The Annals of Statistics
2010, Vol. 38, No. 4, 2388–2421
DOI: 10.1214/09-AOS789

SIMULTANEOUS NONPARAMETRIC INFERENCE OF TIME SERIES[1]

BY WEIDONG LIU AND WEI BIAO WU

University of Pennsylvania and University of Chicago

We consider kernel estimation of marginal densities and regression functions of stationary processes. It is shown that for a wide class of time series, with proper centering and scaling, the maximum deviations of kernel density and regression estimates are asymptotically Gumbel. Our results substantially generalize earlier ones which were obtained under independence or beta mixing assumptions. The asymptotic results can be applied to assess patterns of marginal densities or regression functions via the construction of simultaneous confidence bands for which one can perform goodness-of-fit tests. As an application, we construct simultaneous confidence bands for drift and volatility functions in a dynamic short-term rate model for the U.S. Treasury yield curve rates data.

1. Introduction.

Consider the nonparametric time series regression model

$$Y_i = \mu(X_i)\,dt + \sigma(X_i)\eta_i, \tag{1.1}$$

where $\mu(\cdot)$ [resp., $\sigma^2(\cdot)$] is an unknown regression (resp., conditional variance) function to be estimated, (X_i, Y_i) is a stationary process and η_i are unobserved independent and identically distributed (i.i.d.) errors with $\mathsf{E}\eta_i = 0$ and $\mathsf{E}\eta_i^2 = 1$. Let the regressor X_i be a stationarity causal process

$$X_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{1.2}$$

where ε_i are i.i.d. and the function G is such that X_i exists. Assume that η_i is independent of $(\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Hence, η_i and $(\mu(X_i), \sigma(X_i))$ are independent. As a special case of (1.1), a particularly interesting example is the nonlinear autoregressive model

$$Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\eta_i, \tag{1.3}$$

where $X_i = Y_{i-1}$ and $\varepsilon_i = \eta_{i-1}$. Many nonlinear time series models are of form (1.3) with different choices of $\mu(\cdot)$ and $\sigma(\cdot)$. If the form of $\mu(\cdot)$ is not known, we

Received August 2009; revised December 2009.
[1]Supported in part by NSF Grant DMS-04-78704.
AMS 2000 subject classifications. Primary 62H15; secondary 62G10.
Key words and phrases. Gumbel distribution, kernel density estimation, linear process, maximum deviation, nonlinear time series, nonparametric regression, simultaneous confidence band, stationary process, treasury bill data.

$\hat{E}[Y|X=x] = \frac{\sum K_h(x-x_i)Y_i}{\sum K_h(x-x_i)}$ where $\hat f(x,y) = \frac{1}{n}\sum_{i=1}^{n} K_h(x_i-x)\, K_h(y-y_i)$

$\hat f(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x-x_i)$

can use the Nadaraya–Watson estimator

$$\mu_n(x) = \frac{1}{nbf_n(x)} \sum_{k=1}^{n} K\left(\frac{X_k - x}{b}\right) Y_k, \tag{1.4}$$

where K is a kernel function with $K(\cdot) \geq 0$ and $\int_{\mathbb{R}} K(u)\,du = 1$, the bandwidths $b = b_n \to 0$ and $nb_n \to \infty$, and

$$f_n(x) = \frac{1}{nb} \sum_{k=1}^{n} K\left(\frac{X_k - x}{b}\right)$$

$X_i = G(\ldots \xi_{i-1}, \xi_i)$

is the kernel density estimate of f, the marginal density of X_i. Asymptotic properties of nonparametric estimates for time series have been widely discussed under various strong mixing conditions; see Robinson (1983), Györfi et al. (1989), Tjøstheim (1994), Bosq (1996), Doukhan and Louhichi (1999) and Fan and Yao (2003), among others.

Under appropriate dependence conditions [see, e.g., Robinson (1983), Wu and Mielniczuk (2002), Fan and Yao (2003) and Wu (2005)], we have the central limit theorem

$$\sqrt{nb}[f_n(x) - \mathsf{E} f_n(x)] \Rightarrow N(0, \lambda_K f(x)) \qquad \text{where } \lambda_K = \int_{\mathbb{R}} K^2(u)\,du.$$

The above result can be used to construct point-wise confidence intervals of $f(x)$ at a fixed x. To assess shapes of density functions so that one can perform goodness-of-fit tests, however, one needs to construct *uniform or simultaneous confidence bands* (SCB). To this end, we need to deal with the maximum absolute deviation over some interval $[l, u]$:

$$\Delta_n := \sup_{l \leq x \leq u} \frac{\sqrt{nb}}{\sqrt{\lambda_K f(x)}} |f_n(x) - \mathsf{E} f_n(x)|. \tag{1.5}$$

