The University of Chicago
Department of Statistics
TECHNICAL REPORT SERIES

Confidence Bands in Nonparametric Time Series Regression

Zhibiao Zhao and Wei Biao Wu

TECHNICAL REPORT NO. 574

Departments of Statistics
The University of Chicago
Chicago, Illinois 60637

November 2006

Confidence Bands in Nonparametric Time Series Regression

Zhibiao Zhao and Wei Biao Wu

Department of Statistics, The University of Chicago

Abstract: We consider nonparametric estimation of mean regression and volatility functions in nonlinear stochastic regression models. Simultaneous confidence bands are constructed and the coverage probabilities are shown to be asymptotically correct. The imposed dependence structure allows applications in many linear and nonlinear autoregressive processes. The results are applied to the IBM stock data.

Keywords: Long-range dependence, model validation, moderate deviation, nonlinear time series, nonparametric regression, short-range dependence.

1 Introduction

There are two popular approaches in time series analysis: parametric and nonparametric methods. In the literature various parametric models have been proposed, including the classical ARMA, threshold AR (TAR, Tong 1990), exponential AR (EAR, Haggan and Ozaki 1981) and AR with conditional heteroscedasticity (ARCH, Engle 1982) among others. Those models are widely used in practice. An attractive feature of parametric models is that they can provide explanatory insights into the dynamical characteristics of the underlying data-generating mechanism.

However, a parametric model has good performance only when it is indeed the true model or a good approximation of it. Thus, for parametric models, modelling bias may arise and there is a risk of mis-specification that can lead to misunderstanding of the truth and wrong conclusions. One way out is to use nonparametric techniques which let the data "speak for themselves" by imposing no specific structures on the underlying regression functions other than smoothness assumptions. See Fan and Yao (2003) for an extensive exposition of nonparametric time series analysis.

Nonparametric estimates can suggest parametric models and it is related to the model validation problem. The nonparametric model validation under dependence is important but very difficult. Fan and Yao (2003) dealt with this deep problem for time series data by using the idea of generalized likelihood ratio test (Fan, Zhang and Zhang 2001), which is developed for independent data. Fan and Yao (2003) pointed out that there have been virtually no theoretical development on nonparametric model validations under dependence, despite the importance of the latter problem since dependence is an intrinsic characteristic in time series.

In this paper we shall consider the model validation problem for the stochastic regression model

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{1}$$

where ε_i are independent and identically distributed (iid) unobserved random noises and (X_i, Y_i) are observations. The functions μ and σ are mean regression and volatility functions, respectively. As a special case, let $X_i = Y_{i-1}$. Then (1) is reduced to the nonlinear autoregressive process $Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\varepsilon_i$ and it includes many parametric time series models. For example, if $\mu(x) = ax$, or $\mu(x) = a\max(x, 0) + b\min(x, 0)$, or $\mu(x) = [a + b\exp(-cx^2)]x$, where a, b, c are real parameters, then it becomes AR, TAR, EAR processes, respectively. For the ARCH process, $\mu = 0$ and $\sigma(x) = (a^2 + b^2x^2)^{1/2}$.

We shall address the model validation problem of (1) by constructing nonparametric simultaneous confidence bands (SCB) for μ and σ. SCB is useful in testing whether μ and

σ are of certain parametric forms. For example, in model (1), interesting problems include testing whether μ is linear, quadratic or of other patterns and whether σ is non-constant, namely, the existence of conditional heteroscedasticity. The mean regression function μ can be non-parametrically estimated by kernel, local linear, spline and wavelet methods. To construct an asymptotic SCB for $\mu(x)$ over the interval $x \in \mathcal{T} = [T_1, T_2]$ with level $100(1-\alpha)\%$, $\alpha \in (0,1)$, we need to find two functions $l(\cdot) = l_n(\cdot)$ and $u(\cdot) = u_n(\cdot)$ based on the data $(X_i, Y_i)_{i=1}^n$ such that

$$\lim_{n\to\infty} \mathbb{P}\left\{l(x) \leq \mu(x) \leq u(x) \text{ for all } x \in \mathcal{T}\right\} = 1 - \alpha. \tag{2}$$

It is certainly more desirable to have (2) in a non-asymptotic sense, namely the probability in (2) is exactly $1-\alpha$. However the latter problem is intractable since it is difficult to establish a finite sample distributional theory for nonparametric regression estimates. With the SCB, we can test whether μ is of certain parametric form: $H_0 : \mu = \mu_\theta$, where $\theta \in \Theta$ and Θ is a parameter space. For example, to test whether $\mu(x) = \beta_0 + \beta_1 x$, we can apply the linear regression method and obtain an estimate $(\hat{\beta}_0, \hat{\beta}_1)$ of (β_0, β_1) from the data $(X_i, Y_i)_{i=1}^n$, and then check whether $l(x) \leq \hat{\beta}_0 + \hat{\beta}_1 x \leq u(x)$ holds for all $x \in \mathcal{T}$. If so, then we accept at level α the null hypothesis that μ is linear. Otherwise H_0 is rejected.

The construction of SCB l and u satisfying (2) has been a difficult problem if dependence is present. Assuming that (X_i, Y_i) are independent random samples from a bivariate population, Johnston (1982) obtained an asymptotic distributional theory for $\sup_{0\leq x\leq 1} |\hat{\mu}(x) - \mathbb{E}[\hat{\mu}(x)]|$, where $\hat{\mu}(x)$ is the Nadaraya-Watson estimate of the mean regression function $\mu(x) = \mathbb{E}(Y|X = x)$. Johnston applied his limit theorem and constructed asymptotic SCB for μ. Since his result is no longer valid if dependence is present, Johnston's procedure is not applicable in the time series setting. A key tool in Johnston's approach is Bickel and Rosenblatt's (1973) asymptotic theory for maximal deviations of kernel density estimators. Bickel and Rosenblatt applied a deep result in probability theory, strong approximation, which asserts that normalized empirical processes of independent random variables can be approximated by Brownian bridges. Such a result generally does not exist under dependence. For other contributions under the independence assumption see Härdle (1989), Knafl, Sacks and Ylvisaker (1982, 1985), Hall and Titterington (1988), Härdle and Marron (1991), Eubank and Speckman (1993), Sun and Loader (1994), Xia (1998), Cummins, Filloon and Nychka (2001) and Dümbgen (2003) among others.

In the fixed design case with $X_i = i/n$, by applying Kolmós et al's (1975) strong invariance principle for partial sums, Eubank and Speckman (1993) constructed SCB for μ with asymptotically correct coverage probabilities. Their method was extended to the time series setting by Wu and Zhao (2006). However, Wu and Zhao's result is not applicable here since it heavily relies on the fixed design assumption.

In this paper, we shall consider a variant of (2) and construct SCB over a subset $\mathcal{T}_n$ of $\mathcal{T}$ with $\mathcal{T}_n$ becoming denser as $n \to \infty$. It is shown that our SCB has asymptotically correct coverage probabilities under a general dependence structure on (X_i, Y_i) which allows applications in many popular linear and nonlinear processes. Our method can be used to deal with statistical inference problems in time series including goodness-of-fit, hypothesis testing and others. In the development of our asymptotic theory, we apply a deep martingale moderate deviation principle by Grama and Haeusler (2006).

We now introduce some notation. Throughout the paper denote by $\mathcal{T} = [T_1, T_2]$ a fixed bounded interval for some $T_1 < T_2$. For a random variable Z write $Z \in \mathcal{L}^p, p > 0$, if $\|Z\|_p := [\mathbb{E}(|Z|^p)]^{1/p} < \infty$, and $\|Z\| = \|Z\|_2$. For $a, b \in \mathbb{R}$ let $a \wedge b = \min(a, b), a \vee b = \max(a, b)$ and $\lfloor a \rfloor = \sup\{k \in \mathbb{Z} : k \leq a\}$. Let $\{a_n\}$ and $\{b_n\}$ be two real sequences. We write $a_n \asymp b_n$ if $|a_n/b_n|$ is bounded away from 0 and ∞ for large enough n. For a set $\mathcal{S} \subset \mathbb{R}$ denote by $\mathcal{C}^p(\mathcal{S}) = \{g(\cdot) : \sup_{x\in\mathcal{S}} |g^{(k)}(x)| < \infty, k = 0, 1, \ldots, p\}$ be the set of functions having bounded derivatives on $\mathcal{S}$ up to order $p \geq 1$, and by $\mathcal{C}^0(\mathcal{S})$ the set of continuous functions on $\mathcal{S}$. For $\epsilon > 0$ let $\mathcal{S}^\epsilon = \cup_{y\in\mathcal{S}}\{x : |x-y| \leq \epsilon\}$ be the ϵ-neighborhood of $\mathcal{S}$. Let F_X and F_ε be the distribution functions of X_0 and ε_0, respectively, and let $f_X = F'_X$ and $f_\varepsilon = F'_\varepsilon$ be their densities.

The rest of the paper is structured as follows. We introduce our dependence structure on $(X_i, Y_i, \varepsilon_i)$ in Section 2. We present our main results in Section 3, where SCBs for $\mu(\cdot)$ and $\sigma^2(\cdot)$ with asymptotically correct coverage probabilities are constructed in Sections 3.1 and 3.2, respectively. In Section 4, applications are made to two important cases of (1): nonlinear time series and linear processes, where we consider both short-range dependent and long-range dependent processes. In Section 5, we discuss some implementation issues, including bootstrap procedure, and then perform a simulation study. Section 6 contains an application to the IBM stock data. We collect the proofs in Section 7.

2 Dependence structure

In (1), assume that $\varepsilon_i, i \in \mathbb{Z}$, are iid and that X_i is a stationary process

$$X_i = G(\mathcal{F}_i), \text{ where } \mathcal{F}_i = (\ldots, \eta_{i-1}, \eta_i). \tag{3}$$

Here $\eta_i, i \in \mathbb{Z}$, are iid and G is a measurable function such that X_i is well defined. The framework (3) is very general (Tong 1990; Stine 2006; Wu 2005). Assume that ε_i is independent of $\mathcal{F}_i$ and η_i is independent of $\varepsilon_j, j \leq i-2$.

For a random variable $Z \in \mathcal{L}^1$ define the projections $\mathcal{P}_k Z = \mathbb{E}(Z|\mathcal{F}_k) - \mathbb{E}(Z|\mathcal{F}_{k-1}), k \in \mathbb{Z}$. Let $F_X(x|\mathcal{F}_i) = \mathbb{P}(X_{i+1} \leq x|\mathcal{F}_i)$, $i \in \mathbb{Z}$, be the conditional distribution function of X_{i+1}

given $\mathcal{F}_i$ and $f_X(x|\mathcal{F}_i) = \partial F_X(x|\mathcal{F}_i)/\partial x$ the conditional density. Define

$$\theta_i = \sup_{x\in\mathbb{R}} \|\mathcal{P}_0 f_X(x|\mathcal{F}_i)\| + \sup_{x\in\mathbb{R}} \|\mathcal{P}_0 f'_X(x|\mathcal{F}_i)\|, \tag{4}$$

where $f'_X(x|\mathcal{F}_i) = \partial f_X(x|\mathcal{F}_i)/\partial x$. For $n \in \mathbb{N}$ define

$$\Theta_n = \sum_{i=1}^{n} \theta_i \quad \text{and} \quad \Xi_n = n\Theta_{2n}^2 + \sum_{k=n}^{\infty} (\Theta_{n+k} - \Theta_k)^2. \tag{5}$$

Roughly speaking, θ_i measures the contribution of ε_0 in predicting X_{i+1} (Wu 2005). If $\Theta_\infty < \infty$, then the cumulative contribution of ε_0 in predicting future values is finite, thus implying short-range dependence (SRD). In this case $\Xi_n = O(n)$. Our setting also allows long-range dependence (LRD) or strong dependence. For example, let $\theta_i = i^{-\beta}\ell(i)$, where $\beta > 1/2$ and $\ell(\cdot)$ is a slowly varying function, namely $\lim_{x\to\infty} \ell(\lambda x)/\ell(x) = 1$ for all $\lambda > 0$. Note that $\bar{\ell}(n) = \sum_{i=1}^{n} |\ell(i)|/i$ is also a slowly varying function. By Karamata's theorem,

$$\Xi_n = O(n), \quad O[n^{3-2\beta}\ell^2(n)] \quad \text{or} \quad O\{n[\bar{\ell}(n)]^2\}, \tag{6}$$

under $\beta > 1$ (SRD case), $\beta < 1$ (LRD case) or $\beta = 1$, respectively (see Wu 2003). In the LRD case Ξ_n grows faster than n. In Section 4 we shall give bounds on Ξ_n for SRD and LRD linear processes and some nonlinear time series.

