**The University of Chicago**
**Department of Statistics**
**TECHNICAL REPORT SERIES**

# KERNEL QUANTILE REGRESSION FOR NONLINEAR STOCHASTIC MODELS

*Zhibiao Zhao and Wei Biao Wu*

TECHNICAL REPORT NO. 572

Departments of Statistics
The University of Chicago
Chicago, Illinois 60637

October, 2006

# KERNEL QUANTILE REGRESSION FOR NONLINEAR STOCHASTIC MODELS

By Zhibiao Zhao and Wei Biao Wu
*Department of Statistics, The University of Chicago*

*Abstract:* We consider kernel quantile estimates for drift and scale functions in nonlinear stochastic regression models. Under a general dependence setting, we establish asymptotic point-wise and uniform Bahadur representations for the kernel quantile estimates. Based on those asymptotic representations, central limit theorems are obtained. Applications to nonlinear autoregressive models and linear processes are made. Simulation studies show that the estimates have good performance. The results are applied to the Pound/USD exchange rates data.

# 1 Introduction

Consider the nonlinear stochastic regression model

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

where $\varepsilon_i, i \in \mathbb{Z}$, are independent and identically distributed (iid) random variables and $(X_i, Y_i), i \in \mathbb{Z}$, is a stationary process. Here $\mu(\cdot)$ and $\sigma(\cdot) \geq 0$ are measurable and they represent drift and scale functions, respectively. For statistical inference of (1), our goal is to estimate $\mu(\cdot)$ and $\sigma(\cdot)$ based on the observations $(X_i, Y_i), 1 \leq i \leq n$.

In (1), if we let $Y_i = X_{i+1} - X_i$ and assume that $\varepsilon_i$ are standard normals, then (1) can be viewed as the discretized version of the stochastic diffusion model

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{2}$$

where $\{W_t\}$ is a standard Brownian motion. Many well-known financial models are special cases of (2); see Fan (2005) and references therein. As a second example, (1) can also be used to model time series driven by processes other than Brownian motions, for example, by Lévy processes. In this case, instead of calling $\sigma(\cdot)$ a volatility function, it seems more sensible to call it a scale function since $\varepsilon_i$ has infinite variance.

Due to the flexibility of the forms in $\mu$ and $\sigma$, (1) allows nonlinearity and conditional heteroscedasticity and it becomes a popular model in financial econometrics. As an important special case, we let $X_i = Y_{i-1}$ and then (1) becomes the conditional heteroskedastic autoregressive nonlinear (CHARN) process (Bossaerts *et al.*, 1996); see Section 4.1. CHARN includes many popular nonlinear time series models, for example, threshold AR (TAR) models (Tong, 1990), AR with conditional heteroscedasticity (ARCH, Engle, 1982) and exponential AR (EAR) models (Haggan and Ozaki, 1981) among others. Here in our setting we allow nonparametric forms in $\mu$ and $\sigma$.

For model (1) and its various special cases, there is a vast literature on parametric and nonparametric estimation of the drift function $\mu$ and the scale function $\sigma$. In the parametric scenario, the main focus is on ARCH (Engle, 1982), linear GARCH (Bollerslev, 1986) and exponential GARCH (Nelson, 1991) among others. For nonparametric estimation, Gourieroux and Monfort (1992) considered pseudo-maximum likelihood estimation for qualitative threshold ARCH models; Mckeague and Zhang (1994) proposed histogram type estimates of $\mu$ and $\sigma$; Chen and Tsay (1993) and Chen and Liu (2001) studied the functional-coefficient autoregressive models. Härdle and Tsybakov (1997) established asymptotic normality for local polynomial estimates (LPE) of $\mu$ and $\sigma$ for both the AR case $X_i = Y_{i-1}$ and the general regression case $(X_i, Y_i)$. In the latter case, they assumed that $(X_i, Y_i)$ are iid. Under various mixing conditions on $(X_i, Y_i)$, Masry and Fan (1997) and Fan and Yao (1998) considered LPE of $\mu$ and $\sigma$, Jiang and Mack (2001) studied robust LPE, Ziegelmann (2002) considered local exponential estimates of $\sigma$ with known $\mu$. For difference-based estimates of $\sigma$, see Müller and Stadtmüller (1987) and Hall *et al.* (1990) among others. In all the aforementioned papers, it is assumed that finite second or higher order moments exist.

