, we suggest a smoothing-parameter choice procedure to select h and b jointly based on some testing oriented criteria. In the parametric setting, Andrews (1991) and Newey and West (1994) choose the smoothing parameter to minimize the asymptotic MSE and propose parametric and nonparametric plug-in implementations. Sun and Phillips (2009) and Sun (2014a) consider testing oriented criteria and parametric plug-in methods. For the choice of h, undersmoothing is suggested for inference. However, no explicit formula or data driven procedure is available in the literature. Our methods are testing oriented in that we minimize the type II error after controlling the type I error or the size distortion. As h and b are chosen jointly, it is very difficult to express the type I and II errors as functions of these two smoothing parameters. To solve this problem, we suggest a simulation-based calibration procedure. For example, in the case of kernel density estimation and inference, we first use an AR model to calibrate the temporal dependence in the data, and then simulate the type I and type II errors based on this model. While Gao and Gijbels (2008) suggest a testing-optimal bandwidth parameter for nonparametric kernel testing, they assume that the error term is *iid*.

The remainder of the paper is as follows. Section 2 gives an overview of the problem in kernel density estimation. Sections 3-5 focus on kernel density estimation and inference. We present the main ideas with greater details in those sections. In particular, Section 3 motivates the HAR variance estimator from two different perspectives: the fixed-bandwidth perspective and the pre-asymptotic perspective. In the conventional small-b asymptotic framework, this section also establishes the consistency, rate of convergence and approximate MSE of the HAR variance estimator. In this small-b framework, we accommodate both the new fixed-bandwidth specification where h is held fixed and the conventional small-bandwidth specification where h decays to zero with the sample size T but hT still diverges. Section 4 develops the asymptotic theory of the HAR variance estimator and the associated test statistics under the fixed-b asymptotics.

Section 5 introduces testing oriented criteria for the choice of the smoothing parameters and proposes a simulation-based calibration approach for implementation. Section 6 extends our testing procedures to the local polynomial estimator. The Monte Carlo simulation results are reported in Section 7. The last section concludes. Proofs of the results are given in the Appendix.

2 Kernel Density Inference: Overview of the Problem

Given a strictly stationary time series $\{X_t\}_{t=1}^T$, we are interested in estimating its marginal probability density function $f(x)$ for a point $x \in \mathcal{X}$, the support of X. The kernel estimator $\hat{f}(x)$ of $f(x)$ is

$$\hat{f}(x) = \frac{1}{hT}\sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right) = \frac{1}{T}\sum_{t=1}^{T} K_h(X_t - x),$$

where $K(\cdot)$ is a real-valued kernel function, $K_h(u) = 1/hK(u/h)$, and h is the smoothing parameter.

Define the α-mixing coefficient:

$$\alpha(\ell) = \sup_{A\in\mathcal{F}_{-\infty}^t,\ B\in\mathcal{F}_{t+\ell}^{\infty}} |P(A\cap B) - P(A)P(B)| \text{ with } \ell > 0,$$

where $\mathcal{F}_a^b$ is the σ-field generated by $\{X_t\}_{t=a}^b$.

To establish the asymptotic properties of $\hat{f}(x)$, we maintain the following assumptions, which are typical in the literature.

Assumption 1 *$f(\cdot)$ is continuously differentiable up to order $(r+1)$ in a neighborhood around x.*

Assumption 2 *(i) $K(u)$ is continuous with compact support. (ii) $\int K(u)\,du = 1$, $\int u^j K(u)\,du = 0$ for $j = 1, \ldots, q-1$ and $\int u^q K(u)\,du \neq 0$ where $q = r+1$.*

Assumption 3 *(i) $\{X_t\}_{t=1}^T$ is a strictly stationary and mixing process with α-mixing coefficient satisfying $|\alpha(\ell)| \leq C\ell^{-\beta}$ for some $\beta > 2$. (ii) The probability density $g_{r-s}(u,v)$ of X_r and X_s for $r \neq s$ is uniformly bounded in the sense that $\sup_\ell \sup_{u\in\mathcal{X},v\in\mathcal{X}} g_\ell(u,v) \leq C$ for some $C \in (0,\infty)$.*

The compact support assumption in Assumption 2 is made for convenience. Our asymptotic results remain valid if this assumption is replaced by an assumption that regulates the tail behaviors of $K(\cdot)$, but the proofs will be longer and more tedious.

Define

$$Z_{t,h}(x) := \sqrt{h}K_h(X_t - x) - E\sqrt{h}K_h(X_t - x).$$

Theorem 1 *Let Assumptions 1-3 hold. If $h \to 0, Th \to \infty$ and $Th^{1+2q} \to 0$ as $T \to \infty$, then*

$$\begin{aligned} Var\left[\sqrt{Th}\hat{f}(x)\right] &= Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{t,h}(x)\right) = Var\left[Z_{t,h}(x)\right](1+o(1)) \\ &\to f(x)\left(\int K^2(u)\,du\right) := V(x), \end{aligned} \tag{1}$$

and

$$\sqrt{Th}\left(\hat{f}\left(x\right)-f\left(x\right)\right)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{h}\left[K_{h}\left(X_{t}-x\right)-f(x)\right]\rightarrow^{d}\mathcal{N}\left(0,V\left(x\right)\right) \tag{2}$$

for any interior point x in the support of $f(\cdot)$.

A proof can be found in many papers, for example, Robinson (1983).

The asymptotic distribution in (2) is exactly the same as what we would obtain with *iid* data. It says that temporal dependence does not affect the first order asymptotic distribution of $\hat{f}(x)$. This "dependence irrelevant" result is due to the localization property of kernel smoothing. $\hat{f}(x)$ can be regarded as the relative and weighted frequency of $\{X_t\}_{t=1}^{T}$ in the neighborhood $[x-h/2, x+h/2]$ of x, where the weight is adjusted by the kernel function. For consistency, we require $h \rightarrow 0$ as $T \rightarrow \infty$ so that smoothing is taken in a small and shrinking neighborhood. More importantly, smoothing is taken in the domain of x with no regard for time. Observations that are close in time and hence correlated may receive very different weights with consequential smaller correlation after kernel transformation. For these reasons,

$$\frac{1}{T}\sum_{t\neq\tau}Cov\left[Z_{t,h}\left(x\right),Z_{\tau,h}\left(x\right)\right]\rightarrow 0 \text{ as } h\rightarrow 0 \text{ and } T\rightarrow\infty.$$

As a result, the variance term becomes dominant.

Conventional testing procedures are designed on the basis of the asymptotic result in (2). By having a sequence of h such that $h \rightarrow 0, Th \rightarrow \infty$ and $Th^{1+2q} \rightarrow 0$ as $T \rightarrow \infty$, the standardized statistic $t_{0T}(x)$ converges in distribution to the standard normal as follows:

$$t_{0T}\left(x\right):=\frac{\sqrt{Th}\left(\hat{f}\left(x\right)-f\left(x\right)\right)}{\sqrt{\hat{V}\left(x\right)}}\rightarrow^{d}\mathcal{N}\left(0,1\right),\text{ where }\hat{V}\left(x\right)=\hat{f}\left(x\right)\left(\int K^{2}\left(u\right)du\right). \tag{3}$$

In finite samples, however, h is a positive number, and the effect of temporal dependence on the distribution of $\hat{f}(x)$ does not vanish completely. In particular, if the time series is highly persistent and/or the sample size is not large enough, inference based on the asymptotic result in (3) may suffer from serious size distortion. Robinson (1983) also points out that the accuracy of the asymptotic approximation in finite samples depends critically on the choice of h.

Figure 1 presents some simulation evidence. It plots the empirical type I error of the conventional 5% test under different degrees of temporal dependence. The data is generated from

$$X_t = \rho X_{t-1} + \varepsilon_t \tag{4}$$

with $X_0 \sim \mathcal{N}(0,1)$ and $\varepsilon_t \sim^{iid} \mathcal{N}(0,1)$. To minimize the effect of initialization, we generate a time series of length $2T$ and drop the first T observations, and so our observations are $\{X_t\}_{t=T+1}^{2T}$. The coefficient in the AR(1) process ρ represents the strength of dependence. The marginal density function $f(x)$ of $X_t \sim \mathcal{N}\left(0, 1/\left(1-\rho^2\right)\right)$ is nonparametrically estimated using the Gaussian kernel. Undersmoothing is taken by choosing $h = 1.06\hat{\sigma}_x T^{-1/3}$, where $\hat{\sigma}_x^2$ is the sample variance of X_t. The vertical axis denotes the empirical type I errors, and the horizontal axis denotes the values of x at which $f(x)$ is estimated. The number of simulation replications is 5000. The figure clearly shows a significant discrepancy between the empirical type I error and the nominal level when the process is highly persistent. For example, when $\rho = 0.95$ and $T = 200$, the empirical

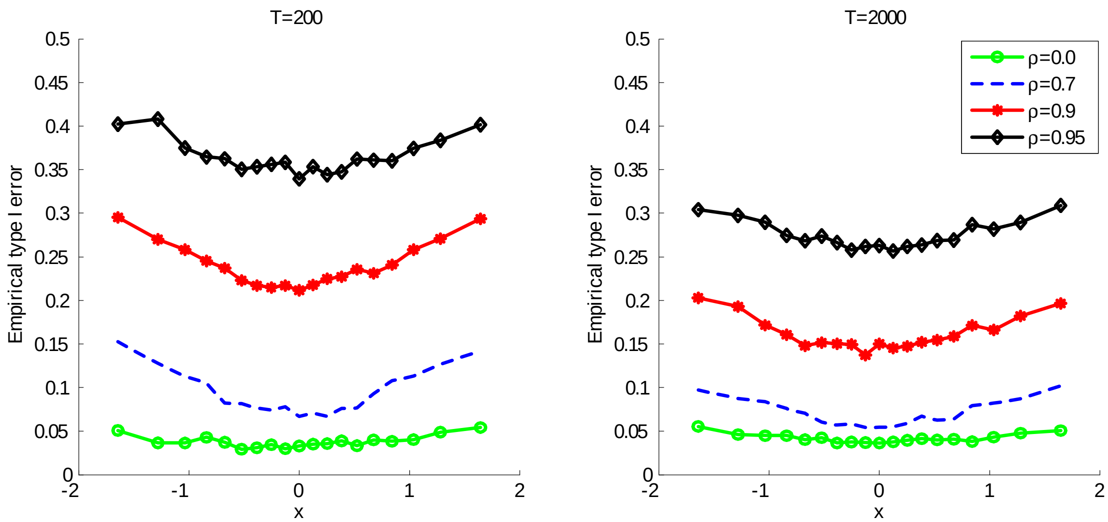

Figure 1: Empirical type I error for the 5% asymptotic normal test based on the asymptotic variance estimator

type I error at $x = 0$ is about 0.35, and it gets even higher when x is near the tails. Increasing the sample size to a high but empirically relevant value is not enough to eliminate this problem. We see some improvement with $T = 2000$, but the size distortion is still apparent. The results are consistent with the findings in Pritsker (1998) and motivate us to develop new testing procedures that take temporal dependence into account.

3 HAR Inference: Motivation and Small-b Asymptotics

In this section, we provide two different motivations for the HAR variance estimator that accommodates temporal dependence. We examine its asymptotic properties under the small-b asymptotics.

3.1 Fixed-h Asymptotic HAR Variance Estimator

A key condition behind the "dependence irrelevant" result is that $h \to 0$ as $T \to \infty$. The decaying rate of h ensures that $\hat{f}(x)$ is consistent. Mathematically, the decaying rate of h also makes $K_h(\cdot)$ behave like the Dirac delta function so that the transformed time series $\{K_h(X_t - x)\}$ has weaker dependence as h decreases. Had we held h fixed as $T \to \infty$, the temporal dependence would remain in the asymptotic approximation. We call the asymptotics under which h is held fixed as $T \to \infty$ the fixed-h asymptotics or the fixed-bandwidth asymptotics.

Theorem 2 *Let Assumptions 2(i) and 3(i) hold. Assume that*

$$V_h(x) = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_{t,h}(x)\right) > 0 \textit{ for each fixed } h > 0.$$

Then for any point x in the support of $f(\cdot)$,

$$\sqrt{Th}\left(\hat{f}(x) - E\hat{f}(x)\right) \to^d \mathcal{N}(0, V_h(x)) \tag{5}$$

as $T \to \infty$ for a fixed h.

For each h, $V_h(x)$ is the long run variance of $\{Z_{t,h}(x)\}$:

$$\begin{aligned} V_h(x) &= Var\left[Z_{t,h}(x)\right] + 2 \lim_{T\to\infty} \sum_{\ell=1}^{T-1} \left(1 - \frac{\ell}{T}\right) Cov\left[Z_{1,h}(x), Z_{\ell+1,h}(x)\right] \\ &= Var\left[Z_{t,h}(x)\right] + 2\sum_{\ell=1}^{\infty} Cov\left[Z_{1,h}(x), Z_{\ell+1,h}(x)\right]. \end{aligned}$$

For a given h, there is no reason to expect the second term, which involves the autocovariances of all lags, to be of smaller order than the first term, which is the variance term. So, in general, the long run variance $V_h(x)$ will not become degenerate. For a general nonparametric estimation problem like this, hT is often regarded as the effective sample size. The nonparametric nature of the problem as $h \to 0$ is reflected in the slower rate of divergence of the effective sample size compared to the actual sample size T. However, if h is fixed, then the effective sample size diverges at the same rate as the actual sample size. To some extent, holding h fixed reduces the nonparametric problem locally to a parametric one. It is well-known that temporal dependence usually affects the asymptotic variance in a parametric setting. In view of this likening of our problem to a parametric one, Theorem 2 is quite intuitive.

Compared with Theorem 1, Theorem 2 centers $\hat{f}(x)$ at its expected value instead of the true density function $f(x)$. In essence, we have ignored the bias of the kernel density estimator. The extra conditions that are needed for Theorem 1 ensure that the bias is asymptotically negligible and the "dependence irrelevant" result holds. The change of centering releases us from worrying about the bias effect.

An alternative view of our strategy may be helpful here. After we decompose the estimation error into a sampling error and a nonsampling error (i.e., the bias), we can free ourselves by allowing potentially different asymptotic specifications when developing approximations to these two types of errors. Conventionally, we insist on using a single asymptotic specification dictated by the asymptotically unbiased requirement. As a result, we obtain the "dependence irrelevant" result that is too optimistic in many economic applications where the underlying time series has high positive autocorrelation. Here we look at the sampling error separately from the nonsampling error and choose an asymptotic specification that may deliver a more accurate approximation.

In principle, we should use $f(x)$ as the center so that we can make inferences on the true density function. However, using $f(x)$ as the center, coupled with the MSE optimal bandwidth, leads to a confidence interval (CI) for $f(x)$ that is asymptotically invalid in that the asymptotic coverage probability differs from the intended coverage probability. To construct an asymptotically valid CI, we often employ undersmoothing with an ad hoc choice of the bandwidth. Another approach is to estimate the bias and make inferences based on the bias-reduced kernel density estimator. However, it is often very difficult to estimate the bias, as it involves high order derivatives of the density function. So no satisfactory solution exists. For this reason, empirical researchers often ignore the bias in constructing the CI's. But this effectively means that the CI's are based on Theorem 2 above.

While Theorem 2 requires that h be fixed, we do not have to fix h at a given value for all sample sizes and data generating processes. We think of the fixed-h asymptotic specification as an asymptotic device to establish a more accurate asymptotic approximation. In fact, in empirical applications, the sample size T is usually given beforehand and the bandwidth parameter needs to be determined using *a priori* information and/or information obtained from the data. Very often, the selected h is smaller for a larger sample size but the effective sample size is still large. So the empirical situations appear to be more compatible with the requirement that $h \to 0$ and $Th \to \infty$. However, once h is chosen, it takes a particular value for the sample under consideration. We can use this value as the fixed value and plug it into the fixed-h asymptotics to conduct inferences. As the fixed value h becomes smaller, we can show that $V_h(x)$ becomes closer to $V(x)$. So $V_h(x)$ is a more robust measure of the sampling variation than $V(x)$.

Following the standard practice in the HAR literature, we can estimate $V_h(x)$ by

$$\hat{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\hat{Z}_{t,h}(x)\hat{Z}_{\tau,h}(x), \tag{6}$$

where

$$\hat{Z}_{t,h}(x) = \sqrt{h}\left[K_h(X_t - x) - \frac{1}{T}\sum_{s=1}^{T}K_h(X_s - x)\right] = \sqrt{h}\left[K_h(X_t - x) - \hat{f}(x)\right],$$

$W(\cdot)$ is a covariance weighting function, and $S_T = S_T(x)$ is the truncation (lag) parameter. To avoid possible confusion, we restrain from referring to $W(\cdot)$ as a kernel function and S_T as the bandwidth parameter, as these two terms have been used for $K(\cdot)$ and h, respectively. If we reparametrize S_T in terms of the ratio $b = S_T/T$, then we have

$$\hat{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{bT}\right)\hat{Z}_{t,h}(x)\hat{Z}_{\tau,h}(x), \tag{7}$$

where b is now the smoothing parameter for covariance weighting.

3.2 Pre-asymptotic HAR Variance Estimator

In the conventional asymptotic framework where $h \to 0$ and $hT \to \infty$ as $T \to \infty$, the estimator in (6) can be motivated by a "pre-asymptotic" argument. We compute the exact finite sample variance of $\sqrt{Th}\hat{f}(x)$ as follows:

$$\begin{aligned} V_{T,h}(x) = Var\left(\sqrt{Th}\hat{f}(x)\right) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} E\left[Z_{t,h}(x) Z_{\tau,h}(x)\right] \\ &= Var\left[Z_{t,h}(x)\right] + 2\sum_{\ell=1}^{T-1}\left(1 - \frac{\ell}{T}\right) Cov\left[Z_{1,h}(x), Z_{\ell+1,h}(x)\right]. \end{aligned} \tag{8}$$

The above calculation involves no asymptotics and thus holds for any value of h and T. It is the variance expression that we can get our hands on before developing any asymptotic approximation. For this reason, we call $V_{T,h}(x)$ the "pre-asymptotic" variance.

Compared to $V_h(x)$, $V_{T,h}(x)$ is equal to $V_h(x)$ except that it involves only the autocovariances up to order $T-1$. So it can be regarded as a finite sample version of $V_h(x)$. Compared to

$V(x)$, $V_{T,h}(x)$ involves further autocovariance terms. Though the second term in (8) vanishes asymptotically under weak dependence as $h \to 0$, it can be nontrivial when T is not large and h is not small. While $V_{T,h}(x)$ is the exact variance of $\sqrt{Th}\hat{f}(x)$, $V(x)$ is an approximation to $V_{T,h}(x)$. In finite samples, it is reasonable to estimate $V_{T,h}(x)$ directly instead of its large sample approximation $V(x)$.

In view of the definition of $V_{T,h}(x)$, a natural estimator is then given by $\hat{V}_h(x)$ in (6). This gives an alternative motivation for $\hat{V}_h(x)$. Regardless of the asymptotic devices we are comfortable working with, the use of a HAR variance can be justified.

3.3 Fixed-h Asymptotic Properties

To examine the asymptotic properties of $\hat{V}_h(x)$, we define its infeasible version as

$$\tilde{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right) Z_{t,h}(x) Z_{\tau,h}(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{bT}\right) Z_{t,h}(x) Z_{\tau,h}(x). \quad (9)$$

Under some conditions, we will show that the difference between $\hat{V}_h(x)$ and $\tilde{V}_h(x)$ is of smaller order than the squared root of the variance of $\tilde{V}_h(x)$. Hence, we can obtain an approximate measure of the mean squared error (MSE) of $\hat{V}_h(x)$ using the asymptotic MSE of $\tilde{V}_h(x)$.

