Published in final edited form as:
J Multivar Anal. 2015 January 1; 133: 277–290. doi:10.1016/j.jmva.2014.09.017.

Nonparametric Functional Central Limit Theorem for Time Series Regression with Application to Self-normalized Confidence Interval

Seonjin Kim[a], **Zhibiao Zhao**[b], and **Xiaofeng Shao**[c]

Seonjin Kim: kims20@miamioh.edu; Zhibiao Zhao: zuz13@stat.psu.edu; Xiaofeng Shao: xshao@illinois.edu

[a]Department of Statistics, Miami University, 311 Upham Hall, Oxford, OH 45056

[b]Department of Statistics, Penn State University, 326 Thomas, University Park, PA 16802

[c]Department of Statistics, University of Illinois, 725 South Wright Street, Urbana, IL 61801

Abstract

This paper is concerned with the inference of nonparametric mean function in a time series context. The commonly used kernel smoothing estimate is asymptotically normal and the traditional inference procedure then consistently estimates the asymptotic variance function and relies upon normal approximation. Consistent estimation of the asymptotic variance function involves another level of nonparametric smoothing. In practice, the choice of the extra bandwidth parameter can be difficult, the inference results can be sensitive to bandwidth selection and the normal approximation can be quite unsatisfactory in small samples leading to poor coverage. To alleviate the problem, we propose to extend the recently developed self-normalized approach, which is a bandwidth free inference procedure developed for parametric inference, to construct point-wise confidence interval for nonparametric mean function. To justify asymptotic validity of the self-normalized approach, we establish a functional central limit theorem for recursive nonparametric mean regression function estimates under primitive conditions and show that the limiting process is a Gaussian process with non-stationary and dependent increments. The superior finite sample performance of the new approach is demonstrated through simulation studies.

Keywords

Conditional heteroscedasticity; Functional central limit theorem; Nonparametric regression; Self-normalization; Time series

Correspondence to: Zhibiao Zhao, zuz13@stat.psu.edu.

1. Introduction

Nonparametric methods are useful complements to the traditional well developed parametric counterparts. They allow the users to entertain model flexibility while reducing modeling bias, and partly due to this reason, nonparametric inference has been extensively studied. This paper concerns a new way of addressing nonparametric inference in the time series setting. There is a huge literature about the use of nonparametric methods in time series analysis, and asymptotic theory for nonparametric estimators and tests has been quite well developed for weakly dependent time series data. We refer the reader to chapters 5-10 in Fan and Yao (2003) for a nice introduction of some basic ideas and results.

Given stationary time series $\{(X_i, Y_i)\}_{i=1}^n$, we focus on inference for the conditional mean function $\mu(x) = \mathbb{E}(Y_i|X_i = x)$; see Section 4 for some possible extensions to other nonparametric functions. Let $\hat{\mu_n}(x)$ be a nonparametric estimate of $\mu(x)$ based on the full sample. Under suitable regularity and weak dependence conditions, we have

$$\sqrt{nb_n}\frac{\hat{\mu}_n(x) - \mu(x) - b_n^2 r(x)}{s(x)} \xrightarrow{d} N(0,1), \quad (1)$$

where b_n is an appropriate bandwidth, $b_n^2 r(x)$ is the bias term, $s^2(x)$ is the asymptotic variance function, and $\xrightarrow{d}$ stands for convergence in distribution. To construct a point-wise confidence interval for $\mu(x)$, the traditional approach involves consistent estimation of $s^2(x)$ through an extra nonparametric smoothing procedure which inevitably introduces estimation error. The latter issue becomes even more serious when $s(x) \approx 0$ so that the left hand side of (1) is very sensitive to the estimation error of $s(x)$. In particular, even if the absolute estimation error is small, the relative estimation error can be large, which leads to poor coverage in the constructed confidence interval. Thus, one needs to deal with the unpleasant phenomenon that, the smaller $s(x)$ (i.e. lower noise level), the more difficult to carry out statistical inference. Furthermore, nonparametric estimation of $s(x)$ involves extra bandwidth parameter(s). Two users using two different bandwidths in estimating $s(x)$ for the same data set may get quite different results.

To alleviate the above-mentioned problem in the traditional inference procedure, we propose to extend the recently developed self-normalized (SN, hereafter) approach (Shao, 2010) to nonparametric setting. The SN approach was developed for a finite dimensional parameter of a stationary time series and it has the nice feature of being bandwidth free. The basic idea of the SN approach, when applied to nonparametric setting, is to use estimates of $\mu(x)$ on the basis of recursive subsamples to form a self-normalizer that is an inconsistent estimator of $s(x)$. Although it is inconsistent, the self-normalizer is proportional to $s(x)$, and the limiting distribution of the self-normalized quantity is pivotal. The literature on the SN approach and related methods (Lai and Siegmund, 1983; Lobato, 2001; Kiefer et al., 2000; Kiefer and Vogelsang, 2005; Shao, 2010, 2012; Shao and Zhang, 2010; Zhou and Shao, 2013) has been growing recently, but most of the work is limited to parametric inference, where the parameter of interest is finite dimensional and the method of estimation does not involve smoothing. Kim and Zhao (2013) studied SN approach for the nonparametric mean function

in longitudinal models, but the data are essentially independent due to the independent subjects. To the best of our knowledge, the SN-based extension to nonparametric time series inference seems new.

An important theoretical contribution of this article is that we establish non-parametric functional central limit theorem (FCLT, hereafter) of some recursive estimates of $\mu(\cdot)$ under primitive conditions. To be specific, denote by $\hat{\mu_m}(x)$ the nonparametric estimate of $\mu(x)$ using data $\{(X_i, Y_i)\}_{i=1}^m$ up to time m and bandwidth b_m. Throughout, denote by $\lfloor \upsilon \rfloor$ the integer part of υ. We show that, due to the sample-size-dependent bandwidths, the process $\{\hat{\mu_{\lfloor nt \rfloor}}(x) - \mu(x)\}$ indexed by t, after proper normalization, converges weakly to a Gaussian process $\{G_t\}$ with non-stationary and dependent increments. Such a result is very different from the FCLT required for the SN approach in the parametric inference problems, where the limiting process is a Brownian motion with stationary and independent increments.

Throughout, we write $\xi \in \mathcal{L}^p$ ($p \geq 1$) if $\|\xi\|_p := (\mathbb{E}|\xi|^p)^{1/p} < \infty$. The symbols $O_p(1)$ and $o_p(1)$ signify being bounded in probability and convergence to zero in probability, respectively. For sequences $\{a_n\}$ and $\{c_n\}$, write $a_n \asymp c_n$ if $a_n/c_n \to 1$. The article is organized as follows. Section 2 presents the main results, including the FCLT for nonparametric recursive estimates and the self-normalization based confidence interval. Simulation results are presented in Section 3. Section 4 concludes and technical details are gathered in Section 5.

2. Main results

We consider the nonparametric mean regression model:

$$Y_i = \mu(X_i) + e_i, \quad (2)$$

where $\mu(\cdot)$ is the nonparametric mean function of interest and $\{e_i\}$ are noises. As an important special case, let $X_i = Y_{i-1}$ and $e_i = \sigma(X_i)\varepsilon_i$ for innovations $\{\varepsilon_i\}$ and a scale function $\sigma(\cdot)$, then we have the nonparametric autoregressive (AR) model $Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\varepsilon_i$, which includes many nonlinear time series models, such as linear AR, threshold AR, exponential AR, and AR with conditional heteroscedasticity; see Fan and Yao (2003). We assume that $\{(X_i, Y_i)\}_{i=1}^n$ are stationary time series observations so that they have a natural ordering in time, i.e., (X_i, Y_i) is the observation at time i.

2.1. Nonparametric FCLT for recursive estimates

Throughout let x be a fixed interior point in the support of X_i. Denote by $\hat{\mu_m}(x)$ the nonparametric estimate of $\mu(x)$ based on data $\{(X_i, Y_i)\}_{i=1}^m$ up to time m. In this paper we consider the local linear kernel smoothing estimator (Fan and Gijbels, 1996) of $\mu(x)$:

$$\hat{\mu}_m(x) = \hat{a}_0,\ (\hat{a}_0, \hat{a}_1) = \underset{(a_0,a_1)}{\operatorname{argmin}} \sum_{i=1}^m [Y_i - a_0 - a_1(X_i - x)]^2 K\left(\frac{X_i - x}{b_m}\right), \quad (3)$$

where $K(\cdot)$ is a kernel function and $b_m > 0$ is the bandwidth. By elementary calculation,

$$\hat{\mu}_m(x)=\frac{M_m(2)N_m(0)-M_m(1)N_m(1)}{M_m(2)M_m(0)-M_m(1)^2}, \quad (4)$$

where, for $j = 0, 1, 2$,

$$M_m(j)=\sum_{i=1}^{m}(X_i-x)^j K\left(\frac{X_i-x}{b_m}\right), N_m(j)=\sum_{i=1}^{m}(X_i-x)^j Y_i K\left(\frac{X_i-x}{b_m}\right). \quad (5)$$

Let $c \in (0, 1)$ be a fixed small constant. With $m = \lfloor cn \rfloor, \lfloor cn \rfloor + 1, \ldots, n$, we can obtain the recursive estimates $\hat{\mu}_{\lfloor cn \rfloor}(x), \hat{\mu}_{\lfloor cn \rfloor+1}(x), \ldots, \hat{\mu}_n(x)$ of the same quantity $\mu(x)$. In this section we establish a nonparametric FCLT for the process $\{\hat{\mu}_{\lfloor nt \rfloor}(x)\}_{c\le t\le 1}$. If $b_m = \infty$ and we drop the linear term $a_1(X_i - x)$ from (3) (i.e., we consider the local constant estimation), then $\hat{\mu}_m(x)=m^{-1}\sum_{i=1}^{m} Y_i$ reduces to the partial sum process of $\{Y_i\}_{i=1}^{m}$, which has been the focus of classical FCLT.