3 Main results

Let $\mathcal{T} = [T_1, T_2]$ be a bounded interval. With Theorems 1 and 2, we can construct SCBs for $\mu(\cdot)$ and $\sigma^2(\cdot)$ with asymptotically correct coverage probabilities on $\mathcal{T}_n$, which gets denser in $\mathcal{T}$ as $n \to \infty$. We assume hereafter without loss of generality (WLOG) that

$\mathbb{E}(\varepsilon_0) = 0$ and $\mathbb{E}(\varepsilon_0^2) = 1$ since otherwise model (1) can be re-parameterized by letting $\bar{\mu}(x) = \mu(x) + \sigma(x)\mathbb{E}(\varepsilon_0)$, $\bar{\sigma}(x) = c\sigma(x)$ and $\bar{\varepsilon}_i = [\varepsilon_i - \mathbb{E}(\varepsilon_i)]/c$, where $c^2 = \mathbb{E}(\varepsilon_0^2) - [\mathbb{E}(\varepsilon_0)]^2$.

3.1 Simultaneous confidence bands for μ

There exists a vast literature on nonparametric estimation of the regression function μ. Here we use the Nadaraya-Watson estimator

$$\hat{\mu}_{b_n}(x) = \frac{1}{nb_n\hat{f}_X(x)}\sum_{i=1}^{n} K_{b_n}(x - X_i)Y_i, \text{ where } \hat{f}_X(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K_{b_n}(x - X_i). \tag{7}$$

Here $K_{b_n}(u) = K(u/b_n)$, K is a kernel function with $\int_{\mathbb{R}} K(u) = 1$ and the bandwidth $b_n \to 0$ satisfies $nb_n \to \infty$. In Definition 1 below, some regularity conditions on K are imposed. Proposition 1 asserts a central limit theorem (CLT) for $\hat{\mu}_{b_n}(x)$, which can be used to construct point-wise confidence intervals for $\mu(x)$.

Definition 1. *Let $\mathcal{K}$ be the set of kernels which are bounded, symmetric, and have bounded derivative and bounded support. Let $\psi_K = \int_{\mathbb{R}} u^2K(u)du/2$ and $\varphi_K = \int_{\mathbb{R}} K^2(u)du$.*

Proposition 1. *Let $x \in \mathbb{R}$ be fixed and $K \in \mathcal{K}$. Assume that $f_X(x) > 0, \sigma(x) > 0$ and $f_X, \mu \in \mathcal{C}^4(\{x\}^\epsilon)$ for some $\epsilon > 0$. Further assume that*

$$nb_n^9 + \frac{1}{nb_n} + \Xi_n\left[\frac{b_n^3}{n} + \frac{1}{n^2}\right] \to 0, \tag{8}$$

Let $\rho_\mu(x) = \mu''(x) + 2\mu'(x)f'_X(x)/f_X(x)$. Then

$$\frac{\sqrt{nb_n}}{\sigma(x)\sqrt{\varphi_K}}\sqrt{\hat{f}_X(x)}\left[\hat{\mu}_{b_n}(x) - \mu(x) - b_n^2\psi_K\rho_\mu(x)\right] \Rightarrow N(0,1).$$

Theorem 1. *Let $\varepsilon_0 \in \mathcal{L}^3$, $\mathcal{T} = [T_1, T_2]$ and let $K \in \mathcal{K}$ have support $[-1, 1]$. Assume that $\inf_{x\in\mathcal{T}} f_X(x) > 0, \inf_{x\in\mathcal{T}} \sigma(x) > 0$ and $f_X, \mu \in \mathcal{C}^4(\mathcal{T}^\epsilon), \sigma \in \mathcal{C}^2(\mathcal{T}^\epsilon)$ for some $\epsilon > 0$. Further assume*

$$nb_n^9 \log n + \frac{(\log n)^3}{nb_n^3} + \Xi_n\Big[\frac{b_n^3 \log n}{n} + \frac{(\log n)^2}{n^2 b_n^{4/3}}\Big] \to 0. \tag{9}$$

Let $\rho_\mu(x)$ be as in Proposition 1. For $n \geq 2$ define

$$B_n(z) = \sqrt{2\log n} - \frac{1}{\sqrt{2\log n}}\Big[\frac{1}{2}\log\log n + \log(2\sqrt{\pi})\Big] + \frac{z}{\sqrt{2\log n}}. \tag{10}$$

Let $\mathcal{T}_n = \{x_j = T_1 + 2b_n j, j = 0, 1, \ldots, m_n\}$ and $m_n = \lfloor (T_2 - T_1)/(2b_n)\rfloor$. Then

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{\sqrt{nb_n}}{\sqrt{\varphi_K}} \sup_{x\in\mathcal{T}_n} \frac{[\hat{f}_X(x)]^{1/2}}{\sigma(x)}\Big|\hat{\mu}_{b_n}(x) - \mu(x) - b_n^2\psi_K\rho_\mu(x)\Big| \leq B_{m_n}(z)\right\} = e^{-2e^{-z}}.$$

Observe that $\mathcal{T}_n$ becomes denser in $\mathcal{T}$ as $b_n \to 0$. Since $b_n \to 0$, if the regression function μ is sufficiently smooth, then $\{\mu(x) : x \in \mathcal{T}\}$ can be well approximated by $\{\mu(x) : x \in \mathcal{T}_n\}$ for large n. Theorem 1 is useful to construct SCB in an approximate version of (2):

$$\lim_{n\to\infty} \mathbb{P}\{l(x) \leq \mu(x) \leq u(x) \text{ for all } x \in \mathcal{T}_n\} = 1 - \alpha. \tag{11}$$

In Theorem 1, (9) imposes conditions on the bandwidth b_n and the strength of the dependence. The first part $nb_n^9 \log n \to 0$ aims to control the bias with b_n being not too large, while the second one $(\log n)^3/(nb_n^3) \to 0$ suggests that b_n should not be too small, thus ensuring the validity of the moderate deviation principle (see the proof of Theorem 3). The third part suggests that the dependence should not be too strong. For SRD processes, we have $\Xi_n = O(n)$ and it is easily seen that the third term in (9) becomes redundant.

Interestingly, (9) also allows long-range dependent processes; see Section 4.2. If $b_n \asymp n^{-\beta}$ with $\beta \in (1/9, 1/3)$, then the first two terms in (9) approach 0 as $n \to \infty$. In particular, (9) allows $\beta = 1/5$, which corresponds to the MSE-optimal bandwidth.

Let $\hat{\sigma}_n(x)$ (resp. $\hat{\rho}_\mu(x)$) be an estimate of $\sigma(x)$ (resp. $\rho_\mu(x)$) such that $\sup_{x \in \mathcal{T}} |\hat{\sigma}_n(x) - \sigma(x)| = o_{\mathrm{p}}[(\log n)^{-1}]$ and $\sup_{x \in \mathcal{T}} |\hat{\rho}_\mu(x) - \rho_\mu(x)| = o_{\mathrm{p}}[(n b_n^5 \log n)^{-1/2}]$. By Slutsky's theorem, Theorem 1 still holds if σ and ρ_μ therein are replaced by $\hat{\sigma}_n$ and $\hat{\rho}_\mu$, respectively. By Theorem 1, an asymptotic $100(1-\alpha)\%$ SCB for μ can be constructed as

$$\hat{\mu}_{b_n}(x) - b_n^2 \psi_K \hat{\rho}_\mu(x) \pm \frac{\sqrt{\varphi_K} \hat{\sigma}_n(x)}{\sqrt{n b_n \hat{f}_X(x)}} B_{m_n}(z_\alpha) \text{ and } z_\alpha = -\log\log[(1-\alpha)^{-1/2}]. \tag{12}$$

In (12), $\rho_\mu(x)$ can not be easily estimated since it involves unknown functions μ'', μ' and f'_X. Following Wu and Zhao (2006), we adopt the simple jackknife-type bias correction procedure which avoids estimating μ'', μ' and f'_X:

$$\hat{\mu}^*_{b_n}(x) = 2\hat{\mu}_{b_n}(x) - \hat{\mu}_{\sqrt{2}b_n}(x). \tag{13}$$

Using (13) is equivalent to using the 4th order kernel $K^*(u) = 2K(u) - K(u/\sqrt{2})/\sqrt{2}$. Obviously, $K^* \in \mathcal{K}$ has support $[-\sqrt{2}, \sqrt{2}]$ and $\psi_{K^*} = 0$. Let $m_n^* = \lfloor (T_2 - T_1)/(2\sqrt{2} b_n) \rfloor$ and $\mathcal{T}_n^* = \{x_j^* = T_1 + 2\sqrt{2} b_n j, j = 0, 1, \ldots, m_n^*\}$. Then Theorem 1 still holds with $\hat{\mu}_{b_n}$ (resp. $K, m_n, \mathcal{T}_n$) replaced by $\hat{\mu}^*_{b_n}$ (resp. $K^*, m_n^*, \mathcal{T}_n^*$).

3.2 Simultaneous confidence bands for σ^2

Let $\hat{\mu}^*_{b_n}$ be as in (13). Since $\mathbb{E}(\varepsilon_i^2) = 1$ and $\mathbb{E}\{[Y_i - \mu(X_i)]^2 | X_i = x\} = \sigma^2(x)$, a natural estimate of $\sigma^2(x)$ is

$$\hat{\sigma}^2_{h_n}(x) = \frac{1}{nh_n \tilde{f}_X(x)} \sum_{i=1}^{n} [Y_i - \hat{\mu}^*_{b_n}(X_i)]^2 \tilde{K}_{h_n}(x - X_i), \tag{14}$$
$$\text{where} \quad \tilde{f}_X(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \tilde{K}_{h_n}(x - X_i).$$

Here $\tilde{K}_{h_n}(u) = \tilde{K}(u/h_n)$ for some kernel $\tilde{K}$ and bandwidth $h_n > 0$. Note that $\tilde{K}$ and h_n can be different from K and b_n.

The asymptotic behavior of $\hat{\sigma}^2_{h_n}$ depends on the relative magnitudes of b_n and h_n. Namely, the asymptotic distribution of $\hat{\sigma}^2_{h_n}$ and the speed of the convergence can be different for the three cases: (i) $h_n/b_n \to 0$, (ii) $h_n/b_n \to \infty$ and (iii) $h_n \asymp b_n$. See Zhao and Wu (2006) for more discussion on this in the context of kernel quantile regression. Under (iii) we have the oracle property that $\sigma^2(\cdot)$ can be estimated with the same convergence rate as if μ were known.

Proposition 2. *Let $K, \tilde{K} \in \mathcal{K}, \varepsilon_0 \in \mathcal{L}^6$ and $h_n \asymp b_n$. Assume that $\inf_{x\in\mathcal{T}} f_X(x) > 0$, $\inf_{x\in\mathcal{T}} \sigma(x) > 0$ and $f_X, \mu \in \mathcal{C}^4(\mathcal{T}^\epsilon)$ for some $\epsilon > 0$. Further assume that*

$$h_n^{3/2} \log n + \frac{1}{n^2 h_n^5} + \frac{\Xi_n}{n^2} \to 0. \tag{15}$$

Then

$$\sup_{x\in\mathcal{T}} |\hat{\sigma}^2_{h_n}(x) - \sigma^2(x)| = O_{\mathrm{p}}\Big\{h_n^2 + \frac{1}{nh_n^{5/2}} + \Big[\frac{\log n}{nh_n}\Big]^{1/2} + \Big[\frac{\log n}{n^3 h_n^7}\Big]^{1/4} + \frac{\Xi_n^{1/2} h_n}{n}\Big\}.$$

Proposition 2 provides a uniform error bound for the estimate $\hat{\sigma}^2_{h_n}(\cdot)$. From the proof of

Proposition 2 and Theorem 2, it is easily seen that, as in Proposition 1, one can establish a CLT for $\hat\sigma^2_{h_n}(x)$ for each fixed x and the optimal bandwidth $h_n \asymp n^{-1/5}$. We omit the details. In Proposition 2, if one uses the optimal bandwidth h_n, then $\sup_{x\in\mathcal{T}} |\hat\sigma^2_{h_n}(x) - \sigma^2(x)| = O_{\mathrm{p}}[n^{-2/5}(\log n)^{1/2} + \Xi_n^{1/2} n^{-6/5}]$. The first part $O_{\mathrm{p}}[n^{-2/5}(\log n)^{1/2}]$ in the error bound is optimal in nonparametric curve estimation for independent data. The second part accounts for dependence, and it can be absorbed into the first one if $\Xi_n = O(n^{8/5})$.