In an influential paper, Bickel and Rosenblatt (1973) obtained an asymptotic distributional theory for Δ_n under the assumption that X_i are i.i.d. It is a very challenging problem to generalize their result to stationary processes where dependence is the rule rather than the exception. In their paper Bickel and Rosenblatt applied the very deep embedding theorem of approximating empirical processes of independent random variables by Brownian bridges with a reasonably sharp rate [Brillinger (1969), Komlós, Major and Tusnády (1975, 1976)]. For stationary processes, however, such an approximation with similar rates can be extremely difficult to obtain. Doukhan and Portal (1987) obtained a weak invariance principle for empirical distribution functions. In 1998, Neumann (1998) made a breakthrough and proved a very useful result for β-mixing processes whose mixing rates decay exponentially quickly. Such processes are very weakly dependent. For mildly weakly dependent processes, the asymptotic problem of Δ_n remains open. Fan and Yao [(2003), page 208] conjectured that similar results hold for stationary processes under certain mixing conditions. Here we shall solve this open problem and establish an asymptotic theory for both short- and long-range dependent processes. It is shown that, for a wide class of short-range dependent processes, we can have a similar asymptotic distributional theory as Bickel and Rosenblatt (1973). However, for long-range dependent processes, the asymptotic behavior can be sharply different. One observes the dichotomy phenomenon: the asymptotic properties depend on the interplay between the strength of dependence and the size of bandwidths. For small bandwidths, the limiting distribution is the same as the one under independence. If the bandwidths are large, then the limiting distribution is half-normal [cf. (2.9)].

A closely related problem is to study the asymptotic uniform distributional theory for the Nadaraya–Watson estimator $\mu_n(x)$. Namely, one needs to find the asymptotic distribution for $\sup_{x\in T}|\mu_n(x)-\mu(x)|$, where $T=[l,u]$. With the latter result, one can construct an asymptotic $(1-\alpha)$ SCB, $0<\alpha<1$, by finding two functions $\mu_n^{\text{lower}}(x)$ and $\mu_n^{\text{upper}}(x)$, such that

$$\lim_{n\to\infty} \mathsf{P}\big(\mu_n^{\text{lower}}(x) \le \mu(x) \le \mu_n^{\text{upper}}(x) \text{ for all } x\in T\big) = 1-\alpha. \tag{1.6}$$

The SCB can be used for model validation: one can test whether $\mu(\cdot)$ is of certain parametric functional form by checking whether the fitted parametric form lies in the SCB. Following the work of Bickel and Rosenblatt (1973), Johnston (1982) derived the asymptotic distribution of $\sup_{0\le x\le 1}|\mu_n(x)-\mathsf{E}[\mu_n(x)]|$, assuming that (X_i, Y_i) are independent random samples from a bivariate population. Johnston's derivation is no longer valid if dependence is present. For other work on regression confidence bands under independence see Knafl, Sacks and Ylvisaker (1985), Hall and Titterington (1988), Härdle and Marron (1991), Sun and Loader (1994), Xia (1998), Cummins, Filloon and Nychka (2001) and Dümbgen (2003), among others. Recently Zhao and Wu (2008) proposed a method for constructing SCB for stochastic regression models which have asymptotically correct coverage probabilities. However, their confidence band is over an increasingly dense grid of points instead of over an interval [see also Bühlmann (1998) and Knafl, Sacks and Ylvisaker (1985)]. Here we shall also solve the latter problem and establish a uniform asymptotic theory for the regression estimate $\mu_n(x)$, so that one can construct a genuine SCB for regression functions. A similar result will be derived for $\sigma(\cdot)$ as well.

The rest of the paper is organized as follows. Main results are presented in Section 2. Proofs are given in Sections 4 and 5. Our results are applied in Section 3 to the U.S. Treasury yield rates data.

2. Main results.

Before stating our theorems, we first introduce dependence measures. Assume $X_k\in\mathcal{L}^p$, $p>0$. Here for a random variable W, we write $W\in\mathcal{L}^p$ $(p>0)$, if $\|W\|_p := (\mathsf{E}|W|^p)^{1/p} < \infty$. Let $\{\varepsilon'_j\}_{j\in\mathbb{Z}}$ be an i.i.d. copy of $\{\varepsilon_j\}_{j\in\mathbb{Z}}$; let $\xi_n = (\ldots,\varepsilon_{n-1},\varepsilon_n)$ and

$$X'_n = G(\xi'_n) \qquad \text{where } \xi'_n = (\xi_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_n).$$

$X_i = G(\xi \ldots)$

Here X_n' is a coupled process of X_n with ε_0 in the latter replaced by an i.i.d. copy ε_0'. Following Wu (2005), define the physical dependence measure

$$\theta_{n,p} = \|X_n - X_n'\|_p.$$

Let $\theta_{n,p} = 0$ if $n < 0$. A similar quantity can be defined if we couple the whole past: let $\xi_{k,n}^\star = (\ldots, \varepsilon_{k-n-2}', \varepsilon_{k-n-1}', \xi_{k-n,k})$, $k \geq n$, where $\xi_{i,j} = (\varepsilon_i, \varepsilon_{i+1}, \ldots, \varepsilon_j)$, and define

$$\Psi_{n,p} = \|G(\xi_n) - G(\xi_{n,n}^\star)\|_p. \tag{2.1}$$

Our conditions on dependence will be expressed in terms of $\theta_{n,p}$ and $\Psi_{n,p}$.