The extension of parametric FCLT to nonparametric setting is far from trivial and the main complication lies in the following two aspects:

- **i.** The bandwidth b_m depends on the sample size m and it has an impact on the asymptotic behavior of $\hat{\mu}_m(x)$. It is well known that the optimal bandwidth of $\hat{\mu}_n(x)$ is $b_n = C(K, r, s)n^{-1/5}$ for some constant $C(K, r, s)$ that depends only on the kernel $K(\cdot)$, the bias function $r(\cdot)$, and the asymptotic variance function $s^2(\cdot)$. Therefore, the optimal bandwidth b_m for sample size m satisfies $b_m = b_n(n/m)^{1/5}$, where b_n is the bandwidth chosen on the basis of full sample $\{(X_i, Y_i)\}_{i=1}^{n}$. For example, we can use the plug-in method or the cross-validation method (Li and Racine, 2007) to choose b_n and then set $b_m = b_n(n/m)^{1/5}$.
- **ii.** For parametric FCLT, the limiting process is typically a scaled Brownian motion in the weakly dependent setting. By contrast, due to the sample-size-dependent bandwidths, the limit for the partial sum process in (4) is unknown and a careful investigation is needed.

Next, we introduce some technical assumptions.

Assumption 1 (Dependence condition)—In (2), $\mathbb{E}(e_i|X_i, X_{i-1}, \ldots, X_1, e_{i-1}, e_{i-2}, \ldots, e_1) = 0$. Moreover, $\{(X_i, e_i)\}_{i\in\mathbb{N}}$ is stationary and α-mixing with mixing coefficient $\alpha_k < C\rho^k$, $k \in \mathbb{N}$, for some constants $C < \infty$ and $\rho \in (0, 1)$.

Assumption 1 implies $\mathbb{E}(e_i|X_i) = \mathbb{E}[\mathbb{E}(e_i|X_i, \ldots, X_1, e_{i-1}, \ldots, e_1)|X_i] = 0$, which ensures the identifiability of $\mu(x)$ through the conditional mean regression $\mathbb{E}(Y_i|X_i = x) = \mu(x)$. The α-mixing framework is widely used in time series analysis; see Fan and Yao (2003).

Definition 1—Let $p_X(\cdot)$ be the density function of X_i. Throughout we assume that $p_X(\cdot)$ is bounded. Recall that x is a given point. For $q > 0$, define

$$\ell_q(\upsilon)=p_X(x+\upsilon)\mathbb{E}(|e_i|^q|X_i=x+\upsilon),\quad \upsilon\in\mathbb{R}\ . \quad (6)$$

Define the set of functions

$$C(q)=\left\{f(\cdot):\int_{\mathbb{R}}|f(u)|^q du<\infty,\lim_{\varepsilon\to 0}\int_{\mathbb{R}}|\ell_q(u\varepsilon)-\ell_q(0)\parallel f(u)|^q du=0\right\}. \quad (7)$$

Intuitively, (7) asserts that $\ell_q(\upsilon)$ is continuous at $\upsilon=0$ under the norm induced by $|f(\cdot)|^q$. If $f(\cdot)$ has a bounded support and $\ell_q(\cdot)$ is continuous at $\upsilon=0$, then $f\in\mathcal{C}(q)$.

Assumption 2 (Regularity condition)—(i) For some $\delta>0$, $e_i\in\mathcal{L}^{4+\delta}$. (ii) $p_X''(\cdot)$ is continuous at x and $p_X(x)>0$. (iii) $\mu'''(\cdot)$ is bounded and continuous at x. (iv) $K(\cdot)$ is symmetric, has bounded derivative, and (recall $\mathcal{C}(q)$ in Definition 1)

$$\begin{aligned}&\sup_{u\in\mathbb{R}}(1+|u|+|u|^2)|K(u)|<\infty,\ \int_{\mathbb{R}}K(u)du=1,\ \int_{\mathbb{R}}|u^3K(u)|du<\infty,\\ &g(u)\in\mathscr{C}(2)\cap\mathscr{C}(4+\delta)\text{ with }g(u)=|K(u)|+\sup_{t\geq c^{1/5}}|uK'(tu)|.\end{aligned} \quad (8)$$

To establish the asymptotic normality of $\hat{\mu}_{\lfloor nt\rfloor}(x)$ at a fixed t, it suffices to assume $e_i\in\mathcal{L}^{2+\delta}$ for some $\delta>0$, see Theorem 2.22 in Fan and Yao (2003); for FCLT, we need the stronger moment assumption $e_i\in\mathcal{L}^{4+\delta}$ to establish the tightness of the process $\{\hat{\mu}_{\lfloor nt\rfloor}(x)\}_{c\leq t\leq 1}$. For nonparametric kernel smoothing estimation, it is typically assumed that the kernel has bounded support and bounded derivative. For this type of kernel function, if $\ell_2(\cdot)$ and $\ell_{4+\delta}(\cdot)$ are continuous at $\upsilon=0$, then (8) trivially holds. On the other hand, Assumption 2(iv) allows the kernel function to have an unbounded support with sufficiently thin tails, such as the standard Gaussian kernel.

Theorem 1—In (3), let $b_m=b_n(n/m)^{1/5}$ and $b_n\propto n^{-1/5}$. Suppose Assumptions 1−2 hold. Then the following weak convergence holds in the Skorokhod space (Billingsley, 2009)

$$\left\{\frac{\sqrt{nb_np_X(x)}}{\sigma(x)}t^{4/5}\left[\hat{\mu}_{\lfloor nt\rfloor}(x)-\mu(x)-b_{\lfloor nt\rfloor}^2r(x)\right]\right\}_{c\leq t\leq 1}\Rightarrow\{G_t\}_{c\leq t\leq 1}, \quad (9)$$

where $r(x)=\mu''(x)\int_{\mathbb{R}}u^2K(u)du/2$, $\sigma^2(x)=\mathbb{E}(e_i^2|X_i=x$, $c\in(0,1)$ is any given constant, and $\{G_t\}_{c\leq t\leq 1}$ is a centered Gaussian process with autocovariance function given by

$$\sum(t,t')=\mathrm{cov}(G_t,G_{t'})=\min(t,t')\int_{\mathbb{R}}K(t^{1/5}u)K(t'^{1/5}u)du. \quad (10)$$

The asymptotic normality in (1) is a direct application of Theorem 1 with $t=1$. For parametric inference, FCLT often admits Brownian motion, which has stationary and independent increments, as its limit. By contrast, in the nonparametric context, the properly standardized nonparametric recursive estimates converge to a Gaussian process $\{G_t\}$ with non-stationary and dependent increments, owing to the sample-size-dependent bandwidth.

The covariance function of the Gaussian process depends on the kernel function as shown in (10). If we use the standard normal density as the kernel function, i.e., $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$, then $\sum(t,t') = \min(t,t')/\sqrt{2\pi(t^{2/5}+t'^{2/5})}$.

2.2. Self-normalization based confidence interval

In this section, we focus on point-wise confidence interval construction for $\mu(x)$ using the SN approach. First, we provide a brief discussion of the traditional approach. By Theorem 1 with $t = 1$, (1) holds with

$$s^2(x) = \frac{\sigma^2(x)}{p_X(x)} \int_{\mathbb{R}} K^2(u)du, \text{ where } \sigma^2(x) = \mathbb{E}(e_i^2 | X_i = x). \quad (11)$$

In the traditional approach, one would construct a consistent estimate of $s^2(x)$ by using consistent estimates for $p_X(x)$ and $\sigma^2(x)$. For $p_X(x)$, we can use a nonparametric kernel density estimate with bandwidth τ_n:

$$\hat{p}_X(x) = \frac{1}{n\tau_n} \sum_{i=1}^{n} K\left(\frac{X_i - x}{\tau_n}\right). \quad (12)$$

Let $\hat{e}_i = Y_i - \hat{\mu_n}(X_i)$ be the residuals, then one can estimate $\sigma^2(x)$ by applying the local constant kernel smoothing procedure to $(X_1, \hat{e}_1^2), \ldots, (X_n, \hat{e}_n^2)$ with another bandwidth h_n:

$$\hat{\sigma}_n^2 = \left\{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)\right\}^{-1} \sum_{i=1}^{n} \hat{e}_i^2 \left(\frac{X_i - x}{h_n}\right). \quad (13)$$

Finally, $s^2(x)$ can be estimated by plugging estimates $\hat{\sigma}_n^2$ and $\hat{p_X}(x)$ into (11). Therefore, traditional approach requires the selection of two additional bandwidths h_n and τ_n.

A distinctive feature of the SN approach (Shao, 2010) is that it does not involve any bandwidth parameter. When applied to the nonparametric inference problem at hand, the key idea of the SN approach is to construct an inconsistent estimator of $s^2(x)$ using recursive estimates of $\mu(x)$ and form a self-normalized quantity. Such an inconsistent self-normalizer is proportional to $s^2(x)$, which can then be canceled out in the limiting distribution of the self-normalized quantity. The SN approach was developed in the context of parametric inference and its generalization to nonparametric inference requires nontrivial modifications. First, we need to deal with the bias in nonparametric estimation. Instead of estimating the bias explicitly, we propose to use a higher order kernel to make the bias asymptotically negligible. In particular, we can impose the following assumption:

Assumption 3—In (9), assume without loss of generality that $r(x) = 0$.