To construct SCB for $\sigma^2(x)$ on the interval $\mathcal{T} = [T_1, T_2]$, as in the case of μ, we assume WLOG that $\tilde K \in \mathcal{K}$ has support $[-1, 1]$. Let $\tilde m_n = \lfloor (T_2 - T_1)/(2h_n) \rfloor$ and $\tilde{\mathcal{T}}_n = \{\tilde x_j = T_1 + 2h_n j : j = 0, 1, \ldots, \tilde m_n\}$ be the grid points on $\mathcal{T}$.

Theorem 2. *Let the conditions in Proposition 2 be fulfilled. Assume that* $\sigma \in \mathcal{C}^4(\mathcal{T}^\epsilon)$ *for some* $\epsilon > 0$ *and*

$$nh_n^9 \log n + \frac{\log n}{nh_n^4} + \Xi_n\Big[\frac{h_n^3 \log n}{n} + \frac{(\log n)^2}{n^2 h_n^{4/3}}\Big] \to 0. \tag{16}$$

Let $B_n(z)$ *be as in (10). Then*

$$\lim_{n\to\infty} \mathbb{P}\left\{ \frac{\sqrt{nh_n}}{\sqrt{\varphi_{\tilde K}\nu_\varepsilon}} \sup_{x\in\tilde{\mathcal{T}}_n} \frac{[\tilde f_X(x)]^{1/2}}{\hat\sigma^2_{h_n}(x)} \Big| \hat\sigma^2_{h_n}(x) - \sigma^2(x) - b_n^2 \psi_{\tilde K} \rho_\sigma(x) \Big| \le B_{\tilde m_n}(z) \right\} = e^{-2e^{-z}},$$

where $\nu_\varepsilon = \mathbb{E}(\varepsilon_0^4) - 1 > 0$ *and*

$$\rho_\sigma(x) = 2\sigma'(x)^2 + 2\sigma(x)\sigma''(x) + 4\sigma(x)\sigma'(x) f_X'(x)/f_X(x).$$

If μ were known and we use $\hat\sigma^2_{h_n}$ in (14) to estimate σ^2 with $\hat\mu^*_{b_n}$ therein replaced by the true function μ, then Theorem 2 is still applicable. So Theorem 2 implies the oracle property that, under the specified conditions, the construction of SCB for σ^2 does not rely on the estimation of μ. As in (13), we propose the bias-corrected estimate

$$\hat{\sigma}^{2*}_{h_n}(x) = 2\hat{\sigma}^2_{h_n}(x) - \hat{\sigma}^2_{\sqrt{2}h_n}, \tag{17}$$

which is equivalent to using the 4th order kernel $\tilde{K}^*(u) = 2\tilde{K}(u) - \tilde{K}(u/\sqrt{2})/\sqrt{2}$. Similarly as in Section 3.1, we can define $\tilde{m}^*_n$ and $\tilde{\mathcal{T}}^*_n$ accordingly and Theorem 2 still holds with $\hat{\sigma}^2_{h_n}$ (resp. $\tilde{K}, \tilde{m}_n, \tilde{\mathcal{T}}_n$) replaced by $\hat{\sigma}^{2*}_{h_n}$ (resp. $\tilde{K}^*, \tilde{m}^*_n, \tilde{\mathcal{T}}^*_n$).

3.3 Estimation of ν_ε in Theorem 2

To apply Theorem 2, one needs to estimate $\nu_\varepsilon = \mathbb{E}(\varepsilon_0^4) - 1$. Here we estimate ν_ε by

$$\hat{\nu}_\varepsilon = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^4 \mathbf{1}_{X_i \in \mathcal{T}}}{\sum_{i=1}^n \mathbf{1}_{X_i \in \mathcal{T}}} - 1, \quad \text{where} \quad \hat{\varepsilon}_i = \frac{Y_i - \hat{\mu}^*_{b_n}(X_i)}{\hat{\sigma}^*_{h_n}(X_i)}, \quad i = 1, 2, \ldots, n. \tag{18}$$

Here $\hat{\varepsilon}_i$ are estimated residuals for model (1). The naive estimate $n^{-1}\sum_{i=1}^n \hat{\varepsilon}_i^4 - 1$ does not have a good practical performance since $\hat{\sigma}^*_{h_n}(x)$ behaves poorly if x is too large or too small. Truncation by $\mathcal{T}$ improves the performance.

Proposition 3. *Assume that the conditions in Proposition 2 are satisfied. Then*

$$\hat{\nu}_\varepsilon - \nu_\varepsilon = O_{\mathrm{p}}\Big\{n^{-1/3} + h_n^4 + \frac{1}{nh_n^{5/2}} + \Big[\frac{\log n}{nh_n}\Big]^{1/2} + \Big[\frac{\log n}{n^3 h_n^7}\Big]^{1/4} + \frac{\Xi_n^{1/2}}{n}\Big\}. \tag{19}$$

By Proposition 3, when one chooses the mean squares error (MSE) optimal bandwidths $b_n \asymp h_n \asymp n^{-1/5}$ and assume $\Xi_n = O(n^{4/3})$, then $\hat{\nu}_\varepsilon - \nu_\varepsilon = O_{\mathrm{p}}(n^{-1/3})$.

4 Examples

To apply Theorems 1 and 2, we need to deal with Ξ_n defined in (5). Let $(\eta_i')_{i\in\mathbb{Z}}$ be an iid copy of $(\eta_i)_{i\in\mathbb{Z}}$, $\mathcal{F}_i' = (\mathcal{F}_{-1}, \eta_0', \eta_1, \ldots, \eta_i)$. By Theorem 1 in Wu (2005), we have

$$\theta_i \le \varpi_i := \sup_{x\in\mathbb{R}} \|f_X(x|\mathcal{F}_i) - f_X(x|\mathcal{F}_i')\| + \sup_{x\in\mathbb{R}} \|f_X'(x|\mathcal{F}_i) - f_X'(x|\mathcal{F}_i')\|. \tag{20}$$

For many processes there exist simple and easy-to-use bounds for ϖ_i. Here we shall consider linear processes and some popular nonlinear time series models.

4.1 Short-range dependent linear processes

Let $\eta_i, i \in \mathbb{Z}$, be iid. Assume $\eta_0 \in \mathcal{L}^q$, $q > 0$, and $\mathbb{E}(\varepsilon_0) = 0$ if $q \ge 1$. For real sequence $(a_i)_{i\in\mathbb{Z}}$ satisfying $\sum_{i=0}^\infty |a_i|^{q\wedge 2} < \infty$, the linear process

$$X_i = \sum_{j=0}^\infty a_j \eta_{i-j}, \tag{21}$$

is well-defined and stationary. Special cases of (21) include ARMA and fractional ARIMA (FARIMA) models. Assume WLOG that $a_0 = 1$. Let $\bar{X}_i = X_i - \eta_i$ and $\bar{X}_i' = \bar{X}_i + a_i(\eta_0' - \eta_0)$. Then $f_X(x|\mathcal{F}_{i-1}) = f_\eta(x - \bar{X}_i)$ and $f_X(x|\mathcal{F}_{i-1}') = f_\eta(x - \bar{X}_i')$, where f_η is the density function of η_0. Assume that $f_\eta \in \mathcal{C}^2(\mathbb{R})$. Then simple calculation shows that $\theta_i = O(|a_i|^{q'})$, where $q' = (q \wedge 2)/2$; see Proposition 2 in Zhao and Wu (2006). Therefore we have $\Xi_n = O(n)$ if $\sum_{i=1}^\infty |a_i|^{q'} < \infty$. If $q \ge 2$, then the latter condition becomes $\sum_{i=0}^\infty |a_i| < \infty$. For causal ARMA models, $a_i \to 0$ geometrically quickly. Note that our setting allows heavy-tailed innovations.

4.2 Long-range dependent linear processes

Consider the linear process (21) with $a_i = i^{-\alpha}\ell(i)$, where $\alpha > 0$ satisfies $\alpha q' > 1/2, q' = (q\wedge 2)/2$, and $\ell(\cdot)$ is a slowly varying function. The case of $\alpha q' > 1$ is covered by Section 4.1. Assume $\alpha q' \in (1/2, 1]$. If $q \geq 2$ and $\alpha \in (1/2, 1)$, by Karamata's theorem, the covariances $\mathbb{E}(X_0 X_n)$ are of order $n^{1-2\alpha}\ell^2(n)$ and not summable, hence (X_i) is long-range dependent. As in Section 4.1, $\theta_i = O[i^{-\alpha q'}\ell^{q'}(i)]$. By (6), $\Xi_n = O[n^{3-2\alpha q'}\ell^{2q'}(n)]$ if $\alpha q' \in (1/2, 1)$ and $\Xi_n = O\{n[\sum_{i=1}^n |\ell^{q'}(i)|/i]^2\}$ if $\alpha q' = 1$.

In Theorems 1 and 2, let the bandwidths $b_n \asymp h_n \asymp n^{-\beta}$. If $\alpha q' \in (17/26, 1]$, then (9) holds provided that

$$\max\left\{\frac{1}{9}, \frac{2(1-\alpha q')}{3}\right\} < \beta < \min\left\{\frac{1}{3}, \frac{3(2\alpha q' - 1)}{4}\right\}. \tag{22}$$

Replacing 1/3 with 1/4 on the right hand side of (22), we then get a sufficient condition for (16). The constraint $\alpha q' \in (17/26, 1]$ is imposed to ensure the compatibility of (22). It is unclear how to deal with the case $\alpha q' \in (1/2, 17/26]$.

Example 1. Let $\alpha(z) = 1 - \sum_{i=1}^k \alpha_i z^i$ and $\beta(z) = 1 + \sum_{i=1}^p \beta_i z^i$ be two polynomial functions with $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_p \in \mathbb{R}$. Denote by B the backward shift operator defined by $B^j X_n = X_{n-j}$. Consider the FARIMA(k, d, p) process X_n given by $\alpha(B)(1-B)^d X_n = \beta(B)\varepsilon_n, d \in (-1/2, 1/2)$. Let $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ be the gamma function. In the simple case of $p = k = 0$, we have $X_n = \sum_{i=0}^\infty a_i \varepsilon_{n-i}$, where

$$a_n = \frac{\Gamma(n+d)}{\Gamma(n+1)\Gamma(d)} \asymp n^{d-1}. \tag{23}$$

If $d \in (0, 1/2)$, then X_n is long-range dependent. More generally, it can be shown that (23) holds for general FARIMA(k, d, p) processes if $\alpha(z) \neq 0$ for all complex $|z| \leq 1$.