Define

$$W^{(p_0)} = \lim_{\xi \to 0} \frac{1 - W(\xi)}{|\xi|^{p_0}} \text{ for } p_0 \in [0, \infty),$$

and let $p = \max\left\{p_0 : W^{(p_0)} < \infty\right\}$ be the *Parzen characteristic exponent* of $W(\xi)$. The magnitude of p reflects the smoothness of $W(\xi)$ at $\xi = 0$.

We make some mild conditions on the covariance weighting function.

Assumption 4 *(i) The covariance weighting function $W(\cdot)$ satisfies $W(0) = 1$, $|W(\xi)| \leq 1$, $W(\xi) = W(-\xi)$, and $W(\xi) = 0$ for $|\xi| \geq 1$; (ii) For all $\xi_1, \xi_2 \in \mathbb{R}$ there is a constant, $c_L < \infty$, such that*

$$|W(\xi_1) - W(\xi_2)| \leq c_L |\xi_1 - \xi_2|.$$

(iii) The Parzen characteristic exponent p of $W(\cdot)$ is greater than or equal to 1.

Most commonly used covariance weighting functions satisfy this condition. Some examples are the Bartlett, Tukey-Hanning and Parzen kernels. The quadratic spectral (QS) kernel does not satisfy this assumption because it does not truncate. We can generalize the results to include the QS kernel, but this requires much more complicated proofs.

To characterize the asymptotic bias of $\hat{V}_h(x)$, we define, for each fixed h :

$$B_h^{(p)}(x) = 2 \lim_{T\to\infty} \sum_{\ell=1}^{T-1} Cov\left[Z_{1,h}(x), Z_{1+\ell,h}(x)\right] \ell^p, \quad (10)$$

which in general depends on h. By the mixing inequality in Section 1.2.2 of Doukhan (1994), we have

$$Cov\left[Z_{1,h}(x), Z_{1+\ell,h}(x)\right] \leq 4 \|K\|_\infty \alpha(\ell)/h,$$

where $\|K\|_\infty = \sup_{u\in\mathbb{R}} |K(u)|$. It then follows that $\sum_{\ell=1}^{T-1} Cov\left[Z_{1,h}(x), Z_{1+\ell,h}(x)\right] \ell^p$ converges absolutely if $\sum_{\ell=1}^{\infty} \ell^p \alpha(\ell) < \infty$. The latter condition holds if the value of β in Assumptions 3(i) is greater than $p+1$. Hence the condition $\beta > p+1$ guarantees the existence of $B_h^{(p)}(x)$ for each h.

Theorem 3 *Let Assumptions 2(i), 3(i) and 4 hold with* $\beta > \max\{p+1, 3\}$. *Suppose that h is fixed,* $S_T \to \infty$, *and* $S_T/T \to 0$ *as* $T \to \infty$. *Then*

(a) $\lim_{T\to\infty} (T/S_T) Var(\tilde{V}_h(x)) = 2\left(\int_{-1}^{1} W(\xi)^2 d\xi\right)[V_h(x)]^2$;

(b) $\lim_{T\to\infty} S_T^p \left(E\tilde{V}_h(x) - V_h(x)\right) = -W^{(p)} B_h^{(p)}(x)$;

(c) $\sqrt{T/S_T}\left(\hat{V}_h(x) - \tilde{V}_h(x)\right) = o_p(1)$.

The rate conditions "$S_T \to \infty$ and $S_T/T \to 0$" are equivalent to "$b \to 0$ and $bT \to \infty$". For easy reference, we call the asymptotics under $b \to 0$ and $bT \to \infty$ the small-b asymptotics. This is to be compared with the fixed-b asymptotics when b is held fixed as $T \to \infty$ in the next section.

Using Theorem 3 and a Nagar-type (Nagar, 1959) moment approximation, we obtain the approximate MSE (AMSE) of $\hat{V}_h(x)$ as

$$\text{AMSE}\left(\hat{V}_h(x)\right) = \frac{2S_T}{T}\left(\int_{-1}^{1} W(\xi)^2 d\xi\right)[V_h(x)]^2 + \frac{1}{S_T^{2p}}\left[W^{(p)}\right]^2\left(B_h^{(p)}(x)\right)^2.$$

This can be also justified based on the asymptotic truncated MSE criterion proposed and studied by Andrews (1991). Here we are content with the Nagar approximation, as we will not gain much additional insight from the asymptotic truncated MSE calculation.

It follows that the AMSE optimal S_T is given by

$$S_T^* = \left(\frac{p\left[W^{(p)} B_h^{(p)}(x)\right]^2}{\left(\int_{-1}^{1} W(\xi)^2 d\xi\right)[V_h(x)]^2}\right)^{\frac{1}{2p+1}} T^{\frac{1}{2p+1}}. \tag{11}$$

A plug-in implementation such as Andrews (1991) can be employed to obtain a data-driven choice of S_T.

Combining Theorems 2 and 3, we obtain the following corollary.

Corollary 1 *Let Assumptions in Theorems 2 and 3 hold. Suppose that h is fixed,* $S_T \to 0$, *and* $S_T/T \to \infty$ *as* $T \to \infty$. *Then*

$$t_{1T}(x) := \frac{\sqrt{Th}\left(\hat{f}(x) - E\hat{f}(x)\right)}{\sqrt{\hat{V}_h(x)}} \to^d \mathcal{N}(0,1).$$

The corollary provides the usual basis for making inferences on $E\hat{f}(x)$. If we ignore the bias in practical situations, the corollary also provides the basis for making inferences on $f(x)$.

3.4 Small-h Asymptotic Properties

In this subsection, we consider the asymptotic properties of the HAR variance estimator and the studentized t-statistic using the small-h asymptotics under which $h \to 0$ and $hT \to \infty$ as $T \to \infty$.

While the asymptotic properties of $\hat{V}_h(x)$ for a fixed h follow from the standard arguments such as those in Andrews (1991), the asymptotic properties under the small-h specification require new arguments. The reason is that the standard arguments require a finite τ-th moment of $Z_{t,h}(x)$ for some $\tau > 2$. However, when $h \to 0$, any moment of $Z_{t,h}(x)$ of order higher than 2 diverges with h. For this reason, we have to control the higher order dependence using a stronger mixing condition.

Define the fourth order cumulant function of $(Z_{t,h}(x), Z_{t+\ell_1,h}(x), Z_{t+\ell_2,h}(x), Z_{t+\ell_3,h}(x))$ as

$$\begin{aligned} & Q_T(\ell_1,\ell_2,\ell_3) \\ & = E\left[Z_{1,h}(x) Z_{1+\ell_1,h}(x) Z_{1+\ell_2,h}(x) Z_{1+\ell_3,h}(x)\right] - E\left[\tilde{Z}_{1,h}(x)\tilde{Z}_{1+\ell_1,h}(x)\tilde{Z}_{1+\ell_2,h}(x)\tilde{Z}_{1+\ell_3,h}(x)\right], \end{aligned}$$

where $\{\tilde{Z}_{t,h}(x)\}$ is a Gaussian sequence with the same mean and covariance structure as $\{Z_{t,h}(x)\}$.

Lemma 1 *Let Assumptions 2(i) and 3 hold with $\beta > \max\{2(p+1), 5\}$. Assume that the probability densities of $(X_{r_1}, X_{r_2}, X_{r_3})$ and $(X_{r_1}, X_{r_2}, X_{r_3}, X_{r_4})$ are bounded uniformly over mutually different values of r_1, r_2, r_3, r_4. Suppose $h \to 0$ as $T \to \infty$. Then*

(a) $\sum_{\ell_1=0}^{T-1}\sum_{\ell_2=0}^{T-1}\sum_{\ell_3=1}^{T-1}|Q_T(\ell_1,\ell_2,\ell_3)| = O\left(h^{-2/\beta}\right)$;

(b) there exists $B^{(p)}(x) < \infty$ *such that* $2h^{-\nu}\sum_{\ell=1}^{T-1} Cov\left[Z_{1,h}(x), Z_{1+\ell,h}(x)\right]\ell^p \to B^{(p)}(x)$ *where* $\nu = 1 - 2(p+1)/\beta$.

Lemma 1(a) helps control the error when we replace $\{Z_{t,h}(x)\}$ by the Gaussian sequence $\{\tilde{Z}_{t,h}(x)\}$ in computing the asymptotic variance of $\tilde{V}_h(x)$. For parametric models such as those considered in Andrews (1991), it is typical that $\sum_{\ell_1=0}^{T-1}\sum_{\ell_2=0}^{T-1}\sum_{\ell_3=0}^{T-1}|Q_T(\ell_1,\ell_2,\ell_3)| < \infty$. Due to the lack of moment conditions, Lemma 1(a) is weaker on two fronts. First, $Q_T(0,0,0)$ is excluded from the sum, as $E[Z_{t,h}^4(x)]$ diverges under the small-h asymptotics. Second, the sum is not bounded even after excluding $Q_T(0,0,0)$. Instead, it diverges as $h \to 0$ albeit at a manageable rate. Nevertheless, the weaker result with the upper bound given in Lemma 1(a) is sufficient for our asymptotic variance calculation.

Lemma 1(b) facilitates the asymptotic bias calculation. Here $B^{(p)}(x)$ is similar to that of $B_h^{(p)}(x)$ defined in (10). While $B_h^{(p)}(x)$ depends on h, $B^{(p)}(x)$ does not. The difference arises because we normalize $2\sum_{\ell=1}^{T-1} Cov\left[Z_{1,h}(x), Z_{1+\ell,h}(x)\right]\ell^p$ by h^ν to ensure its convergence as $h \to 0$. In the parametric setting, we often assume that $2\sum_{\ell=1}^{T-1} Cov\left[Z_{1,h}(x), Z_{1+\ell,h}(x)\right]\ell^p$ is finite. In our setting, this sum converges to zero when the temporal dependence is weak enough. So our HAR variance estimator is expected to have a smaller bias, reflecting the whitening effect of the kernel transformation.

Using Lemma 1, we can prove Theorem 4, which summarizes the asymptotic properties of $\hat{V}_h(x)$ under the small-h asymptotics.

Theorem 4 *Let Assumptions 2-4 and the extra assumptions in Lemma 1 hold. Suppose $h \to 0, S_T \to \infty$ but $S_T/T \to 0$ and $Th \to \infty$ as $T \to \infty$.*

(a) For $\phi_T = \min\{Th, T/S_T\}$,

$$\lim_{T\to\infty}\phi_T Var\left(\tilde{V}_h(x)\right) = \begin{cases} f(x)\int K(u)^4 du, & \text{if } hS_T \to 0, \\ f(x)\left(\int K(u)^4 du\right)/\gamma + 2\left(\int_{-1}^1 W(\xi)^2 d\xi\right) V(x)^2, & \text{if } hS_T \to \gamma, \\ 2\left(\int_{-1}^1 W(\xi)^2 d\xi\right) V(x)^2, & \text{if } hS_T \to \infty, \end{cases}$$

where $\gamma \in (0, \infty)$.

(*b*) $\lim_{T\to\infty} S_T^p/h^\nu \left[E\tilde{V}_h(x) - V_h(x)\right] = -W^{(p)}B^{(p)}(x)$.

(*c*) $\sqrt{\phi_T}\left(\hat{V}_h(x) - \tilde{V}_h(x)\right) = o_p(1)$.

(*d*) $V_h(x) = V(x) + o(1)$.

From Theorem 4(a) and (b), we can see that the variance and bias of $\tilde{V}_h(x)$ depend not only on the truncation parameter S_T but also on the bandwidth parameter h. This is in contrast to the case of parametric models in which the trade-off is based on the choice of S_T. Theorem 4(a) shows that the variance of $\tilde{V}_h(x)$ has two sources. The first is the fourth moment of the kernel process. As h shrinks to zero as $T \to \infty$, $E[Z_{t,h}(x)^4]$, which is of order $1/h$ and is unbounded, contributes to the variance inflation of $\tilde{V}_h(x)$. In the parametric setting, the fourth moment is assumed to be finite in general, and its contribution to the variance is asymptotically negligible. The second source is the usual variance term of the kernel LRV estimator. As usual, this term increases with S_T. Depending on the relative rate of the two smoothing parameters, one of them may dominate the other.

Theorem 4(b) shows that the bias of $\tilde{V}_h(x)$ decreases as we reduce the degree of smoothing in estimating $f(x)$ or the degree of smoothing in estimating $V_h(x)$. For the former, the degree of smoothing is indicated by h. For the latter, the degree of smoothing is indicated by $1/S_T$. The bias decreases as h decreases because the strength of autocorrelation attenuates as $h \to 0$. The bias decreases as S_T increases because the downweighing and truncating become less severe as $S_T \to \infty$.

It follows from Theorem 4(c) that

$$\sqrt{\phi_T}\left(\hat{V}_h(x) - V_h(x)\right) = \sqrt{\phi_T}\left(\tilde{V}_h(x) - V_h(x)\right) + o_p(1).$$

Assuming that $\phi_T h^{2\nu}/S_T^{2p} \to C \in (0, \infty)$, we can invoke Nagar's argument to obtain the approximate MSE of $\hat{V}_h(x)$:

$$\text{AMSE}\left(\hat{V}_h(x)\right) = \frac{1}{Th} f(x) \int K(u)^4 du + 2\frac{S_T}{T}\left(\int_{-1}^{1} W(\xi)^2 d\xi\right) V(x)^2 + \frac{h^{2\nu}}{S_T^{2p}}\left(W^{(p)}B^{(p)}(x)\right)^2,$$

which is actually equal to the asymptotic MSE of $\tilde{V}_h(x)$. Note that the first term does not depend on S_T. For each given h, we can find the AMSE optimal S_T to be

$$\begin{aligned}\tilde{S}_T^* &= \arg\min_{S_T} \frac{S_T}{T} 2\left(\int_{-1}^{1} W(\xi)^2 d\xi\right) V(x)^2 + \frac{h^{2\nu}}{S_T^{2p}}\left(W^{(p)}B^{(p)}(x)\right)^2 \\ &= \left(\frac{p\left[W^{(p)}h^\nu B^{(p)}(x)\right]^2}{\left(\int_{-1}^{1} W(\xi)^2 d\xi\right)[V(x)]^2}\right)^{\frac{1}{2p+1}} T^{\frac{1}{2p+1}}. \end{aligned} \tag{12}$$

Obviously, $\tilde{S}_T^*$ decreases with h. As h becomes smaller, the kernel transformed time series $\{Z_{t,h}(x)\}$ becomes less correlated, and so a smaller truncation lag parameter is desired.

The formula in (12) is the same as that in (11) except that $B_h^{(p)}(x)$ and $V_h(x)$ are replaced by $h^\nu B^{(p)}(x)$ and $V(x)$ respectively. Note that both $B_h^{(p)}(x)$ and $h^\nu B^{(p)}(x)$ are approximately equal to

$$2\sum_{\ell=1}^{T-1} Cov\left[Z_{1,h}\left(x\right), Z_{1+\ell,h}\left(x\right)\right]\ell^{p}.$$

We can employ a plug-in parametric model to estimate this quantity and use it as the estimate for both $B_h^{(p)}(x)$ and $h^{\nu}B^{(p)}(x)$. Similarly, we can use the same plug-in value, i.e., the implied long run variance from the parametric model, for $V_h(x)$ and $V(x)$. In this case, the data driven $\hat{S}_T$ will be the same regardless of which formula we use. While the two formulae are obtained under different asymptotic specifications and have different justifications, they lead to the same practice in empirical applications.

Under the AMSE optimal $\tilde{S}_T^*$ given above and the assumption that $B^{(p)}(x) \in (0, \infty)$, we have

$$\text{AMSE}\left(\hat{V}_h\left(x\right)\right) \asymp \frac{1}{Th} + \frac{\left(Th^{2\nu}\right)^{\frac{1}{2p+1}}}{T} = \frac{1}{Th} + T^{-\frac{2p}{2p+1}}h^{\frac{2\nu}{2p+1}},$$

where '$\asymp$' signifies that the two sides are of the same order. To balance the two terms, we choose h such that $h \asymp T^{-\frac{1}{2p+2\nu+1}}$ under which

$$\text{AMSE}\left(\hat{V}_h\left(x\right)\right) \asymp T^{-\frac{2(p+\nu)}{2p+2\nu+1}}.$$

The rate of convergence of $\hat{V}_h(x)$ to $V_h(x)$ is then $T^{-\frac{p+\nu}{2p+2\nu+1}}$. If $\nu > 0$, this rate is faster than $T^{-p/(2p+1)}$, which is the convergence rate of the HAR variance estimator in the parametric setting. For commonly used kernel and weighting functions, we have $p = q = 2$. In this case, it is well-known that the MSE optimal h for the point estimation of $f(x)$ satisfies $h \asymp T^{-1/(2q+1)} = T^{-1/5}$. Undersmoothing, which is often suggested for inference, requires that $h = o\left(T^{-1/5}\right)$. Neither rate is compatible with the AMSE optimal rate for the point estimation of $V_h(x)$. Thus, the optimal convergence rate of $\hat{V}_h(x)$ is not achieved. This provides a motivation for alternative rules for smoothing-parameter choice considered in Section 5.

It follows from Theorem 4(d) that $\hat{V}_h(x)$ is consistent for $V(x)$ under the small-h and small-b asymptotics. In the previous subsection, we have shown that $\hat{V}_h(x)$ is consistent for $V_h(x)$ under the fixed-h and small b asymptotics. So no matter whether h is fixed or small, $\hat{V}_h(x)$ converges to the target we want. We may conclude that $\hat{V}_h(x)$ achieves the robustness with respect to the asymptotic specification of h. By design, $\hat{V}_h(x)$ is also robust to the presence of temporal dependence or its absence. Therefore, $\hat{V}_h(x)$ enjoys double robustness.

The corollary below follows from Theorems 1 and 4.

Corollary 2 *Let Assumptions 1-4 and the extra assumptions in Lemma 1 hold. Then, we have*

$$t_{2T}\left(x\right) := \frac{\sqrt{Th}\left(\hat{f}\left(x\right) - f\left(x\right)\right)}{\sqrt{\hat{V}_h\left(x\right)}} \to^d \mathcal{N}\left(0,1\right)$$

if $h \to 0, S_T \to \infty, T/S_T \to \infty, Th \to \infty$ *and* $Th^{1+2q} \to 0$ *as* $T \to \infty$.

Corollary 2 is the same as Corollary 1 except for the centering difference. This centering difference is derived from that between Theorems 1 and 2. So the discussions on the centering issue after Theorem 2 apply here as well.

4 HAR Inference: Fixed-b Asymptotics

In this section, we hold $b = S_T/T$ fixed while letting $T \to \infty$ and develop the so-called fixed-b asymptotic approximations to $t_{1T}(x)$ and $t_{2T}(x)$. To emphasize their dependence on b, we now write $\hat{V}_h(x, b) = \hat{V}_h(x)$, $t_{1T}(x, b) = t_{1T}(x)$, and $t_{2T}(x, b) = t_{2T}(x)$.

Letting $W_b(s) = W(s/b)$, we can write

$$\hat{V}_h(x, b) = \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \tilde{W}_{b,T}\left(\frac{t}{T}, \frac{\tau}{T}\right) Z_{t,h}(x) Z_{\tau,h}(x),$$

where

$$\tilde{W}_{T,b}\left(\frac{t}{T}, \frac{\tau}{T}\right) = W_b\left(\frac{t-\tau}{T}\right) - \frac{1}{T}\sum_{s_1=1}^{T} W_b\left(\frac{t-s_1}{T}\right) - \frac{1}{T}\sum_{s_2=1}^{T} W_b\left(\frac{s_2-\tau}{T}\right) + \frac{1}{T^2}\sum_{s_1=1}^{T}\sum_{s_2=1}^{T} W_b\left(\frac{s_1-s_2}{T}\right)$$

is the demeaned version of $W_b((t-\tau)/T)$. As $T \to \infty$, we obtain the "continuous" version of $\tilde{W}_{T,b}(t/T, \tau/T)$ below:

$$W_b^*(t, \tau) = W_b(t-\tau) - \int_0^1 W_b(t - s_1)\, ds_1 - \int_0^1 W_b(s_2 - \tau)\, ds_2 + \int_0^1 \int_0^1 W_b(s_1 - s_2)\, ds_1 ds_2.$$

By construction, $\int_0^1 W_b^*(s, \tau)\, ds = \int_0^1 W_b^*(t, s)\, ds = 0$ for any t and τ.