Empirical studies, however, have found heavy-tailed distributions in many fields, including finance, economics, telecommunication and others. For example, high-frequency financial time series may have tails heavier than the Student-$t$ distributions (Tsay, 2005). Nolan (2001) found heavy tails in foreign exchange (FX) rates. In the special case

$$Y_i = \alpha Y_{i-1} + \sqrt{\beta + \lambda Y_{i-1}^2}\varepsilon_i, \quad \alpha \in \mathbb{R},\ \beta,\ \lambda > 0, \tag{3}$$

Borkovec and Klüppelberg (2001) showed that the stationary distribution of $Y_i$ may have a heavy tail of the Pareto type and thus $Y_i$ may not have finite second moment even though

$\mathbb{E}(\varepsilon_i^2) < \infty$. For more applications and modelling of heavy tails, See Resnick (1997) and references therein. For heavy-tailed data, the classical least squares (LS) method which requires finite second moment may not be a good choice. Attractive alternatives are the quantile, least absolute deviation or other robust regression methods. Since the seminal work of Koenker and Bassett (1978), quantile regression has become popular in parametric and nonparametric inference and we refer the readers to Yu and Jones (1998), Yu *et al.* (2003), Koenker (2005) for recent developments.

There are very few results on properties of quantile estimates of $\mu(\cdot)$ and $\sigma(\cdot)$ for model (1) under a general dependence structure on $(X_i, Y_i)$. For special cases results have been obtained under stringent conditions by assuming that the process is either iid (cf Samanta (1989), Jones and Hall (1990), Bhattacharya and Gangopadhyay (1990) and Chaudhuri (1991a, 1991b)) or strong mixing (cf Truong and Stone (1992), Shi (1995), Honda (2000), Cai (2002), Franke and Mwita (2003) and Ziegelmann (2005)). Assuming that $(X_i, Y_i)$ are iid, Bhattacharya and Gangopadhyay (1990) obtained a point-wise Bahadur representation of quantile estimates of $\mu(\cdot)$, while Chaudhuri (1991b) derived a local Bahadur representation. Such Bahadur representations provide deep insights into the asymptotic properties of the estimates. Honda (2000) obtained point-wise and uniform Bahadur representations of conditional quantile estimates under very complicated and stringent strong mixing conditions.

In this paper, we shall consider kernel quantile estimates of $\mu(\cdot)$ and $\sigma(\cdot)$ and establish their asymptotic point-wise and uniform Bahadur representations under general dependence structures. The uniform Bahadur representation allows us to apply a jackknife bias reduction technique and, consequently, provide a theoretical justification for Fan and Zhang's (2000) two-step procedure to smooth the jackknifed estimates. It turns out that an asymptotic distributional theory of the jackknifed estimates cannot be established if one only has a point-wise Bahadur representation. Additionally, our dependence conditions have a nice interpretation which is based on nonlinear system theory. In contrast to mixing conditions which may be difficult to work with, our dependence conditions are closely related to the data generating mechanism, and thus are easily verifiable. Also, we impose a very mild moment condition on $\varepsilon_0$.

The rest of the paper is structured as follows. In Section 2, based on kernel regression quantiles, we obtain raw estimates of $\mu(\cdot)$ and $\sigma(\cdot)$. In Section 3, the asymptotic behavior of the raw estimates is studied and refined estimates are proposed. Section 4 concerns two important cases of (1): nonlinear AR models and linear processes. In Section 5, a simulation study is performed and our procedure is applied to the Pound/USD FX rates data. Proofs are given in Section 6.

# 2 Kernel quantile regression estimates

In model (1), we assume that $X_i$ is a causal stationary process of the form

$$X_i = G(\ldots, \eta_{i-1}, \eta_i), \tag{4}$$

where $\eta_i, i \in \mathbb{Z}$, are iid random variables and $G$ is a measurable function such that $X_i$ is properly defined. Framework (4) is very general for stationary processes (cf Tong (1990), Stine (2006) and Wu (2005a)). Let $\mathcal{F}_i = (\ldots, \eta_{i-1}, \eta_i)$ be the one-sided shift process. We assume that $\varepsilon_i$ in (1) is independent of $\mathcal{F}_i$ and $\eta_i$ is independent of $\varepsilon_j, j \leq i-2$.