Assumption 3 assumes $r(x) = 0$; otherwise, we can use a higher order kernel to achieve bias reduction. Note that the constant in $r(x)$ is proportional to $\int_{\mathbb{R}} u^2 K(u)du$. Let

$$K^*(u)=2K(u) - K(u/\sqrt{2})/\sqrt{2}.$$

Then we can easily verify $\int_{\mathbb{R}} u^2 K^*(u)du = 0$ so the second order bias vanishes, i.e., $r(x) = 0$. The idea can be traced back to jackknifing kernel regression estimator of Hardle (1986) and is also used in Wu and Zhao (2007) in the inference of trend with time series errors. In practice, using this higher order kernel is asymptotically equivalent to $\tilde{\mu}_m(x)=2\hat{\mu}_m(x|b_m) - \hat{\mu}_m(x|\sqrt{2}b_m)$, where $\tilde{\mu_m}(x|b_m)$ and $\hat{\mu}_m(x|\sqrt{2}b_m)$ are the estimates of $\mu(x)$ using bandwidth b_m and $\sqrt{2}b_m$, respectively. If we use the latter kernel $K^*(u)$ with $K(u)$ being the standard Gaussian kernel, then the autocovariance function in (10) becomes

$$\sum(t,t')=\frac{\min(t,t')}{\sqrt{2\pi}}\left\{\frac{4+1/\sqrt{2}}{\sqrt{t^{2/5}+t'^{2/5}}}-\frac{2}{\sqrt{t^{2/5}+t'^{2/5}}}-\frac{2}{\sqrt{2t^{2/5}+t'^{2/5}}}\right\}. \quad (14)$$

By stationarity and (1), for each $t \in (0, 1)$, both $\hat{\mu_{\lfloor nt \rfloor}}(x)$ and $\hat{\mu_n}(x)$ have asymptotic variances proportional to $s^2(x)$. Motivated by this feature, we consider certain ratio of $\hat{\mu_n}(x)$ and an aggregated version of the process $\{\hat{\mu_{\lfloor nt \rfloor}}(x), t \in [c,1]\}$ to cancel out $s^2(x)$.

Theorem 2—Suppose that the same assumptions in Theorem 1 hold. Further assume Assumption 3 holds. Then we have

$$\frac{\hat{\mu}_n(x)-\mu(x)}{V_n}\xrightarrow{d}\frac{G_1}{\sqrt{\int_c^1|G_t-t^{4/5}G_1|^2dt}}=\xi,\ where\ V_n=n^{-13/10}\left\{\sum_{m=\lfloor cn\rfloor}^{n}m^{8/5}|\hat{\mu}_m(x)-\hat{\mu}_n(x)|^2\right\}^{1/2}. \quad (15)$$

Here $\{G_t\}$ is the Gaussian process in Theorem 1. Consequently, an asymptotic $100(1-\alpha)\%$ confidence interval for $\mu(x)$ is $\hat{\mu_n}(x)\pm q_{1-\alpha}V_n$, where q_τ is the τ quantile of $|\xi|$.

For a given c and kernel function, the distribution of ξ is pivotal and the quantiles of $|\xi|$ can be obtained through Monte Carlo simulations; see Table 1 for simulated quantiles. In the context of confidence interval construction for finitedimensional parameters, Shao (2010) used a similar self-normalization method with no trimming (i.e. $c = 0$). The use of trimming is also adopted in Zhou and Shao (2013), who proposed an extension of the self-normalized approach to linear regression models with fixed regressors and dependent errors. In our problem, trimming seems necessary as nonparametric estimate of $\mu(x)$ on the basis of a small sample is very unstable, and in the extreme case of only one point (X_1, Y_1), we are unable to carry out the estimation. Throughout the simulation and data illustration, we set $c = 0.1$, which seems to work pretty well. In general large c is not recommended as we lose some efficiency, whereas some recursive estimates may not be stable when c is too small. A similar finding is reported in Zhou and Shao (2013), where $c = 0.1$ is also shown to be a good choice via simulations.

We summarize the procedure to obtain the SN-based confidence interval for $\mu(x)$:

1. Find the optimal bandwidth b_n using the existing bandwidth selection procedure, such as cross-validation or plug-in method; set $b_m = b_n(n/m)^{1/5}$.
2. Calculate the recursive estimates of $\mu(x)$, i.e., $\hat{\mu_m}(x)$ for $m = \lfloor cn \rfloor, \cdots, n$; see (3).
3. For a given nominal level $1 - \alpha$, the SN-based interval is constructed as $\hat{\mu_n}(x) \pm q_{1-\alpha} V_n$; see (15).

Thus the SN approach only involves the choice of a smoothing parameter in the estimation stage, which seems necessary as a good estimator is usually needed for inference. By contrast, the traditional approach requires a consistent nonparametric estimate of $s(x)$, which involves selecting two extra smoothing parameters and always introduces estimation error in a finite sample. Thus, our proposed method provides an easy-to-implement and fully nonparametric inference technique.

The proposed self-normalization based approach is effectively performing inference using an inconsistent estimator for the asymptotic variance, the idea of which has attracted considerable attention recently. Kiefer, Vogelsang and co-authors proposed the fixed-b approach in the context of heteroscedasticityautocorrelation consistent robust testing; see Kiefer et al. (2000), Kiefer and Vogelsang (2002, 2005), Vogelsang (2003), among others. By holding the smoothing parameter or truncation lag as a fixed proportion of sample size, the resulting asymptotic variance matrix estimator is no longer consistent, but is proportional to the asymptotic variance matrix, and consequently the resulting studentized statistic has a pivotal non-normal limiting distribution. Similar ideas can be found in the self-normalization scheme of Lobato (2001) and Shao (2010). For confidence interval construction of finite-dimensional parameters, Shao's selfnormalized approach relies on the functional convergence of standardized recursive estimates based partial sum process to standard Brownian motion. In contrast, due to the sample-size-dependent bandwidth of nonparametric recursive estimates, our self-normalization is based on a different Gaussian process with non-stationary dependent increments. To the best of our knowledge, the functional convergence result and the extension of the self-normalization idea to nonparametric time series context seems new.

Remark 1—In Theorems 1−2, the limiting distribution is the same for any arbitrarily given (but fixed) x, and we can use this result to construct the pointwise confidence interval for $\mu(\cdot)$. On the other hand, if we wish to construct the uniform or simultaneous confidence interval on an interval $x \in [a, b]$, then we must obtain some uniform convergence in $x \in [a, b]$, which is more technically challenging. In this paper we focus on the pointwise confidence interval case, and the uniform confidence interval case will serve as a direction for future research.

3. Numerical results

We compare the finite sample performance of the proposed self-normalization based method in Theorem 2 to that of the traditional method based on asymptotic normality and consistent estimation of the asymptotic variance function. We adopt the bias reduction procedure in Assumption 3 so the effective kernel becomes $K(u) = 2\phi(u) - \phi(u/\sqrt{2})/\sqrt{2}$ where $\phi(\cdot)$ is

the standard normal density, and the covariance function of the Gaussian process $\{G_t\}$ in Theorem 2 is given by (14).

Denote by ℓ_τ the τ-percentile of X's. Let $x_j = \ell_{0.1} + (j-1)(\ell_{0.9}-\ell_{0.1})/20$, $j = 1, \ldots, 21$, be uniform grid points on $[\ell_{0.1}, \ell_{0.9}]$. For each x_j, we construct a 95% confidence interval for $\mu(x_j)$, and denote by p_j the empirical coverage probability, which is computed as the proportion of realizations among 1,000 replications whose confidence interval covers $\mu(x_j)$. Define the average deviation of p_j from the nominal level 95% as

$$\text{Average deviation of coverage probablities} = \frac{1}{21}\sum_{j=1}^{21}|p_j - 95\%|, \quad (16)$$

with a smaller value indicating better overall performance.

First, consider the stochastic regression model with time series errors:

$$\text{Model I: } Y_i = \mu(X_i) + \lambda\sqrt{1+2X_i^2}\varepsilon_i,\ \varepsilon_i = \theta\varepsilon_{i-1} + \sqrt{1-\theta^2}g_i, \quad (17)$$

where X_i are independently distributed with uniform distribution on [0, 1] and g_i are independent standard normal random variables. The model allows conditional heteroscedasticity and dependence in ε_i, with the parameter $\theta \in (0, 1)$ controlling the strength of dependence. We consider $\theta = 0.0, 0.4, 0.8$, representing models ranging from independence to strong dependence. Let $\mu(x) = 0.6x$ be the mean regression function of interest. To investigate the effect of noise level, we let $\lambda = 0.03, 0.06, 0.12, 0.24$, ranging from low noise level to high noise level.

Next, consider the autoregressive conditional heteroscedastic model

$$\text{Model II: } Y_i = \mu(Y_{i-1}, \theta) + \lambda\sqrt{1+2Y_{i-1}^2}\varepsilon_i, \quad (18)$$

where ε_i are independent standard normal errors, $(X_i, Y_i) = (Y_{i-1}, Y_i)$, and $\mu(x;\theta) = \theta x$ is the function of interest. As in Model I, we consider different combinations of $\theta = 0.0, 0.4, 0.8$ and $\lambda = 0.03, 0.06, 0.12, 0.24$.