Assumption 5 *For $b \in (0, 1]$, $W_b(\cdot)$ is symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable on $[-1, 1]$.*

Assumption 5 strengthens Assumption 4 by requiring $W_b(\cdot)$ to be piecewise monotonic and piecewise continuously differentiable. Assumption 5 is still mild and satisfied by commonly used covariance weighting functions such as the Bartlett and Parzen kernels.

Under Assumption 5, $W_b^*(t, \tau)$ has the following representation:

$$W_b^*(t, \tau) = \sum_{n=1}^{\infty} \lambda_n \Psi_{bn}(t) \Psi_{bn}(\tau), \tag{13}$$

where $\{\Psi_{bn}(\cdot)\}$ is a sequence of continuously differentiable functions satisfying $\int_0^1 \Psi_{bn}(r)\, dr = 0$. The right hand side of (13) converges absolutely and uniformly over $(t, \tau) \in [0, 1] \times [0, 1]$. See Sun (2014a) for further discussion.

Define $\Psi_{b0}(\cdot) \equiv 1$. We introduce the following high level assumption.

Assumption 6 *Under both the small-h asymptotic sequence where $h \to 0$ and $Th \to \infty$ as $T \to \infty$ and the fixed-h asymptotic sequence where h is held fixed as $T \to \infty$, the following holds:*

$$P\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{bn}\left(\frac{t}{T}\right) Z_{t,h}(x) \le \zeta \text{ for } n = 0, 1, \ldots, \mathcal{L}\right)$$
$$= P\left(\frac{\sqrt{V_h(x)}}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{bn}\left(\frac{t}{T}\right) e_t \le \zeta \text{ for } n = 0, 1, \ldots, \mathcal{L}\right) + o(1) \text{ as } T \to \infty$$

for every fixed $\mathcal{L}$ where $e_t \sim^{iid} \mathcal{N}(0, 1)$, $b \in (0, 1]$ and $\zeta \in \mathbb{R}$.

Assumption 6 is satisfied if a CLT holds jointly over $n = 0, 1, 2, \ldots, \mathcal{L}$ for

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{bn}\left(\frac{t}{T}\right) Z_{t,h}\left(x\right)$$

under the two asymptotic sequences given in the assumption.

When Assumption 6 holds, we say $T^{-1/2}\sum_{t=1}^{T} \Psi_{bn}\left(t/T\right) Z_{t,h}\left(x\right)$ is asymptotically equivalent in distribution to $\left(V_h\left(x\right)/T\right)^{1/2} \sum_{t=1}^{T} \Psi_{bn}\left(t/T\right) e_t$, and write

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{bn}\left(\frac{t}{T}\right) Z_{t,h}\left(x\right) \sim^a \frac{\sqrt{V_h\left(x\right)}}{\sqrt{T}} \sum_{t=1}^{T} \Psi_{bn}\left(\frac{t}{T}\right) e_t. \tag{14}$$

Some primitive sufficient conditions for Assumption 6 are provided in the web appendix of Sun and Kim (2014).

Proposition 1 *Let Assumptions 2(i), 3, 5 and 6 hold. Then under both the small-h asymptotic sequence and the fixed-h asymptotic sequence, we have, for* $b \in (0, 1]$:

$$\hat{V}_h(x, b) \sim^a V_{T,h}^a(x, b) := \frac{V_h\left(x\right)}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} W_b\left(\frac{t-\tau}{T}\right)\left(e_t - \bar{e}\right)\left(e_\tau - \bar{e}\right)$$

where $\bar{e} = T^{-1}\sum_{s=1}^{T} e_s$ *and the asymptotic equivalence holds jointly with (14).*

Proposition 1 shows that $V_{T,h}^a(x, b)$ is a random variable proportional to $V_h\left(x\right)$. Based on this result, we can derive the asymptotically equivalent distribution of the studentized statistic. Under the fixed-b and fixed-h asymptotics, we have

$$\begin{aligned} t_{1T}\left(x, b\right) &= \frac{T^{-1/2}\sum_{t=1}^{T} Z_{t,h}\left(x\right)}{\sqrt{\hat{V}_h\left(x, b\right)}} \sim^a \frac{T^{-1/2}\sum_{t=1}^{T} Z_{t,h}\left(x\right)/\sqrt{V_h\left(x\right)}}{\sqrt{\hat{V}_h\left(x, b\right)/V_h\left(x\right)}} \\ &\sim^a \frac{T^{-1/2}\sum_{t=1}^{T} e_t}{\sqrt{T^{-1}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b\left((t-\tau)/T\right)\left(e_t - \bar{e}\right)\left(e_\tau - \bar{e}\right)}} := t_T^a\left(x, b\right). \end{aligned} \tag{15}$$

Under the fixed-b and the small-h asymptotics such that $Th^{1+2q} \to 0$, we have

$$t_{2T}\left(x, b\right) = t_{1T}\left(x, b\right) + o_p\left(1\right) \sim^a t_T^a\left(x, b\right). \tag{16}$$

The asymptotically equivalent distribution $t_T^a\left(x, b\right)$ is a function of T *iid* standard normal random variables. The numerator and denominator of $t_T^a\left(x, b\right)$ in (16) are independent, since $Cov(\sum_{t=1}^{T} e_t, e_\tau - \bar{e}) = 0$ for all $\tau = 1, \ldots, T$. $t_T^a\left(x, b\right)$ is pivotal but not a standard normal variable due to the random denominator. The randomness of $\hat{V}_h(x, b)$ is captured by this random denominator. By subtracting $\bar{e}$, $t_T^a\left(x, b\right)$ also captures the demeaning bias of $\hat{V}_h\left(x, b\right)$, which is due to the use of $\hat{Z}_{t,h}\left(x\right)$ instead of $Z_{t,h}\left(x\right)$ in constructing $\hat{V}_h\left(x, b\right)$.

Theorem 5 summarizes the result above.

Theorem 5 *Let Assumptions 2(i), 3, 5 and 6 hold.*

(a) *For a fixed* $b \in (0,1]$ *and a fixed* h*,*

$$P\left(t_{1T}(x,b) < \zeta\right) = P\left(t_T^a(x,b) < \zeta\right) + o(1), \text{ where } \zeta \in \mathbb{R}$$

as $T \to \infty$.

(b) *If in addition Assumptions 1 and 2(ii) hold, then for a fixed* $b \in (0,1]$ *:*

$$P\left(t_{2T}(x,b) < \zeta\right) = P\left(t_T^a(x,b) < \zeta\right) + o(1) \text{ where } \zeta \in \mathbb{R}$$

when $h \to 0, Th \to \infty$ *and* $Th^{1+2q} \to 0$ *as* $T \to \infty$.

As $b \to 0$, $t_T^a(x,b)$ converges to the standard normal distribution. So regardless of whether b is held fixed or allowed to decay to zero, critical values from the distribution of $t_T^a(x,b)$ are asymptotically valid. Our fixed-b asymptotic test has thus achieved triple robustness – it is asymptotically valid regardless of whether the temporal dependence is present or not, whether the kernel smoothing bandwidth is held constant or allowed to decay with the sample size, and whether the truncation lag, S_T, governing the covariance weighting, grows at the same rate as or at a slower rate than the sample size.

We can conduct the fixed-b asymptotic test by simulating $t_T^a(x,b)$. The simulation is not computationally expensive, because we can obtain a realization of $t_T^a(x,b)$ by drawing T *iid* standard normal variables and by plugging them into the simple representation in (15).

If we want to avoid simulating the nonstandard critical values, we can extend Sun (2014a, Theorems 1 and 4) to establish a t-approximation, which approximates $t_T^a(x,b)$ by a Student's t-distribution. Let

$$c_1 = \int_{-\infty}^{\infty} W(\xi)d\xi \text{ and } c_2 = \int_{-\infty}^{\infty} W^2(\xi)\,d\xi.$$

Theorem 6 formalizes the t-approximation.

Theorem 6 *Let Assumption 5 hold. As* $b \to 0$, *we have*

$$P\left(t_T^a(x,b) \le a\right) = P\left(\frac{1}{\sqrt{\kappa}} t(v^*) \le a\right) + o(b),$$

for $\kappa = 1 - bc_1$ *and* $v^* = \lceil 1/bc_2 \rceil$, *where* $\lceil \cdot \rceil$ *is the ceiling function and* $t(v^*)$ *follows the* t-*distribution with degree of freedom* v^*.

Based on Theorem 6, we can perform the test that uses

$$t_T^*(x,b) = \sqrt{\kappa}\, t_{2T}(x,b)$$

as the test statistic and $t(v^*)$ as the reference distribution. The proposed test is as easy to use as the asymptotic normal test and is able to capture the effect of b to the first order.

5 Choice of Smoothing Parameters

An important issue in conducting the proposed test is to choose the two smoothing parameters. While the MSE optimal b based on (11) or (12) may provide some guidance on the selection of b, there is no reason to expect that such a choice is the most appropriate in hypothesis testing.

In addition, the performance of our test depends not only on the choice of b but also on the choice of h. It is desirable to select them jointly using a testing oriented criterion. In this section, we consider selecting (h, b) to minimize the type II error while controlling for the type I error or the size distortion. For empirical implementation, we suggest a simulation-based calibration approach.

5.1 Basic Idea

Let $t^{\alpha}(v^*)$ denote the $(1-\alpha)$th quantile of the $t(v^*)$ distribution. Suppose that we are interested in the following two-sided test:

$$H_0 : f(x) = f_0 \text{ vs. } H_1 : f(x) \neq f_0.$$

The level α test rejects H_0 if $|t_T^*(x,b)| > t^{\alpha/2}(v^*)$, and the type I error of the test is

$$e_I(h,b) = P(|t_T^*(x,b)| > t^{\alpha/2}(v^*)|H_0).$$

To calculate the power of the test, we consider the local alternative hypothesis

$$H_1(\delta_o) : f(x) = f_0 + c/\sqrt{Th},$$

where $c = c(\delta_o)$ for some noncentrality parameter δ_o. Then, the type II error is

$$e_{II}(h,b) = P(|t_T^*(x,b)| \leq t^{\alpha/2}(v^*)|H_1(\delta_o)).$$

Here we have explicitly written e_I and e_{II} as functions of (h,b). For a one-sided test, depending on the direction of the local alternative, the definitions of $e_I(h,b)$ and $e_{II}(h,b)$ need to be modified accordingly.

We choose the pair of smoothing parameters (h,b) to minimize the type II error while controlling for the type I error. More specifically, the optimal (h,b) solves

$$\begin{aligned}(h^*, b^*) = \arg\min_{(h,b)} P(|t_T^*(x,b)| \leq t^{\alpha/2}(v^*)|H_1(\delta_o)) \\ s.t.\ P(|t_T^*(x,b)| > t^{\alpha/2}(v^*)|H_0) \leq \pi\alpha,\end{aligned} \tag{17}$$

where $\pi > 1$ is the tolerance parameter. We allow the type I error to be different from the nominal type I error α but it cannot be larger than $\pi\alpha$. For example, when $\pi = 1.2$ and $\alpha = 5\%$, the upper bound is 6% rather than 5%.

The above approach, which is also used in Sun (2014a), has a decision theoretic basis, as it amounts to selecting the smoothing parameters to minimize a loss function that is a weighted average of type I and type II errors with the weight given by the implied Lagrangian multiplier for the constraint. Sun, Phillips and Jin (2008) consider the loss-function-based approach explicitly.

Another approach to smoothing-parameter choice involves selecting the smoothing parameters to minimize the absolute error in coverage probability of the confidence intervals or the absolute error in rejection probability (ERP) under the null. See for example, Hall (1992) and Sun and Phillips (2009). For a given α, the absolute ERP is $|e_I(h,b) - \alpha|$. The ERP minimization approach involves selecting the smoothing parameters to minimize $|e_I(h,b) - \alpha|$ while ignoring the type II error. To allow for some flexibility in the ERP control and take the type II error into consideration, we propose to solve the following problem:

$$
\begin{aligned}
(h^*, b^*) &= \arg\min_{(h,b)} P(|t_T^*(x,b)| \leq t^{\alpha/2}(v^*)|H_1(\delta_o)) \\
s.t.\ & \left| P(|t_T^*(x,b)| > t^{\alpha/2}(v^*)|H_0) - \alpha \right| \leq (\pi - 1)\,\alpha
\end{aligned}
\tag{18}
$$

where as before $\pi > 1$ is the tolerance parameter. This problem is the same as (17) except that we now constrain the type I error to be close to the nominal significance level α. This is in contrast to (17) where the type I error can not be too large compared to α but is allowed to be smaller than α to the maximum extent.

Our choice of smoothing parameters involves the noncentrality parameter δ_o. This parameter cannot be consistently estimated from the data. We may choose δ_o to reflect a value of scientific interest if such a value is available. In the absence of such a value, we recommend choosing δ_o such that the first order power of the test, as measured by $1 - G_{1,\delta_o^2}(\chi_1^\alpha)$, is 75%, where $G_{1,\delta_o^2}(\cdot)$ is the cdf of the noncentral chi-squared distribution with the degree of freedom 1 and χ_1^α is the $(1-\alpha)$th quantile of the χ_1^2 distribution. That is, we obtain δ_o by solving $1 - G_{1,\delta_o^2}(\chi_1^\alpha) = 0.75$ for the significance level α.

The constrained minimization problems in (17) and (18) are not operational, as we do not know how type I and II errors depend on (h, b). In the next subsection, we propose a simulation-based calibration method to implement the testing-optimal smoothing parameters.

5.2 Implementation: Simulation-based Calibration

Our simulation-based calibration approach involves the following steps. Let $\mathcal{H}_T = \{h^{(1)}, \ldots, h^{(L)}\}$ and $\mathcal{B}_T = \{b^{(1)}, \ldots, b^{(M)}\}$ be the sets of reasonable h and b values given T.

1. Fit an AR(d) model to X_t

$$X_t = \rho_0 + \rho_1 X_{t-1} + ... + \rho_d X_{t-d} + \epsilon_t, \quad t = d+1, ..., T.$$

   We suggest using either Akaike's Information Criterion (AIC) or Schwarz's Bayesian Information Criterion (BIC) to select the AR order d^*. We may also use the parsimonious AR(1) model.

2. Generate $\mathcal{J}$ pseudo samples $\{X_t^j\}$ according to the fitted AR(d^*) process. We draw the error $\{\epsilon_t^j\}$ independently either from the normal distribution $\mathcal{N}(0, \hat{\sigma}_\epsilon^2)$ or from the empirical distribution of $\{\hat{\epsilon}_t\}$ with $\hat{\epsilon}_t = X_t - (\hat{\rho}_0 + \hat{\rho}_1 X_{t-1} + ... + \hat{\rho}_d X_{t-d^*})$.

3. For each pseudo sample $\{X_t^j\}$, pick $h^{(1)}$ and $b^{(1)}$ to construct the kernel density estimator $\hat{f}^j(x, h^{(1)})$ and variance estimator $\hat{V}_h^j(x, h^{(1)}, b^{(1)})$, where

$$\hat{V}_h^j(x, h^{(1)}, b^{(1)}) = \frac{1}{T} \sum_{t=1}^T \sum_{\tau=1}^T W_{b^{(1)}} \left(\frac{t - \tau}{T} \right) \hat{Z}_{t,h^{(1)}}^j(x)\, \hat{Z}_{\tau,h^{(1)}}^j(x).$$

4. Compute the test statistic $t_{T,j}^*(h^{(1)}, b^{(1)}, c)$

$$t_{T,j}^*(h^{(1)}, b^{(1)}, c) = \sqrt{\kappa} \frac{\sqrt{Th} \left(\hat{f}^j(x, h^{(1)}) - \tilde{E} \hat{f}^j(x, h^{(1)}) \right) + c}{\sqrt{\hat{V}_h^j(x, h^{(1)}, b^{(1)})}},$$

where

$$\tilde{E}\hat{f}^j(x, h^{(1)}) = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} \hat{f}^j(x, h^{(1)}).$$

We choose $c = 0$ to simulate the type I error and $c = \sqrt{\hat{V}(x)}\tilde{c}$ with

$$\tilde{c} = \begin{cases} |\delta_o|, & \text{with probability } 0.5 \\ -|\delta_o|, & \text{with probability } 0.5 \end{cases}$$

to simulate the type II error.

5. Compute

$$\hat{e}_I(h^{(1)}, b^{(1)}) = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} 1\left(\left|t^*_{T,j}(h^{(1)}, b^{(1)}, 0)\right| > t^{\alpha/2}(v^*)\right),$$

and

$$\hat{e}_{II}(h^{(1)}, b^{(1)}) = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} 1\left(\left|t^*_{T,j}(h^{(1)}, b^{(1)}, c))\right| \leq t^{\alpha/2}(v^*)\right),$$

where $1(\cdot)$ is the indicator function.

6. Repeat steps $3 \sim 5$ for each $(h, b) \in \mathcal{H}_T \otimes \mathcal{B}_T$.

7. Find the value of (h, b) that solves

$$\min_{h \in \mathcal{H}_T,\ b \in \mathcal{B}_T} \hat{e}_{II}(h, b), \quad \text{s.t. } \hat{e}_I(h, b) \leq \pi\alpha$$

or

$$\min_{h \in \mathcal{H}_T,\ b \in \mathcal{B}_T} \hat{e}_{II}(h, b), \quad \text{s.t. } |\hat{e}_I(h, b) - \alpha| \leq (\pi - 1)\alpha.$$

The calibration approach via simulation is related to the conventional plug-in procedure based on the MSE criterion in Andrews (1991) or the more recent plug-in approach based on a testing-oriented criterion in Sun, Phillips and Jin (2008). Both procedures involve first deriving an optimal formula, followed by a plug-in implementation using an approximating parametric model. In contrast, for our simulation-based calibration approach, we fit an approximating parametric model first and then simulate the type I and II errors from it. The essential difference is whether the approximating model is used to implement the optimal formula or to simulate the optimal smoothing parameters.

6 Extension to the Kernel Regression

In this section, we extend the proposed robust testing procedures to kernel regression estimators. For notational economy we use some of the same notations as before, but they may stand for different objects in this section. This should not cause any confusion.

6.1 Overview of the Problem

Given the observations $\{(Y_t, X_t)\}_{t=1}^{T}$, we are interested in estimating the regression function or conditional mean function $m(x) = E(Y_t|X_t = x)$. Putting this in a regression form, we have

$$Y_t = m(X_t) + \varepsilon_t,$$

where the error term ε_t satisfies

$$E(\varepsilon_t|X_t = x) = 0,\ Var(\varepsilon_t|X_t = x) = \sigma^2(x).$$

For the simplicity of exposition, we assume that X_t is a scalar process. It is straightforward to generalize our results to a vector process.