To ensure identifiability of the estimates, we need to impose assumptions on the innovations $\varepsilon_i$. For a random variable $\varepsilon$ denote its median by $\text{med}(\varepsilon) = \inf\{t : \mathbb{P}(\varepsilon \leq t) \geq 1/2\}$. We can assume without loss of generality (WLOG) that $a = \text{med}(\varepsilon_0) = 0$ and $b = \text{med}(|\varepsilon_0 - a|) = \text{med}(|\varepsilon_0|) = 1$ since otherwise (1) can be re-parameterized by letting $\bar{\mu}(x) = \mu(x) + a\sigma(x), \bar{\sigma}(x) = b\sigma(x)$ and $\bar{\varepsilon}_i = (\varepsilon_i - a)/b$.

Under the above conventions, the conditional median of $Y_i$ given $X_i = x$ is $\mu(x)$. It is well-known that $\mu(x)$ is a solution to the minimization problem $\text{argmin}_\theta \mathbb{E}[(|Y_i - \theta| - |Y_i|)|X_i = x]$. Here we put $|Y_i|$ into the expectation to guarantee that $|Y_i - \theta| - |Y_i|$ always has a finite mean. A weighted sample analog is

$$\operatorname*{argmin}_\theta \sum_{i=1}^n w_i[|Y_i - \theta| - |Y_i|] \ \text{ or equivalently } \ \operatorname*{argmin}_\theta \sum_{i=1}^n w_i|Y_i - \theta|, \tag{5}$$

where $w_i = w_i(x)$ are non-negative weight functions such that $\sum_{i=1}^n w_i = 1$. Here we use the normalized kernel weights with

$$w_i(x) = \frac{K_{b_n}(x - X_i)}{\sum_{i=1}^n K_{b_n}(x - X_i)} \ \text{ and } \ K_{b_n}(u) = K(u/b_n), \tag{6}$$

where $K$ is a kernel function and $b_n > 0$ is a bandwidth. If the optimization problem (5) has a unique solution, we denote the minimizer by $\hat{\mu}_{b_n}(x)$; otherwise, we let $\hat{\mu}_{b_n}(x)$ be an arbitrary solution. Asymptotic properties of $\hat{\mu}_{b_n}(x)$ is studied in Section 3.1.

Assume that a consistent estimate $\tilde{\mu}_{b_n}(x)$ of $\mu(x)$ is obtained. The next question is to estimate the scale function $\sigma(\cdot)$. Let $e_i = |Y_i - \tilde{\mu}_{b_n}(x)|$ be the estimated conditional absolute residuals given $X_i = x$. Since the median of $|\varepsilon_0|$ is 1, as in (5), we can estimate $\sigma(x)$ by the solution, denoted by $\hat{\sigma}_{h_n}(x)$, to the optimization problem

$$\underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \tilde{w}_i |e_i - \theta|, \quad \tilde{w}_i = \frac{\tilde{K}_{h_n}(x - X_i)}{\sum_{i=1}^{n} \tilde{K}_{h_n}(x - X_i)}, \tag{7}$$

where $\tilde{K}_{h_n}(u) = \tilde{K}(u/h_n)$ for some kernel $\tilde{K}$ and bandwidth $h_n > 0$. Here $\tilde{K}$ and bandwidth $h_n$ may be different from $K$ and $b_n$ that are used to estimate $\mu$.