2.1. *Kernel density estimates.* We first consider a special case of (1.2) in which X_n has the form

$$X_n = a_0\varepsilon_n + g(\ldots, \varepsilon_{n-2}, \varepsilon_{n-1}) = a_0\varepsilon_n + g(\xi_{n-1}), \tag{2.2}$$

$X_i := G(\cdots \xi_{i-1}, \xi_i)$

where g is a measurable function and $a_0 \neq 0$. Then the coupled process $X_n' = a_0\varepsilon_n + g(\xi_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_{n-1})$. We need the following conditions:

(C1). There exists $0 < \delta_2 \leq \delta_1 < 1$ such that $n^{-\delta_1} = O(b_n)$ and $b_n = O(n^{-\delta_2})$. n

(C2). Suppose that $X_1 \in \mathcal{L}^p$ for some $p > 0$. Let $p' = \min(p, 2)$ and $\Theta_n = \sum_{i=0}^n \theta_{i,p'}^{p'/2}$. Assume $\Psi_{n,p'} = O(n^{-\gamma})$ for some $\gamma > \delta_1/(1-\delta_1)$ and

b_n

$$\mathcal{Z}_n bn^{-1} = o(\log n) \qquad \text{where } \mathcal{Z}_n = \sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2. \tag{2.3}$$

(C3). The density function f_ε of ε_1 is positive and

$$\sup_{x \in \mathsf{R}} [f_\varepsilon(x) + |f_\varepsilon'(x)| + |f_\varepsilon''(x)|] < \infty.$$

(C4). The support of K is $[-A, A]$, where K is differentiable over $(-A, A)$, the right (resp., left) derivative $K'(-A)$ [resp., $K'(A)$] exists, and $\sup_{|x| \leq A} |K'(x)| < \infty$. The Lebesgue measure of the set $\{x \in [-A, A] : K(x) = 0\}$ is zero. Let $\lambda_K = \int K^2(y)\,dy$, $K_1 = [K^2(-A) + K^2(A)]/(2\lambda_K)$ and $K_2 = \int_{-A}^A (K'(t))^2\,dt/(2\lambda_K)$.

THEOREM 2.1. *Let $l, u \in \mathsf{R}$ be fixed and X_n be of form* (2.2). *Assume* (C1)–(C4). *Then we have for every $z \in \mathsf{R}$,*

$$\mathsf{P}((2\log \bar{b}^{-1})^{1/2}(\Delta_n - d_n) \leq z) \to e^{-2e^{-z}}, \tag{2.4}$$

where $\bar{b} = b/(u-l)$,

$$d_n = (2\log \bar{b}^{-1})^{1/2} + \frac{1}{(2\log \bar{b}^{-1})^{1/2}} \left\{ \log \frac{K_1}{\pi^{1/2}} + \frac{1}{2} \log\log \bar{b}^{-1} \right\},$$

if $K_1 > 0$, and otherwise

$$d_n = (2\log \bar{b}^{-1})^{1/2} + \frac{1}{(2\log \bar{b}^{-1})^{1/2}} \log \frac{K_2^{1/2}}{2^{1/2}\pi}.$$

We now discuss conditions (C1)–(C4). The bandwidth condition (C1) is fairly mild. In (C2), the quantity Θ_n measures the cumulative dependence of $X_0, \dots, X_n$ on ε_0, and, with (C1), it gives sufficient dependence and bandwidth conditions for the asymptotic Gumbel convergence (2.4). For short-range dependent linear process $X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j}$ with $\mathsf{E}\varepsilon_1 = 0$ and $\mathsf{E}\varepsilon_1^2 = 1$, (C2) is satisfied if $\sum_{j=0}^{\infty} |a_j| < \infty$ and $\sum_{j=n}^{\infty} a_j^2 = O(n^{-\gamma})$ for some $\gamma > 2\delta_1/(1-\delta_1)$. The latter condition can be weaker than $\sum_{j=0}^{\infty} |a_j| < \infty$ if $\delta_1 < 1/3$. Interestingly, (C2) also holds for some long-range dependent processes; see Theorem 2.3. With (C3), it is easily seen that X_i does have a density. If (C3) is violated, then X_i may not have a density. For example, if ε_i are i.i.d. Bernoulli with $\mathsf{P}(\varepsilon_i = 0) = \mathsf{P}(\varepsilon_i = 1) = 1/2$, then $X_0 = \sum_{i=0}^{\infty} \rho^i \varepsilon_{-i}$, where $\rho = (\sqrt{5}-1)/2$, does not have a density [Erdös (1939)]. The kernel condition (C4) is quite mild and it is satisfied by many popular kernels. For example, it holds for the Epanechnikov kernel $K(u) = 0.75(1-u^2)\mathbf{1}_{|u|\le 1}$.