To select the bandwidth b_n, we use Ruppert et al. (1995)'s plug-in method, implemented using the `R` command `dpill` in the package `KernSmooth`. To implement the traditional method, we select h_n in (13) using the latter plug-in method, and consider the following five popular choices for the nonparametric kernel density bandwidth τ_n in (12):

- **i.** The normal reference rule-of-thumb method with factor 0.9, i.e., $\tau_n = 0.9 * n^{-1/5} * \min\{\text{sd}(X), \text{IQR}(X)\}$, where sd($X$) and IQR($X$) are, respectively, the standard deviation and interquartile of $X_1, \ldots, X_n$. This method is implemented using the `R` command `bw.nrd0` in the stats `package`.

ii. The normal reference rule-of-thumb method with factor 1.06, implemented using the `R` command `bw.nrd` in the `stats` package.

iii. The unbiased cross-validation bandwidth method, implemented using the `R` command `bw.ucv` in the `stats` package.

iv. The biased cross-validation bandwidth method, implemented using the `R` command `bw.bcv` in the `stats` package.

v. Sheather and Jones (1991)'s method by minimizing estimates of the mean integrated squared error, implemented using the `R` command `bw.SJ` in the `stats` package.

In all settings, we use sample size $n = 300$.

The results are presented in Table 2. For the asymptotic normality methods with the five different bandwidth selection methods, there is a substantial deviation between the actual coverage probability and the nominal level. For Model I, the deviation becomes clearly more severe as the dependence increases. The latter can be explained by the fact that stronger positive dependence corresponds to a smaller effective sample size, which results in larger estimation error and worse coverage. What appears intriguing is that for Model I, as the noise level λ decreases from 0.24 to 0.03, the traditional methods perform even worse. This phenomenon is presumably due to the fact that the relative estimation error of estimating $s(x)$ in (1) is more severe when $s(x)$ becomes smaller, especially when $s(x)$ is close to 0. For Model II, the performance of the traditional methods is relatively consistent across dependence and noise level, with average deviations around 6%-7%. By contrast, the proposed self-normalization based method delivers much more accurate coverage probabilities and is fairly robust with respect to the magnitude of dependence and noise level.

4. Discussions and conclusions

This article proposes an extension of the self-normalized approach (Shao, 2010) to nonparametric inference in a time series context. The new approach overcomes the drawbacks of the traditional approach, where consistent estimation of the asymptotic variance function is needed with an extra smoothing procedure. The finite sample performance convincingly demonstrates that the proposed methodology delivers substantially more accurate coverage than the traditional approach. The new inference method does not require any additional bandwidth parameters other than b_n, which seems necessary for the estimation of the nonparametric mean function.

The work presented here seems to be the first attempt to generalize the selfnormalization based methods to nonparametric inference problems in the time series setting. We limit our framework to nonparametric mean regression with one covariate variable and our theory is developed for time series data. There are certainly room for further extensions of our methodology to nonparametric and semiparametric problems with multiple covariates and to dependent data of other types, such as longitudinal data or spatial data. The key difficulty would be to establish FCLT for certain recursive estimates based on some natural ordering

of the data. For example, for data on a squared lattice, we can construct the m-th recursive estimate from data on expanding squares $-m \le i, j \le m$, but the corresponding FCLT is more challenging. It is also worth noting that nonparametric estimation and inference have been well studied for iid data, and an application of the SN approach developed in this article to iid setting encounters a practical problem because there is no natural ordering with iid data. With different ordering, the SN approach may deliver different results. In view of this practical drawback, it would be interesting to develop a new SN-based approach that does not depend on the ordering of the data.

Another direction for possible extension is to consider self-normalization based inferences for a general nonparametric function, denoted by $\mu(\cdot)$, such as conditional mean function, conditional quantile function, nonparametric density function, and conditional distribution function. Consider recursive estimates $\hat{\mu_m}(x)$ of $\mu(\cdot)$ using data up to time m. Assume that there exists some functions $r(\cdot)$ and $H(\cdot,\cdot)$ such that the following asymptotic Bahadur type representation holds

$$\hat{\mu}_m(x) - \mu(x) = b_m^2 r(x) + \frac{1}{m b_m p_X(x)} \sum_{i=1}^{m} H(X_i, Y_i) K\left(\frac{X_i - x}{b_m}\right) + R_m(x), \quad (19)$$

uniformly over $\lfloor cn \rfloor \le m \le n$, where $R_m(x)$ is the negligible remainder term. Under conditions similar to Assumptions 1 and 2, we can establish similar FCLT as in Theorem 1, which can be used to construct self-normalized confidence interval for $\mu(x)$ as in Theorem 2. However, it can be challenging to obtain the uniform representation (19). As an example, consider nonparametric quantile regression and denote by $\mu(x|\tau)$, $\tau \in (0, 1)$, the conditional τ-quantile of Y_i given $X_i = x$. We can estimate $\mu(x|\tau)$ by the local linear quantile regression

$$(\hat{\mu}(x|\tau), \hat{\theta}) = \operatorname*{argmin}_{\mu,\theta} \sum_{i=1}^{n} \rho_\tau(Y_i - \mu - \theta(X_i - x)) K\left(\frac{X_i - x}{b_n}\right),$$

where $\rho_\tau(t) = |t| + (2\tau - 1)t$ is the check function or quantile loss function at quantile τ. By Honda (2000), (19) holds for each fixed m, but it requires significantly more work to establish uniform representations and show the uniform negligibility of the remainder terms $R_m(x)$. We leave these possible extensions for future work.

Acknowledgments

We are grateful to the Editor and two anonymous referees for their constructive comments. Shao acknowledges partial financial support from NSF grants DMS-0804937 and DMS-1104545; Zhao acknowledges partial financial support from NSF grant DMS-1309213 and NIDA grant P50-DA10075. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NSF.

5. Appendix: Technical proofs

Throughout this section C, c_1, c_2, … are generic constants that may vary from line to line.

5.1. Some results on mixing process

In this section we present some results on mixing process which may be of independent interest for other nonparametric inference problems involving dependent data.

Lemma 1 (Proposition 2.5 in Fan and Yao (2003))

Let U and V be two random variables such that $U \in \mathcal{L}^{p_1}$ and $V \in \mathcal{L}^{p_2}$ for some $p_1 > 1$, $p_2 > 1$, and $1/p_1 + 1/p_2 < 1$. Then

$$|\mathrm{cov}(U, V)| \le 8\alpha(U, V)^{1-1/p_1-1/p_2}\|U\|_{p_1}\|V\|_{p_2}.$$

Here $\alpha(U, V)$ is the α-mixing coefficient between the two σ-algebras generated by U and V.

In Lemmas 2–4 below, let $\{Z_i\}_{i\in\mathbb{Z}}$ be a stationary α-mixing process with mixing coefficient $a_k \le C\rho^k$ for some constants $C < \infty$ and $\rho \in (0, 1)$. Lemma 2 presents an exponential inequality for the tail probability of $\sum_{i=1}^n Z_i$, Lemma 3 establishes a uniform convergence result with optimal rate (up to a logarithm factor) for partial sum process of functions of $\{Z_i\}$, and Lemma 4 presents a moment inequality for $\mathbb{E}(Z_0 Z_i Z_r^2)$.

Lemma 2

Assume $\mathbb{E}(Z_0) = 0$ and $\mathbb{P}\{|Z_0| \le b\} = 1$ for some b. For $\ell \le \lfloor n/2 \rfloor$ and $z > 0$,

$$\mathbb{P}\left\{\left|\sum_{i=1}^n Z_i\right| > z\right\} \le 4\exp\left\{-\frac{z^2\ell}{144n^2\mathbb{E}(Z_0^2)+4bzn}\right\} + 22\ell\alpha_{\lfloor n/(2\ell)\rfloor}\sqrt{1+\frac{4bn}{z}}.$$

Proof. Let $s = n/(2\ell)$. By Theorem 2.18 in Fan and Yao (2003),

$$\mathbb{P}\left\{\left|\sum_{i=1}^n Z_i\right| > z\right\} \le 4\exp\left\{-\frac{z^2\ell}{16n^2\Gamma_s/s^2+4bzn}\right\} + 22\ell\alpha_{\lfloor s\rfloor}\sqrt{1+\frac{4bn}{z}},$$

where $\Gamma_s = \max_{0\le j\le 2\ell-1}\mathbb{E}\{(\lfloor js\rfloor+1-js)Z_1+Z_2+\cdots+Z_r+(js+s-\lfloor js+s\rfloor)Z_r+1\}^2$ and $r = \lfloor (j+1)s\rfloor - \lfloor js\rfloor$. The result then follows from the Cauchy-Schwarz inequality

$$\Gamma_s \le (r+1)\mathbb{E}(Z_1^2+\cdots+Z_{r+1}^2) \le \lfloor s+2\rfloor^2\mathbb{E}(Z_0^2) \le 9s^2\mathbb{E}(Z_0^2).$$