4.3 Nonlinear AR models

Consider the following model

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \quad X_i = \tilde{\mu}(X_{i-1}) + \tilde{\sigma}(X_{i-1})\eta_i. \tag{24}$$

As a special case, if $X_i = Y_{i-1}, \eta_i = \varepsilon_{i-1}, \tilde{\mu} = \mu$ and $\tilde{\sigma} = \sigma$, then (24) becomes the nonlinear AR model

$$Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\varepsilon_i. \tag{25}$$

Special cases of (25) include linear AR, ARCH, TAR and EAR processes. Denote by f_η the density of η_0. Assume $\eta_0 \in \mathcal{L}^q$ and $\sup_{x\in\mathbb{R}}(1+|x|)[|f'_\eta(x)| + |f''_\eta(x)|] < \infty$. As in Zhao and Wu (2006), we have $\theta_i = O(r^i)$ with $r \in (0,1)$, and hence $\Xi_n = O(n)$, provided that

$$\inf_{x\in\mathbb{R}} \tilde{\sigma}(x) > 0, \quad \sup_{x\in\mathbb{R}}[|\tilde{\mu}'(x)| + |\tilde{\sigma}'(x)|] < \infty, \quad \sup_{x\in\mathbb{R}} \|\tilde{\mu}'(x) + \tilde{\sigma}'(x)\eta_0\|_q < 1. \tag{26}$$

Example 2. Consider the ARCH model $X_n = \eta_n\sqrt{a^2 + b^2X_{n-1}^2}$, where $\eta_i, i \in \mathbb{Z}$, are iid and a, b are real parameters. If $\eta_0 \in \mathcal{L}^q$ and $|b|\|\eta_0\|_q < 1$, then (26) holds.

5 A simulation study

In this section we shall present a simulation study for the performance of our SCBs constructed in Section 3. Let $\varepsilon_i, i \in \mathbb{Z}$, be iid standard normal variables. We shall consider the following two models

$$\text{Model 1: AR(1)} \quad Y_i = \mu(Y_{i-1}) + s\varepsilon_i, \quad i = 1, 2, \ldots, n.$$

$$\text{Model 2: ARCH(1)} \quad Y_i = \sigma(Y_{i-1})\varepsilon_i, \quad i = 1, 2, \ldots, n.$$

Here μ and σ are functions of interest and $s > 0$ is the scale parameter. Model 1 is a nonlinear AR model and Model 2 is an ARCH model.

By Proposition 1, the MSE-optimal bandwidth b_n of $\hat{\mu}_{b_n}$ is of order $cn^{-1/5}$ for some constant c. In practice, it is non-trivial to find a c that has good performance. On one hand, the bias correction (13) allows one to choose relatively larger bandwidth b_n. On the other hand, however, larger bandwidth b_n results in relatively fewer grid points in $\mathcal{T}_n$, and consequently a less accurate approximation of $\{\mu(x) : x \in \mathcal{T}_n\}$ to $\{\mu(x) : x \in \mathcal{T}\}$. In our simulations, we tried different bandwidths and different sets $\mathcal{T}_{(30)}, \mathcal{T}_{(50)}$ and $\mathcal{T}_{(100)}$ of grid points to access the performance of our SCBs. Here $\mathcal{T}_{(k)}, k \in \mathbb{N}$, denotes the set containing k grid points evenly spaced over $\mathcal{T}$, regardless of the bandwidth. Since Nadaraya-Watson estimate suffers the boundary problem, we employ the local linear estimate (Fan and Gijbels 1996) in all our subsequent data analysis.

5.1 A bootstrap-based procedure

By Theorem 1 and the discussions there, the asymptotic distribution involved in the construction of SCBs for μ does not depend on the underlying process. However, the convergence in Theorems 1 and 2 is quite slow. Here we propose a bootstrap procedure to obtain the cutoff values based on iid standard normals. We shall illustrate the ideas by constructing SCBs for μ in Model 1.

(i) Choose an appropriate bandwidth b_n and $\mathcal{T}_{(k)}$, $k = 30, 50$ or 100.

(ii) Generate iid standard normals U_i and $Z_i, 1 \le i \le n$, and compute

$$D = \sup_{x \in \mathcal{T}_{(k)}} [\hat{f}_U^{\diamond}(x)]^{1/2} |2\hat{\mu}_{b_n}^{\diamond}(x) - \hat{\mu}_{\sqrt{2}b_n}^{\diamond}(x)|, \text{ where}$$

$$\hat{f}_U^\diamond(x) = \frac{1}{nb_n}\sum_{i=1}^n K_{b_n}(x-U_i) \text{ and } \hat{\mu}_b^\diamond(x) = \frac{\sum_{i=1}^n K_b(x-U_i)Z_i}{\sum_{i=1}^n K_b(x-U_i)}.$$

(iii) Repeat step (ii) for 10^4 (say) times and obtain the 95% quantile $\hat{q}_{0.95}$ of these Ds.

(iv) Compute $\hat{\mu}^*_{b_n}$ and $\hat{f}_X$ as in (13) and (7), respectively, and estimate s^2 by $\hat{s}^2 = \sum_{i=1}^n [Y_i - \hat{\mu}^*_{b_n}(X_i)]^2 \mathbf{1}_{X_i\in\mathcal{T}} / \sum_{i=1}^n \mathbf{1}_{X_i\in\mathcal{T}}$.

(v) The 95% SCB is constructed as $\hat{\mu}^*_{b_n}(x) \pm \hat{s}\hat{q}_{0.95}[\hat{f}_X(x)]^{-1/2}$.

To assess the performance of our SCB, we generate 10^4 realizations of $(\hat{\mu}^*_{b_n}(x), \hat{f}_X(x), \hat{s})$ from Model 1. For each realization, if μ lies within the band $\hat{\mu}^*_{b_n}(x) \pm \hat{s}\hat{q}_{0.95}[\hat{f}_X(x)]^{-1/2}$ for all $x \in \mathcal{T}_{(k)}$, namely $\max_{x\in\mathcal{T}_{(k)}}[\hat{f}_X(x)]^{-1/2}|\hat{\mu}^*_{b_n}(x) - \mu(x)|/\hat{s} \le \hat{q}_{0.95}$, then we say that the SCB covers μ. The simulated coverage probabilities is the proportion of these 10^4 SCBs that cover μ.

The case of $\sigma^2(\cdot)$ can be similarly treated. Let (U_i, Z_i) be as in step (ii) and note that $\mathbb{E}(Z_0^4) = 3$. Based on 10^4 realizations of

$$V = 2^{-1/2} \sup_{x\in\mathcal{T}_{(k)}} [\tilde{f}_U^\diamond(x)]^{1/2} |1 - 1/[2\hat{\sigma}^{2\diamond}_{h_n}(x) - \hat{\sigma}^{2\diamond}_{\sqrt{2}h_n}(x)]|,$$

where

$$\tilde{f}_U^\diamond(x) = \frac{1}{nh_n}\sum_{i=1}^n \tilde{K}_{h_n}(x-U_i) \quad \text{and} \quad \hat{\sigma}_h^{2\diamond}(x) = \frac{\sum_{i=1}^n \tilde{K}_h(x-U_i)Z_i^2}{\sum_{i=1}^n \tilde{K}_h(x-U_i)},$$

we obtain the estimated 95% quantile $\hat{q}_{0.95}$ of these Vs. For the construction of SCB for $\sigma^2(\cdot)$, we estimate ν_ε by $\hat{\nu}_\varepsilon$ in (18) with $\hat{\varepsilon}_i = Y_i/\hat{\sigma}^*_{h_n}(X_i)$, where $\hat{\sigma}^*_{h_n}(X_i)$ is as in (17).

5.2 Coverage probabilities

For Model 1, we let $n = 2000$, $\mu(x) = 0.7x$ and $s = 0.6$. Then (26) holds. Under this setting, simulations show that about 1800-1900 (90-95%) of the Ys lie within the interval $[-1.5, 1.5]$. Thus we take $\mathcal{T} = [-1.5, 1.5]$ and $\mathcal{T}_{(k)} = \{-1.5+3j/(k-1) : j = 0, 1, \ldots, k-1\}$, $k = 30, 50, 100$. To study how bandwidth affects the coverage probabilities, we tried 13 bandwidths $b_n = 0.03, 0.04, \ldots, 0.10, 0.12, \ldots, 0.20$. When applying the simulation procedure in Section 5.1, we adopt the following technique for fitting $\hat{\mu}_{b_n}$ at $Y_i, 0 \le i \le n$: we fit 500 grid points evenly spaced on the range of Y_i's and use the fitted value of the nearest grid point to Y_i as $\hat{\mu}_{b_n}(Y_i)$. Doing this allows one to gain better smoothness since the original series Y may be irregularly spaced. In Model 2, we let $n = 2000$ and $\sigma(x) = (0.4+0.2x^2)^{1/2}$. Then (26) holds. We take $\mathcal{T} = [-1, 1]$ and $\mathcal{T}_{(k)} = \{-1 + 2j/(k-1) : j = 0, 1, \ldots, k-1\}$, $k = 30, 50, 100$. We apply the same procedure as in Model 1. Tables 1 and 2 show that the coverage probabilities of SCBs for μ and σ^2 are very close to the nominal level 95% and they are relatively insensitive to the choice of bandwidths.

Insert Table 1 and Table 2 about here

6 Application to the IBM stock data

The dataset contains 2336 records, $S_0, S_1, \ldots, S_{2335}$, of IBM's weekly adjusted closing price during the period December 31st 1969 to October 9th 2006. Let $Y_i = \log(S_{i+1}/S_i), i = 0, 1, \ldots, 2334$, be the log returns. Since 2302 out of the 2335 (98.6%) Y's lie within the range -0.1 and 0.1, we deleted the other 33 Y's in our subsequent analysis. Furthermore, among these 2302 Y's, 2161 (93.9%) of them lie within the band $[-0.06, 0.06]$, so we choose the "interior" interval $\mathcal{T} = [-0.06, 0.06]$ and construct SCB for $\mu(\cdot)$ and $\sigma^2(\cdot)$ in model (1) with $(X_i, Y_i) = (Y_{i-1}, Y_i)$ for grid points $\mathcal{T}_{(50)} = \{-0.06 + 0.12j/49 : j = 0, 1, \ldots, 49\}$. An alternative approach is to keep only those Y's that are within $\mathcal{T}$ while completely deleting other Y's that are outside $\mathcal{T}$. We do not recommend the latter approach since it may cause the boundary problem due to the insufficiency of points around the two boundaries -0.06 and 0.06. In contrast, the first approach can alleviate the boundary effect by keeping those 141 points that are within $[-0.10, -0.06]$ or $[0.06, 0.10]$.

To construct SCB, we apply the bootstrap procedure described in Section 5.1 with some modifications to obtain more accurate cutoff values. In step (ii) therein, to better mimic the structure in the original data, we generate $U_1, U_2, \ldots, U_{2302}$ as a mixture of 2161 uniform random variables on $\mathcal{T} = [-0.06, 0.06]$, 64 uniform variables on $[-0.10, -0.06]$ and 77 uniform variables on $[0.06, 0.10]$, where the numbers 2161, 64 and 77 represent the counts of the original data points that lie within the three corresponding intervals, respectively. We adopt the automatic bandwidth selector (function `dpill` in the `R` package `KernSmooth`) of Ruppert et al (1995) with the specified range.x $= \mathcal{T}$ and obtain the optimal bandwidths $b_n = 0.020$ and $h_n = 0.013$. The estimated 95% cutoff quantiles for constructing SCB of $\mu(\cdot)$ and $\sigma^2(\cdot)$ are 0.386 and 0.548, respectively, and the estimated $\hat{\nu}_\varepsilon = 4.29$.

Interestingly, the 95% SCB for μ and σ^2 in Figure 1 suggests that we can accept the two null hypotheses that the regression function μ is linear and that the squared volatility function is quadratic. The fitted linear equation is $\hat{\mu}_{\text{linear}}(x) = 0.000958 + 0.0474x$ and the fitted quadratic curve is $\hat{\sigma}^2_{\text{quadratic}}(x) = 0.00094 - 0.000205x + 0.197367x^2$. We conclude that the following AR(1)-ARCH(1) model is an adequate fit for IBM weekly log returns:

$$Y_i = 0.000958 + 0.0474 Y_{i-1} + \sqrt{0.00094 - 0.000205 Y_{i-1} + 0.197367 Y_{i-1}^2}\,\varepsilon_i. \qquad (27)$$

Insert Figure 1 about here

7 Appendix

Recall that $\mathcal{F}_i = (\ldots, \eta_{i-1}, \eta_i)$. Let $\mathcal{G}_i = (\ldots, \eta_i, \eta_{i+1}; \varepsilon_i, \varepsilon_{i-1}, \ldots)$. By the assumption in Section 2, ε_i is independent of $\mathcal{G}_{i-1}$. In the sequel, with a slight abuse of notation we refer $\mathcal{F}_i$ (resp. $\mathcal{G}_i$) as the sigma field generated by $\mathcal{F}_i$ (resp. $\mathcal{G}_i$). Recall (5) for Ξ_n.