In Theorem 2.2 below, we do not assume the special form (2.2). We need regularity conditions on conditional density functions. For jointly distributed random vectors ξ and η, let $F_{\eta|\xi}(\cdot)$ be the conditional distribution function of η given ξ; let $f_{\eta|\xi}(x) = \partial F_{\eta|\xi}(x)/\partial x$ be the conditional density. For function g with $\mathsf{E}|g(\eta)| < \infty$, let $\mathsf{E}(g(\eta)|\xi) = \int g(x)\, dF_{\eta|\xi}(x)$ be the conditional expectation of $g(\eta)$ given ξ.

Conditions (C2) and (C3) are replaced, respectively, by:

(C2)$'$. Suppose that $X_1 \in \mathcal{L}^p$ and $\theta_{n,p} = O(\rho^n)$ for some $p > 0$ and $0 < \rho < 1$.

(C3)$'$. The density function f is positive and there exists a constant $B < \infty$ such that

$\sup[f_\xi(x) + |f'_\xi(x)| + |f''_\xi(x)|] < \infty$

$$\sup_x [|f_{X_n|\xi_{n-1}}(x)| + |f'_{X_n|\xi_{n-1}}(x)| + |f''_{X_n|\xi_{n-1}}(x)|] \le B \qquad \text{almost surely.}$$

THEOREM 2.2. *Under* (C1), (C2)$'$, (C3)$'$ *and* (C4), *we have* (2.4).

Many nonlinear time series models (e.g., ARCH models, bilinear models, exponential AR models) satisfy (C2)$'$; see Shao and Wu (2007). If (X_i) is a Markov chain of the form $X_i = R(X_{i-1}, \varepsilon_i)$, where $R(\cdot,\cdot)$ is a bivariate measurable function, then $f_{X_i|\xi_{i-1}}(\cdot)$ is the conditional density of X_i given X_{i-1}. Consider the ARCH model $X_i = \varepsilon_i(a^2 + b^2 X_{i-1}^2)^{1/2}$, where $a > 0, b > 0$ are real parameters and ε_i has density function f_ε, then $f_{X_i|X_{i-1}}(x) = f_\varepsilon(x/H_i)/H_i$, where $H_i = (a^2 + b^2 X_{i-1}^2)^{1/2}$. So (C3)$'$ holds if $\sup_x [f_\varepsilon(x) + |f'_\varepsilon(x)| + |f''_\varepsilon(x)|] < \infty$ [cf. (C3)]. For more general ARCH-type processes see Doukhan, Madre and Rosenbaum (2007).

For short-range dependent processes for which

$$\Theta_\infty = \sum_{i=0}^{\infty} \theta_{i,p'}^{p'/2} < \infty, \tag{2.5}$$

we have $\mathcal{Z}_n = O(n)$ and (2.3) of condition (C2) trivially holds. For long-range dependent processes, (2.5) can be violated. A popular model for long-range dependence is the fractionally integrated auto-regressive moving average process [Granger and Joyeux (1980), Hosking (1981)]. Here we consider the more general form of linear processes with slowly decaying coefficients:

$$X_n = \sum_{j=0}^{\infty} a_j \varepsilon_{n-j} \quad \text{where } a_j = j^{-\beta}\ell(j), 1/2 < \beta < 1. \tag{2.6}$$

Here $a_0 = 1$, $\ell(\cdot)$ is a slowly varying function and ε_i are i.i.d. with $\mathsf{E}\varepsilon_i = 0$ and $\mathsf{E}\varepsilon_i^2 = 1$.

THEOREM 2.3. *Assume* (2.6). *Let $l, u \in \mathsf{R}$ be fixed.* (i) *Assume* (C1), (C3), (C4), $\delta_1/(1-\delta_1) < \beta - 1/2$ *and*

$$b_n^{1/2} n^{1-\beta}\ell(n) = o(\log^{-1/2} n). \tag{2.7}$$

① < ② Small b

Then (2.4) *holds.* (ii) *Assume* (C1), (C3), (C4), $\sup_x |f_\varepsilon'''(x)| < \infty$ *and*

$$\log^{1/2} n = o(b_n^{1/2} n^{1-\beta}\ell(n)). \tag{2.8}$$

① > ② large b

Let $c_\beta = \int_0^\infty (x + x^2)^{-\beta}\, dx / [(3-2\beta)(1-\beta)]$. *Then*

$$\frac{\Delta_n}{b_n^{1/2} n^{1-\beta}\ell(n)} \Rightarrow |N(0,1)| \frac{\sqrt{c_\beta}}{\sqrt{\lambda_K}} \max_{l \le x \le u} \frac{|f'(x)|}{\sqrt{f(x)}}. \tag{2.9}$$

half-normal

Theorem 2.3 reveals the interesting dichotomy phenomenon for the maximum deviation Δ_n: if the bandwidth b_n is small such that (2.7) holds, then the asymptotic distribution is the same as the one under short-range dependence. However, if b_n is large, then both the normalizing constant and the asymptotic distribution change. Let $b_n = n^{-\delta}\ell_1(n)$, where ℓ_1 is another slowly varying function. Simple algebra shows that, if $\max((1+\delta)/(1-\delta), 2-\delta) < 2\beta$, then the bandwidth condition in Theorem 2.3(i) holds. The latter inequality requires $\beta > \sqrt{3}/2 = 0.866025, \ldots$. If $\beta < 1 - \delta/2$, then (2.8) holds. Theorem 2.3(ii) is similar to Theorem 3.1 in Ho and Hsing (1996), with our result having a wider range of β.