Lemma 3

Let $\{\theta_m\}_{m\in\mathbb{N}}$, be a sequence of deterministic parameters and $h(\cdot,\cdot)$ a bivariate function such that $\mathbb{E}[h(Z_0, \theta_m)] = 0$ and $\mathbb{P}\{|h(Z_0, \theta_m)| \le b\} = 1$ for a constant b. Define

$$H_m=\sum_{i=1}^{m} h(Z_i,\theta_m),\ m\in\mathbb{N}.$$

Let $c \in (0, 1)$ be any fixed constant. Suppose there exist σ_m and a constant c_1 such that

$$\mathbb{E}[h^2(Z_0,\theta_m)]\le\sigma_m^2 \text{ and } \sqrt{m}\sigma_m>c_1,\ m=\lfloor cn\rfloor,\dots,n,$$

with sufficiently large n. Then as $n\to\infty$,

$$\max_{\lfloor cn\rfloor\le m\le n}|H_m|=O_p(\chi_n),\ \text{where } \chi_n=\sqrt{n}\log^3 n\max_{\lfloor cn\rfloor\le m\le n}\sigma_m.$$

Proof. Let $c_2 > 0$ be a constant to be determined later. Applying Lemma 2 with $z=c_2\sqrt{m}\sigma_m\log^3 m$ and $\ell = m/(2\lfloor\log^2 m\rfloor)$, we obtain

$$\mathbb{P}(|H_m|\ge c_2\sqrt{m}\sigma_m\log^3 m)\le 4\exp\left(-\frac{z^2\ell}{144m^2\sigma_m^2+4bzm}\right)+22\ell\alpha_{\lfloor\log^2 m\rfloor}\sqrt{1+\frac{4bm}{z}}.$$

Notice that, for $\lfloor cn\rfloor \le m \le n$ with large enough n,

$$\frac{z^2\ell}{144m^2\sigma_m^2+4bzm}=\frac{c_2^2\log^3 m/\lfloor\log^2 m\rfloor}{288/\log^3 m+8bc_2/(\sqrt{m}\sigma_m)}\ge\frac{c_2^2\log m}{1+8bc_2/c_1}.$$

As $m\to\infty$, $m\to\infty$, $a_{\lfloor\log^2 m\rfloor} = O(\rho^{\log^2 m}) = o(m^{-4})$. Therefore, for large m,

$$\begin{aligned}\mathbb{P}(|H_m|\ge c_2\sqrt{m}\sigma_m\log^3 m) &\le 4\exp\left(-\tfrac{c_2^2\log m}{1+8bc_2/c_1}\right)+O(m^2)\alpha_{\lfloor\log^2 m\rfloor}\\ &=O\{m^{-c_2^2/(1+8bc_2/c_1)}+m^{-2}\}.\end{aligned}\quad(20)$$

Choose c_2 such that $c_2^2/[1+8bc_2/c_1]\ge 2$. By (20),

$$\begin{aligned}\mathbb{P}\left(\max_{\lfloor cn\rfloor\le m\le n}|H_m|\ge c_2\chi_n\right) &\le \sum_{m=\lfloor cn\rfloor}^{n}\mathbb{P}(|H_m|\ge c_2\chi_n)\\ &\le \sum_{m=\lfloor cn\rfloor}^{n}\mathbb{P}(|H_m|\ge c_2\sqrt{m}\sigma_m\log^3 m)\to 0,\end{aligned}$$

completing the proof.

Lemma 4

Assume $Z_0 \in \mathcal{L}^{4+\delta}$, $\mathbb{E}(Z_0) = 0$, and $\mathbb{E}(Z_0 Z_i) = 0$, $i \geq 1$. Then for $r \geq 1$,

$$\max_{i=1,2,\ldots,r} |\mathbb{E}(Z_0 Z_i Z_r^2)| \leq 8C^{\delta/(4+\delta)} \rho^{r\delta/[2(4+\delta)]} \|Z_0\|_{4+\delta}^4.$$

Proof. Write $p = (4 + \delta)/3$. From $\mathbb{E}(Z_0) = 0$, $\mathbb{E}(Z_0 Z_i Z_r^2) = \mathrm{cov}(Z_0, Z_i Z_r^2)$.

By Lemma 1,

$$\begin{aligned} |\mathbb{E}(Z_0 Z_i Z_r^2)| = |\mathrm{cov}(Z_0, Z_i Z_r^2)| &\leq 8\alpha_i^{\delta/(4+\delta)} \|Z_0\|_{3p} \cdot \|Z_i Z_r^2\|_p \\ &\leq 8\alpha_i^{\delta/(4+\delta)} \|Z_0\|_{4+\delta}^4. \end{aligned} \quad (21)$$

Here the second "≤" follows from the Hölder's inequality

$$\|Z_i |Z_r^2\|_p^p = \mathbb{E}(|Z_i|^p |Z_r|^p |Z_r|^p) \leq \||Z_i|^p\|_3 \cdot \||Z_r|^p\|_3 \cdot \||Z_r|^p\|_3 = \|Z_0\|_{3p}^{3p}.$$

Furthermore, by $\mathbb{E}(Z_0 Z_i) = 0$ and Lemma 1,

$$|\mathbb{E}(Z_0 Z_i Z_r^2)| = |\mathrm{cov}(Z_0 Z_i, Z_r^2)| \leq 8\alpha_{r-i}^{\delta/(4+\delta)} \|Z_0 Z_i\|_{3p/2} \cdot \|Z_r^2\|_{3p/2}. \quad (22)$$

Notice that $\|Z_r^2\|_{3p/2} = \|Z_r\|_{3p}^2$. Also, by the Cauchy-Schwarz inequality,

$$\|Z_0 Z_i\|_{3p/2}^{3p/2} = \mathbb{E}(|Z_0|^{3p/2} |Z_i|^{3p/2}) \leq \||Z_0|^{3p/2}\|_2 \cdot \||Z_i|^{3p/2}\|_2 = \|Z_0\|_{3p}^{3p}.$$

Therefore, by (22), we have

$$|\mathbb{E}(Z_0 Z_i Z_r^2)| = |\mathrm{cov}(Z_0 Z_i, Z_r^2)| \leq 8\alpha_{r-i}^{\delta/(4+\delta)} \|Z_0\|_{4+\delta}^4. \quad (23)$$

Combining (21) and (23), we obtain $|\mathbb{E}(Z_0 Z_i Z_r^2)| \leq 8 \min\{\alpha_i, \alpha_{r-i}\}^{\delta/(4+\delta)} \|Z_0\|_{4+\delta}^4$. The desired result then follows from $\min\{\alpha_i, \alpha_{r-i}\} \leq \sqrt{\alpha_i \alpha_{r-i}} \leq C\rho^{r/2}$.

5.2. Proof of Theorems 1-2

Lemma 5

Recall $\ell_q(\cdot)$ and $\mathcal{C}(q)$ in Definition 1. Then for any $f(\cdot) \in \mathcal{C}(q)$ and $b_n \to 0$,

$$(i)\ \|e_i f\{(X_i-x)/b_n\}\|_q^q = b_n \ell_q(0)\int_{\mathbb{R}}|f(u)|^q du + o(b_n);$$
$$(ii)\ \textit{For all } p \le q,\ \|e_i f\{(X_i-x)/b_n\}\|_p^p = O(b_n^{p/q}).$$

Proof. (i) Conditioning on X_i and then using the double-expectation formula, we obtain

$$\begin{aligned}\left\|e_i f\left(\tfrac{X_i-x}{b_n}\right)\right\|_q^q &= \int_{\mathbb{R}}\mathbb{E}\left[|e_i|^q|X_i=v\right]\left|f\left(\tfrac{v-x}{b_n}\right)\right|^q p_X(v)dv \\ &= b_n\int_{\mathbb{R}}\ell_q(ub_n)|f(u)|^q du \\ &= b_n\left[\ell_q(0)\int_{\mathbb{R}}|f(u)|^q du + o(1)\right],\end{aligned} \tag{24}$$

where the second "=" follows from the change-of-variable $u = (v-x)/b_n$ and the third "=" follows from (7). (ii) It follows from (i) and the Jensen's inequality

$$\|e_i f\{(X_i-x)/b_n\}\|_p \le \|e_i f\{(X_i-x)/b_n\}\|_q = O(b_n^{1/q}).$$

Lemma 6

Suppose the conditions in Theorem 1 hold. Write $\sigma^2(x)=\mathbb{E}(e_i^2|X_i=x)$,

$$W_n(t) = \frac{1}{\sigma(x)\sqrt{nb_n p_X(x)}}\sum_{i=1}^{\lfloor nt\rfloor}\zeta_i(t),\ \textit{where}\ \zeta_i(t) = e_i K\left(\frac{X_i-x}{b_{\lfloor nt\rfloor}}\right).$$

Then $\{W_n(t)\}_{c\le t\le 1} \Rightarrow \{G_t\}_{c\le t\le 1}$ with $\{G_t\}_{c\le t\le 1}$ being the Gaussian process in Theorem 1.