When $m(\cdot)$ is smooth enough in a local neighborhood of x, we can approximate $m(\breve{x})$ for $\breve{x}$ near x by a polynomial in $\breve{x}$:

$$\begin{aligned} m(\breve{x}) &\approx m(x) + m^{(1)}(x)(\breve{x} - x) + ... + \frac{m^{(r)}(x)}{r!}(\breve{x} - x)^r \\ &\equiv \theta_0 + \theta_1\left(\frac{\breve{x} - x}{h}\right) + ... + \theta_r\left(\frac{\breve{x} - x}{h}\right)^r \end{aligned}$$

where

$$\theta = (\theta_0, ..., \theta_r)' = \left(m(x), m^{(1)}(x)h, ..., \frac{m^{(r)}(x)h^r}{r!}\right)'$$

is the vector of local parameters, and h is the bandwidth parameter. Fitting this polynomial locally around x leads to the weighted LS problem:

$$\hat{\theta} = \arg\min_{\tilde{\theta}} \sum_{t=1}^{T} K_h(X_t - x)\left\{Y_t - P_t(x;h)'\tilde{\theta}\right\}^2, \tag{19}$$

where

$$P_t(x) := P_t(x;h) = \begin{pmatrix} 1, & (X_t - x)/h, & ... & ((X_t - x)/h)^r \end{pmatrix}',$$

and, as before, $K(\cdot)$ is the kernel function. The local polynomial estimator of $m(x)$ is then given by $\hat{m}(x) \equiv \hat{\theta}_0$. We can also back out the derivative $m^{(j)}$ of $m(x)$ from $\hat{\theta}_j h^{-j}$ for $j = 1, ..., r$, but our focus here is on the function $m(x)$ itself. In the special case that $r = 0$, we obtain the Nadaraya-Watson estimator, i.e.,

$$\hat{m}(x) = \frac{\sum_{t=1}^{T} K_h(X_t - x)Y_t}{\sum_{t=1}^{T} K_h(X_t - x)}.$$

Let

$$P_x = \begin{pmatrix} P_1'(x;h) \\ ... \\ P_T'(x;h) \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}, \hat{\theta} = \begin{pmatrix} \hat{\theta}_0 \\ \vdots \\ \hat{\theta}_r \end{pmatrix}, \text{ and } \Omega_x = diag(K_h(X_t - x)),$$

then $\hat{\theta} = (P_x'\Omega_x P_x)^{-1} P_x'\Omega_x Y$, and

$$\hat{m}(x) = \hat{\theta}_0 = e_1'\left(P_x'\Omega_x P_x\right)^{-1} P_x'\Omega_x Y = \frac{1}{Th}\sum_{t=1}^{T}\omega_T\left(\frac{X_t - x}{h}\right)Y_t,$$

where $e_1 = (1, 0, ..., 0)' \in \mathbb{R}^{r+1}$ and

$$\omega_T(u) = e_1' \left(\frac{P_x' \Omega_x P_x}{T} \right)^{-1} [1, u, ..., u^r]' K(u).$$

$\omega_T(u)$ is often referred to as the *effective* kernel underlying the local polynomial regression. The *effective* kernel enjoys the finite sample high order property, as by construction we have

$$\sum_{t=1}^{T} \frac{1}{h} \omega_T \left(\frac{X_t - x}{h} \right) \left(\frac{X_t - x}{h} \right)^{\kappa} = 1 \{ \kappa = 0 \} \text{ for } \kappa = 0, ..., r.$$

Let J be the $q \times q$ matrix whose (i,j)-th element is given by $\int K(u) u^{i+j-2} du$. Under the small-h asymptotics, $P_x' \Omega_x P_x / T$ converges to $Jf(x)$. As a result, $\omega_T(u)$ converges to $K^*(u)/f(x)$ where

$$K^*(u) = e_1' J^{-1} [1, u, ..., u^r]' K(u) \equiv (a_0 + a_1 u + ... + a_r u^r) K(u)$$

for some constants $a_0, ..., a_r$. $K^*(u)$ is the *equivalent* kernel underlying the local polynomial regression.

To study the asymptotic properties of $\hat{m}(x)$, we maintain the following assumption.

Assumption 7 *(i) $K(u)$ is continuous with a compact support $[-1, 1]$.*

(ii) The conditional densities $f_{X_1|Y_1}(x_1|y_1)$, $f_{X_1, X_{\ell+1}|Y_1, Y_{\ell+1}}(x_1, x_{\ell+1}|y_1, y_{\ell+1})$ are bounded uniformly over $\ell \geq 1$ and $x_1, y_1, x_{1+\ell}, y_{1+\ell}$.

(iii) (X_t, Y_t) is strictly stationary and α-mixing with the α-mixing coefficients satisfying

$$\sum_{\ell=1}^{\infty} \ell^a \alpha(\ell)^{(\epsilon-1)/\epsilon} < \infty \text{ and } E|\varepsilon_1|^{4\epsilon} < \infty$$

for some $a \geq 2$ and $\epsilon > 1$.

(iv) There exists $d_T \to \infty$ and $d_T = o(\sqrt{Th})$ such that $\sqrt{n/h}\alpha(d_T) \to 0$.

(v) $\sigma^2(x)$ and $f(x)$ are continuous at the point x and $f(x) > 0$.

(vi) $m(\cdot)$ is continuously differentiable up to order $q = (r+1)$ in a neighborhood around x.

Theorem 7 *Let Assumption 7 hold. Under the small-h asymptotics where $h \to 0, Th \to \infty$ and $Th^{1+2q} \to 0$ as $T \to \infty$, we have*

$$\sqrt{Th}(\hat{m}(x) - m(x)) \to^d \mathcal{N}(0, V(x))$$

where

$$V(x) = \frac{\sigma^2(x)}{f(x)} \int [K^*(u)]^2 du.$$

For a proof of the theorem, see Masry and Fan (1997) or Section 6.6.2 of Fan and Yao (2003). The compact support of $K(u)$ is taken to be $[-1, 1]$ without loss of generality. Assumption 7 is the same as what is given in Section 6.6.2 of Fan and Yao (2003) except that Assumption 7(iii) is stronger. The stronger version is needed for the theorems in the next subsection.

As in the case of kernel density estimation, the asymptotic variance is identical to what we would obtain for *iid* data. Under the small-h asymptotics, temporal dependence has no effect on the asymptotic variance. Using the standard sandwich formula, we can estimate $V(x)$ by

$$\hat{V}(x) = e_1' \left(\frac{P_x' \Omega_x P_x}{T} \right)^{-1} \left(\frac{h}{T} P_x' \Omega_x \hat{\Sigma}_x \Omega_x P_x \right) \left(\frac{P_x' \Omega_x P_x}{T} \right)^{-1} e_1$$

where $\hat{\Sigma}_x = diag(\hat{\varepsilon}_{tx}^2)$ and $\hat{\varepsilon}_{tx} = Y_t - P_t(x;h)'\hat{\theta}$. Statistical inferences can then be made based on the following asymptotic normality result:

$$t_{0T}(x) = \frac{\sqrt{Th}\left(\hat{m}(x) - m(x)\right)}{\sqrt{\hat{V}(x)}} \to^d \mathcal{N}(0,1).$$

While the small-h asymptotic theory is neat and elegant, the asymptotic approximation may not be accurate in finite samples. In practical situations when the sample size is not very large, temporal dependence may have a large effect on the sampling variation of the local polynomial estimator. Ignoring the temporal dependence will lead to misleading inferences. As discussed in Robinson (1983), the "dependence irrelevant" result is not to be taken too seriously. To alleviate this problem, we can take either the fixed-h approach or the pre-asymptotic approach. This is entirely analogous to the approaches employed for kernel density estimation and inference. In the next subsection, we consider the fixed-h approach as an example.

6.2 The Fixed-bandwidth Solution

For each h, we define the fixed-h probability limit θ_h of $\hat{\theta}$ as

$$\theta_h \equiv (\theta_{0h}, \theta_{1h}, ..., \theta_{rh})' = \operatorname*{plim}_{T \to \infty} \hat{\theta}.$$

In general, $\theta_h \neq \theta$, reflecting the asymptotic bias of $\hat{\theta}$ under the fixed-h asymptotics. Let $J_h(x)$ be the fixed-h probability limit of $(P_x' \Omega_x P_x)/T$. Then

$$\begin{aligned} &\sqrt{Th}\left[\hat{m}(x) - \theta_{0,h}\right] \\ &= e_1' \left(\frac{P_x' \Omega_x P_x}{T} \right)^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} P_t(x;h) \sqrt{h} K_h(X_t - x) \left[Y_t - P_t(x;h)'\theta_h\right] \\ &:= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{t,h}(x)\left(1 + o_p(1)\right) \end{aligned}$$

where

$$Z_{t,h}(x) = e_1' J_h^{-1}(x) P_t(x;h) \sqrt{h} K_h(X_t - x) \varepsilon_{tx} \text{ and } \varepsilon_{tx} = Y_t - P_t(x;h)'\theta_h.$$

Using this representation, we can prove the theorem below.

Theorem 8 *Let Assumption 7(i) and (iii) hold. Assume that (i) $J_h(x)$ is nonsingular, (ii) $E|m(X_t)|^{4\epsilon} 1\{|X_t - x| \leq h\} < \infty$ for some $\epsilon > 1$, and (iii)*

$$V_h(x) = \lim_{T \to \infty} Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_{t,h} \right) > 0 \text{ for each fixed } h > 0.$$

Then for a fixed h as $T \to \infty$,

$$\sqrt{Th}\left[\hat{m}\left(x\right)-\theta_{0,h}\right] \to^{d} \mathcal{N}(0, V_h\left(x\right)).$$

Barring the difference in the definition of $Z_{t,h}$, the asymptotic variance $V_h\left(x\right)$ is defined in the same manner as in Theorem 2. In general, the effect of the temporal dependence in $\{Z_{t,h}\left(x\right)\}$ will not vanish in $V_h\left(x\right)$. As before, $V_h\left(x\right)$ can also be motivated from the pre-asymptotic perspective.

To estimate $V_h\left(x\right)$, we use the formulae given in (6) and (9) to obtain the estimator $\hat{V}_h(x)$ and its infeasible version $\tilde{V}_h(x)$:

$$\hat{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\hat{Z}_{t,h}\left(x\right)\hat{Z}_{\tau,h}\left(x\right) \tag{20}$$

$$\tilde{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right) Z_{t,h}\left(x\right) Z_{\tau,h}\left(x\right) \tag{21}$$

where

$$\hat{Z}_{t,h}\left(x\right) = \left[e_1'\left(\frac{P_x'\Omega_x P_x}{T}\right)^{-1} P_t(x;h)\right] K_h\left(X_t - x\right)\hat{\varepsilon}_{tx} \text{ and } \hat{\varepsilon}_{tx} = Y_t - P_t(x;h)'\hat{\theta}.$$

The results in subsection 3.3 continue to hold with modified conditions. The theorem below gives results analogous to Theorem 3.

Theorem 9 *Let Assumptions 4, 7(i) and (iii) hold with $a \geq p$. Suppose that h is fixed, $S_T \to \infty$, and $S_T/T \to 0$ as $T \to \infty$. Then*

(a) $\lim_{T\to\infty}\left(T/S_T\right) Var(\tilde{V}_h(x)) = 2\left(\int_{-1}^{1} W\left(\xi\right)^2 d\xi\right)\left[V_h(x)\right]^2$;

(b) $\lim_{T\to\infty} S_T^{p}\left(E\tilde{V}_h(x) - V_h\left(x\right)\right) = -W^{(p)}B_h^{(p)}\left(x\right)$;

(c) $\sqrt{T/S_T}\left(\hat{V}_h\left(x\right) - \tilde{V}_h\left(x\right)\right) = o_p\left(1\right)$.

It follows from Theorems 8 and 9 that

$$\frac{\sqrt{Th}\left[\hat{m}\left(x\right)-\theta_{0,h}\right]}{\sqrt{\hat{V}_h\left(x\right)}} \to^{d} \mathcal{N}(0,1).$$

While $\theta_{0,h} \neq m(x)$ for a fixed h, we have $\theta_{0,h} \to m(x)$ as $h \to 0$, and so the above result can be used to make inferences on $m\left(x\right)$ with the same qualification as discussed in the case of kernel density estimation.

Distributional approximations that are more accurate than the normal approximation can be obtained under the fixed-b asymptotics. Operationally, we can estimate the parameter θ_h and conduct inferences as if we are in a locally parametric world with

$$Y_t = P_t(x;h)'\theta_h + \varepsilon_{tx}.$$

For each given h, the fixed-b inference procedures in the literature, such as Kiefer and Vogelsang (2005) and Sun (2014a), can be directly applied. In particular, if the uniform kernel is used, then we can proceed as if we have a linear regression model and a subsample of observations whose X_t belongs to $[x-h, x+h]$. Standard software packages can then be used for estimation and inference. For brevity, we do not present the rigorous conditions underlying the fixed-b asymptotics here.

6.3 Choice of Smoothing Parameters

We employ the same idea as in the kernel density case to select the smoothing parameters. The only remaining issue here is that we have to come up with a parametric approximating model. We propose to fit a model of the form

$$Y_t = \gamma_0 + \sum_{j=1}^{\mathcal{M}_{pilot}} \gamma_j \phi_j (X_t) + e_t \tag{22}$$

to the data by OLS, where $\phi_j(\cdot)$ are some basis functions such as Fourier bases and spline bases. We select $\mathcal{M}_{pilot}$ by AIC or BIC. After the model fitting, we obtain the residual $\hat{e}_t$. We then fit a univariate AR model to each of the time series X_t and $\hat{e}_t$ in order to capture their dynamics. Again, the order of the AR lags can be chosen by AIC or BIC. Essentially, we assume that the nonparametric function $m(\cdot)$ is given by the series regression fitting, and X_t and e_t follow univariate AR(d) processes with the innovation covariance chosen to be the empirical covariance. This completely pins down the approximating data generating process. On the basis of this, we can calibrate the type I and type II errors by simulation. The rest of the smoothing-parameter choice is entirely the same as in the case of kernel density estimation and inference.

7 Monte Carlo Simulation

This section presents some simulation evidence on the finite sample properties of our testing procedures. Both kernel density and regression are considered.

7.1 Kernel Density

We conduct inference on the marginal density function $f(x)$ using the following data generating process:

$$X_t = \rho X_{t-1} + \sqrt{\frac{1-\rho^2}{1-2\rho\theta+\theta^2}}(\varepsilon_t - \theta \cdot \varepsilon_{t-1}), \quad t = 2, \ldots, T \tag{23}$$

with $T = 200$, $X_1 \sim \mathcal{N}(0,1)$ and $\varepsilon_t \sim^{iid} \mathcal{N}(0,1)$. By design, the marginal density is the pdf of the standard normal distribution for all θ and ρ such that $|\rho| < 1$. To compare the size and power properties, we take $\theta = 0, -0.4, 0.4$, $\rho = 0.0, 0.6, 0.9$ and evaluate $f(x)$ at $x = Q_{0.5}, Q_{0.7}$ and $Q_{0.9}$, where Q_a denotes the ath quantile of $\mathcal{N}(0,1)$. We use the Gaussian kernel to construct $\hat{f}(x)$. Results for other kernels are qualitatively similar. The number of simulation replications is 2000.

We compare five different tests. The first one denoted by "asymptotic normal" uses

$$\hat{\mathcal{V}}(x) = \frac{1}{T}\sum_{t=1}^{T} \hat{Z}_{t,h}(x)^2$$

to construct the t-statistic and $\mathcal{N}(0,1)$ to obtain the critical values. Since $\hat{\mathcal{V}}(x)$ is based on $Var[Z_{t,h}(x)]$, it does not allow for the effect of temporal dependence as $\hat{V}_h(x)$. We employ $\hat{\mathcal{V}}(x)$ instead of $\hat{V}_h(x)$ in order to examine the benefit of accommodating the autocovariance terms in $\hat{V}_h(x)$. Note that $\hat{\mathcal{V}}(x)$ is different from $\hat{V}(x)$ given in (3) which relies on deep asymptotics. Our preliminary simulation shows that $\hat{\mathcal{V}}(x)$ delivers a more accurate test than $\hat{V}(x)$. The second procedure denoted by "pre-asymptotic normal" employs $\hat{V}_h(x)$ and the standard normal approximation. We use the Parzen kernel for the covariance weighting function in $\hat{V}_h(x)$. The third method denoted by "pre-asymptotic hybrid" is based on $\hat{V}_h(x)$ and t-approximation. For these three tests, we estimate $f(x)$ with the bandwidth $\tilde{h} = 1.06\hat{\sigma}_x T^{-1/3}$, where $\hat{\sigma}_x^2$ is the sample variance of $\{X_t\}$. Since $\sigma_x = 1$, $\tilde{h}$ is around 0.18. Hall (1992, p. 223) shows that $h \sim cT^{-1/(1+q)}$ minimizes the coverage error for two-sided tests, but no explicit formula is available for the constant c. Our choice of $\tilde{h}$ seems to be reasonable in that the empirical sizes of the asymptotic normal test are close to the nominal level $\alpha = 0.05$ when $(\rho, \theta) = (0, 0)$. See Table 1 below. To implement the pre-asymptotic normal and pre-asymptotic hybrid tests, we choose S_T based on the MSE criterion in (12). We use the parametric plug-in method with the AR(1) model.

The last two tests are the tests proposed in this paper. They are based on $\hat{V}_h(x)$ and the t-approximation and use the calibrated smoothing parameters. The test with the type I error constraint is denoted "calibrated t_I", and the test with the absolute ERP constraint is denoted "calibrated t_{ERP}". We let $\mathcal{H}_T = \{\tilde{h}, \tilde{h} \pm 0.03\hat{\sigma}_x, \tilde{h} \pm 0.06\hat{\sigma}_x\}$ and $\mathcal{B}_T = \{0.01, 0.02, \ldots, 0.3\}$. We also require $h \geq 0.05$ to avoid h being too small. For each testing-oriented criterion considered, if the constraint is violated for all $(h, b) \in \mathcal{H}_T \otimes \mathcal{B}_T$, we choose the combination of (h, b) that has the smallest type I error or ERP. We consider two tolerance parameters $\pi = 1.1$ and 1.2 but report only the representative case with $\pi = 1.2$.

In order to simulate the type II error, we consider the local alternative with $\delta_o = 1.1503$, which solves $1 - G_{1,\delta_o^2}(\chi_1^\alpha) = 0.75$. The AIC is used to select d in the AR(d) model used to calibrate the dependence structure. The number of simulation replications $\mathcal{J}$ used in calibration is 1000.

Table 1 reports the empirical sizes of different 5% tests. From the table, we first observe that temporal dependence does affect the sampling distribution of $\hat{f}(x)$. The asymptotic normal test, which does not account for temporal dependence, tends to suffer from size distortion in the presence of temporal dependence. The size distortion can be severe when the process is highly persistent. For example, when $(\rho, \theta) = (0.9, 0)$ and $f(x)$ is estimated at $x = Q_{0.5}$, the empirical type I error of the asymptotic normal test is 0.279. The empirical type I error is even higher for higher quantile points of interest. Second, we can improve the test accuracy by employing the kernel HAR variance estimator that includes the autocovariance terms. The pre-asymptotic normal test is shown to alleviate the degree of size distortion. Third, comparison between the pre-asymptotic normal test and pre-asymptotic hybrid test shows that the fixed-b asymptotic approximation improves the size accuracy of the test, but the improvement is limited when S_T is chosen based on the MSE criterion. Fourth, the fixed-b asymptotic approximation combined with the proposed smoothing-parameter choice achieves remarkable size accuracy. Except for the extreme cases when $\rho = 0.9$ and $f(x)$ is estimated at $Q_{0.9}$, the empirical type I errors are very close to the nominal level. Finally, the fixed-b asymptotic test performs as well as the asymptotic normal test when $(\rho, \theta) = (0, 0)$, which implies that there is virtually no cost of using the proposed test in terms of size accuracy.

Figure 2 compares the size adjusted power of the three types of tests: the pre-asymptotic normal test, the calibrated t_I test, and the calibrated t_{ERP} test. While the pre-asymptotic normal test employs a rule of thumb to select h and the MSE criterion to select S_T, the t_I and t_{ERP} tests use calibrated smoothing parameters. We use the DGP in (23) but consider the following local alternative hypothesis:

$$H_1(\delta) : f(x) = f_0 + c/\sqrt{Th}$$

where $c = (V(x))^{1/2}\,\tilde{c}$ and

$$\tilde{c} = \begin{cases} \delta, & \text{with probability } 0.5 \\ -\delta, & \text{with probability } 0.5. \end{cases}$$

We compute the power using the 5% empirical critical values from the null distributions. For brevity, Figure 2 reports only a subset of the cases in Table 1. The horizontal and vertical axes represent δ and the size adjusted power respectively. It is clear that the size adjusted powers of the calibrated t_I and t_{ERP} tests are comparable to that of the pre-asymptotic normal test. For some cases the former two tests are more powerful, but for other cases the latter test is more powerful. Regardless of the scenarios, the power difference is not large. We may conclude that our proposed tests improve the size accuracy without sacrificing power.