Recall that $f_X(x|\mathcal{F}_{i-1})$ is the conditional density of X_i at x given $\mathcal{F}_{i-1}$. Define

$$I_n(x) = \sum_{i=1}^{n} [f_X(x|\mathcal{F}_{i-1}) - \mathbb{E} f_X(x|\mathcal{F}_{i-1})], \quad x \in \mathbb{R}. \tag{28}$$

Lemma 1. *Let $T > 0$ be fixed. Then $\| \sup_{|x| \le T} |I_n(x)| \| = O(\Xi_n^{1/2})$.*

Proof. By Theorem 1 in Wu (2006), $\sup_{x \in \mathbb{R}} [\|I_n(x)\| + \|I'_n(x)\|] = O(\Xi^{1/2})$. Then Lemma 1 easily follows in view of $\sup_{|x| \le T} |I_n(x) - I_n(-T)| \le \int_{-T}^{T} |\partial I_n(x)/\partial x| dx$. ♢

7.1 A general CLT and a maximal deviation result

Let g and h be measurable functions such that $h(\varepsilon_0) \in \mathcal{L}^2$ and $\vartheta_h^2 = \mathrm{Var}[h(\varepsilon_0)] > 0$. For $K \in \mathcal{K}$ define

$$S_n(x) = \sum_{i=1}^{n} \xi_i(x), \text{ where } \xi_i(x) = \frac{g(X_i)[h(\varepsilon_i) - \mathbb{E}(h(\varepsilon_i))] K_{b_n}(x - X_i)}{\vartheta_h g(x) \sqrt{n b_n \varphi_K f_X(x)}}. \tag{29}$$

In Proposition 4 and Theorem 3 below, we shall establish a general central limit theorem and a maximal deviation result for $S_n(x)$. These results are of independent interest and they are essential to the proof of our main results in Section 3.

Proposition 4. *Let $x \in \mathbb{R}$ be fixed, $K \in \mathcal{K}$ and $h(\varepsilon_0) \in \mathcal{L}^2$. Assume that $f_X(x) > 0$, $g(x) \neq 0$, and $f_X, g \in \mathcal{C}^0(\{x\}^\epsilon)$ for some $\epsilon > 0$. Further assume that $b_n \to 0$, $n b_n \to \infty$ and $\Xi_n / n^2 \to 0$. Then $S_n(x) \Rightarrow N(0, 1)$.*

Proof. Since ε_i is independent of $\mathcal{G}_{i-1}$, $\{\xi_i(x)\}_{i=1}^n$ form martingale differences with respect to $\mathcal{G}_i$. By the martingale central limit theorem, it suffices to verify the convergence of conditional variance and the Lindeberg condition. Let $\gamma_i = g^2(X_i)K_{b_n}^2(x - X_i)$, $u_i = \gamma_i - \mathbb{E}(\gamma_i|\mathcal{F}_{i-1})$ and $v_i = \mathbb{E}(\gamma_i|\mathcal{F}_{i-1}) - \mathbb{E}(\gamma_i)$. Write

$$\sum_{i=1}^n [\gamma_i - \mathbb{E}(\gamma_i)] = M_n + R_n, \quad \text{where} \quad M_n = \sum_{i=1}^n u_i \quad \text{and} \quad R_n = \sum_{i=1}^n v_i. \tag{30}$$

Hereafter we shall call (30) M/R-decomposition. Since $\{u_i\}_{i=1}^n$ are martingale differences with respect to $\mathcal{F}_i$, we have $M_n = O_{\mathrm{p}}(\sqrt{nb_n})$. Recall (28) for $I_n(x)$. By Lemma 1,

$$\begin{aligned} \|R_n\| &= b_n \left\| \int_{\mathbb{R}} K^2(u)g^2(x - ub_n)I_n(x - ub_n)du \right\| \\ &\leq b_n \int_{\mathbb{R}} K^2(u)g^2(x - ub_n)\|I_n(x - ub_n)\|du = O(\Xi_n^{1/2} b_n). \end{aligned} \tag{31}$$

Since $b_n \to 0$, $nb_n \to \infty$ and $\Xi_n/n^2 \to 0$, simple calculations show that

$$\sum_{i=1}^n \mathbb{E}[\xi_i^2(x)|\mathcal{G}_{i-1}] = \frac{M_n + R_n}{nb_n\varphi_K f_X(x)g^2(x)} + \frac{\sum_{i=1}^n \mathbb{E}(\gamma_i)}{nb_n\varphi_K f_X(x)g^2(x)} \xrightarrow{p} 1. \tag{32}$$

Since K is bounded and has bounded support and $g \in \mathcal{C}^0(\{x\}^\epsilon)$, we have for sufficiently large n that $\sup_u |g(u)K_{b_n}(x - u)| \leq c$ for some constant c. Let $\lambda = \vartheta_h g(x)[\varphi_K f_X(x)]^{1/2}$ and $\bar{h}(\varepsilon_0) = h(\varepsilon_0) - \mathbb{E}[h(\varepsilon_0)]$. For any $s > 0$, by the independence of X_0 and ε_0,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[\xi_i^2(x)\mathbf{1}_{|\xi_i(x)|\geq s}] &= \frac{1}{\lambda^2 b_n}\mathbb{E}[g^2(X_0)K_{b_n}^2(x - X_0)\bar{h}^2(\varepsilon_0)\mathbf{1}_{|g(X_0)K_{b_n}(x-X_0)\bar{h}(\varepsilon_0)|\geq\lambda s\sqrt{nb_n}}] \\ &\leq \frac{1}{\lambda^2 b_n}\mathbb{E}[g^2(X_0)K_{b_n}^2(x - X_0)\bar{h}^2(\varepsilon_0)\mathbf{1}_{|\bar{h}(\varepsilon_0)|\geq\lambda s\sqrt{nb_n}/c}] \\ &= \frac{1}{\lambda^2 b_n}\mathbb{E}[g^2(X_0)K_{b_n}^2(x - X_0)] \times \mathbb{E}[\bar{h}^2(\varepsilon_0)\mathbf{1}_{|\bar{h}(\varepsilon_0)|\geq\lambda s\sqrt{nb_n}/c}] \to 0 \end{aligned}$$

in view of $nb_n \to \infty$ and $\bar{h}(\varepsilon_0) \in \mathcal{L}^2$. So the Lindeberg condition holds. $\diamondsuit$

Recall Theorem 1 for the definitions of m_n and $\mathcal{T}_n$. Theorem 3 below provides a maximal deviation result for $\sup_{x\in\mathcal{T}_n}|S_n(x)|$. Results of this type are essential to the construction of SCB (cf. Bickel and Rosenblatt 1973; Johnston 1982; Eubank and Speckman 1993 among others). To obtain a maximal deviation result under dependence, we shall apply Grama and Haeusler's (2006) martingale moderate deviation theorem.

Theorem 3. *Let $K \in \mathcal{K}$ have support $[-1,1]$ and $h(\varepsilon_0) \in \mathcal{L}^3$. Assume $\inf_{u\in\mathcal{T}} f_X(u) > 0$, $g(x) \neq 0, x \in \mathcal{T}$, and f_X, $g \in \mathcal{C}^2(\mathcal{T}^\epsilon)$ for some $\epsilon > 0$. Further assume that*

$$b_n^{4/3}\log n + \frac{(\log n)^3}{nb_n^3} + \frac{\Xi_n(\log n)^2}{n^2 b_n^{4/3}} \to 0. \tag{33}$$

Let $B_n(z)$ be as in (10). Then

$$\lim_{n\to\infty}\mathbb{P}\Big\{\sup_{x\in\mathcal{T}_n}|S_n(x)| \le B_{m_n}(z)\Big\} = e^{-2e^{-z}}. \tag{34}$$

Proof. Recall (29) for $\xi_i(x)$. For fixed integer $k \in \mathbb{N}$ and mutually different integers $0 \le j_1, j_2, \ldots, j_k \le m_n$, let the k-dimensional vector $\zeta_i = [\xi_i(x_{j_1}), \xi_i(x_{j_2}), \ldots, \xi_i(x_{j_k})]^T$ and $S_{n,k} = \sum_{i=1}^n \zeta_i = [S_n(x_{j_1}), S_n(x_{j_2}), \ldots, S_n(x_{j_k})]^T$. Here ^{T} denotes transpose. Then $\{\zeta_i\}_{i=1}^n$ are k-dimensional martingale differences with respect to $\mathcal{G}_i$. Let Q denote the quadratic characteristic matrix of $S_{n,k}$, i.e.

$$Q = \sum_{i=1}^n \mathbb{E}(\zeta_i\zeta_i^T|\mathcal{G}_{i-1}) := (Q_{rr'})_{1\le r,r'\le k}. \tag{35}$$

Let $\tau_{rr'} = \varphi_K g(x_{j_r})g(x_{j_{r'}})[f_X(x_{j_r})f_X(x_{j_{r'}})]^{1/2}$ and write

$$Q_{rr'} = \sum_{i=1}^n \mathbb{E}[\xi_i(x_{j_r})\xi_i(x_{j_{r'}})|\mathcal{G}_{i-1}] = \frac{1}{nb_n\tau_{rr'}}\sum_{i=1}^n g^2(X_i)K_{b_n}(x_{j_r} - X_i)K_{b_n}(x_{j_{r'}} - X_i).$$

For $r \neq r'$, since $|x_{j_r} - x_{j_{r'}}| \geq 2b_n$ and K has support $[-1, 1]$, $Q_{rr'} = 0$. For $r = r'$, we use the M/R-decomposition technique in (30). Define

$$\begin{aligned} \alpha_i(r) &= g^2(X_i)K^2_{b_n}(x_{j_r} - X_i) - \mathbb{E}[g^2(X_i)K^2_{b_n}(x_{j_r} - X_i)|\mathcal{F}_{i-1}], \\ \beta_i(r) &= \mathbb{E}[g^2(X_i)K^2_{b_n}(x_{j_r} - X_i)|\mathcal{F}_{i-1}] - \mathbb{E}[g^2(X_i)K^2_{b_n}(x_{j_r} - X_i)]. \end{aligned}$$

Since $\{\alpha_i(r)\}_{i=1}^n$ form martingale differences with respect to $\mathcal{F}_i$, we have

$$\Big\| \sum_{i=1}^n \alpha_i(r) \Big\|_2 = \Big[\sum_{i=1}^n \|\alpha_i(r)\|_2^2 \Big]^{1/2} = O(\sqrt{nb_n}), \tag{36}$$

uniformly over r. By Schwarz's inequality and Lemma 1, as in (31), we have

$$\begin{aligned} \Big\| \sum_{i=1}^n \beta_i(r) \Big\|_2 &= b_n \Big\| \int_{\mathbb{R}} K^2(u) g^2(x_{j_r} - ub_n) I_n(x_{j_r} - ub_n) du \Big\|_2 \\ &\leq b_n \int_{\mathbb{R}} K^2(u) g^2(x_{j_r} - ub_n) \|I_n(x_{j_r} - ub_n)\|_2 du = O(\Xi_n^{1/2} b_n), \end{aligned} \tag{37}$$

uniformly over r. Since $f_X, g \in \mathcal{C}^2(\mathcal{T}^\epsilon)$, by Taylor's expansion and the symmetry of K,

$$\Big| \sum_{i=1}^n \mathbb{E}[g^2(X_i)K^2_{b_n}(x_{j_r} - X_i)] - nb_n\tau_{rr} \Big| = O(nb_n^3). \tag{38}$$

Let $\delta_n = (nb_n)^{-1/2} + b_n^2 + \Xi_n^{1/2}/n$. By (36), (37) and (38), we have

$$\begin{aligned} \Big\| Q_{rr} - 1 \Big\|_{3/2} &\leq \frac{1}{nb_n\tau_{rr}} \Big\{ \Big\| \sum_{i=1}^n \alpha_i(r) \Big\|_{3/2} + \Big\| \sum_{i=1}^n \beta_i(r) \Big\|_{3/2} \\ &\quad + \Big\| \sum_{i=1}^n \mathbb{E}[g^2(X_i)K^2_{b_n}(x_{j_r} - X_i)] - nb_n\tau_{rr} \Big\|_{3/2} \Big\} = O(\delta_n) \end{aligned} \tag{39}$$

uniformly over r. Let $\mathbb{I}_k = \text{diag}(1, 1, \ldots, 1) = (u_{rr'})_{1 \leq r, r' \leq k}$ be the $k \times k$ identity matrix. Then $\mathbb{E}|Q_{rr'} - u_{rr'}|^{3/2} = O(\delta_n^{3/2})$, uniformly over $1 \leq r, r' \leq k$. It is easily seen that

$\sum_{i=1}^n \mathbb{E}|\xi_i(x_{j_r})|^3 = O[(nb_n)^{-1/2}]$ uniformly over $1 \le r \le k$. Then $\sum_{i=1}^n \mathbb{E}|\xi_i(x_{j_r})|^3 + \mathbb{E}|Q_{rr'} - u_{rr'}|^{3/2} = O(\Lambda_n)$, where $\Lambda_n = (nb_n)^{-1/2} + b_n^3 + \Xi_n^{3/4}/n^{3/2}$.