Proof. Consider the approximation of $W_n(t)$:

$$U_n(t) = \frac{1}{\sigma(x)\sqrt{nb_n p_X(x)}}\sum_{i=1}^{\lfloor nt\rfloor}\eta_i(t),\ \text{where}\ \eta_i(t) = e_i K\left(\frac{X_i-x}{b_n t^{-1/5}}\right).$$

Let $g(\cdot)$ be defined in (8). Note that for all $c^{1/5} \le s \le s'$, by (8), we have

$$\begin{aligned}|K(su)-K(s'u)| &= |(s-s')uK'(s*u)| \\ &\le |s-s'|\sup_{t\ge c^{1/5}}|uK'(tu)| \le |(s-s')|g(u),\end{aligned} \tag{25}$$

where $s^* \in [s, s']$. Also, by Taylor's expansion $(z_0+z)^a = z_0^a + O(z)$ for $z \to 0$ and any fixed z_0 and a, we have

$$\frac{b_n}{b_{\lfloor nt\rfloor}} - t^{1/5} = \left(\frac{\lfloor nt\rfloor}{n}\right)^{1/5} - t^{1/5} = \left[t + O\left(\frac{1}{n}\right)\right]^{1/5} - t^{1/5} = O\left(\frac{1}{n}\right), \tag{26}$$

uniformly in $t \in [c, 1]$. Thus, by (25) and (26), we have

$$\begin{aligned}\max_{t\in[c,1]}|\zeta_i(t)-\eta_i(t)| &= |e_i|\max_{t\in[c,1]}\left|K\left(\tfrac{b_n}{b_{\lfloor nt\rfloor}}\tfrac{X_i-x}{b_n}\right)-K\left(t^{1/5}\tfrac{X_i-x}{b_n}\right)\right| \\ &= O\left(\tfrac{1}{n}\right)|e_i|g\left(\tfrac{X_i-x}{b_n}\right).\end{aligned} \quad (27)$$

Since $g(\cdot)\in\mathcal{C}(4+\delta)$, by Lemma 5(ii) and (27),

$$\max_{t\in[c,1]}|W_n(t)-U_n(t)|=\frac{O(1)}{n\sqrt{nb_n}}\sum_{i=1}^{n}|e_i|g\left(\frac{X_i-x}{b_n}\right)=\frac{O_p\left\{b_n^{1/(4+\delta)}\right\}}{\sqrt{nb_n}}\xrightarrow{p}0. \quad (28)$$

Thus, it suffices to show the convergence of $\{U_n(t)\}_{c\le t\le 1}$.

We need to establish the finite-dimensional convergence and the tightness of $\{U_n(t)\}_{c\le t\le 1}$; see Billingsley (2009). For finite-dimensional convergence, by the Cramér-Wold device, it suffices to consider linear combinations of $U_n(t)$. Let $t, t'\in[c,1]$. Recall $\ell_q(z)$ in (6). By the same double-expectation argument in (24) and using $\ell_2(0)=p_X(x)\sigma^2(x)$, we have

$$\begin{aligned}&\left|\frac{[\eta_i(t)\eta_i(t')]}{b_n}-p_X(x)\sigma^2(x)\int_{\mathbb{R}}K(t^{1/5}u)K(t'^{1/5}u)du\right| \\ =\ &\left|\int_{\mathbb{R}}[\ell_2(ub_n)-\ell_2(0)]\,K(t^{1/5}u)K(t'^{1/5}u)du\right| \\ \le\ &\int_{\mathbb{R}}|\ell_2(ub_n)-\ell_2(0)|\frac{K(t^{1/5}u)^2+K(t'^{1/5}u)^2}{2}du\to 0,\end{aligned} \quad (29)$$

where the last convergence holds since $K(\cdot)\in\mathcal{C}(2)$.

For $k\in\mathbb{N}$, let $c\le t_1,\ldots,t_k\le 1$ and $w_1,\ldots,w_k\in\mathbb{R}$. Consider the linear combination

$$U_n=\sum_{s=1}^{k}w_sU_n(t_s)=\frac{1}{\sigma(x)\sqrt{nb_np_X(x)}}\sum_{i=1}^{n}\eta_i,\ \text{where}\ \eta_i=\sum_{s=1}^{k}w_s\eta_i(t_s)1_{i\le\lfloor nt_s\rfloor}.$$

Let $\mathcal{F}_i$ be the sigma-algebra generated by $(X_{i+1}, X_i, \ldots, X_1, e_i, e_{i-1}, \ldots, e_1)$. By Assumption 1, $\mathbb{E}[\eta_i(t)\,\mathcal{F}_{i-1}]=0$ for each fixed t, and thus $\{\eta_i\}_{i\in\mathbb{N}}$ are martingale differences with respect to $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$. We shall apply the martingale central limit theorem (CLT) to establish a CLT for U_n. Write $a\wedge b=\min\{a,b\}$. From $\mathbf{1}_{i\le\lfloor nt_s\rfloor}\mathbf{1}_{i\le\lfloor nt_{s'}\rfloor}=\mathbf{1}_{i\le\lfloor n(t_s\wedge t_{s'})\rfloor}$ and (29),

$$\begin{aligned}\frac{(\eta_i^2)}{b_n} &= \sum_{s=1}^{k}\sum_{s'=1}^{k}w_sw_{s'}\mathbf{1}_{i\le\lfloor n(t_s\wedge t_{s'})\rfloor}\frac{[\eta_i(t_s)\eta_i(t_{s'})]}{b_n} \\ &= p_X(x)\sigma^2(x)\sum_{s=1}^{k}\sum_{s'=1}^{k}\mathbf{1}_i\le\lfloor n(t_s\wedge t_{s'})\rfloor\,w_sw_{s'}\int_{\mathbb{R}}K\left(t_s^{1/5}u\right)K\left(t_{s'}^{1/5}u\right)du+o(1).\end{aligned} \quad (30)$$

Therefore, by the orthogonality of martingale differences and (30),

$$\mathrm{var}(U_n)=\frac{\sum_{i=1}^n \mathrm{var}(\eta_i)}{nb_n p_X(x)\sigma^2(x)} \to \sum_{s=1}^k \sum_{s'=1}^k (t_s \wedge t_{s'})\, w_s w_{s'} \int_{\mathbb{R}} K\left(t_s^{1/5}u\right) K\left(t_{s'}^{1/5}u\right) du,$$

which is the variance of $\sum_{s=1}^k w_s G_{t_s}$. Next, we verify the Lindeberg condition. Since k is fixed, it suffices to verify that $(nb_n)^{-1}\sum_{i=1}^n \mathbb{E}[\eta_i^2(t)\mathbf{1}_{|\eta_i(t)|\geq c_1\sqrt{nb_n}}] \to 0$ for any given $t > c$ and $c_1 > 0$. Since $K(\cdot) \in \mathcal{C}(4+\delta)$, by Lemma 5(i), $\mathbb{E}[|\eta_i(t)|^{4+\delta}] = O(b_n)$. Therefore,

$$\begin{aligned} \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}\left[\eta_i^2(t)\mathbf{1}_{|\eta_i(t)|\geq c_1\sqrt{nb_n}}\right] &\leq \frac{1}{nb_n}\sum_{i=1}^n \frac{[|\eta_i(t)|^{4+\delta}]}{\left(c_1\sqrt{nb_n}\right)^{2+\delta}} \\ &= O\left[\left(\sqrt{nb_n}\right)^{-2+\delta}\right] \to 0. \end{aligned}$$

This proves the Lindeberg condition. By martingale CLT, U_n has the desired CLT.

It remains to prove the tightness of $U_n(t)$. Let $c \leq t < t' \leq 1$. By the inequality $(a+b)^4 \leq 16(a^4+b^4)$, we obtain

$$\begin{aligned} \mathbb{E}\left[\{U_n(t)-U_n(t')\}^4\right] &= \frac{1}{(nb_n)^2 p_X^2(x)\sigma^4(x)}\mathbb{E}\left\{\sum_{i=1}^{\lfloor nt\rfloor}\eta_i(t) - \sum_{i=1}^{\lfloor nt'\rfloor}\eta_i(t')\right\}^4 \\ &= \frac{O(I_1+I_2)}{(nb_n)^2}, \end{aligned} \tag{31}$$

where

$$I_1=\mathbb{E}\left\{\sum_{i=1}^{\lfloor nt\rfloor}\left[\eta_i(t)-\eta_i(t')\right]\right\}^4 \text{ and } I_2=\mathbb{E}\left\{\sum_{i=1}^{\lfloor nt\rfloor}\eta_i(t') - \sum_{i=1}^{\lfloor nt'\rfloor}\eta_i(t')\right\}^4.$$

Write $Z_i = \eta_i(t) - \eta_i(t')$. By $\mathbb{E}(\eta_i(t)|\mathcal{F}_{i-1}) = 0$, $\mathbb{E}(Z_i|\mathcal{F}_{i-1}) = 0$. For $i < j < r < s$, $\mathbb{E}(Z_i) = 0$, $\mathbb{E}(Z_iZ_j) = \mathbb{E}[\mathbb{E}(Z_iZ_j|\mathcal{F}_{i-1})] = 0$,

$\mathbb{E}(Z_i^3 Z_j)=\mathbb{E}[\mathbb{E}(Z_i^3 Z_j|\mathcal{F}_{j-1})]=0$, $\mathbb{E}(Z_iZ_j^2Z_r)=\mathbb{E}[\mathbb{E}(Z_iZ_j^2Z_r|\mathcal{F}_{r-1})]=0$. Therefore,

$$\begin{aligned} I_1 &\leq O(1)\left[\sum_i \mathbb{E}(Z_i^4)+\sum_{i<j}\mathbb{E}(Z_i^2Z_j^2)+\left|\sum_{i<j<r}\mathbb{E}(Z_iZ_jZ_r^2)\right|+\left|\sum_{i<j}\mathbb{E}(Z_iZ_j^3)\right|\right] \\ &\leq O(1)\left[n\mathbb{E}(Z_1^4)+n\sum_{r=1}^n \mathbb{E}(Z_1^2Z_{1+r}^2)+\sum_{1\leq i<r\leq n}(r-i)\max_{j=i+1,\ldots,r}|\mathbb{E}(Z_iZ_jZ_r^2)|\right] \end{aligned} \tag{32}$$