2.2. *Estimation of $\mu(\cdot)$ and $\sigma^2(\cdot)$.* Let $\tilde{\xi}_i = (\ldots, \eta_{i-1}, \eta_i, \xi_i)$. For a function h with $\mathsf{E}h^2(\eta_i) < \infty$, write

ξ_i

$$M_n^r(x) = \frac{1}{nb}\sum_{k=1}^{n} K\left(\frac{X_k - x}{b}\right) Z_k \quad \text{where } Z_k = h(\eta_k) - \mathsf{E}h(\eta_k).$$

σ_n

PROPOSITION 2.1. *Let $l, u \in \mathsf{R}$ be fixed. Assume $\sigma^2 = \mathsf{E}Z_1^2$ and $\mathsf{E}|Z_1|^p < \infty$, $p > 2/(1-\delta_1)$.* (i) *Assume* (2.2), (C1), (C3)–(C4) *and $\Psi_{n,q} = O(n^{-\gamma})$ for some $q > 0$ and $\gamma > \delta_1/(1-\delta_1)$. Then for all $z \in \mathsf{R}$,*

$$\mathsf{P}\left(\sqrt{\frac{nb}{\lambda_K}} \sup_{l \le x \le u} \frac{|M_n^r(x)|}{f^{1/2}(x)\sigma} - d_n \le \frac{z}{(2\log \bar{b}^{-1})^{1/2}}\right) \to e^{-2e^{-z}} \tag{2.10}$$

$X_i = a_0 \varepsilon_0 + g_C$

$X = G(\xi, \cdots)$

as $n \to \infty$. (ii) *Assume* (1.2), (C1), (C2)$'$, (C3)$'$ *and* (C4) *hold with* ξ_{n-1} *in* (C2)$'$ *replaced by* $\tilde{\xi}_{n-1}$. *Then* (2.10) *holds*.

Proposition 2.1(i) allows for long-range dependent processes. For (2.6), by Karamata's theorem, $\Psi_{n,2} = O(n^{1/2-\beta}\ell(n))$. So we have $\Psi_{n,2} = O(n^{-\gamma})$ with $\gamma > \delta_1/(1-\delta_1)$ if $\delta_1 < (2\beta-1)/(2\beta+1)$.

For $S \subset \mathsf{R}$, denote by $\mathcal{C}^p(S) = \{g(\cdot) : \sup_{x \in S} |g^{(k)}(x)| < \infty, k = 0, \ldots, p\}$ the set of functions having bounded derivatives on S up to order $p \geq 1$. Let $S^\epsilon = \bigcup_{y \in S}\{x : |x-y| \leq \epsilon\}$ be the ϵ-neighborhood of S, $\epsilon > 0$.

THEOREM 2.4. *Let* $l, u \in \mathsf{R}$ *be fixed and* K *be symmetric. Assume that the conditions in Proposition* 2.1 *hold with* $Z_n = \eta_n$, $f_\varepsilon(\cdot), \mu(\cdot) \in \mathcal{C}^4(T^\epsilon)$ *for some* $\epsilon > 0$, *where* $T = [l, u]$, *and that* b *satisfies*

$$0 < \delta_1 < 1/3, \qquad nb^9 \log n = o(1) \quad \text{and} \quad \mathcal{Z}_n b^3 = o(n \log n).$$

Let $\psi_K = \int u^2 K(u)\,du/2$ *and* $\rho_\mu(x) = \mu''(x) + 2\mu'(x) f'(x)/f(x)$. *Then*

$$\mathsf{P}\left(\sqrt{\frac{nb}{\lambda_K}} \sup_{l \leq x \leq u} \frac{\sqrt{f_n(x)}|\mu_n(x) - \mu(x) - b^2\psi_K \rho_\mu(x)|}{\sigma(x)} - d_n \leq \frac{z}{(2\log \bar{b}^{-1})^{1/2}}\right) \to e^{-2e^{-z}}. \tag{2.11}$$

$\mu_n = \mu_n(x) - b^2 \psi_K \rho(x)$

Note that $\sigma^2(x) = \mathsf{E}[(Y_k - \mu(X_k))^2 | X_k = x]$. It is natural to use the Nadaraya–Watson method to estimate $\sigma^2(x)$ based on the residuals $\hat{e}_k = Y_k - \mu_n(X_k)$:

$$\sigma_n^2(x) = \frac{1}{nhf_{n1}(x)} \sum_{k=1}^n K\left(\frac{X_k - x}{h}\right)[Y_k - \mu_n(X_k)]^2,$$

where the bandwidths $h = h_n \to 0$ and $nh_n \to \infty$, and

$$f_{n1}(x) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{X_k - x}{h}\right).$$