Under (33), elementary calculations show that $[1 + B_{m_n}(z)]^4 \exp[B_{m_n}^2(z)/2]\Lambda_n \to 0$ for fixed z. Denote by A_j the event $\{|S_n(x_j)| > B_{m_n}(z)\}$ and by $[N_1, N_2, \ldots, N_k]^T$ a k-dimensional centered normal random vector with the identity covariance matrix $\mathbb{I}_k$. By Theorem 1 in Grama and Haeusler (2006),

$$\mathbb{P}\Big[\bigcap_{r=1}^k A_{j_r}\Big] = \mathbb{P}\Big[\bigcap_{r=1}^k \{|N_r| > B_{m_n}(z)\}\Big][1 + o(1)] = \Big(\frac{2e^{-z}}{m_n}\Big)^k [1 + o(1)], \tag{40}$$

in view of the independence of $N_r, 1 \le r \le k$, and $\mathbb{P}(N_1 > x) = [1 + o(1)]\phi(x)/x$ as $x \to \infty$, where ϕ is the standard normal density function. Notice that $\mathbb{P}(\sup_{x\in\mathcal{T}_n} |S_n(x)| > B_{m_n}(z)) = \mathbb{P}(\bigcup_{j=0}^{m_n} A_j)$. By the inclusion-exclusion inequality, we have, for large enough n,

$$\begin{aligned}
\mathbb{P}\Big[\bigcup_{j=0}^{m_n} A_j\Big] &\le \sum_{j=0}^{m_n} \mathbb{P}[A_j] - \sum_{j_1<j_2\le m_n} \mathbb{P}\Big[A_{j_1} \bigcap A_{j_2}\Big] + \ldots + \sum_{j_1<j_2<\ldots<j_{2k-1}\le m_n} \mathbb{P}\Big[\bigcap_{r=1}^{2k-1} A_{j_r}\Big] \\
&= \sum_{r=1}^{2k-1} (-1)^{r-1} \binom{m_n + 1}{r} \Big(\frac{2e^{-z}}{m_n}\Big)^r [1 + o(1)] \\
&= -\sum_{r=1}^{2k-1} \frac{(-2e^{-z})^r}{r!}[1 + o(1)].
\end{aligned} \tag{41}$$

Thus,

$$\limsup_{n\to\infty} \mathbb{P}(\bigcup_{j=1}^{m_n} A_j) \le -\sum_{r=1}^{2k-1} \frac{[-2\exp(-z)]^r}{r!}.$$

Similarly, $\liminf_{n\to\infty} \mathbb{P}(\bigcup_{j=1}^{m_n} A_j) \ge -\sum_{r=1}^{2k} [-2\exp(-z)]^r/r!$. Letting $k \to \infty$, we have $\lim_{n\to\infty} \mathbb{P}(\bigcup_{j=1}^{m_n} A_j) = -\sum_{r=1}^{\infty} [-2\exp(-z)]^r/r! = 1 - \exp[-2\exp(-z)]$. So (34) follows. $\diamondsuit$

7.2 Proof of the results in Section 3

Before proving Proposition 1 and Theorem 1, we first note that, by the definition of $\hat{\mu}_{b_n}(x)$ and $\hat{f}_X(x)$ in (7),

$$\frac{\sqrt{nb_n}}{\sigma(x)\sqrt{\varphi_K}}\sqrt{\hat{f}_X(x)}\Big\{\hat{\mu}_{b_n}(x)-\mu(x)-\omega_n(x)U_n(x)\Big\}=V_n(x)\sqrt{\omega_n(x)}, \tag{42}$$

where $\omega_n(x)=f_X(x)/\hat{f}_X(x)$,

$$U_n(x) = \frac{1}{nb_nf_X(x)}\sum_{i=1}^{n}K_{b_n}(x-X_i)[\mu(X_i)-\mu(x)], \tag{43}$$

$$V_n(x) = \frac{1}{\sigma(x)\sqrt{nb_n\varphi_K f_X(x)}}\sum_{i=1}^{n}\sigma(X_i)\varepsilon_i K_{b_n}(x-X_i). \tag{44}$$

In (42), we can view $U_n(x)$ (resp. $V_n(x)$) as the bias (resp. stochastic) part of $\hat{\mu}_{b_n}(x)-\mu(x)$. The stochastic part $V_n(x)$ is treated in Proposition 4 and Theorem 3. The following Lemma 2 concerns $\omega_n(x)$ and the bias part $U_n(x)$.

Lemma 2. *Let $K\in\mathcal{K}$. (i) Recall the definition of $\hat{f}_X(x)$ in (7). Assume that $f_X\in\mathcal{C}^2(\mathcal{T}^\epsilon)$ for some $\epsilon>0$. Then*

$$\sup_{x\in\mathcal{T}}|\hat{f}_X(x)-f_X(x)|=O_{\mathrm{p}}(q_n),\ \text{ where } q_n=\sqrt{\log n/(nb_n)}+b_n^2+\Xi_n^{1/2}/n. \tag{45}$$

(ii) Recall the definition of $U_n(x)$ in (43). Let $\rho_\mu(x)$ be as in Theorem 1. Assume that $f_X,\mu\in\mathcal{C}^4(\mathcal{T}^\epsilon)$ for some $\epsilon>0$ and $\inf_{x\in\mathcal{T}}f_X(x)>0$. Then

$$\begin{aligned}&\sup_{x\in\mathcal{T}}|U_n(x)-b_n^2\psi_K\rho_\mu(x)|=O_{\mathrm{p}}(r_n),\\ &\text{where}\quad r_n=\sqrt{b_n\log n/n}+b_n^4+\Xi_n^{1/2}b_n/n.\end{aligned} \tag{46}$$

(iii) Let g, h be measurable functions such that $h(\varepsilon_0) \in \mathcal{L}^q$ for some $q \ge 2$ and $g \in \mathcal{C}^0(\mathcal{T}^\epsilon)$ for some $\epsilon > 0$. Then

$$\sup_{x\in\mathcal{T}} \frac{1}{nh_n}\Big|\sum_{i=1}^n K_{b_n}(x-X_i)g(X_i)[h(\varepsilon_i) - \mathbb{E}h(\varepsilon_i)]\Big| = O_{\rm p}[\chi_n(q)],$$
$$\text{where} \quad \chi_n(q) = \sqrt{\log n/(nb_n)} + n^{-q/4}b_n^{-q/4-1}(\log n)^{q/4-1/2}, q \ge 2. \tag{47}$$

Proof. (i) We use the M/R-decomposition technique in (30). By the argument in Lemma 4 in Zhao and Wu (2006), we can show that

$$\sup_{x\in\mathcal{T}}\Big|\sum_{i=1}^n \Big\{K_{b_n}(x-X_i) - \mathbb{E}[K_{b_n}(x-X_i)|\mathcal{F}_{i-1}]\Big\}\Big| = O_{\rm p}(\sqrt{nb_n\log n}). \tag{48}$$

By Lemma 1, since K has bounded support, we have

$$\sum_{i=1}^n \Big\{\mathbb{E}[K_{b_n}(x-X_i)|\mathcal{F}_{i-1}] - \mathbb{E}[K_{b_n}(x-X_i)]\Big\}$$
$$= b_n\int_{\mathbb{R}} K(u)f_X(x-ub_n)I_n(x-ub_n)du = O_{\rm p}(\Xi_n^{1/2}b_n). \tag{49}$$

So (i) follows from (48), (49) and the Taylor expansion $\sum_{i=1}^n \mathbb{E}[K_{b_n}(x-X_i)] - nb_nf_X(x) = O(nb_n^3)$. Similarly, we can show (ii).

(iii) We shall only consider the special case of $h(u) = u$ since other cases follow similarly. Let $c_n = (nb_n/\log n)^{1/2}$ and define

$$\overline{D}_n(x) = \sum_{i=1}^n \overline{d}_i(x), \text{ where } \overline{d}_i(x) = K_{b_n}(x-X_i)g(X_i)[\varepsilon_i\mathbf{1}_{|\varepsilon_i|>c_n} - \mathbb{E}(\varepsilon_i\mathbf{1}_{|\varepsilon_i|>c_n})],$$
$$\underline{D}_n(x) = \sum_{i=1}^n \underline{d}_i(x), \text{ where } \underline{d}_i(x) = K_{b_n}(x-X_i)g(X_i)[\varepsilon_i\mathbf{1}_{|\varepsilon_i|\le c_n} - \mathbb{E}(\varepsilon_i\mathbf{1}_{|\varepsilon_i|\le c_n})]/c_n.$$

Note that, for each fixed x, $\{\overline{d}_i(x)\}_{i=1}^n$ and $\{\underline{d}_i(x)\}_{i=1}^n$ form martingale differences with respect to $\mathcal{G}_i$, and $\mathbb{E}(\varepsilon_i^2 \mathbf{1}_{|\varepsilon_i|>c_n}) \le \mathbb{E}(|\varepsilon_i|^q)/c_n^{q-2} = O(c_n^{2-q})$. Simple calculations show that $\|\overline{D}_n(x)\|^2 = \sum_{i=1}^n \|\overline{d}_n(x)\|^2 = O(nb_n c_n^{2-q})$ and $\sup_{x\in\mathcal{T}} \|\partial\overline{D}_n(x)/\partial x\|^2 = O(nc_n^{2-q}/b_n)$, uniformly over $x \in \mathcal{T}$. Since $\sup_{x\in\mathcal{T}} |\overline{D}_n(x) - \overline{D}_n(T_1)| \le \int_{\mathcal{T}} |\partial\overline{D}_n(u)/\partial u| du$, by Schwarz's inequality, we have $\mathbb{E}[\sup_{x\in\mathcal{T}} |\overline{D}_n(x)|^2] = O(nc_n^{2-q}/b_n)$. Since $\{\underline{d}_i(x)\}_{i=1}^n$ are uniformly bounded martingale differences, by the argument in the proof of Lemma 4 in Zhao and Wu (2006), $\sup_{x\in\mathcal{T}} |\underline{D}_n(x)| = O_{\mathrm{p}}[(nb_n \log n)^{1/2}/c_n]$. Note that $\mathbb{E}(\varepsilon_i) = 0$. Then $|\sum_{i=1}^n K_{b_n}(x - X_i)g(X_i)\varepsilon_i| \le |\overline{D}_n(x)| + c_n|\underline{D}_n(x)|$. So (iii) follows. $\diamondsuit$