7.2 Kernel Regression

We consider the following DGP

$$Y_t = \tilde{m}_0(\tilde{X}_t) + \varepsilon_t,\ \ t = 1, 2, ..., T,$$

where

$$\tilde{X}_t = \rho\tilde{X}_{t-1} + \sqrt{1-\rho^2}e_{xt} \text{ and } \varepsilon_t = \rho\varepsilon_{t-1} + \sqrt{1-\rho^2}e_{\varepsilon t},$$

are scaler AR(1) processes with the same AR parameter ρ, and $(e_{xt}, e_{\varepsilon t})'$ are *iid* $\mathcal{N}(0, I_2)$. We set $\tilde{m}_0(\tilde{X}_t) = \sin(\tilde{X}_t)$ or $\cos(\tilde{X}_t)$. The DGP is similar to those of Chen, Liao and Sun (2014, CLS hereafter). The only difference is that we consider a purely nonparametric model here while CLS consider a partially linear model. We focus on $\tilde{h}_0(\tilde{X}_t) = \cos(\tilde{X}_t)$ below as it is harder to be approximated by a linear function around the center of the distribution of $\tilde{X}_t$, but the qualitative results are the same for the case that $\tilde{m}_0(\tilde{X}_t) = \sin(\tilde{X}_t)$.

In order to use sine and cosine basis functions to approximate the unknown function, we first transform $\tilde{X}_t$ into $[0, 1]$ using the transformation:

$$X_t = \frac{1}{1 + \exp(-\tilde{X}_t)}$$

or equivalently $\tilde{X}_t = \ln\left[X_t/(1 - X_t)\right]$. Then

$$\tilde{m}_0(\tilde{X}_t) = \cos\left(\ln\left(\frac{X_t}{1 - X_t}\right)\right) \equiv m_0\left(X_t\right).$$

The new function $m_0(\cdot)$, which is a highly nonlinear function, is our estimand.

We approximate $m_0(\cdot)$ using the sine and cosine functions as follows:

$$m_0(x) = \gamma_0 + \sum_{j=1}^{\mathcal{M}_{pilot}} \gamma_j\phi_j(x),$$

where $\mathcal{M}_{pilot}$ is even, $\phi_{2j-1}(x) = \sqrt{2}\cos(2j\pi x)$, and $\phi_{2j} = \sqrt{2}\sin(2j\pi x)$ for $j = 1, ..., \mathcal{M}_{pilot}/2$. Setting the upper bound for $\mathcal{M}_{pilot}$ to be 20 for $T = 200$, we use the AIC to select the best $\hat{\mathcal{M}}_{pilot}$.

Conditional on $\hat{\mathcal{M}}_{pilot}$, we estimate $\left\{\gamma_j, j=0,...,\hat{\mathcal{M}}_{pilot}\right\}$ by the OLS and the error variance σ_ε^2 by

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{T-\hat{\mathcal{M}}_{pilot}-1}\sum_{t=1}^T \hat{\varepsilon}_t^2 \text{ where } \hat{\varepsilon}_t = Y_t - \hat{\gamma}_0 - \sum_{j=1}^{\hat{\mathcal{M}}_{pilot}} \hat{\gamma}_j \phi_j\left(X_t\right).$$

For the bandwidth choice, we first use Mallows' C_p method, which entails selecting h to minimize $C_p(h)$:

$$C_p(h) = \frac{1}{T}\sum_{t=1}^T [Y_t - \hat{m}_h(X_t)]^2 + \frac{2\hat{\sigma}_\varepsilon^2}{T} tr(L_h).$$

Here $\hat{m}_h(\cdot)$ is the kernel estimator under the bandwidth h and L_h is the $T\times T$ linear smoothing matrix such that $\left(\hat{m}_h\left(X_1\right),...,\hat{m}_h\left(X_T\right)\right)' = L_h(Y_1,...,Y_T)'$. We use the uniform kernel as a representative example. Based on the selected h, we construct the "asymptotic normal" test which ignores the temporal dependence and the "pre-asymptotic normal" test which takes the temporal dependence into consideration and uses the HAR variance estimator. The truncation lag parameter is chosen by the AR(1) plug-in implementation of the MSE-optimal truncation lag as given in Andrews (1991). The Parzen kernel is used as the covariance weighting function. Both tests employ the normal approximations. For the pre-asymptotic test, we also use the t-approximation, leading to the "pre-asymptotic hybrid" test.

The above three tests serve as the benchmark for our proposed tests. To save space, we present only the "calibrated t_I" test results. For this test, we use $\hat{\gamma}_0 + \sum_{j=1}^{\hat{\mathcal{M}}_{pilot}} \hat{\gamma}_j\phi_j\left(x\right)$ as the plug-in model for the nonparametric function, and we use two univariate AR(d) processes fitted to $\{X_t\}$ and $\{\hat{\varepsilon}_t\}$ respectively to gauge the dynamics in X_t and ε_t, and the AR parameters are allowed to be different for the two fitted AR processes. Each AR lag order is selected to minimize the AIC criterion over $1, 2, ..., 20$.

Given the nonparametric function and the specifications of temporal dependence in X_t and ε_t, we can simulate the type I and type II errors of our proposed HAR t test and select h and S_T to optimize our testing oriented criterion. The mechanics is the same as in the case of kernel density estimation and inference. Considering the computation cost, we choose the grid $\mathcal{H}_T\otimes\mathcal{B}_T$ with $\mathcal{H}_T = [0.10 : 0.01 : 0.8]$ and $\mathcal{B}_T = \{0.01, 0.02, \ldots, 0.3\}$.

Tables 2 and 3 report the empirical null rejection probabilities of the four tests under consideration when the tolerance parameter $\pi = 1.2$. It is clear from the tables that the "asymptotic normal" test can have serious size distortion in the presence of strong autocorrelation. The "pre-asymptotic normal" test reduces the size distortion and the "pre-asymptotic hybrid" test reduces the size distortion further. Our proposed "calibrated t_I" test is the least size distorted. In terms of size accuracy, the ranking of the four tests is the same as in the kernel density case. We note that our preferred "calibrated t_I" test can still have considerable size distortion when the temporal dependence is extremely strong. This phenomenon is not unique to the kernel method in the nonparametric setting. The same phenomenon happens to the nonparametric sieve method (e.g., CLS) and parametric methods. In a parametric setting, Sun (2014c) has developed the fixed-smoothing asymptotics in the presence of strong autocorrelation and has shown that the near-unity fixed-smoothing asymptotic approximation can help reduce the size distortion. It will be interesting to extend Sun (2014c) to nonparametric settings. We leave this to future research as it is beyond the scope of the current paper.

We have also examined the power properties of the four tests. We omit the power figures, but we comment on them briefly. We find that the powers for the "pre-asymptotic normal" test and the "calibrated t_I" test are comparable. There is no clear advantage of either method. This is encouraging, as there is often some cost involved in achieving higher size accuracy.

8 Conclusion

Kernel smoothing is one of the most popular nonparametric methods and has been widely studied in both statistics and econometrics. An interesting feature of this nonparametric method is that the distribution of a kernel estimator with weakly dependent data is asymptotically equivalent to that with *iid* data under the conventional asymptotics. Many empirical papers, particularly in finance, have conducted kernel based nonparametric tests with time series data using this asymptotic result. However, the conventional nonparametric tests tend to suffer from serious size distortion and lead to wrong conclusions because temporal dependence does affect the sampling distribution of a kernel estimator in finite samples.

In this paper, we develop new testing procedures for the kernel methods which are robust to the temporal dependence of unknown forms. Both kernel density estimation and local polynomial regression are considered. The proposed tests are based on a kernel HAR variance estimator and the fixed-b asymptotics. We motivate the kernel HAR variance estimator from two different perspectives: the fixed-bandwidth asymptotics and the pre-asymptotic argument. For easy implementation, we establish the validity of a t-approximation to the fixed-b asymptotics. For the choice of the smoothing parameters, we propose the simulation-based calibration approach that optimizes some testing-oriented criterion. A simulation study shows that the proposed tests are much more accurate in size than the conventional tests and have comparable power.

The ideas of the paper, including the kernel HAR variance estimator and the calibration approach to smoothing-parameter choice, are more widely applicable. The fixed-bandwidth asymptotics and the pre-asymptotic argument can be used to justify the use of the HAR variance in any nonparametric model that involves temporal dependence of unknown forms. For example, in a partially linear model, our procedures can be adopted to make more accurate inferences on the nonparametric part of this model. From an operational point of view, we may proceed as if we are in a parametric world, albeit only locally, and use a well-researched parametric analogue to conduct inferences.

Table 1: Empirical type I error of different tests of kernel density with sample size $T = 200$, significance level $\alpha = 0.05$, and ARMA(1,1) errors

	$Q_{0.5}$	$Q_{0.7}$	$Q_{0.9}$	$Q_{0.5}$	$Q_{0.7}$	$Q_{0.9}$	$Q_{0.5}$	$Q_{0.7}$	$Q_{0.9}$
	$(\rho, \theta) = (0, 0)$			$(\rho, \theta) = (0, 0.4)$			$(\rho, \theta) = (0, -0.4)$		
Asymp Normal	0.057	0.060	0.060	0.072	0.040	0.045	0.064	0.070	0.077
Pre-asymp Normal	0.056	0.061	0.063	0.066	0.045	0.057	0.062	0.066	0.071
Pre-asymp Hybrid	0.052	0.059	0.058	0.061	0.042	0.053	0.059	0.062	0.066
Calibrated t_{ERP}	0.068	0.061	0.054	0.068	0.052	0.056	0.062	0.048	0.054
Calibrated t_I	0.062	0.065	0.061	0.058	0.044	0.046	0.053	0.050	0.053
	$(\rho, \theta) = (0.6, 0)$			$(\rho, \theta) = (0.6, 0.4)$			$(\rho, \theta) = (0.6, -0.4)$		
Asymp Normal	0.096	0.102	0.152	0.072	0.078	0.091	0.118	0.136	0.171
Pre-asymp Normal	0.082	0.081	0.115	0.074	0.077	0.079	0.081	0.092	0.103
Pre-asymp Hybrid	0.078	0.076	0.103	0.069	0.074	0.077	0.073	0.083	0.093
Calibrated t_{ERP}	0.046	0.064	0.055	0.042	0.046	0.058	0.064	0.060	0.076
Calibrated t_I	0.087	0.069	0.065	0.063	0.064	0.052	0.056	0.056	0.052
	$(\rho, \theta) = (0.9, 0)$			$(\rho, \theta) = (0.9, 0.4)$			$(\rho, \theta) = (0.9, -0.4)$		
Asymp Normal	0.279	0.327	0.426	0.213	0.246	0.347	0.298	0.351	0.445
Pre-asymp Normal	0.137	0.166	0.228	0.158	0.181	0.253	0.128	0.142	0.201
Pre-asymp Hybrid	0.137	0.145	0.203	0.151	0.172	0.237	0.107	0.123	0.184
Calibrated t_{ERP}	0.063	0.074	0.087	0.054	0.054	0.060	0.056	0.062	0.092
Calibrated t_I	0.076	0.072	0.114	0.069	0.064	0.087	0.058	0.065	0.096

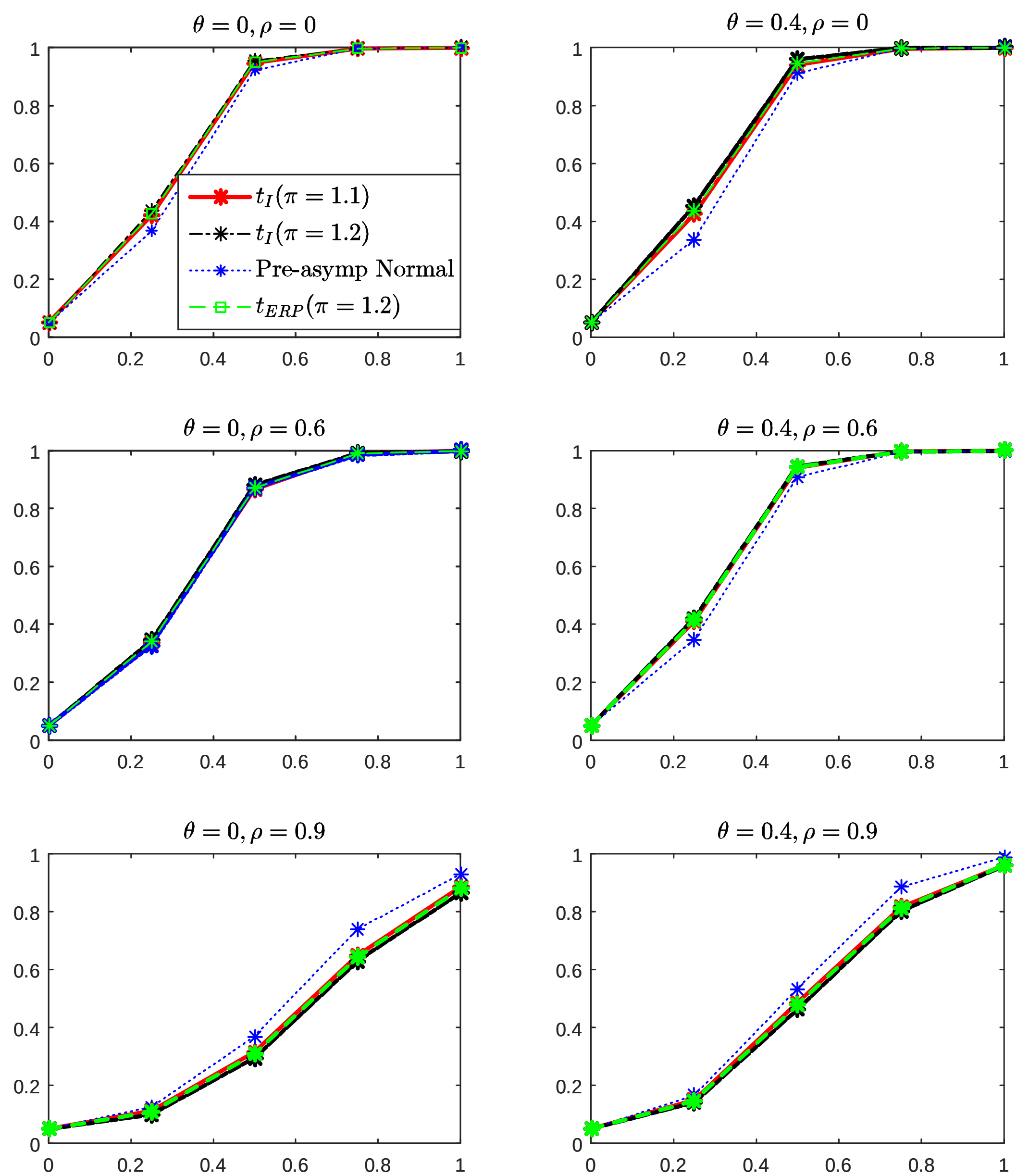

Figure 2: Size-adjusted power of different testing procedures with $T = 200$, ARMA(1,1) errors for $\theta = 0, 0.4$ and $\rho = 0, 0.6, 0.9$, and $x = Q_{0.5}$

Table 2: Empirical type I error of different tests of the nonparametric function with sample size $T = 200$ and significance level $\alpha = 0.05$

	$\tilde{x} = -1$	$\tilde{x} = -0.5$	$\tilde{x} = 0$	$\tilde{x} = 0.5$	$\tilde{x} = 1$
			$\rho = 0.0$		
Asymptotic Normal	0.078	0.052	0.064	0.058	0.072
Pre-asymptotic Normal	0.076	0.062	0.066	0.054	0.076
Pre-asymptotic Hybrid	0.076	0.061	0.064	0.054	0.075
Calibrated t_I	0.052	0.040	0.042	0.037	0.051
			$\rho = 0.25$		
Asymptotic Normal	0.080	0.072	0.092	0.088	0.100
Pre-asymptotic Normal	0.052	0.060	0.068	0.068	0.088
Pre-asymptotic Hybrid	0.053	0.059	0.063	0.064	0.081
Calibrated t_I	0.031	0.042	0.041	0.044	0.055
			$\rho = 0.3$		
Asymptotic Normal	0.118	0.096	0.106	0.130	0.112
Pre-asymptotic Normal	0.104	0.066	0.070	0.116	0.090
Pre-asymptotic Hybrid	0.093	0.063	0.068	0.101	0.074
Calibrated t_I	0.069	0.042	0.044	0.073	0.059
			$\rho = 0.5$		
Asymptotic Normal	0.156	0.164	0.182	0.190	0.176
Pre-asymptotic Normal	0.112	0.078	0.098	0.122	0.118
Pre-asymptotic Hybrid	0.096	0.071	0.082	0.107	0.099
Calibrated t_I	0.075	0.053	0.065	0.068	0.074

Table 3: Empirical type I error of different tests of the nonparametric function with sample size $T = 200$ and significance level $\alpha = 0.05$

	$\tilde{x} = -1$	$\tilde{x} = -0.5$	$\tilde{x} = 0$	$\tilde{x} = 0.5$	$\tilde{x} = 1$
			$\rho = 0.6$		
Asymptotic Normal	0.212	0.216	0.208	0.190	0.208
Pre-asymptotic Normal	0.122	0.106	0.080	0.106	0.126
Pre-asymptotic Hybrid	0.086	0.079	0.069	0.078	0.083
Calibrated t_I	0.073	0.068	0.045	0.064	0.082
			$\rho = 0.7$		
Asymptotic Normal	0.254	0.264	0.268	0.278	0.240
Pre-asymptotic Normal	0.132	0.112	0.100	0.108	0.132
Pre-asymptotic Hybrid	0.097	0.088	0.084	0.086	0.096
Calibrated t_I	0.076	0.071	0.062	0.065	0.088
			$\rho = 0.75$		
Asymptotic Normal	0.300	0.352	0.344	0.302	0.330
Pre-asymptotic Normal	0.144	0.150	0.166	0.128	0.178
Pre-asymptotic Hybrid	0.101	0.114	0.129	0.108	0.139
Calibrated t_I	0.085	0.097	0.101	0.082	0.113
			$\rho = 0.9$		
Asymptotic Normal	0.522	0.526	0.596	0.538	0.500
Pre-asymptotic Normal	0.286	0.226	0.270	0.246	0.286
Pre-asymptotic Hybrid	0.254	0.186	0.210	0.205	0.233
Calibrated t_I	0.173	0.124	0.163	0.138	0.181

9 Appendix

In the appendix, we write $Z_t = Z_{t,h}(x)$ and $\hat{Z}_t = \hat{Z}_{t,h}(x)$ to simplify the notation. We let C be a generic constant.