THEOREM 2.5. *Let* $l, u \in \mathsf{R}$ *be fixed and* K *be symmetric. Assume* $\nu_\eta = \mathsf{E}\eta_1^4 - 1 < \infty$. *Further assume that the conditions in Proposition* 2.1 *hold with* $Z_n = \eta_n^2 - 1$, $f(\cdot), \sigma(\cdot) \in \mathcal{C}^4(T^\epsilon)$ *for some* $\epsilon > 0$, *where* $T = [l, u]$, *and that* $h \asymp b$ *satisfies*

$$0 < \delta_1 < 1/4, \qquad nb^9 \log n = o(1)$$

and

$$\mathcal{Z}_n b^3 = o(n \log n).$$

Let $\rho_\sigma(x) = 2\sigma'^2(x) + 2\sigma(x)\sigma''(x) + 4\sigma(x)\sigma'(x)f'(x)/f(x)$. *Then*

$$\mathsf{P}\left(\sqrt{\frac{nh}{\lambda_K \nu_\eta}} \sup_{l \le x \le u} \frac{\sqrt{f_{n1}(x)}|\sigma_n^2(x) - \sigma^2(x) - h^2\psi_K\rho_\sigma(x)|}{\sigma^2(x)} - d_n \le \frac{z}{(2\log \bar{h}^{-1})^{1/2}}\right) \to e^{-2e^{-z}}, \tag{2.12}$$

where d_n *is defined as in Theorem* 2.1 *by replacing* $\bar{b}$ *with* $\bar{h} = h/(u-l)$.

We now compare the SCBs constructed based on Theorem 1 in Zhao and Wu (2008) and Theorem 2.4. Assume $l = 0$ and $u = 1$. The former is over the grid point $T_n = \{2b_n j, j = 0, 1, \ldots, J_n\}$ with $J_n = \lceil 1/(2b_n) \rceil$, while the latter is a genuine SCB in the sense that it is over the whole interval $T = [0, 1]$. Let $\hat{\rho}_\mu(\cdot)$ [resp., $\hat{\sigma}(\cdot)$] be a consistent estimate of $\rho_\mu(\cdot)$ [resp., $\sigma(\cdot)$] and $z_\alpha = -\log\log(1-\alpha)^{-1/2}$, $0 < \alpha < 1$. By Theorem 2.4, we can construct the $1-\alpha$ SCB for $\mu(x)$ over $x \in [0, 1]$ as

$$\mu_n(x) - b^2\psi_K\hat{\rho}_\mu(x) \pm l_1\hat{\sigma}(x)\sqrt{\frac{\lambda_K}{nbf_n(x)}} \quad \text{where } l_1 = \frac{z_\alpha}{(2\log b^{-1})^{1/2}} + d_n. \tag{2.13}$$

Similarly, using Theorem 1 in Zhao and Wu (2008), the $1-\alpha$ confidence band for $\mu(x)$ over $x \in T_n$ is also of form (2.13) with l_1 replaced by

$$l_2 = \frac{z_\alpha}{(2\log J_n)^{1/2}} + (2\log J_n)^{1/2} - \frac{1/2\log\log J_n + \log(2\sqrt{\pi})}{(2\log J_n)^{1/2}}.$$

Elementary calculations show that, interestingly, l_1 and l_2 are quite close: $l_1 - l_2 = (\log\log b^{-1})/(2\log b^{-1})^{1/2}(1 + o(1))$ if $K_1 > 0$.

3. Application to the treasury bill data. There is a huge literature on models for short-term interest rates. Let R_t be the interest rate at time t. Assume that R_t follows the diffusion model

$$dR_t = \mu(R_t)\,dt + \sigma(R_t)\,d\mathbb{B}(t), \tag{3.1}$$

where $\mathbb{B}$ is the standard Brownian motion, $\mu(\cdot)$ is the instantaneous return or drift function and $\sigma(\cdot)$ is the volatility function. Black and Scholes (1973) considered the model with $\mu(x) = \alpha x$ and $\sigma(x) = \sigma x$. Vasicek (1977) assumed that $\mu(x) = \alpha_0 + \alpha_1 x$ and $\sigma(x) \equiv \sigma$, where α_0, α_1 and σ are unknown constants. Cox, Ingersoll and Ross (1985) and Courtadon (1982) assumed that $\sigma(x) = \sigma x^{1/2}$ and $\sigma(x) = \sigma x$, respectively. Both models are generalized by Chan et al. (1992) to the form $\sigma(x) = \sigma x^\gamma$, with σ and γ being unknown parameters. Stanton (1997), Fan and

$3.1 \quad dR_t = \mu(R_t)dt + \sigma(R_t)dB(t)$

Yao (1998), Chapman and Pearson (2000) and Fan and Zhang (2003) considered the nonparametric estimation of $\mu(\cdot)$ and $\sigma(\cdot)$ in (3.1); see also Aït-Sahalia (1996a, 1996b). Stanton (1997) constructed *point-wise* confidence intervals which serve as a tool for suggesting which parametric models to use. Zhao (2008) gave an excellent review of parametric and nonparametric approaches of (3.1). See also the latter paper for further references.