Proof of Proposition 1. Applying the M/R-decomposition technique in (30), we can show that, for fixed x, $\hat{f}_X(x) - f_X(x) = O_{\mathrm{p}}[(nb_n)^{-1/2} + b_n^2 + \Xi_n^{1/2}/n]$ and $U_n(x) - b_n^2\psi_K\rho_\mu(x) = O_{\mathrm{p}}[(b_n/n)^{1/2} + b_n^4 + \Xi_n^{1/2}b_n/n]$. Thus, $\omega_n(x) = 1 + O_{\mathrm{p}}[(nb_n)^{-1/2} + b_n^2 + \Xi_n^{1/2}/n]$. Proposition 1 then follows from Slutsky's theorem and Proposition 4. $\diamondsuit$

Proof of Theorem 1. By Lemma 2, $\sup_{x\in\mathcal{T}} |\omega_n(x) - 1| = O_{\mathrm{p}}(q_n)$ and $\sup_{x\in\mathcal{T}} |U_n(x) - b_n^2\psi_K\rho_\mu(x)| = O_{\mathrm{p}}(r_n)$, where q_n and r_n are as in (45) and (46), respectively. Under condition (9), simple calculations show that $q_n \log n + (nb_n \log n)^{1/2}(r_n + q_n b_n^2) \to 0$ and that (33) holds. So Theorem 1 follows from the decomposition (42) in view of Slutsky's theorem and Theorem 3. $\diamondsuit$

Proof of Proposition 2 and Theorem 2. Let q_n, r_n and $\chi_n(q)$ be as in (45), (46) and (47) in Lemma 2, respectively. Accordingly, we define $\tilde{q}_n, \tilde{r}_n$ and $\tilde{\chi}_n(q)$ with b_n in q_n, r_n and $\chi_n(q)$ replaced by h_n. For example, $\tilde{q}_n = [\log n/(nh_n)]^{1/2} + h_n^2 + \Xi_n^{1/2}/n$. Recall the definition of $\omega_n(x)$ and $U_n(x)$ in (42). Define

$$W_n(x) = \sum_{i=1}^n \frac{\sigma(X_i)\varepsilon_i K_{b_n}(x - X_i)}{nb_n f_X(x)}, \quad W_n^*(x) = \sum_{i=1}^n \frac{\sigma(X_i)\varepsilon_i K_{b_n}^*(x - X_i)}{nb_n f_X(x)},$$

where $K^*(u) = 2K(u) - K(u/\sqrt{2})/\sqrt{2}$. Applying Lemma 2 (iii) with $h(u) = u$ and $q = 6$, we have $\sup_{x\in\mathcal{T}} |W_n^*(x)| = O_{\rm p}[\chi_n(6)]$. So, by Lemma 2 (i) and (ii), elementary calculations show that, uniformly over $x \in \mathcal{T}$,

$$\hat{\mu}_{b_n}(x) - \mu(x) = \omega_n(x)U_n(x) + \omega_n(x)W_n(x) = b_n^2\psi_K\rho_\mu(x) + W_n(x) + O_{\rm p}(\Delta_n),$$

where $\Delta_n = r_n + q_n[b_n^2 + \chi_n(6)]$. Consequently, $\hat{\mu}_{b_n}^*(x) = \mu(x) + W_n^*(x) + O_{\rm p}(\Delta_n)$. Let $\tilde{f}_X$ be as in (14). By definition,

$$\begin{aligned}\hat{\sigma}_{h_n}^2(x) &= \frac{1}{nh_n\tilde{f}_X(x)}\sum_{i=1}^n \Big[\sigma(X_i)\varepsilon_i - W_n^*(X_i) + O_{\rm p}(\Delta_n)\Big]^2 \tilde{K}_{h_n}(x-X_i)\\ &= \frac{T_n(x)}{nh_n\tilde{f}_X(x)} + O_{\rm p}\Big[\frac{L_n(x)}{nh_n} + \frac{\Delta_n}{nh_n}J_n(x) + \chi_n^2(6) + \Delta_n^2\Big], \qquad (50)\end{aligned}$$

where

$$\begin{aligned}T_n(x) &= \sum_{i=1}^n \sigma^2(X_i)\varepsilon_i^2\tilde{K}_{h_n}(x-X_i),\\ L_n(x) &= \sum_{i=1}^n \sigma(X_i)\varepsilon_i W_n^*(X_i)\tilde{K}_{h_n}(x-X_i),\\ J_n(x) &= \sum_{i=1}^n \sigma(X_i)|\varepsilon_i|\tilde{K}_{h_n}(x-X_i).\end{aligned}$$

By the argument in Lemma 2 (i), we can show that $\sum_{i=1}^n \sigma(X_i)\tilde{K}_{h_n}(x-X_i) = O_{\rm p}[nh_n(1+\tilde{q}_n)]$, uniformly over $x \in \mathcal{T}$. By Lemma 2 (iii) with $h(u) = |u|$ and $q = 6$, we have uniformly over $x \in \mathcal{T}$ that

$$\begin{aligned}J_n(x) &= \sum_{i=1}^n \sigma(X_i)[|\varepsilon_i| - \mathbb{E}|\varepsilon_i|]\tilde{K}_{h_n}(x-X_i) + \mathbb{E}|\varepsilon_0|\sum_{i=1}^n \sigma(X_i)\tilde{K}_{h_n}(x-X_i)\\ &= O_{\rm p}\{nh_n[1+\tilde{q}_n+\tilde{\chi}_n(6)]\}. \qquad (51)\end{aligned}$$

Since ε_i is independent of $\mathcal{G}_{i-1}$ and $\mathbb{E}(\varepsilon_i) = 0$, simple calculation shows that

$$\|L_n(x)\|^2 = \mathbb{E}\Big[\sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j \frac{\sigma(X_i)\sigma(X_j)}{nb_n f_X(X_i)} K^*_{b_n}(X_i - X_j)\tilde{K}_{h_n}(x - X_i)\Big]^2 = O(h_n/b_n^2),$$

uniformly over $x \in \mathcal{T}$. Likewise, $\|\partial L_n(x)/\partial x\|^2 = O[1/(h_n b_n^2)]$. Since $\sup_{x\in\mathcal{T}} |L_n(x) - L_n(T_1)| \le \int_{\mathcal{T}} |\partial L_n(u)/\partial u| du$, by Schwarz's inequality, $\sup_{x\in\mathcal{T}} |L_n(x)| = O_{\mathrm{p}}[1/(h_n^{1/2} b_n)]$. Recall that $\mathbb{E}(\varepsilon_0^2) = 1$. Write

$$\frac{T_n(x)}{nh_n \tilde{f}_X(x)} - \sigma^2(x) = \frac{D_n + E_n(x)}{nh_n \tilde{f}_X(x)}, \tag{52}$$

where

$$\begin{aligned} D_n(x) &= \sum_{i=1}^{n} [\sigma^2(X_i) - \sigma^2(x)]\tilde{K}_{h_n}(x - X_i), \\ E_n(x) &= \sum_{i=1}^{n} \sigma^2(X_i)[\varepsilon_i^2 - \mathbb{E}(\varepsilon_i^2)]\tilde{K}_{h_n}(x - X_i). \end{aligned}$$

Let $\tilde{\omega}_n(x) = f_X(x)/\tilde{f}_X(x)$. By Lemma 2 (i), $\sup_{x\in\mathcal{T}} |\tilde{\omega}_n(x) - 1| = O_{\mathrm{p}}(\tilde{q}_n)$. Let $\rho_\sigma(x)$ be as in Theorem 2. As in Lemma 2 (ii), we can show that

$$\sup_{x\in\mathcal{T}} \Big| \frac{D_n(x)}{nh_n \tilde{f}_X(x)} - h_n^2 \psi_{\tilde{K}} \rho_\sigma(x) \Big| = O_{\mathrm{p}}(\tilde{r}_n). \tag{53}$$

Thus, by (50), (51), (52) and (53), we have

$$\begin{aligned} &\hat{\sigma}^2_{h_n}(x) - \sigma^2(x) - h_n^2 \psi_{\tilde{K}} \rho_\sigma(x) = \frac{E_n(x)}{nh_n \tilde{f}_X(x)} + O_{\mathrm{p}}(\ell_n), \\ &\text{where} \quad \ell_n = (nh_n^{3/2} b_n)^{-1} + \tilde{r}_n + [1 + \tilde{q}_n + \tilde{\chi}_n(6)]\Delta_n + \chi_n^2(6) + \Delta_n^2. \end{aligned} \tag{54}$$

Applying Lemma 2 (iii) with $h(x) = x^2$ and $q = 3$, we have $\sup_{x\in\mathcal{T}} |E_n(x)| = O_{\mathrm{p}}[nh_n \tilde{\chi}_n(3)]$.

When $h_n \asymp b_n$ and condition (15) is satisfied, Proposition 2 follows from (54) by simplifying $\ell_n + h_n^2 + \tilde{\chi}_n(3)$. The calculations involved are tedious and thus are omitted.

For the proof of Theorem 2, let $\varsigma_n(x) = \sigma^2(x)/\hat{\sigma}^2_{h_n}(x)$. Since $\mathbb{E}(\varepsilon_0) = 0, \mathbb{E}(\varepsilon_0^2) = 1$ and ε_0 has continuous density, we have $\nu_\varepsilon = \mathbb{E}(\varepsilon^4) - 1 > 0$. By (54),

$$\begin{aligned} & \frac{\sqrt{nh_n}}{\sqrt{\varphi_{\tilde{K}}\nu_\varepsilon}} \frac{[\tilde{f}_X(x)]^{1/2}}{\hat{\sigma}^2_{h_n}(x)} [\hat{\sigma}^2_{h_n}(x) - \sigma^2(x) - h_n^2 \psi_{\tilde{K}} \rho_\sigma(x)] \\ & = \varsigma_n(x)\sqrt{\tilde{\omega}_n(x)} \frac{E_n(x)}{\sigma^2(x)\sqrt{nh_n \nu_\varepsilon \varphi_{\tilde{K}} f_X(x)}} + O_{\mathrm{p}}(\sqrt{nh_n}\ell_n). \end{aligned} \quad (55)$$

By Proposition 2 and (16), $\sup_{x\in\mathcal{T}} |\varsigma_n(x) - 1| = o_{\mathrm{p}}(1/\log n)$. Also, it is easy to check that $\sup_{x\in\mathcal{T}} |\tilde{\omega}_n(x) - 1| = o_{\mathrm{p}}(1/\log n)$ and $(nh_n \log n)^{1/2}\ell_n \to 0$. Thus, Theorem 2 follows from Theorem 3 via Slutsky's theorem. $\diamondsuit$

Proof of Proposition 3. As shown in the proof of Proposition 2, we have $\sup_{x\in\mathcal{T}} |\hat{\mu}^*_{b_n} - \mu(x)| = O_{\mathrm{p}}[\Delta_n + \chi_n(6)]$. By (54), we have $\sup_{x\in\mathcal{T}} |\hat{\sigma}^{2*}_{h_n}(x) - \sigma^2(x)| = O_{\mathrm{p}}[\tilde{\chi}_n(3) + \ell_n]$. Therefore,

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i \mathbf{1}_{X_i\in\mathcal{T}} &= \sum_{i=1}^n \Big[\frac{\sigma(X_i)\varepsilon_i + \mu(X_i) - \hat{\mu}^*_{b_n}(X_i)}{\hat{\sigma}^*_{h_n}(X_i)}\Big]^4 \mathbf{1}_{X_i\in\mathcal{T}} \\ &= \sum_{i=1}^n \varepsilon_i^4 \mathbf{1}_{X_i\in\mathcal{T}} + O\{n[\tilde{\chi}_n(3) + \chi_n(6) + \ell_n + \Delta_n]\}. \end{aligned} \quad (56)$$

By the independence of ε_i and $\mathcal{G}_{i-1}$, $\{[\varepsilon_i^4 - \mathbb{E}(\varepsilon_i^4)]\mathbf{1}_{X_i\in\mathcal{T}}\}_{i=1}^n$ form martingale differences with respect to $\mathcal{G}_i$. Since $\varepsilon_i \in \mathcal{L}^6$, $\| \sum_{i=1}^n [\varepsilon_i^4 - \mathbb{E}(\varepsilon_i^4)]\mathbf{1}_{X_i\in\mathcal{T}}\|_{3/2} = O(n^{2/3})$. Furthermore, apply the M/R-decomposition technique in (30), we can show that

$$\begin{aligned} \sum_{i=1}^n [\mathbf{1}_{X_i\in\mathcal{T}} - \mathbb{E}(\mathbf{1}_{X_i\in\mathcal{T}})] &= \sum_{i=1}^n [\mathbf{1}_{X_i\in\mathcal{T}} - \mathbb{E}(\mathbf{1}_{X_i\in\mathcal{T}}|\mathcal{F}_{i-1})] + \sum_{i=1}^n [\mathbb{E}(\mathbf{1}_{X_i\in\mathcal{T}}|\mathcal{F}_{i-1}) - \mathbb{E}(\mathbf{1}_{X_i\in\mathcal{T}})] \\ &= O_{\mathrm{p}}(\sqrt{n} + \Xi_n^{1/2}). \end{aligned} \quad (57)$$

Thus, the desired result follows from (56) and (57) via elementary manipulations. $\diamondsuit$

Acknowledgments. We would like to thank Professor Erich Haeusler for clarifications on martingale moderate deviations. The work is supported in part by NSF grant DMS-0478704.