Proof of Theorem 2. In view of the definition of Z_t, we have

$$\sqrt{nh}\left[\hat{f}(x) - E\hat{f}(x)\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t.$$

We invoke Theorem 0 in Bradley (1985) to complete the proof. The theorem combines two classic theorems in Ibraigomov (1962). Assumption 2(i) implies that $K(\cdot)$ is bounded on its support. So $|Z_t| \le C_K/\sqrt{h} = C$ for some constant $C_K > 0$ almost surely. Since any measurable transformation of a strictly stationary process is also strictly stationary, $\{Z_t\}$ is strictly stationary with the same α-mixing coefficients as $\{X_t\}$. Assumption 3(i) implies that $\sum_{\ell=1}^{\infty}\alpha(\ell) < \infty$. So condition (ii) of Bradley (1985, Theorem 0) holds. As a result,

$$\sqrt{Th}\left(\hat{f}(x) - E\hat{f}(x)\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \to^d \mathcal{N}(0, V_h(x)).$$

■

Proof of Theorem 3. Parts (a) and (b) follow from Proposition 1 of Andrews (1991). The only condition that we need to verify is the mixing condition in his Lemma 1, viz. $\sum_{\ell=1}^{\infty}\ell^2\alpha(\ell)^{(\epsilon-1)/\epsilon} < \infty$ for some $\epsilon > 1$. In our setting, we can choose ϵ as large as possible, as Z_t is bounded for each fixed h. Given that $|\alpha(\ell)| \le C\ell^{-\beta}$ and $\beta > 3$, there is a large enough ϵ such that $\sum_{\ell=1}^{\infty}\ell^2\alpha(\ell)^{(\epsilon-1)/\epsilon} < \infty$. To prove part (c), we note that

$$\begin{aligned}
&\hat{V}_h(x) - \tilde{V}_h(x) \\
&= -\frac{2h}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\left\{\frac{1}{T}\sum_{s=1}^{T}\left[K_h(X_s - x) - EK_h(X_s - x)\right]\right\} K_h(X_\tau - x) \\
&+ \frac{h}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\left\{\left[\frac{1}{T}\sum_{s=1}^{T} K_h(X_s - x)\right]^2 - \left[EK_h(X_s - x)\right]^2\right\} \\
&= h\left\{\frac{1}{T}\sum_{s=1}^{T}\left[K_h(X_s - x) - EK_h(X_s - x)\right]\right\} \\
&\times \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\left[\frac{1}{T}\sum_{s=1}^{T} K_h(X_s - x) + EK_h(X_s - x) - 2K_h(X_\tau - x)\right] \\
&= O_p\left(\sqrt{\frac{h}{T}}\right)\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\left[\frac{1}{T}\sum_{s=1}^{T} K_h(X_s - x) - EK_h(X_s - x)\right] \\
&+ O_p\left(\sqrt{\frac{h}{T}}\right)\frac{2}{T}\sum_{\tau=1}^{T}\left[\sum_{t=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\right]\left[EK_h(X - x) - K_h(X_\tau - x)\right].
\end{aligned}$$

The first term in the above equation is

$$\left[O_p\left(\sqrt{\frac{h}{T}}\frac{1}{\sqrt{Th}}\right)\right]\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\left|W\left(\frac{t-\tau}{S_T}\right)\right|$$
$$=\left[O_p\left(\frac{1}{T}\right)\right]\frac{1}{T}\sum_{t=1}^{T}\sum_{\{\tau:|t-\tau|\le S_T\}}\left|W\left(\frac{t-\tau}{S_T}\right)\right|=O_p\left(\frac{S_T}{T}\right).$$

For the second term, we observe that

$$Var\left\{\frac{1}{T}\sum_{\tau=1}^{T}\left[\sum_{t=1}^{T}W\left(\frac{t-\tau}{S_T}\right)\right][EK_h(X-x)-K_h(X_\tau-x)]\right\}$$
$$\le\frac{C}{T^2h^2}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T}\left|\sum_{t=1}^{T}W\left(\frac{t-\tau_1}{S_T}\right)\right|\cdot\left|\sum_{t=1}^{T}W\left(\frac{t-\tau_2}{S_T}\right)\right|\alpha(\tau_1-\tau_2)=O\left(\frac{S_T^2}{Th^2}\right),$$

and so the second term is of order $O_p\left[\sqrt{h/T}\sqrt{S_T^2/(Th^2)}\right]=O_p\left[S_T/\left(T\sqrt{h}\right)\right]$. Therefore

$$\sqrt{\frac{T}{S_T}}\left[\hat{V}_h(x)-\tilde{V}_h(x)\right]=O_p\left[\sqrt{\frac{T}{S_T}}\left(\frac{S_T}{T}+\frac{S_T}{T\sqrt{h}}\right)\right]=o_p(1).$$

■

Proof of Lemma 1. (i) We have

$$\sum_{\ell_1=0}^{T-1}\sum_{\ell_2=0}^{T-1}\sum_{\ell_3=1}^{T-1}|Q(\ell_1,\ell_2,\ell_3)|=\sum_{\ell_1=1}^{T-1}\sum_{\ell_2=1}^{T-1}\sum_{\ell_3=1}^{T-1}|Q(\ell_1,\ell_2,\ell_3)|+\sum_{\ell_2=1}^{T-1}\sum_{\ell_3=1}^{T-1}|Q(0,\ell_2,\ell_3)|$$
$$+\sum_{\ell_1=1}^{T-1}\sum_{\ell_3=1}^{T-1}|Q(\ell_1,0,\ell_3)|+\sum_{\ell_3=1}^{T-1}|Q(0,0,\ell_3)|. \tag{24}$$

As shown in Andrews (1991, Proof of Lemma 1),

$$\sum_{\ell_1=1}^{T-1}\sum_{\ell_2=1}^{T-1}\sum_{\ell_3=1}^{T-1}|Q(\ell_1,\ell_2,\ell_3)|$$
$$\le 6\sum_{0\le\ell_2,\ell_3\le\ell_1}\left(\left|E\left[Z_1\left(Z_{1+\ell_1}Z_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)\right]\right|+\left|E\left[\tilde{Z}_1\left(\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right]\right|\right)$$
$$+6\sum_{0\le\ell_1,\ell_3\le\ell_2}\left(\left|E\left[Z_1Z_{1+\ell_1}\left(Z_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)\right]-E\left[Z_1Z_{1+\ell_1}\right]E\left[Z_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right]\right|\right.$$
$$\left.+\left|E\left[\tilde{Z}_1\tilde{Z}_{1+\ell_1}\left(\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right]-E\left[\tilde{Z}_1\tilde{Z}_{1+\ell_1}\right]E\left[\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right]\right|\right)$$
$$+6\sum_{0\le\ell_1,\ell_2\le\ell_3}\left(\left|E\left[\left(Z_1Z_{1+\ell_1}Z_{1+\ell_1+\ell_2}\right)Z_{1+\ell_1+\ell_2+\ell_3}\right]\right|+\left|E\left[\left(\tilde{Z}_1\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\right)\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right]\right|\right).$$

First,

$$
\begin{aligned}
\mathcal{A}_1 &:= \sum_{0\le \ell_2,\ell_3\le \ell_1} \left|E\left[Z_1\left(Z_{1+\ell_1}Z_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)\right]\right| \\
&= \sum_{\ell_1=1}^{T-1} \left|E\left[Z_1 Z_{1+\ell_1}^3\right]\right| + \sum_{0<\ell_3\le \ell_1} \left|E\left[Z_1 Z_{1+\ell_1}^2 Z_{1+\ell_1+\ell_3}\right]\right| + \sum_{0<\ell_2\le \ell_1} \left|E\left[Z_1 Z_{1+\ell_1} Z_{1+\ell_1+\ell_2}^2\right]\right| \\
&+ \sum_{0<\ell_2,\ell_3\le \ell_1} \left|E\left[Z_1 Z_{1+\ell_1}Z_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right]\right| \\
&= \mathcal{A}_1^{(1)} + \mathcal{A}_1^{(2)} + \mathcal{A}_1^{(3)} + \mathcal{A}_1^{(4)}.
\end{aligned}
$$

For $\mathcal{A}_1^{(1)}$, we have

$$
\begin{aligned}
& E\left[Z_1 Z_{1+\ell_1}^3\right] \\
&= E\left[\left(\frac{1}{\sqrt{h}}K(\frac{X_1-x}{h}) - O(\sqrt{h})\right)\left(\frac{1}{\sqrt{h}}K(\frac{X_{1+\ell_1}-x}{h}) - O(\sqrt{h})\right)^3\right] \\
&= E\left[\left(\frac{1}{\sqrt{h}}K(\frac{X_1-x}{h}) - O(\sqrt{h})\right)\right. \\
&\times \left.\left(\frac{1}{h^{3/2}}K^3(\frac{X_{1+\ell_1}-x}{h}) - O\left(h^{3/2}\right) - \frac{1}{h}K^2(\frac{X_{1+\ell_1}-x}{h})O(\sqrt{h}) + \frac{1}{\sqrt{h}}K(\frac{X_{1+\ell_1}-x}{h})O(h)\right)\right] \\
&= E\left[\frac{1}{h^2}K(\frac{X_1-x}{h})K^3(\frac{X_{1+\ell_1}-x}{h})\right] + E\left[\frac{1}{h}K(\frac{X_1-x}{h})\right]O\left(h^2\right) \\
&+ E\left[\frac{1}{h}K(\frac{X_1-x}{h})\frac{1}{h}K^2(\frac{X_{1+\ell_1}-x}{h})\right]O(h) + E\left[\frac{1}{\sqrt{h}}K(\frac{X_1-x}{h})\frac{1}{\sqrt{h}}K(\frac{X_{1+\ell_1}-x}{h})\right]O(h) \\
&+ E\left[\frac{1}{h}K^3(\frac{X_{1+\ell_1}-x}{h})\right]O(1) + O(h^2) + E\left[\frac{1}{h}K^2(\frac{X_{1+\ell_1}-x}{h})\right]O(h) + \left[EK(\frac{X_{1+\ell_1}-x}{h})\right]O(h) \\
&= O(1) + O(h^2) + O(h) + O(h^2) + O(1) + O(h^2) + O(h) + O(h^2) = O(1). \qquad (25)
\end{aligned}
$$

Using a result in Section 1.2.2 of Doukhan (1994), we also get

$$
\left|E\left[Z_1 Z_{1+\ell_1}^3\right]\right| \le 4\alpha(\ell_1)\max\left(|Z_1|\right)\max\left(\left|Z_{1+\ell_1}^3\right|\right) = 4\alpha(\ell_1)\|K\|_\infty^4/h^2. \qquad (26)
$$

By (25) and (26), we have, for some d_T :

$$
\mathcal{A}_1^{(1)} = \sum_{\ell_1=1}^{d_T} O(1) + \frac{C}{h^2}\sum_{\ell_1=d_T+1}^{T-1}\frac{1}{\ell_1^\beta} = O\left(d_T\right) + O\left(\frac{1}{d_T^{\beta-1}h^2}\right).
$$

Upon choosing d_T such that $d_T \asymp (d_T^{\beta-1}h^2)^{-1}$, we obtain $\mathcal{A}_1^{(1)} = O\left(h^{-2/\beta}\right)$.

For $\mathcal{A}_1^{(2)}$, since

$$
\begin{aligned}
& E\left[Z_1 Z_{1+\ell_1}^2 Z_{1+\ell_1+\ell_3}\right] \\
&= E\left\{\left[\frac{1}{h}K(\frac{X_1-x}{h}) + O(1)\right]\left[\frac{1}{h}K^2(\frac{X_{1+\ell_1}-x}{h}) + K(\frac{X_{1+\ell_1}-x}{h})O(1) + O(h)\right]\right. \\
&\times \left.\left[\frac{1}{h}K(\frac{X_{1+\ell_1+\ell_3}-x}{h}) + O(1)\right]\times h\right\} = O(h),
\end{aligned}
$$

where we have used the boundedness of the density of $(X_1, X_{1+\ell_1}, X_{1+\ell_1+\ell_3})$, and

$$\left|E\left[Z_1 Z_{1+\ell_1}^2 Z_{1+\ell_1+\ell_3}\right]\right| \le 4\alpha(\ell_1)\|K\|_\infty^4/h^2,$$

we have

$$\begin{aligned}\mathcal{A}_1^{(2)} &= \sum_{\ell_1=1}^{T-1}\sum_{\ell_3=1}^{\ell_1}\left|E\left[Z_1 Z_{1+\ell_1}^2 Z_{1+\ell_1+\ell_3}\right]\right| = \sum_{\ell_1=1}^{d_T}\ell_1 O(h) + C\sum_{\ell_1=d_T+1}^{T-1}\frac{1}{\ell_1^{\beta-1}h^2}\\ &= O\left(d_T^2 h\right) + O\left(\frac{1}{d_T^{\beta-2}h^2}\right).\end{aligned}$$

Upon choosing d_T such that $d_T^2 h \asymp (d_T^{\beta-2}h^2)^{-1}$, we have $\mathcal{A}_1^{(2)} = O\left(h^{1-6/\beta}\right)$. Using the same procedure, we can show that $\mathcal{A}_1^{(3)} = O\left(h^{1-6/\beta}\right)$.

For $\mathcal{A}_1^{(4)}$, using the boundedness of the density of $(X_1, X_{1+\ell_1}, X_{1+\ell_1+\ell_2}, X_{1+\ell_1+\ell_2+\ell_3})$, we have

$$|E[Z_1 Z_{1+\ell_1} Z_{1+\ell_1+\ell_2} Z_{1+\ell_1+\ell_2+\ell_3}]| = O\left(h^2\right).$$

In addition,

$$|E[Z_1 Z_{1+\ell_1} Z_{1+\ell_1+\ell_2} Z_{1+\ell_1+\ell_2+\ell_3}]| \le 4\alpha(\ell_1)\|K\|_\infty^4/h^2.$$

So we have

$$\begin{aligned}\mathcal{A}_1^{(4)} &= \sum_{\ell_1=1}^{T-1}\sum_{\ell_2=1}^{\ell_1}\sum_{\ell_3=1}^{\ell_1}|E[Z_1 Z_{1+\ell_1} Z_{1+\ell_1+\ell_2} Z_{1+\ell_1+\ell_2+\ell_3}]|\\ &= \sum_{\ell_1=1}^{d_T}\ell_1^2 O\left(h^2\right) + C\sum_{\ell_1=d_T+1}^{T-1}\frac{1}{\ell_1^{\beta-2}h^2} = O\left(d_T^3 h^2\right) + O\left(\frac{1}{d_T^{\beta-3}h^2}\right).\end{aligned}$$

Letting $d_T = Ch^{-4/\beta}$ for $C > 0$, we have $\mathcal{A}_1^{(4)} = O\left(h^{2-12/\beta}\right)$.

Combining the results for $\mathcal{A}_1^{(1)} \sim \mathcal{A}_1^{(4)}$, we have

$$\mathcal{A}_1 = O\left(h^{-2/\beta}\right) + O\left(h^{1-6/\beta}\right) + O\left(h^{2-12/\beta}\right) = O\left(h^{-2/\beta}\right)$$

where the second equality uses the assumption that $\beta > 5$.

Using the similar arguments, we can show that

$$\begin{aligned}\mathcal{A}_2 &:= \sum_{0\le\ell_1,\ell_3\le\ell_2}|E[Z_1 Z_{1+\ell_1}(Z_{1+\ell_1+\ell_2} Z_{1+\ell_1+\ell_2+\ell_3})] - E[Z_1 Z_{1+\ell_1}]E[Z_{1+\ell_1+\ell_2} Z_{1+\ell_1+\ell_2+\ell_3}]|\\ &= O\left(h^{-2/\beta}\right),\end{aligned}$$

and

$$\mathcal{A}_3 := \sum_{0\le\ell_1,\ell_2\le\ell_3}|E[(Z_1 Z_{1+\ell_1} Z_{1+\ell_1+\ell_2}) Z_{1+\ell_1+\ell_2+\ell_3}]| = O\left(h^{-2/\beta}\right).$$

Next we consider the terms that involve the Gaussian sequence $\{\tilde{Z}_t\}$. We have

$$\begin{aligned}
&\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|E\left[\tilde{Z}_1\left(\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right]\right| \\
&=\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(E\tilde{Z}_1\tilde{Z}_{1+\ell_1}\right)\left(E\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right| \\
&+\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(E\tilde{Z}_1\tilde{Z}_{1+\ell_1+\ell_2}\right)\left(E\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right| \\
&+\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(E\tilde{Z}_1\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\left(E\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\right)\right|.
\end{aligned}$$

Since

$$\left(EZ_1Z_{1+\ell_1}\right)\left(EZ_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)=\begin{cases} O(1), & \text{if } (\ell_1,\ell_3)=(0,0)\\ O(h), & \text{if } (\ell_1,\ell_3)=(0,1)\\ O(h), & \text{if } (\ell_1,\ell_3)=(1,0)\\ O(h^2), & \text{otherwise}\end{cases}$$

$$\left(EZ_1Z_{1+\ell_1}\right)\left(EZ_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)=O\left(\frac{1}{h^2}\alpha\left(\ell_1\right)\alpha\left(\ell_3\right)\right),$$

we have, for some d_{1T} and d_{2T} :

$$\begin{aligned}
&\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(E\tilde{Z}_1\tilde{Z}_{1+\ell_1}\right)\left(E\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right| \\
&=\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(EZ_1Z_{1+\ell_1}\right)\left(EZ_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)\right| \\
&=\sum_{\ell_1=1}^{T-1}\sum_{0<\ell_3\le\ell_1}\sum_{0\le\ell_2\le\ell_1}\left|\left(EZ_1Z_{1+\ell_1}\right)\left(EZ_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2+\ell_3}\right)\right| \\
&+\sum_{\ell_1=1}^{T-1}\sum_{0\le\ell_2\le\ell_1}\left|\left(EZ_1Z_{1+\ell_1}\right)\left(EZ_{1+\ell_1+\ell_2}Z_{1+\ell_1+\ell_2}\right)\right| \\
&=\left(\sum_{\ell_1=1}^{d_{1T}}\ell_1^2O(h^2)+\frac{1}{h^2}\sum_{\ell_1=d_{1T}+1}^{T-1}\sum_{\ell_3=0}^{\ell_1}\ell_1\alpha\left(\ell_1\right)\alpha\left(\ell_3\right)\right)+\left(\sum_{\ell_1=1}^{d_{2T}}\ell_1O(h)+\frac{1}{h^2}\sum_{\ell_1=d_{2T}+1}^{T-1}\ell_1\alpha\left(\ell_1\right)\right) \\
&=O\left(d_{1T}^3h^2+\frac{1}{h^2}\sum_{\ell_1=d_{1T}+1}^{T-1}\ell_1\alpha\left(\ell_1\right)\right)+O\left(d_{2T}^2h+\frac{1}{h^2}\sum_{\ell_1=d_{2T}+1}^{T-1}\ell_1\alpha\left(\ell_1\right)\right) \\
&=O\left(d_{1T}^3h^2+\frac{1}{h^2}\frac{1}{d_{1T}^{\beta-2}}\right)+O\left(d_{2T}^2h+\frac{1}{h^2}\frac{1}{d_{2T}^{\beta-2}}\right)=O(h^{-2/\beta})
\end{aligned}$$

upon choosing d_{1T} and d_{2T} appropriately. Similarly, we can show that

$$\begin{aligned}
\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(E\tilde{Z}_1\tilde{Z}_{1+\ell_1+\ell_2}\right)\left(E\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right|&=O(h^{-2/\beta}),\\
\sum_{0\le\ell_2,\ell_3\le\ell_1}\left|\left(E\tilde{Z}_1\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\left(E\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\right)\right|&=O(h^{-2/\beta}).
\end{aligned}$$

As a result, $\sum_{0\le \ell_2,\ell_3\le \ell_1}\left|E\left[\tilde{Z}_1\left(\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right]\right| = O(h^{-2/\beta})$.