Here we shall consider the U.S. six-month treasury yield rates data from January 2nd, 1990 to July 31st, 2009. The data can be downloaded from the U.S. Treasury department's website http://www.ustreas.gov/. It has 4900 daily rates and a plot is given in Figure 1. Let $X_i = R_{t_i}$ be the rate at day $i = 1, \dots, 4900$. For the daily data, since one year has 250 transaction days, $t_i - t_{i-1} = 1/250$. Let $\Delta = 1/250$. As a discretized version of (3.1), we consider the model

$dB(t)$

$$Y_i = \mu(X_i)\Delta + \sigma(X_i)\Delta^{1/2}\eta_i, \tag{3.2}$$

where $Y_i = R_{t_{i+1}} - R_{t_i} = X_{i+1} - X_i$ and $\eta_i = (\mathbb{B}(t_{i+1}) - \mathbb{B}(t_i))/\Delta^{1/2}$ are i.i.d. standard normal. For convenience of applying Theorem 2.4, in the sequel we shall write $\mu(X_i)\Delta$ [resp., $\sigma(X_i)\Delta^{1/2}$] in (3.2) as $\mu(X_i)$ [resp., $\sigma(X_i)$]. So (3.2) is rewritten as

$$Y_i = \mu(X_i) + \sigma(X_i)\eta_i. \tag{3.3}$$

FIG. 1. *U.S. six-month treasury yield curve rates data from January* 2*nd*, 1990 *to July* 31*st*, 2009. *Source*: *U.S. Treasury department's website* *http://www.ustreas.gov/*.

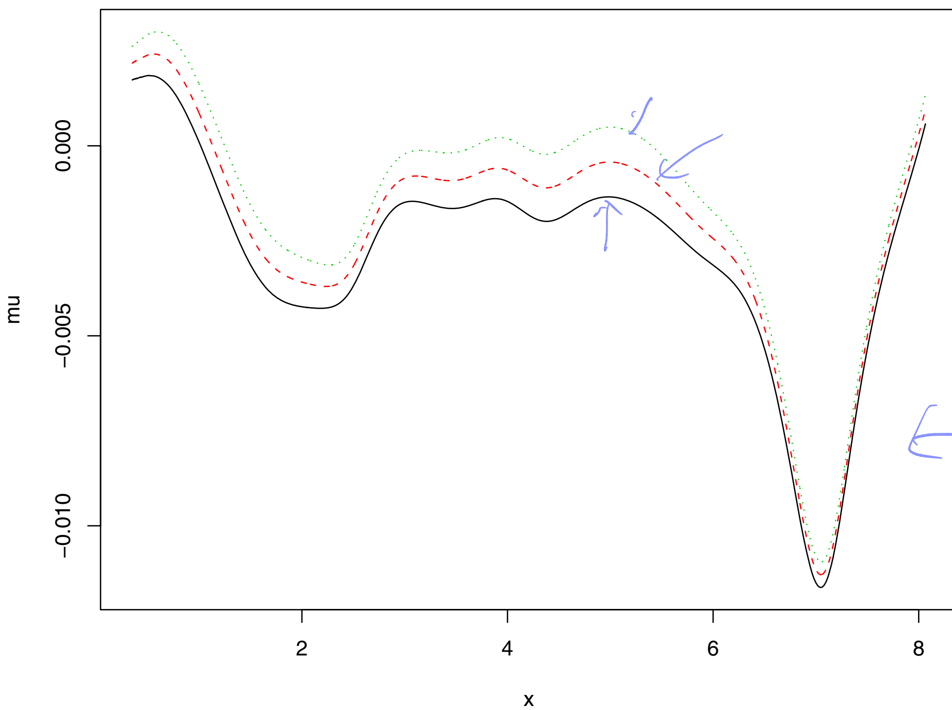

FIG. 2. *95% SCB of the regression function $\mu(\cdot)$ over the interval $[l, u] = [0.35, 8.06]$. The dashed curve in the middle is $\mu_n(x) - b^2\psi_K\hat{\rho}(x)$, the bias-corrected estimate of μ.*