Recall $\ell_q(\cdot)$ in (6). By the same argument in the derivation of (24), for $q \in (0, 4+\delta]$,

$$\begin{aligned}\mathbb{E}\left(|Z_i|^q\right) &= b_n \int_{\mathbb{R}} \ell_q(ub_n)\left|K(t^{1/5}u) - K(t'^{1/5}u)\right|^q du \\ &= O(b_n)|t' - t|^q \int_{\mathbb{R}} \ell_q(ub_n)|g(u)|^q du,\end{aligned} \quad (33)$$

where the last "=" follows from (25) and the inequality $|t'^{1/5} - t^{1/5}| \le c^{-4/5}|t' - t|/5$ for $c \le t < t' \le 1$. By (33) and the definition of $\mathcal{C}(q)$, if $g(\cdot) \in \mathcal{C}(q)$ so that $\int_{\mathbb{R}} \ell_q(ub_n)|g(u)|^q du = \ell_q(0) \int_{\mathbb{R}} |g(u)|^q du + o(1) = O(1)$, we have $\mathbb{E}(|Z_i|^q) = O(b_n|t' - t|^q)$. Therefore, under the condition $g(\cdot) \in \mathcal{C}(2) \cap \mathcal{C}(4 + \delta)$, we obtain

$$\|Z_i\|_2 = O\left[b_n^{1/2}|t' - t|\right] \text{ and } \|Z_i\|_{4+\delta} = O\left[b_n^{1/(4+\delta)}|t' - t|\right]. \quad (34)$$

An application of Lemma 1 with $p_1 = p_2 = (4 + \delta)/2$ gives

$$\begin{aligned}|\mathrm{cov}(Z_1^2, Z_{1+r}^2)| &\le 8\alpha_r^{\delta/(4+\delta)} \|Z_1^2\|^2_{(4+\delta)/2} = 8\alpha_r^{\delta/(4+\delta)} \|Z_i\|^4_{4+\delta} \\ &= O\left[\rho^{r\delta/(4+\delta)} b_n^{4/(4+\delta)}|t' - t|4\right].\end{aligned}$$

Combining the above two expressions, we obtain

$$\mathbb{E}\left(Z_1^2 Z_{1+r}^2\right) = \mathrm{cov}(Z_1^2, Z_{1+r}^2) + \left[\mathbb{E}(Z_1^2)\right]^2 = |t' - t|^4 O\left[\rho^{r\delta/(4+\delta)} b_n^{4/(4+\delta)} + b_n^2\right]. \quad (35)$$

Applying Lemma 4 and (34), we have

$$\mathbb{E}(Z_i Z_j Z_r^2) = O\left\{\rho^{(r-i)\delta/[2(4+\delta)]} b_n^{4/(4+\delta)}|t' - t|^4\right\}, \quad j = i+1, \ldots, r. \quad (36)$$

Furthermore, by (34) and Jensen's inequality, $\mathbb{E}(Z_i^4) \le \|Z_i\|^4_{4+\delta} = O[b_n^{4/(4+\delta)}|t' - t|^4]$. Therefore, by (32) and (34)-(36), it is easy to see

$$I_1 = O\left\{|t' - t|^4 \left[n b_n^{4/(4+\delta)} + n^2 b_n^2\right]\right\}. \quad (37)$$

For I_2, by the same argument in I_1, we can show

$$I_2 = \mathbb{E}\left\{\sum_{i=\lfloor nt \rfloor + 1}^{\lfloor nt' \rfloor} \eta_i(t')\right\}^4 = O\left[n b_n^{4/(4+\delta)}|t' - t| + n^2 b_n^2 |t' - t|^2\right]. \quad (38)$$

Therefore, by (31) and (37)–(38),

$$\mathbb{E}\left[\left\{U_n(t)-U_n(t')\right\}^4\right]=O\left[\frac{|t'-t|+|t'-t|^4}{nb_n^{(4+2\delta)/(4+\delta)}}+|t'-t|^2+|t'-t|^4\right],$$

completing the tightness of $U_n(t)$ in view of $\left[nb_n^{(4+2\delta)/(4+\delta)}\right]^{-1}=O(n^{-3/5})$ under the condition $b_n \propto n^{-1/5}$; see condition A1 and Remark 2.1 of Shao and Yu (1996).

Lemma 7

Suppose $p_X(\cdot)$ and $g(\cdot)$ are bounded and continuous at x. Then for any integrable function $f(\cdot)$, we have

$$\max_{\lfloor cn\rfloor\le m\le n}\left|\mathbb{E}\left[g(X_i)f\left(\frac{X_i-x}{b_m}\right)\right]-b_m p_X(x)g(x)\int_{\mathbb{R}}f(u)\,du\right|=o(b_n).$$

Proof. Note that $g(x)p_X(x)$ is continuous at x because $g(x)$ and $p_X(x)$ are continuous at x and $p_X(x)$ is bounded. Observe that

$$\begin{aligned}\mathbb{E}\left[g(X_i)f\left(\tfrac{X_i-x}{b_m}\right)\right] &= \int_{\mathbb{R}} f\left(\tfrac{v-x}{b_m}\right)g(v)p_X(v)dv\\ &= b_m\int_{\mathbb{R}} f(u)g(x+ub_m)p_X(x+ub_m)\,du.\end{aligned}$$

Note that $b_m \propto b_n$ for $\lfloor cn\rfloor \le m \le n$. The result then follows from

$$\begin{aligned}&\max_{\lfloor cn\rfloor\le m\le n}\left|\int_{\mathbb{R}} f(u)\left[\,g(x+ub_m)p_X(x+ub_m)-g(x)p_X(x)\right]du\right|\\ &\le \int_{\mathbb{R}}|f(u)|\max_{\lfloor cn\rfloor\le m\le n}|g(x+ub_m)p_X(x+ub_m)-g(x)p_X(x)|du\to 0.\end{aligned}$$

Here the last convergence follows from the dominated convergence theorem in view of the continuity of $g(\cdot)p_X(\cdot)$ at x, the boundedness of $g(\cdot)p_X(\cdot)$, and the integrality of $f(\cdot)$.

Lemma 8

Suppose the conditions in Theorem 1 hold. For $M_m(j)$ defined in (5), we have

$$\max_{\lfloor cn\rfloor\le m\le n}\left|\frac{M_m(j)}{mb_m^j}-b_m p_X(x)\int_{\mathbb{R}}u^j K(u)\,du\right|=o(b_n),\quad j=0,1,2.$$

Proof. We abbreviate "uniformly in $\lfloor cn\rfloor \le m \le n$" as "uniformly in m". Define

$$h_j(X_i,b_m)=\left(\frac{X_i-x}{b_m}\right)^j K\left(\frac{X_i-x}{b_m}\right).$$

Let $c_1 = \sup_u(1 + |u| + |u|^2)|K(u)|$. By (8), $c_1 < \infty$ and thus $h_j(X_i, b_m)$ is bounded for $j = 0, 1, 2$. By the integrability of $|u^3K(u)|$ [see (8)], $|u^jK(u)|$ is integrable for $j = 0, 1, 2$, which implies that $|u^{2j}K^2(u)| \le c_1|u^jK(u)|$ is also integrable for $j = 0, 1, 2$. By Lemma 7, $\mathbb{E}[h_j(X_i, b_m)^2] = O(b_n)$ uniformly in m. Therefore, applying Lemma 3, we obtain $\sum_{i=1}^{m}\{h_j(X_i,b_m) - \mathbb{E}[h_j(X_i,b_m)]\} = O_p\{\sqrt{nb_n}(\log n)^3\}$ uniformly in m. Furthermore, by Lemma 7, $\mathbb{E}[h_j(X_i, b_m)] = b_m p_X(x)\int_{\mathbb{R}} u^jK(u)du + o(b_n)$ uniformly in m. Thus,

$$\begin{aligned}\frac{M_m(j)}{mb_m^j} &= \mathbb{E}[h_j(X_1,b_m)] + \frac{1}{m}\sum_{i=1}^{m}\{h_j(X_i,b_m) - \mathbb{E}[h_j(X_i,b_m)]\}\\ &= b_m p_X(x)\int_{\mathbb{R}} u^jK(u)du + o(b_n) + O_p\{\sqrt{b_n/n}(\log n)^3\},\end{aligned}$$

uniformly in m. Since $b_n \propto n^{-1/5}$, $\sqrt{b_n/n}(\log n)^3 = o(b_n)$. This completes the proof.

Proof of Theorem 1—By (4), we can easily obtain the following decomposition

$$\hat{\mu}_m(x) - \mu(x) = \frac{M_m(2)[N_m(0) - \mu(x)M_m(0)] - M_m(1)[N_m(1) - \mu(x)M_m(1)]}{M_m(2)M_m(0) - M_m(1)^2}. \quad (39)$$

Define

$$B_m(j) = \sum_{i=1}^{m}[\mu(X_i) - \mu(x)](X_i - x)^j K\left(\frac{x - X_i}{b_m}\right),\ j = 0, 1.$$

By (39), we can derive the decomposition

$$\begin{aligned}\hat{\mu}_m(x) - \mu(x) &= \frac{M_m(2)B_m(0) - M_m(1)B_m(1)}{M_m(2)M_m(0) - M_m(1)^2}\\ &+ \frac{M_m(2)}{M_m(2)M_m(0) - M_m(1)^2}\sum_{i=1}^{m} e_i K\left(\frac{x - X_i}{b_m}\right)\\ &- \frac{b_m M_m(1)}{M_m(2)M_m(0) - M_m(1)^2}\sum_{i=1}^{m} e_i\left(\frac{X_i - x}{b_m}\right)K\left(\frac{X_i - x}{b_m}\right).\end{aligned} \quad (40)$$

Below we consider the three terms on the right hand side of (40) separately. We shall show that the first term is the bias term, the second term is the stochastic component that determines the asymptotic distribution, and the third term is negligible.