REFERENCES

Bickel, P.J., and Rosenblatt, M. (1973), "On some global measures of the deviations of density function estimates," *The Annals of Statistics,* 1, 1071–1095.

Cummins, D.J., Filloon, T.G., and Nychka, D. (2001), "Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage," *Journal of the American Statistical Association,* 96, 233–246.

Dümbgen, L. (2003), "Optimal confidence bands for shape-restricted curves," *Bernoulli*, 9, 423–449.

Engle, R.F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation," *Econometrica*, 50, 987–1008.

Eubank, R.L. and Speckman, P.L. (1993), "Confidence bands in nonparametric regression," *Journal of the American Statistical Association,* 88, 1287–1301.

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

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

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized likelihood ratio statistics and Wilks phenomenon," *The Annals of Statistics*, 29, 153–193.

Grama, I.G., and Haeusler, E. (2006), "An Asymptotic Expansion for Probabilities of Moderate Deviations for Multivariate Martingales," *Journal of Theoretical Probability*, 19, 1–44.

Haggan, V., and Ozaki, T. (1981), "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model," *Biometrika*, 68, 189–196.

Hall, P., and Titterington, D.M. (1988), "On confidence bands in nonparametric density estimation and regression," *Journal of Multivariate Analysis*, 27, 228–254.

Härdle, W. (1989), "Asymptotic maximal deviation of M-smoothers," *Journal of Multivariate Analysis*, 29, 163–179.

Härdle, W., and Marron, J.S. (1991), "Bootstrap simultaneous error bars for nonparametric regression," *The Annals of Statistics*, 19, 778–796.

Johnston, G.J. (1982), "Probabilities of maximal deviations for nonparametric regression function estimates," *Journal of Multivariate Analysis*, 12, 402–414.

Knafl, G., Sacks, J., and Ylvisaker, D. (1982), "Model robust confidence intervals," *Journal of Statistical Planning and Inference*, 6, 319–334.

Knafl, G., Sacks, J., and Ylvisaker, D. (1985), "Confidence bands for regression functions," *Journal of the American Statistical Association*, 80, 683–691.

Komlós, J., Major, P., and Tusnády, G. (1975), "An approximation of partial sums of independent RV's and the sample DF. I," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 32, 111–131.

Stine, R.A. (2006), "Nonlinear time series," In *Encyclopedia of Statistical Sciences*, Wiley, 2nd Edition, Edited by S. Kotz, C. B. Read, N. Balakrishnan, and B. Vidakovic, pp. 5581–5588.

Sun, J., and Loader, C.R. (1994), "Simultaneous confidence bands for linear regression and smoothing," *The Annals of Statistics*, 22, 1328–1345.

Tong, H. (1990), *Nonlinear time series analysis: A dynamic approach,* Oxford University Press, Oxford.

Wu, W.B. (2003), "Empirical processes of long-memory sequences," *Bernoulli*, 9, 809–831.

Wu, W.B. (2005), "Nonlinear system theory: Another look at dependence," *Proceedings of the National Academy of Sciences, USA* , 102, 14150–14154.

Wu, W.B. (2006), "Strong invariance principles for dependent random variables," *To appear, Annals of Probability.*

Wu and Zhao (2006), "Inference of trends in time series," *To appear, Journal of the Royal*

Statistical Society, Ser. B.

Xia, Y. (1998), "Bias-corrected confidence bands in nonparametric regression," *Journal of the Royal Statistical Society, Ser. B*, 60, 797–811.

Zhao and Wu (2006), "Kernel quantile regression for nonlinear stochastic models," *Technical report*, Department of Statistics, University of Chicago.

Table 1. Coverage probabilities of SCB for μ in Model 1.

b_n	$\mathcal{T}_{(30)}$		$\mathcal{T}_{(50)}$		$\mathcal{T}_{(100)}$	
	$\hat{q}_{0.95}$	coverage	$\hat{q}_{0.95}$	coverage	$\hat{q}_{0.95}$	coverage
0.03	0.195	0.9424	0.257	0.9406	0.319	0.9368
0.04	0.202	0.9422	0.236	0.9440	0.304	0.9406
0.05	0.191	0.9445	0.228	0.9496	0.290	0.9412
0.06	0.179	0.9440	0.222	0.9510	0.276	0.9446
0.07	0.174	0.9444	0.218	0.9517	0.261	0.9469
0.08	0.172	0.9434	0.212	0.9485	0.247	0.9451
0.09	0.169	0.9447	0.206	0.9478	0.234	0.9448
0.10	0.168	0.9467	0.200	0.9485	0.224	0.9457
0.12	0.163	0.9450	0.188	0.9478	0.208	0.9492
0.14	0.159	0.9474	0.178	0.9501	0.194	0.9504
0.16	0.152	0.9473	0.169	0.9494	0.183	0.9527
0.18	0.148	0.9477	0.160	0.9470	0.173	0.9521
0.20	0.143	0.9485	0.156	0.9537	0.163	0.9472

Table 2. Coverage probabilities of SCB for σ^2 in Model 2.

	$\mathcal{T}_{(30)}$		$\mathcal{T}_{(50)}$		$\mathcal{T}_{(100)}$	
h_n	$\hat{q}_{0.95}$	coverage	$\hat{q}_{0.95}$	coverage	$\hat{q}_{0.95}$	coverage
0.03	0.354	0.9458	0.474	0.9491	0.934	0.9397
0.04	0.306	0.9443	0.443	0.9460	0.801	0.9399
0.05	0.293	0.9431	0.434	0.9464	0.663	0.9403
0.06	0.282	0.9410	0.415	0.9474	0.581	0.9469
0.07	0.277	0.9425	0.392	0.9480	0.517	0.9506
0.08	0.272	0.9453	0.365	0.9486	0.463	0.9515
0.09	0.266	0.9466	0.341	0.9500	0.400	0.9453
0.10	0.257	0.9476	0.321	0.9506	0.381	0.9501
0.12	0.247	0.9483	0.278	0.9445	0.312	0.9449
0.14	0.229	0.9504	0.252	0.9484	0.286	0.9546
0.16	0.211	0.9525	0.230	0.9494	0.252	0.9511
0.18	0.203	0.9559	0.215	0.9535	0.240	0.9591
0.20	0.192	0.9561	0.202	0.9578	0.213	0.9560

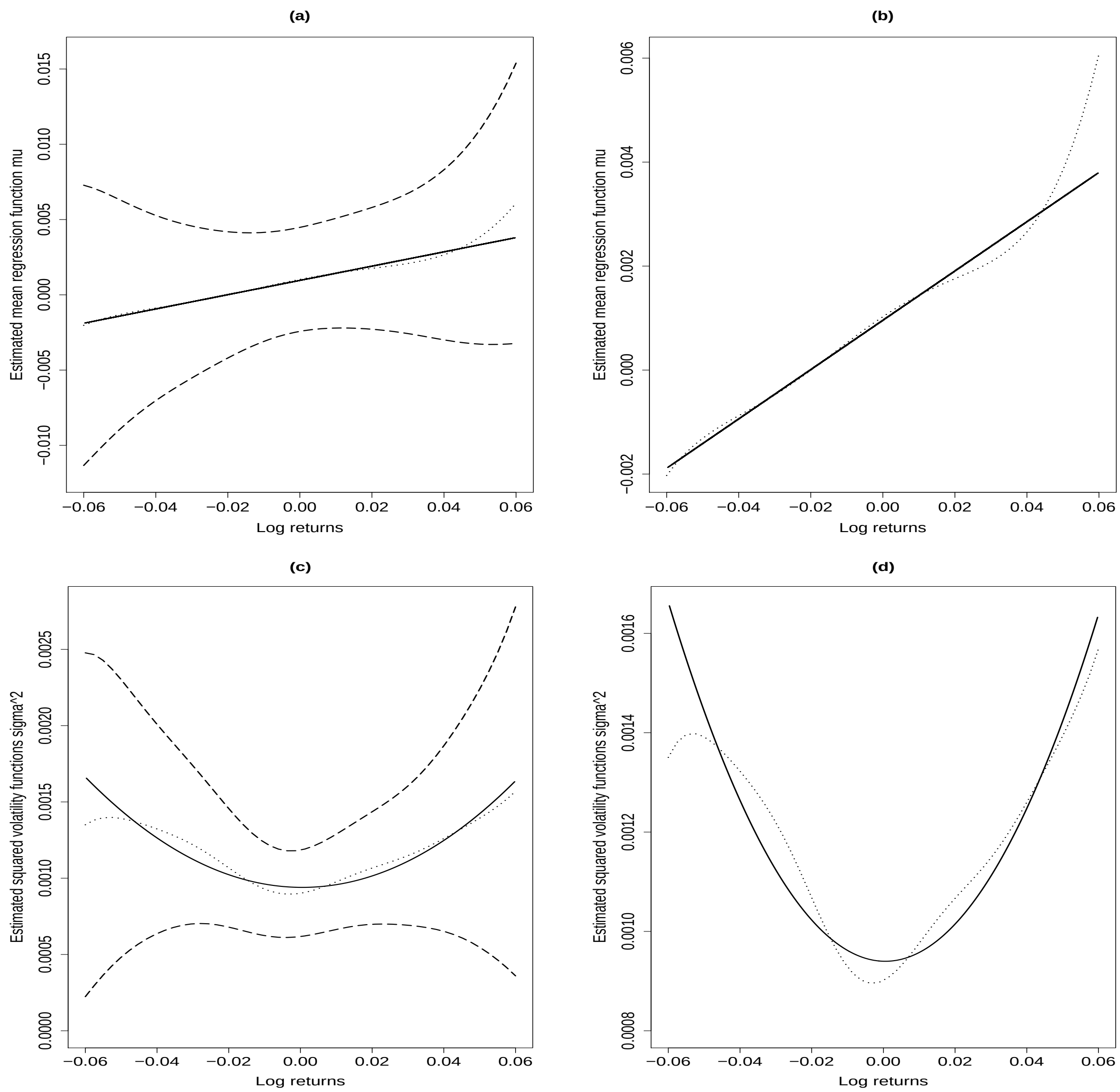

Figure 1: (a): SCB for regression function μ; dotted, long-dashed and solid lines are the estimated curve $\hat{\mu}_{b_n}$, SCB and the fitted linear curve $\hat{\mu}_{\text{linear}}(x) = 0.000958 + 0.0474x$, respectively. (b): Zoom-in version of $\hat{\mu}_{b_n}$ and $\hat{\mu}_{\text{linear}}$ in (a). (c): SCB for squared volatility function σ^2; dotted, long-dashed and solid lines are the estimated curve $\hat{\sigma}^2_{h_n}$, SCB and the fitted quadratic curve $\hat{\sigma}^2_{\text{quadratic}}(x) = 0.00094 - 0.000205x + 0.197367x^2$, respectively. (d): Zoom-in version of $\hat{\sigma}^2_{h_n}$ and $\hat{\sigma}^2_{\text{quadratic}}$ in (c).