Using the same argument, we can show that

$$\sum_{0\le \ell_1,\ell_3\le \ell_2}\left(E\left[\tilde{Z}_1\tilde{Z}_{1+\ell_1}\left(\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right)\right]-E\left[\tilde{Z}_1\tilde{Z}_{1+\ell_1}\right]E\left[\tilde{Z}_{1+\ell_1+\ell_2}\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right]\right)$$
$$= O(h^{-2/\beta}),$$

and

$$\sum_{0\le \ell_1,\ell_2\le \ell_3}\left|E\left[\left(\tilde{Z}_1\tilde{Z}_{1+\ell_1}\tilde{Z}_{1+\ell_1+\ell_2}\right)\tilde{Z}_{1+\ell_1+\ell_2+\ell_3}\right]\right| = O(h^{-2/\beta}).$$

Combining the above results, we have

$$\sum_{\ell_1=1}^{T-1}\sum_{\ell_2=1}^{T-1}\sum_{\ell_3=1}^{T-1}\left|Q\left(\ell_1,\ell_2,\ell_3\right)\right| = O\left(h^{-2/\beta}\right).$$

Similarly, we can show that the rest of the three terms in (24) are all of the order $O\left(h^{-2/\beta}\right)$. Therefore,

$$\sum_{\ell_1=0}^{T-1}\sum_{\ell_2=0}^{T-1}\sum_{\ell_3=1}^{T-1}\left|Q\left(\ell_1,\ell_2,\ell_3\right)\right| = O\left(h^{-2/\beta}\right).$$

(ii) Let $g_\ell\left(u,v\right)$ be the pdf of $\left(x_1,x_{\ell+1}\right)$, then

$$\begin{aligned}\left|Cov(Z_1,Z_{\ell+1})\right| &\le \left|h\int\int K_h\left(u-x\right)K_h\left(v-x\right)g_\ell\left(u,v\right)dudv\right|\\ &+\left|h\left(\int K_h\left(u-x\right)f\left(u\right)du\right)\left(\int K_h\left(v-x\right)f\left(v\right)dv\right)\right|\\ &= h\left\|g_\ell\right\|_\infty + hC = O(h).\end{aligned}$$

By a result in Section 1.2.2 of Doukhan (1994), we also have

$$Cov(Z_1,Z_{\ell+1}) \le 4\alpha\left(\ell\right)\max\left(\left|Z_1\right|\right)\max\left(\left|Z_{\ell+1}\right|\right) = 4\alpha\left(\ell\right)\left\|K\right\|_\infty^2/h. \tag{27}$$

Using the above two results, we obtain, for some d_T,

$$\begin{aligned}\sum_{\ell=1}^{T-1}h^{-\nu}\left|Cov(Z_1,Z_{\ell+1})\right|\ell^p &= \sum_{\ell=1}^{d_T}h^{-\nu}\left|Cov(Z_1,Z_{\ell+1})\right|\ell^p + \sum_{\ell=d_T+1}^{T-1}h^{-\nu}\left|Cov(Z_1,Z_{\ell+1})\right|\ell^p\\ &= O\left(d_T^{p+1}h^{1-\nu}\right)+O\left(\frac{1}{h^{1+\nu}}\sum_{\ell=d_T+1}^{\infty}\ell^{p-\beta}\right)\\ &= O\left(d_T^{p+1}h^{1-\nu}\right)+O\left(\frac{1}{h^{1+\nu}}\frac{1}{d_T^{\beta-p-1}}\right).\end{aligned}$$

Let us take a d_T such that

$$d_T^{p+1}h^{1-\nu} \asymp \frac{1}{h^{1+\nu}}\frac{1}{d_T^{\beta-p-1}},$$

that is, $d_T = h^{-2/\beta}$. Then

$$\sum_{\ell=1}^{T-1} h^{-\nu} \left| Cov(Z_1, Z_{\ell+1}) \right| \ell^p = O\left(h^{1-2(p+1)/\beta-\nu} \right),$$

So if $\beta \geq 2(p+1)$, then $\sum_{\ell=1}^{T-1} h^{-\nu} \left| Cov(Z_1, Z_{\ell+1}) \right| \ell^p = O(1)$ for $\nu = 1 - 2(p+1)/\beta$. This ensures the existence of $B^{(p)}(x)$ such that

$$B^{(p)}(x) = 2 \lim_{T\to\infty} \sum_{\ell=1}^{T-1} h^{-\nu} Cov(Z_1, Z_{\ell+1}) \ell^p.$$

■

Proof of Theorem 4.

(a) Asymptotic variance

Let $\mathcal{I} = \left\{ (t, \tau, r, s) \in \left\{ (1, \ldots, T)^4 \right\} \setminus \{ t = \tau = r = s \} \right\}$. Then

$$\begin{aligned}
&Var\left(\tilde{V}_h(x) \right) \\
&= Var\left(\frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T} \right) Z_t Z_\tau \right) \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \sum_{r=1}^{T} \sum_{s=1}^{T} W\left(\frac{t-\tau}{S_T} \right) W\left(\frac{r-s}{S_T} \right) \left\{ E\left[Z_t Z_\tau Z_r Z_s \right] - E\left[Z_t Z_\tau \right] E\left[Z_r Z_s \right] \right\} \\
&= A + B, \qquad (28)
\end{aligned}$$

where

$$\begin{aligned}
A &= \frac{1}{T} \left\{ E\left[Z_t^4 \right] - \left[E Z_t^2 \right]^2 \right\}, \\
B &= \frac{1}{T^2} \sum_{(t,\tau,r,s)\in\mathcal{I}} W\left(\frac{t-\tau}{S_T} \right) W\left(\frac{r-s}{S_T} \right) \left\{ E\left[Z_t Z_\tau Z_r Z_s \right] - E\left[Z_t Z_\tau \right] E\left[Z_r Z_s \right] \right\}.
\end{aligned}$$

For A, we have

$$\begin{aligned}
A_1 &= \frac{1}{T} E\left[Z_t^4 \right] = \frac{1}{Th} \int h^3 \left\{ K_h(v-x) - E K_h(X_t - x) \right\}^4 f(v)\, dv \\
&= \frac{h^2}{T} \int \left\{ K_h(v-x)^4 f(v) - 4 K_h(v-x)^3 E K_h(X_t - x) \right\} f(v)\, dv \\
&+ \frac{h^2}{T} \int \left\{ 6 K_h(v-x)^2 \left[E K_h(X_t - x) \right]^2 - 4 K_h(v-x) \left[E K_h(X_t - x) \right]^3 \right\} f(v)\, dv \\
&+ \frac{h^2}{T} \left[E K_h(X_t - x) \right]^4 = \frac{1}{Th} f(x) \int K(u)^4\, du + O\left(\frac{1}{T} \right) \qquad (29)
\end{aligned}$$

and

$$
\begin{aligned}
A_2 &= \frac{1}{T}\left[EZ_t^2\right]^2 = \frac{1}{T}\left(\int h\left\{K_h\left(v-x\right)-E\left[K_h\left(X_t-x\right)\right]\right\}^2 f\left(v\right)dv\right)^2 \\
&= \frac{1}{T}\left(\int h\left\{K_h\left(v-x\right)^2-2K_h\left(v-x\right)E\left[K_h\left(X_t-x\right)\right]+\left[EK_h\left(X_t-x\right)\right]^2\right\} f\left(v\right)dv\right)^2 \\
&= O\left(\frac{1}{T}\right), \qquad (30)
\end{aligned}
$$

and so $A = \frac{1}{Th}f\left(x\right)\int K\left(u\right)^4 du + O\left(\frac{1}{T}\right).$

For B, we first use Lemma 1(a) to obtain

$$
\begin{aligned}
&\left|\frac{1}{T^2}\sum_{(t,\tau,r,s)\in\mathcal{I}} W\left(\frac{t-\tau}{S_T}\right)W\left(\frac{r-s}{S_T}\right)\left(E\left[Z_tZ_\tau Z_rZ_s\right]-E\left[\tilde{Z}_t\tilde{Z}_\tau\tilde{Z}_r\tilde{Z}_s\right]\right)\right| \\
&= \left|\frac{1}{T^2}\sum_{(t,\tau,r,s)\in\mathcal{I}} W\left(\frac{t-\tau}{S_T}\right)W\left(\frac{r-s}{S_T}\right)\left(E\left[Z_1Z_{1+\tau-t}Z_{1+r-t}Z_{1+s-t}\right]-E\left[\tilde{Z}_1\tilde{Z}_{1+\tau-t}\tilde{Z}_{1+r-t}\tilde{Z}_{1+s-t}\right]\right)\right| \\
&\leq \frac{1}{T}\sum_{\ell=-T+1}^{T-1}\sum_{m=-T+1}^{T-1}\sum_{n=-T+1}^{T-1}\left|E\left[Z_1Z_{1+\ell}Z_{1+m}Z_{1+n}\right]-E\left[\tilde{Z}_1\tilde{Z}_{1+\ell}\tilde{Z}_{1+m}\tilde{Z}_{1+n}\right]\right| 1\left\{(\ell,m,n)\neq(0,0,0)\right\} \\
&= \frac{1}{T}\sum_{\ell=-T+1}^{T-1}\sum_{m=-T+1}^{T-1}\sum_{n=-T+1}^{T-1} Q_T\left(\ell,m,n\right)1\left\{(\ell,m,n)\neq(0,0,0)\right\} \\
&= O\left(\frac{1}{Th^{2/\beta}}\right). \qquad (31)
\end{aligned}
$$

Combining (31) with

$$
E\left[\tilde{Z}_t\tilde{Z}_\tau\tilde{Z}_r\tilde{Z}_s\right] = E\left[Z_tZ_\tau\right]E\left[Z_rZ_s\right]+E\left[Z_tZ_r\right]E\left[Z_\tau Z_s\right]+E\left[Z_tZ_s\right]E\left[Z_\tau Z_r\right],
$$

we have

$$
\begin{aligned}
B &= \frac{1}{T^2}\sum_{(t,\tau,r,s)\ \in\mathcal{I}} W\left(\frac{t-\tau}{S_T}\right)W\left(\frac{r-s}{S_T}\right)\left\{E\left[Z_tZ_\tau Z_rZ_s\right]-E\left[Z_tZ_\tau\right]E\left[Z_rZ_s\right]\right\} \\
&= \frac{1}{T^2}\sum_{(t,\tau,r,s)\ \in\mathcal{I}} W\left(\frac{t-\tau}{S_T}\right)W\left(\frac{r-s}{S_T}\right)\left\{E\left[Z_tZ_r\right]E\left[Z_\tau Z_s\right]+E\left[Z_tZ_s\right]E\left[Z_\tau Z_r\right]\right\}+O\left(\frac{1}{Th^{2/\beta}}\right) \\
&:= B_1+B_2+O\left(\frac{1}{Th^{2/\beta}}\right). \qquad (32)
\end{aligned}
$$

For B_1,

$$\begin{aligned}
B_1 &= \frac{1}{T^2} \sum_{(t,\tau,r,s)\ \in \mathcal{I}} W\left(\frac{t-\tau}{S_T}\right) W\left(\frac{r-s}{S_T}\right) E\left[Z_t Z_r\right] E\left[Z_\tau Z_s\right] \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \sum_{r=1}^{T} \sum_{s=1}^{T} W\left(\frac{t-\tau}{S_T}\right) W\left(\frac{r-s}{S_T}\right) E\left[Z_t Z_r\right] E\left[Z_\tau Z_s\right] + O\left(\frac{1}{T}\right) \\
&= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)^2 E\left[Z_t^2\right] E\left[Z_\tau^2\right] \\
&+ \frac{1}{T^2} \sum_{t=1}^{T} \sum_{r\neq t} \sum_{\tau=1}^{T} \sum_{\tau\neq s} W\left(\frac{t-\tau}{S_T}\right) W\left(\frac{r-s}{S_T}\right) E\left[Z_t Z_r\right] E\left[Z_\tau Z_s\right] \\
&+ \frac{2}{T^2} \sum_{t=1}^{T} \sum_{r\neq t} \sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right) W\left(\frac{r-\tau}{S_T}\right) E\left[Z_t Z_r\right] E\left[Z_\tau^2\right] + O\left(\frac{1}{T}\right) \\
&:= B_{11} + B_{12} + B_{13} + O\left(\frac{1}{T}\right).
\end{aligned}$$

For B_{11}, we have

$$\begin{aligned}
B_{11} &= \frac{S_T}{T} \frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \left(1 - \frac{|\ell|}{T}\right) W\left(\frac{\ell}{S_T}\right)^2 \left(E\left[Z_1^2\right]\right)^2 \\
&= \frac{S_T}{T} \frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} W\left(\frac{\ell}{S_T}\right)^2 \left(E\left[Z_1^2\right]\right)^2 - \frac{S_T}{T} \frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \frac{|\ell|}{T} W\left(\frac{\ell}{S_T}\right)^2 \left(E\left[Z_1^2\right]\right)^2 \\
&= \left(\frac{S_T}{T}\right) \left(\int_{-1}^{1} W(\xi)^2 \, d\xi\right) V(x)^2 + o\left(\frac{S_T}{T}\right) + O\left(\frac{S_T^2}{T^2}\right) \qquad (33)
\end{aligned}$$

because

$$\begin{aligned}
&\left| \frac{S_T}{T} \frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \frac{|\ell|}{T} W\left(\frac{\ell}{S_T}\right)^2 \left(E\left[Z_1^2\right]\right)^2 \right| \\
&\leq \frac{S_T}{T} \frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \frac{|\ell|}{T} \left[E Z_1^2\right]^2 \leq \frac{S_T}{T} \frac{1}{S_T} \frac{S_T\left(2S_T+1\right)}{T} \left[E Z_1^2\right]^2 = O\left(\frac{S_T^2}{T^2}\right).
\end{aligned}$$

For B_{12}, we have

$$\begin{aligned}
|B_{12}| &= \left| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{r \neq t} \sum_{\tau=1}^{T} \sum_{s \neq \tau} W\left(\frac{t-\tau}{S_T}\right) W\left(\frac{r-s}{S_T}\right) E\left[Z_t Z_r\right] E\left[Z_\tau Z_s\right] \right| \\
&= \left| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{r \neq t} \sum_{\ell=\max\{-S_T, t-T\}}^{\min\{S_T, t-1\}} \sum_{k=\max\{-S_T, r-T\}}^{\min\{S_T, r-1\}} W\left(\frac{\ell}{S_T}\right) W\left(\frac{k}{S_T}\right) \right. \\
&\quad \left. \times E\left[Z_t Z_r\right] E\left[Z_{t-\ell} Z_{r-k}\right] 1\left\{t-\ell \neq r-k\right\} \right| \\
&\leq \frac{S_T}{T} \frac{1}{T} \sum_{t=1}^{T} \sum_{r \neq t} \left|E\left[Z_t Z_r\right]\right| \left(\frac{1}{S_T} \sum_{\ell=\max\{-S_T, t-T\}}^{\min\{S_T, t-1\}} \sum_{k=\max\{-S_T, r-T\}}^{\min\{S_T, r-1\}} \left|\Gamma_{Z, k-r-\ell+t}\right| 1\left\{\ell \neq k-r+t\right\} \right) \\
&\leq \frac{S_T}{T} \frac{1}{T} \sum_{t=1}^{T} \sum_{r \neq t} \left|E\left[Z_t Z_r\right]\right| \left(\frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \sum_{k=\max\{-S_T-r+t, -T+t\}}^{\min\{S_T-r+t, -1+t\}} \left|\Gamma_{Z, k-\ell}\right| 1\left\{k \neq \ell\right\} \right) \\
&\leq \frac{S_T}{T} \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{r \neq t} \left|E\left[Z_t Z_r\right]\right| \right) \left(\frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \sum_{k=-T+1}^{T-1} \left|\Gamma_{Z, k-\ell}\right| 1\left\{k \neq \ell\right\} \right) \\
&\leq 2 \frac{S_T}{T} \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{r \neq t} \left|E\left[Z_t Z_r\right]\right| \right) \frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \sum_{k=0}^{T-1} \left|\Gamma_{z, k-\ell}\right| 1\left\{k \neq \ell\right\}, \qquad (34)
\end{aligned}$$

where $\Gamma_{Z,s}$ is defined to the $|s|$-th order autocovariance of $\{Z_t\}$. But with $d_T = h^{-2/\beta}$ we have

$$\begin{aligned}
\sum_{k=0}^{T-1} \left|\Gamma_{Z, k-\ell}\right| 1\left\{k \neq \ell\right\} \leq \sum_{k=1}^{\infty} \left|Cov(Z_1, Z_{k+1})\right| &= \sum_{k=1}^{d_T} O(h) + \sum_{k=d_T+1}^{\infty} \frac{1}{h} \alpha(k) \\
&= O\left(d_T h + \frac{1}{d_T^{\beta-1} h} \right) = O\left(h^{1-2/\beta} \right) = o\left(1\right)
\end{aligned}$$

uniformly over ℓ. Similarly, $\sum_{r \neq t} \left|E\left[Z_t Z_r\right]\right| = o(1)$ uniformly over t. Combining these with (34) yields $B_{12} = o(S_T/T)$.

For B_{13}, we have

$$\begin{aligned}
|B_{13}| &= \left| \frac{2}{T^2} \sum_{\tau=1}^{T} \sum_{t=1}^{T} \sum_{r \neq t} W\left(\frac{\tau-t}{S_T}\right) W\left(\frac{\tau-r}{S_T}\right) E\left[Z_t Z_r\right] E\left[Z_\tau^2\right] \right| \\
&= \left| \frac{2}{T^2} \sum_{\tau=1}^{T} \sum_{\ell=\max\{-S_T, \tau-T\}}^{\min\{S_T, \tau-1\}} \sum_{k=\max\{-S_T, \tau-T\}}^{\min\{S_T, \tau-1\}} W\left(\frac{\ell}{S_T}\right) W\left(\frac{k}{S_T}\right) E\left[Z_{\tau-\ell} Z_{\tau-k}\right] E\left[Z_\tau^2\right] \right| 1\left\{k \neq \ell\right\} \\
&\leq \frac{S_T}{T} \frac{2}{T} \sum_{\tau=1}^{T} E\left[Z_\tau^2\right] \left(\frac{1}{S_T} \sum_{\ell=-S_T}^{S_T} \sum_{k=-T+1}^{T-1} \left|\Gamma_{Z, k-\ell}\right| 1\left\{k \neq \ell\right\} \right) = o\left(\frac{S_T}{T}\right). \qquad (35)
\end{aligned}$$

In view of (33), (34) and (35), we have

$$B_1 = \left(\frac{S_T}{T}\right) \left(\int_{-1}^{1} W\left(\xi\right)^2 d\xi \right) V\left(x\right)^2 + o\left(\frac{S_T}{T}\right) + O\left(\frac{S_T^2}{T^2}\right).$$

Using $W(\xi) = W(-\xi)$ in Assumption 4(i), it is easy to show that $B_1 = B_2$. Thus, we have

$$B = \left(\frac{S_T}{T}\right) 2 \left(\int_{-1}^{1} W(\xi)^2 d\xi\right) [V(x)]^2 + o\left(\frac{S_T}{T}\right) + O\left(\frac{1}{Th^{2/\beta}}\right). \tag{36}$$

To sum up, we have proved that

$$Var\left(\tilde{V}_h(x)\right) = \frac{1}{Th} f(x) \int K(u)^4 du + \left(\frac{S_T}{T}\right) 2 \left(\int_{-1}^{1} W(\xi)^2 d\xi\right) [V(x)]^2 + o\left(\frac{S_T}{T}\right) + o\left(\frac{1}{Th}\right).$$

So under the asymptotic sequence: $S_T \to \infty, h \to 0$ but $Th \to \infty$ and $T/S_T \to \infty$, as $T \to \infty$, we have the following:

(i) if $hS_T \to 0$,

$$\lim_{T\to\infty} ThVar\left(\tilde{V}_h(x)\right) = f(x) \int K(u)^4 du;$$

(ii) if $hS_T \to \gamma$,

$$\lim_{T\to\infty} \frac{T}{S_T} Var\left(\tilde{V}_h(x)\right) = f(x) \int K(u)^4 du/\gamma + 2\left(\int_{-1}^{1} W(\xi)^2 d\xi\right) V(x)^2;$$

(iii) if $hS_T \to \infty$,

$$\lim_{T\to\infty} \frac{T}{S_T} Var\left(\tilde{V}_h(x)\right) = 2\left(\int_{-1}^{1} W(\xi)^2 d\xi\right) V(x)^2.$$

(b) Asymptotic bias

Using Lemma 1(b) and the dominated convergence theorem, we have

$$\frac{S_T^p}{h^\nu}\left(E\tilde{V}_h(x) - V_h(x)\right) = \frac{S_T^p}{h^\nu} \frac{1}{T} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \left(W\left(\frac{t-\tau}{S_T}\right) - 1\right) E(Z_t Z_\tau)$$

$$= \frac{2}{h^\nu} \sum_{\ell=1}^{S_T} \left(\frac{W(\ell/S_T) - 1}{(\ell/S_T)^p}\right) \left(1 - \frac{\ell}{T}\right) E(Z_1 Z_{1+\ell}) \ell^p$$

$$= -W^{(p)} B^{(p)}(x) + o(1),$$

as desired.