Figure 2 shows the estimated 95% simultaneous confidence band for the regression function $\mu(\cdot)$ over the interval $T = [l, u] = [0.35, 8.06]$, which includes 96% of the daily rates X_i. To select the bandwidth, we use the R program `bw.nrd` which gives $b = 0.37$. Then we use the R program `locpoly` for local polynomial regression. The Nadaraya–Watson estimate is a special case of the local polynomial regression with degree 0. The function $\rho(x)$ in the bias term $b^2\psi_K\rho(x)$ in Theorem 2.4 involves the first and second order derivatives μ', f' and μ''. The program `locpoly` can also be used to estimate derivatives μ' and μ'', where we use the bigger bandwidth $2b = 0.74$. For f, we use the R program `density`, and estimate f' by differentiating the estimated density. Then we can have the bias-corrected estimate $\tilde{\mu}_n(x) = \mu_n(x) - b^2\psi_K\hat{\rho}(x)$ for μ, which is plotted in the the middle curve in Figure 2. To estimate $\sigma(\cdot)$, as in Stanton (1997), we shall make use of the estimated residuals $\hat{e}_i = Y_i - \tilde{\mu}_n(X_i)$, and perform the Nadaraya–Watson regression of $\hat{e}_i^2$ versus X_i with the bandwidth b. In our data analysis the boundary problem of the Nadaraya–Watson regression raised in Chapman and Pearson (2000) is not severe since we focus on the interval $T = [0.35, 8.06]$, while the whole range is $[\min X_i, \max X_i] = [0.14, 8.49]$.

The Gumbel convergence in Theorem 2.4 can be quite slow, so the SCB in (2.13) may not have a good finite-sample performance. To circumvent this problem, we

$\rightarrow \boxed{e^{-2e^{-z}}}$

shall adopt a simulation based method. Let

$$\Pi_n = \sup_{x \in T} \frac{|\sum_{k=1}^{n} K(X_k^*/b - x/b)\eta_k^*|}{nbf^{1/2}(x)},$$

where X_k^* are i.i.d. with density f, η_k^* are i.i.d. with $\mathsf{E}\eta_n = 0$, $\mathsf{E}\eta_n^2 = 1$ and $\mathsf{E}|\eta_1|^p < \infty$, and (X_k^*) and (η_k^*) are independent. As in Theorem 2.4, let

$$\Pi_n' = \sup_{x \in T} \frac{\sqrt{f(x)}|\mu_n(x) - \mu(x) - b^2\psi_K\rho(x)|}{\sigma(x)}.$$

By Theorem 2.4 and Proposition 2.1, with proper centering and scaling, Π_n and Π_n' have the same asymptotic Gumbel distribution. So the cutoff value, the $(1-\alpha)$th quantile of Π_n', can be estimated by the sample $(1-\alpha)$th quantile of many simulated Π_n's. For the U.S. Treasury bill data, we simulated 10,000 Π_n's and obtained the 95% sample quantile 0.39. Then the SCB is constructed as $\tilde{\mu}_n(x) \pm 0.39\hat{\sigma}(x)/f_n^{1/2}(x)$; see the upper and lower curves in Figure 2.

We now apply Theorem 2.5 to construct SCB for $\sigma^2(\cdot)$. We choose $h = b$, which has a reasonably satisfactory performance in our data analysis. By Theorem 2.5,

$$\Pi_n'' = \frac{1}{\sqrt{v_\eta}} \sup_{x \in T} \frac{\sqrt{f(x)}|\sigma_n^2(x) - \sigma^2(x) - b^2\psi_K\rho_\sigma(x)|}{\sigma^2(x)}$$

has the same asymptotic distribution as Π_n and Π_n'. Based on the above simulation, we choose the cutoff value 0.39. As in the treatment of μ' and μ'' in the bias term of μ_n, we use a similar estimate, noting that $\rho_\sigma(x) = (\sigma^2(x))'' + 2(\sigma^2(x))'f'(x)/f(x)$ has the same form as $\rho_\mu(x)$. The 95% SCB of $\sigma^2(\cdot)$ is presented in Figure 3.

Based on the 95% SCB of $\mu(\cdot)$, we conclude that the linear drift function hypothesis $H_0 : \mu(x) = \alpha_0 + \alpha_1 x$ for some α_0 and α_1 is rejected at the 5% level. Other simple parametric forms do not seem to exist. Similar claims can be made for $\sigma^2(\cdot)$, and none of the parametric forms previously mentioned seems appropriate. This suggests that the dynamics of the treasury yield rates might be far more complicated than previously speculated.

4. Proofs of Theorems 2.1–2.3. Throughout the proofs C denotes constants which do not depend on n and b_n. The values of C may vary from place to place. Let $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ be the floor and ceiling functions, respectively. Without loss of generality, we assume $l = 0$, $u = 1$ in (1.5) and $A = 1$ in condition (C4). Write

$$\frac{\sqrt{nb}}{\sqrt{\lambda_K f(bt)}}[f_n(bt) - \mathsf{E}f_n(bt)] = M_n(t) + N_n(t),$$

where $M_n(t)$ has summands of martingale differences

$$M_n(t) = \frac{1}{\sqrt{nb\lambda_K f(bt)}} \sum_{k=1}^{n} \{K(X_k/b - t) - \mathsf{E}[K(X_k/b - t)|\xi_{k-1}]\},$$