The symmetry of $K(\cdot)$ implies $\int_{\mathbb{R}} uK(u)du = 0$. By Lemma 8, we can easily obtain

$$M_m(2)M_m(0) - M_m(1)^2 = m^2b_m^4 p_X^2(x)\int_{\mathbb{R}} u^2K(u)du + o(n^2b_n^4), \quad (41)$$

uniformly in $\lfloor cn \rfloor \le m \le n$ (hereafter, abbreviated as "uniformly in m"). Note that, by the same argument in Lemma 8, we can show $\sum_{i=1}^{m}|[(X_i - x)/b_m]^3 K\{(X_i - x)/b_m\}| = O(nb_n)$ uniformly in m. Thus, since $\mu(\cdot)$ has

bounded third derivative, by the Taylor's expansion $\mu(X_i) - \mu(x) = (X_i - x)\,\mu'(x) + (X_i - x)^2\mu''(x)/2 + O[(X_i - x)^3]$, we can obtain

$$B_m(0)=M_m(1)\mu'(x)+M_m(2)\mu''(x)/2+O(nb_n^3).$$

Similarly, $B_m(1)=M_m(2)\mu'(x)+O(nb_n^4)$ uniformly in m. Combining the latter two approximations with (41) and Lemma 8, after some algebra we see that the first term in (40) has the approximation

$$\frac{M_m(2)^2\mu''(x)/2+o(n^2b_n^6)}{M_m(2)M_m(0)-M_m(1)^2}=b_m^2\frac{\mu''(x)}{2}\int_{\mathbb{R}}u^2K(u)\,du+o(b_n^2),$$

uniformly in m. This gives the asymptotic bias. From (41), Lemma 8, and the FCLT in Lemma 6, after proper normalization the second term in (40) has the desired FCLT. By Lemma 8 and the same argument in Lemma 6, the third term in (40) satisfies a FCLT with a faster convergence rate and thus is negligible. This completes the proof.

Proof of Theorem 2—For a function f, denote its L_2 norm by $L_2(f)=\{\int_c^1|f(t)|^2dt\}^{1/2}$. Note that $|n^{-4/5}\lfloor nt\rfloor^{4/5}-t^{4/5}|\le n^{-4/5}$ uniformly for $t\in[c,1]$, and $\{t^{4/5}[\hat{\mu}_{\lfloor nt\rfloor}(x)-\hat{\mu}_n(x)]\}_{t\in[c,1]}$ and $\{n^{-4/5}\lfloor nt\rfloor^{4/5}[\hat{\mu}_{\lfloor nt\rfloor}(x)-\hat{\mu}_n(x)]\}_{t\in[c,1]}$ are asymptotically equivalent. Thus, by Theorem 1 and the continuous mapping theorem,

$$\frac{\hat{\mu}_n(x)-\mu(x)}{L_2\left\{n^{-4/5}\lfloor nt\rfloor^{4/5}(\hat{\mu}_{\lfloor nt\rfloor}(x)-\hat{\mu}_n(x)),t\in[c,1]\right\}}\xrightarrow{d}\xi.$$

Since $\lfloor nt\rfloor$ is piecewise constant, it is easy to verify $L_2\{n^{-4/5}\lfloor nt\rfloor^{4/5}[\hat{\mu}_{\lfloor nt\rfloor}(x)-\hat{\mu}_n(x)],\,t\in[c,1]\}=V_n$, completing the proof.

References

Billingsley, P. Convergence of probability measures. Vol. 493. John Wiley & Sons; 2009.

Fan, J.; Gijbels, I. Local polynomial modelling and its applications: monographs on statistics and applied probability 66. Vol. 66. CRC Press; 1996.

Fan, J.; Yao, Q. Nonlinear time series. Vol. 2. Springer; 2003.

Hardle W. A note on jackknifing kernel regression function estimators (corresp.). Information Theory, IEEE Transactions on. 1986; 32(2):298–300.

Honda T. Nonparametric estimation of a conditional quantile for α-mixing processes. Annals of the Institute of Statistical Mathematics. 2000; 52(3):459–70.

Kiefer NM, Vogelsang TJ. Heteroskedasticity-autocorrelation robust testing using bandwidth equal to sample size. Econometric Theory. 2002; 18(06):1350–66.

Kiefer NM, Vogelsang TJ. A new asymptotic theory for heteroskedasticityautocorrelation robust tests. Econometric Theory. 2005; 21(06):1130–64.

Kiefer NM, Vogelsang TJ, Bunzel H. Simple robust testing of regression hypotheses. Econometrica. 2000; 68(3):695–714.

Kim S, Zhao Z. Unified inference for sparse and dense longitudinal models. Biometrika. 2013; 100(1): 203–12. [PubMed: 24966413]

Lai T, Siegmund D. Fixed accuracy estimation of an autoregressive parameter. The Annals of Statistics. 1983; 11(2):478–85.

Li, Q.; Racine, JS. Nonparametric econometrics: Theory and practice. Princeton University Press; 2007.

Lobato IN. Testing that a dependent process is uncorrelated. Journal of the American Statistical Association. 2001; 96(455):1066–76.

Ruppert D, Sheather SJ, Wand MP. An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association. 1995; 90(432):1257–70.

Shao QM, Yu H. Weak convergence for weighted empirical processes of dependent sequences. The Annals of Probability. 1996:2098–127.

Shao X. A self-normalized approach to confidence interval construction in time series. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2010; 72(3):343–66.

Shao X. Parametric inference in stationary time series models with dependent errors. Scandinavian Journal of Statistics. 2012; 39(4):772–83.

Shao X, Zhang X. Testing for change points in time series. Journal of the American Statistical Association. 2010; 105(491)

Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society series B. 1991; 53(3):683–90.

Vogelsang TJ. Testing in gmm models without truncation. Advances in Econometrics. 2003; 17:199–233.

Wu WB, Zhao Z. Inference of trends in time series. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2007; 69(3):391–410.

Zhou Z, Shao X. Inference for linear models with dependent errors. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2013; 75(2):323–43.

Table 1

Simulated quantiles of $|\xi|$ [cf. (15)] at $c = 0.1$ based on 10^6 replications and approximation of the Gaussian process $\{G_t\}$ on 10^3 evenly spaced grid points.

τquantile	50%	60%	70%	80%	90%	95%	97.5%	99%	99.5%	99.9%
q_τ	1.74	2.22	2.81	3.63	4.99	6.37	7.70	9.50	10.83	13.88

Table 2

Empirical average deviation of coverage probabilities [cf. (16)] for the mean regression function for Models I and II in (17)–(18). Columns SN stands for the self-normalization based method. For the asymptotic normality based method, we consider five different bandwidth methods for τ_n in (12) as described in Section 3.

Model I							
			Traditional method				
λ	θ	SN	bw.nrd0	bw.nrd	bw.ucv	bw.bcv	bw.SJ
0.03	$\theta = 0.0$	0.005	0.092	0.089	0.073	0.089	0.062
	$\theta = 0.4$	0.009	0.134	0.131	0.114	0.131	0.088
	$\theta = 0.8$	0.005	0.317	0.314	0.297	0.314	0.259
0.06	$\theta = 0.0$	0.006	0.076	0.074	0.061	0.075	0.053
	$\theta = 0.4$	0.005	0.114	0.112	0.096	0.112	0.070
	$\theta = 0.8$	0.005	0.284	0.282	0.263	0.282	0.228
0.12	$\theta = 0.0$	0.006	0.061	0.059	0.048	0.059	0.046
	$\theta = 0.4$	0.006	0.100	0.098	0.082	0.097	0.056
	$\theta = 0.8$	0.006	0.254	0.251	0.233	0.250	0.198
0.24	$\theta = 0.0$	0.004	0.057	0.055	0.041	0.055	0.040
	$\theta = 0.4$	0.006	0.092	0.090	0.076	0.090	0.051
	$\theta = 0.8$	0.006	0.224	0.221	0.203	0.222	0.166

Model II							
			Traditional method				
λ	θ	SN	bw.nrd0	bw.nrd	bw.ucv	bw.bcv	bw.SJ
0.03	$\theta = 0.0$	0.006	0.073	0.073	0.054	0.072	0.048
	$\theta = 0.4$	0.007	0.070	0.069	0.051	0.069	0.048
	$\theta = 0.8$	0.007	0.073	0.072	0.062	0.072	0.051
0.06	$\theta = 0.0$	0.007	0.066	0.065	0.063	0.065	0.053
	$\theta = 0.4$	0.006	0.061	0.060	0.062	0.061	0.054
	$\theta = 0.8$	0.006	0.080	0.078	0.075	0.078	0.065

Model II							
			Traditional method				
λ	θ	SN	bw.nrd0	bw.nrd	bw.ucv	bw.bcv	bw.SJ
0.12	$\theta = 0.0$	0.005	0.068	0.067	0.074	0.067	0.068
	$\theta = 0.4$	0.006	0.069	0.068	0.069	0.067	0.065
	$\theta = 0.8$	0.006	0.082	0.081	0.080	0.081	0.080
0.24	$\theta = 0.0$	0.007	0.065	0.064	0.064	0.064	0.071
	$\theta = 0.4$	0.006	0.069	0.069	0.070	0.069	0.075
	$\theta = 0.8$	0.008	0.082	0.082	0.082	0.081	0.087