(c) Approximation error: $\sqrt{\phi_T}\left(\hat{V}_h(x) - \tilde{V}_h(x)\right) = o_p(1)$.

We use the same decomposition as in the proof of Theorem 3 to get

$$\left[\hat{V}_h(x) - \tilde{V}_h(x)\right] = O_p\left(\frac{S_T}{T}\right) + O_p\left(\sqrt{\frac{1}{T}}\right) \frac{1}{T} \sum_{\tau=1}^{T} \left[\sum_{t=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\right] Z_\tau$$

but we will establish a tighter bound for the second term. We have

$$
\begin{aligned}
&Var\left\{\frac{1}{T}\sum_{\tau=1}^{T}\left[\sum_{t=1}^{T}W\left(\frac{t-\tau}{S_T}\right)\right]Z_\tau\right\}\\
&=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{\tau_1=1}^{T}\sum_{\tau_2=1}^{T}W\left(\frac{t_1-\tau_1}{S_T}\right)W\left(\frac{t_2-\tau_2}{S_T}\right)Cov(Z_{\tau_1},Z_{\tau_2})\\
&\leq\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{\tau_1:|\tau_1-t_1|\leq S_T}\sum_{\tau_2:|\tau_2-t_2|\leq S_T}|\Gamma_{Z,\tau_1-\tau_2}|\leq\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{\tau_1=-S_T}^{S_T}\sum_{\tau_2=-S_T}^{S_T}|\Gamma_{Z,\tau_1-\tau_2+t_1-t_2}|\\
&\leq\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{|t_2-t_1|\leq 3S_T}\sum_{\tau_1=-S_T}^{S_T}\sum_{\tau_2=-S_T}^{S_T}|\Gamma_{Z,\tau_1-\tau_2+t_1-t_2}|+\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{|t_2-t_1|>3S_T}\sum_{\tau_1=-S_T}^{S_T}\sum_{\tau_2=-S_T}^{S_T}|\Gamma_{Z,\tau_1-\tau_2+t_1-t_2}|\\
&=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{|t_2-t_1|\leq 3S_T}\sum_{\tau_1=-S_T}^{S_T}\sum_{\tau_2=-S_T}^{S_T}|\Gamma_{Z,\tau_1-\tau_2+t_1-t_2}|+\frac{1}{T^2h}\sum_{t_1=1}^{T}\sum_{|t_2-t_1|>3S_T}\sum_{\tau_1=-S_T}^{S_T}\sum_{\tau_2=-S_T}^{S_T}S_T^{-\beta}\\
&=\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{|t_2-t_1|\leq 3S_T}\sum_{\tau_1=-S_T}^{S_T}\sum_{\tau_2=-S_T}^{S_T}|\Gamma_{Z,\tau_1-\tau_2+t_1-t_2}|+O\left(\frac{1}{hS_T^{\beta-2}}\right)\\
&=C\frac{S_T}{T^2}\sum_{t_1=1}^{T}\sum_{|t_2-t_1|\leq 3S_T}^{T}[Var(Z_\tau)+o(1)]+O\left(\frac{1}{hS_T^{\beta-2}}\right)\\
&=O\left(\frac{S_T^2}{T}\right)+O\left(\frac{1}{hS_T^{\beta-2}}\right).
\end{aligned}
$$

So

$$
\hat{V}_h(x)-\tilde{V}_h(x)=O_p\left(\frac{S_T}{T}\right)+O_p\left(\sqrt{\frac{1}{T}}\sqrt{\frac{S_T^2}{T}}+\frac{1}{\sqrt{Th}}\frac{1}{S_T^{\beta/2-1}}\right)=O_p\left(\frac{S_T}{T}+\frac{1}{\sqrt{Th}}\frac{1}{S_T^{\beta/2-1}}\right).
$$

Noting that $\phi_T\leq T/S_T$ and $\phi_T\leq Th$, we have

$$
\sqrt{\phi_T}[\hat{V}_h(x)-\tilde{V}_h(x)]=O_p\left(\sqrt{\frac{T}{S_T}}\frac{S_T}{T}+\sqrt{Th}\frac{1}{\sqrt{Th}}\frac{1}{S_T^{\beta/2-1}}\right)=O_p\left(\sqrt{\frac{S_T}{T}}+\frac{1}{S_T^{\beta/2-1}}\right)=o_p(1).
$$

(d) Note that

$$
|cov(Z_1,Z_{j+1})|=O(h)\text{ and }|cov(Z_1,Z_{j+1})|\leq 4\alpha(j)\|K_\infty\|^2/h.
$$

we have, for some $d_T>1$:

$$
\begin{aligned}
V_h(x)&=Var(Z_{t,h})+2\sum_{j=1}^{\infty}Cov(Z_1,Z_{1+j})\\
&=V(x)+o(1)+2\left[\sum_{j=1}^{d_T}O(h)+\sum_{j=d_T+1}^{\infty}\frac{1}{h}\alpha(d_T)\right]\\
&=V(x)+o(1)+O\left(d_Th+\frac{1}{d_T^{\beta-1}h}\right).
\end{aligned}
$$

Taking $d_T = h^{-2/\beta}$ yields

$$V_h(x) = V(x) + o(1) + O\left(h^{1-2/\beta}\right) = V(x) + o(1).$$

■

Proof of Proposition 1. Let

$$\Delta_T(\tau) = \frac{1}{T}\sum_{s=1}^{T} W_b\left(\frac{s}{T} - \tau\right) - \int_0^1 W_b(s-\tau)\,ds,$$

$$\Delta_T = \frac{1}{T}\sum_{\tau=1}^{T}\sum_{r=1}^{T} W_b\left(\frac{\tau - r}{T}\right) - \int_0^1\int_0^1 W_b(\tau - r)d\tau dr.$$

Assumption 5 implies that $\Delta_T(\tau) = O(1/T)$ uniformly over τ and $\Delta_T = O(1/T)$. Then

$$\begin{aligned}
\hat{V}_h(x,b) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b\left(\frac{t-\tau}{T}\right)\hat{Z}_t\hat{Z}_\tau = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\tilde{W}_{T,b}\left(\frac{t}{T},\frac{\tau}{T}\right)Z_tZ_\tau \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right)Z_tZ_\tau - \left[\frac{1}{T}\sum_{t=1}^{T}Z_t\right]\left[\frac{1}{T}\sum_{\tau=1}^{T}\Delta_T(\tau)Z_\tau\right] \\
&- \left[\frac{1}{T}\sum_{t=1}^{T}\Delta_T(t)Z_t\right]\left[\frac{1}{T}\sum_{\tau=1}^{T}Z_\tau\right] + \Delta_T\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T}Z_tZ_\tau \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right)Z_tZ_\tau + O_p\left(\frac{1}{\sqrt{T}}\right)\left[\frac{1}{T}\sum_{\tau=1}^{T}\Delta_T(\tau)Z_\tau\right] \\
&+ \left[\frac{1}{T}\sum_{t=1}^{T}\Delta_T(t)Z_t\right]O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{T}\right).
\end{aligned}$$

But

$$\begin{aligned}
&Var\left[\frac{1}{\sqrt{T}}\sum_{\tau=1}^{T}\Delta_T(\tau)Z_\tau\right] \\
&= \Delta_T(\tau)^2 Var[Z_\tau] + \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau\neq t}\Delta_T(\tau)\Delta_T(t)Cov(Z_\tau, Z_t) \\
&= O\left(\frac{1}{T^2}\right) + \frac{1}{T}\sum_{t=1}^{T}\sum_{1\le|t-\tau|\le d_T}\Delta_T(\tau)\Delta_T(t)Cov(Z_\tau, Z_t) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\sum_{|t-\tau|>d_T}\Delta_T(\tau)\Delta_T(t)Cov(Z_\tau, Z_t) \\
&= O\left(\frac{1}{T^2}\right) + O\left(\frac{1}{T^2}\right)\left[hd_T + \frac{1}{hd_T^{\beta-1}}\right] = O(\frac{h^{1-2/\beta}}{T^2}) = O\left(\frac{1}{T^2}\right).
\end{aligned} \tag{37}$$

upon taking $d_T \asymp h^{-2/\beta}$, and so

$$\hat{V}_h(x,b) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right) Z_t Z_\tau + O_p\left(\frac{1}{T}\right),$$

where the $O_p(\cdot)$ term holds under both the fixed h and small h asymptotics. Using the same proof as that for Lemma 1 in Sun (2014b), we can show that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right) Z_t Z_\tau \sim^a V_h(x)\,\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right) e_t e_\tau,$$

and so

$$\hat{V}_h(x,b) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} \tilde{W}_{T,b}\left(\frac{t}{T},\frac{\tau}{T}\right) Z_t Z_\tau \sim^a V_h(x)\,\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right) e_t e_\tau.$$

Similarly,

$$V_h(x)\,\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} \tilde{W}_{T,b}\left(\frac{t}{T},\frac{\tau}{T}\right) e_t e_\tau \sim^a V_h(x)\,\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b^*\left(\frac{t}{T},\frac{\tau}{T}\right) e_t e_\tau.$$

Combining the above two results leads to

$$\begin{aligned}\hat{V}_h(x,b) &\sim^a V_h(x)\,\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} \tilde{W}_{T,b}\left(\frac{t}{T},\frac{\tau}{T}\right) e_t e_\tau \\ &= \frac{V_h(x)}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W_b\left(\frac{t-\tau}{T}\right)(e_t-\bar{e})(e_\tau-\bar{e})\end{aligned}$$

under both the fixed h and small h asymptotics. ■

Proof of Theorem 8. Note that

$$\begin{aligned}&\left|K_h(X_t-x)\left(\frac{X_t-x}{h}\right)^{i+j-2}\right| \\ &= \left|\frac{1}{h}K\left(\frac{X_t-x}{h}\right)\left(\frac{X_t-x}{h}\right)^{i+j-2}\right| 1\{|X_t-x|\le h\} \le \|K\|_\infty/h.\end{aligned}$$

By the mixing inequality of Doukhan (1994) for almost surely bounded variables, we have

$$Cov\left(K_h(X_t-x)\left(\frac{X_t-x}{h}\right)^{i+j-2}, K_h(X_s-x)\left(\frac{X_s-x}{h}\right)^{i+j-2}\right) \le 8\|K\|_\infty \alpha(t-s)/h.$$

Given that $\sum_{\ell=1}^{\infty}\alpha(\ell)<\infty$, we have, for a fixed h,

$$\frac{1}{T}\sum_{t=1}^{T} K_h(X_t-x)\left(\frac{X_t-x}{h}\right)^{i+j-2} \to^p E\left[K_h(X_t-x)\left(\frac{X_t-x}{h}\right)^{i+j-2}\right].$$

This implies that $P_x'\Omega_x P_x/T \to^p J_h(x)$.

As a consequence,

$$\sqrt{Th}\left[\hat{m}(x)-\theta_{0,h}\right]=e_1'\left(\frac{P_x'\Omega_x P_x}{T}\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}P_t(x)\sqrt{h}K_h(X_t-x)\left[Y_t-P_t(x)'\theta_h\right]$$

$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t(1+o_p(1))$$

where

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_1'J_h^{-1}(x)P_t(x)\sqrt{h}K_h(X_t-x)\left[\varepsilon_t+m(X_t)-P_t(x)'\theta_h\right].$$

In view of

$$\theta_h=\operatorname*{plim}_{T\to\infty}\left(\frac{P_x'\Omega_x P_x}{T}\right)^{-1}\left(\frac{P_x'\Omega_x Y}{T}\right)=J_h^{-1}(x)E\left[P_t(x)K_h(X_t-x)m(X_t)\right],$$

we obtain

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_1'J_h^{-1}(x)P_t(x)\sqrt{h}K_h(X_t-x)\varepsilon_t$$
$$+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_1'J_h^{-1}(x)\left\{P_t(x)\sqrt{h}K_h(X_t-x)m(X_t)-E\left[P_t(x)\sqrt{h}K_h(X_t-x)m(X_t)\right]\right\}.$$

Define

$$\tilde{K}^*(u)=e_1'J_h^{-1}(x)\left[1,u,...,u^r\right]'K(u)=\left(a_{0,hx}+a_{1,hx}u+...+a_{r,hx}u^r\right)K(u)$$

where $(a_{0,hx},a_{1,hx},...,a_{r,hx})=e_1'J_h^{-1}(x)$ depends on x and h but not on T. Given that $K(u)$ is continuous with a compact support $[-1,1]$, $\tilde{K}^*(u)$ is also continuous and hence bounded with the same compact support. In terms of $\tilde{K}^*(\cdot)$, we can write Z_t as

$$Z_t=\sqrt{h}\left\{\tilde{K}_h^*(X_t-x)\varepsilon_t+\tilde{K}_h^*(X_t-x)m(X_t)-E\left[\tilde{K}_h^*(X_t-x)m(X_t)\right]\right\}.$$

So $E|Z_t|^{2\epsilon}<\infty$ and

$$Cov\left[Z_1,Z_{1+\ell}\right]\le 8\left[\alpha(\ell)\right]^{(\epsilon-1)/\epsilon}\left\{E\left[Z_1\right]^{2\epsilon}\right\}^{1/\epsilon}=O\left[\alpha(\ell)\right]^{(\epsilon-1)/\epsilon}. \tag{38}$$

These moment and mixing conditions are sufficient for the CLT:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t\to^d\mathcal{N}(0,V_h(x)),$$

where we have used Theorem 2.21 in Fan and Yao (2003). ■

Proof of Theorem 9. We first show that $B_h^{(p)}(x)$ is well-defined. Combining (38) with Assumption 7(i) and (iii) with $a \geq p$, we have

$$\sum_{\ell=1}^{\infty} |Cov[Z_1, Z_{1+\ell}]| \ell^p < \infty$$

and hence $B_h^{(p)}(x) = 2\lim_{T\to\infty} \sum_{\ell=1}^{T-1} Cov[Z_1, Z_{1+\ell}] \ell^p$ is indeed well-defined.

Next we prove the theorem. Parts (a) and (b) follow from Proposition 1 of Andrews (1991). It remains to show Part (c). Let $\varepsilon_{tx}(\tilde{\theta}) = Y_t - P_t(x)'\tilde{\theta}$ and

$$\nu(X_t; \tilde{\theta}) = e_1' J_h^{-1}(x) P_t(x) K_h(X_t - x) \varepsilon_{tx}(\tilde{\theta}).$$

Then

$$\hat{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\nu(X_t;\hat{\theta})\nu(X_\tau;\hat{\theta}) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\tilde{V}_h(x) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} W\left(\frac{t-\tau}{S_T}\right)\nu(X_t;\theta_h)\nu(X_\tau;\theta_h) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

The above representation casts our problem into the same problem considered in the second part of Theorem 1(b) of Andrews (1991). To see this, we only need to replace the notation $V(Z_t;\tilde{\theta})$ in Andrews (1991) by our $\nu(X_t;\tilde{\theta})$. To complete the proof, we verify Andrews' Assumptions B and C in our context. Note that

$$\frac{\partial \nu(X_t;\tilde{\theta})}{\partial \theta} = \left[e_1' J_h^{-1}(x) P_t(x)\right] K_h(X_t - x) P_t(x)$$

which does not depend on $\tilde{\theta}$. Under the assumptions given in the theorem, it is straightforward to show that Andrews' Assumptions B and C are satisfied. ■

References

[1] Andrews, D.W.K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59(3), 817–858.

[2] Bester, C. A., Conley, T. G., Hansen, C. B., and Vogelsang, T. J. (2015). Fixed-*b* Asymptotics for Spatially Dependent Robust Nonparametric Covariance Matrix Estimators, Forthcoming, *Econometric Theory.*

[3] Bradley, Richard C. (1985). On the Central Limit Question under Absolute Regularity, *The Annals of Probability*, 13(4), 1314–1325.

[4] Chen, X., Z. Liao and Y. Sun (2014). Sieve Inference on Possibly Misspecified Seminonparametric Time Series Models, *Journal of Econometrics,* 178, 639–658.

[5] Conley, T.G., L. P. Hansen, W. Liu (1997). Bootstrapping the Long Run, *Macroeconomic Dynamics*, 1, 279–311.

[6] Doukhan, P. (1994). *Mixing: Properties and Examples*, New York: Springer-Verlag.

[7] Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods,* New York: Springer.

[8] Gao, J. and Gijbels, I. (2008). Bandwidth Selection in Nonparametric Kernel Testing, *Journal of the American Statistical Association*, 484, 1584–1594.

[9] Gonçlaves, S. (2011). The Moving Blocks Bootstrap for Panel Linear Regression Models with Individual Fixed Effects, *Econometric Theory*, 27(5), 1048–1082

[10] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

[11] Ibragimov, I. A. (1962). Some Limit Theorems for Stationary Processes, *Theory of Probability & Its Applications*, 7(4), 349–382.

[12] Jansson, M. (2004). The Error in Rejection Probability of Simple Autocorrelation Robust Tests, *Econometrica*, 72(3), 937–946.

[13] Kiefer, N. and Vogelsang, T. (2002a). Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size, *Econometric Theory*, 18(6): 1350–1366.

[14] Kiefer, N. and Vogelsang, T. (2002b). Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation, *Econometrica*, 70(5), 2093–2095.

[15] Kiefer, N. and Vogelsang, T. (2005). A New Asymptotic Theory for Heteroskedasticity-autocorrelation Robust Tests, *Econometric Theory*, 21(06), 1130–1164.

[16] Kiefer, N., Vogelsang, T., and Bunzel, H. (2000). Simple Robust Testing of Regression Hypotheses, *Econometrica*, 68(3), 695–714.

[17] Kim, M. S. and Sun, Y. (2013). Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects, *Journal of Econometrics*, 177, 85–108.

[18] Masry, E. and J. Fan (1997). Local Polynomial Estimation of Regression Functions for Mixing Processes. *Scandinavian Journal of Statistics*, 24, 165–179.

[19] Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, 27(4), 575-595

[20] Newey, W. and West, K. (1987). A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703–708.

[21] Newey, W. and West, K. (1994). Automatic Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, 61(4), 631–654.

[22] Pritsker, M. (1998). Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models. *Review of Financial Studies*, 11, 449–487.

[23] Robinson, P. (1983). Nonparametric Estimators for Time Series, *Journal of Time Series Analysis*, 4(3), 185-207.

[24] Sun, Y. (2014a). Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference, *Journal of Econometrics* 178(3), 659–677.

[25] Sun, Y. (2014b). Fixed-smoothing Asymptotics in a Two-step GMM Framework, *Econometrica*, 82(6), 2327–2370.

[26] Sun, Y. (2014c). Fixed-smoothing Asymptotics in the Present of Strong Autocorrelation, *Advances in Econometrics:* Essays in Honor of Peter C. B. Phillips, 33, 2014, 23–63.

[27] Sun, Y. and Kim, M. S. (2015). Asymptotic F Test in a GMM Framework with Cross Sectional Dependence, *Review of Economics and Statistics*, 97(1), 210–223.

[28] Sun, Y. and Phillips, P. (2009). Bandwidth Choice for Interval Estimation in GMM Regression, Working paper, UCSD.

[29] Sun, Y., Phillips, P., and Jin, S. (2008). Optimal Bandwidth Selection in Heteroskedasticity-autocorrelation Robust Testing, *Econometrica*, 76(1), 175–194.

[30] Vogelsang, T. (2012). Heteroskedasticity, Autocorrelation, and Spatial Correlation Robust Inference in Linear Panel Models with Fixed-effects, *Journal of Econometrics*, 166, 303–319.