We now introduce some notation. Recall $\mathcal{F}_i = (\ldots, \eta_{i-1}, \eta_i)$. Let $(\eta_i')_{i\in\mathbb{Z}}$ be an iid copy of $(\eta_i)_{i\in\mathbb{Z}}$ and let $\mathcal{F}_i^* = (\mathcal{F}_{-1}, \eta_0', \eta_1, \ldots, \eta_i)$. For a random variable $Z$ write $Z \in \mathcal{L}^p, p > 0$, if $\|Z\|_p := [\mathbb{E}(|Z|^p)]^{1/p} < \infty$, and denote $\|Z\| = \|Z\|_2$. Define the projections $\mathcal{P}_k Z = \mathbb{E}(Z|\mathcal{F}_k) - \mathbb{E}(Z|\mathcal{F}_{k-1}), k \in \mathbb{Z}$. For $a, b \in \mathbb{R}$ let $a \wedge b = \min(a, b), a \vee b = \max(a, b)$ and $\lceil a \rceil = \inf\{k \in \mathbb{Z} : k \geq a\}$. Let $\{a_n\}$ and $\{b_n\}$ be two real sequences. We write $a_n \asymp b_n$ if $0 < \liminf_{n\to\infty} |a_n/b_n| \leq \limsup_{n\to\infty} |a_n/b_n| < \infty$. For $\epsilon > 0$ and a set $\mathcal{T} \subset \mathbb{R}$, let $\mathcal{T}_\epsilon = \cup_{y\in\mathcal{T}}\{x : |x - y| \leq \epsilon\}$ be the $\epsilon$-neighborhood of $\mathcal{T}$ and write $x_\epsilon = \{x\}_\epsilon$ for $x \in \mathbb{R}$. Let $\mathcal{C}^p(\mathcal{T}) = \{g(\cdot) : \sup_{x\in\mathcal{T}} |g^{(p)}(x)| < \infty\}$ be the set of functions having bounded $p$-th order derivatives on $\mathcal{T}$. Here $g^{(p)}$ is the $p$-th order derivative of $g$ with the convention $g^{(0)} = g$. Denote by $\mathbf{1}_A$ the indicator function for an event $A$.

# 3 Main results

Denote by $F_X$ and $F_\varepsilon$ the distribution functions of $X_0$ and $\varepsilon_0$, respectively, and by $f_X = F_X'$ and $f_\varepsilon = F_\varepsilon'$ the corresponding densities. For $k, i \in \mathbb{N}$, let $F_k(x|\mathcal{F}_i) = \mathbb{P}(X_{i+k} \leq x|\mathcal{F}_i)$ be the $k$-step ahead conditional distribution function of $X_{i+k}$ given $\mathcal{F}_i$ and $f_k(x|\mathcal{F}_i) = \partial F_k(x|\mathcal{F}_i)/\partial x$ the conditional density. Recall that $\mathcal{F}_i^* = (\mathcal{F}_{-1}, \eta_0', \eta_1, \ldots, \eta_i)$. Let $X_i^* = G(\mathcal{F}_i^*)$ be a coupled process of $X_i$. For $\nu \in (1, 2]$ define

$$\theta_i(\nu) = \sup_{x\in\mathbb{R}} \|f_i(x|\mathcal{F}_0) - f_i(x|\mathcal{F}_0^*)\|_\nu. \tag{8}$$

Then $\theta_i(\nu)$ quantifies the distance between conditional distributions $[X_i|\mathcal{F}_0]$ and $[X_i^*|\mathcal{F}_0^*]$ via the difference between conditional densities. If $f_i(x|\mathcal{F}_0)$ does not depend on $\eta_0$, then $\theta_i(\nu) = 0$. So $\theta_i(\nu)$ measures the contribution of the innovation $\eta_0$ in predicting $X_i$ given $\mathcal{F}_0$

by perturbing the input via coupling. To state our results we need to introduce dependence conditions and regularities conditions.

**Condition 1** ( Dependence conditions )**.** *(i) Let $\nu \in (1, 2]$. Assume that*

$$\Theta(\nu) := \sum_{i=1}^{\infty} \theta_i(\nu) < \infty. \tag{9}$$

*(ii) Let $\theta_i' = \sup_{x\in\mathbb{R}} \|f_i'(x|\mathcal{F}_0) - f_i'(x|\mathcal{F}_0^*)\|$, where $f_i'(x|\mathcal{F}_0) = \partial f_i(x|\mathcal{F}_0)/\partial x$. Assume that*

$$\Theta := \sum_{i=1}^{\infty} (\theta_i + \theta_i') < \infty, \text{ where } \theta_i = \theta_i(2). \tag{10}$$

Note that $\theta_i(\nu)$ measures the contribution of $\eta_0$ in predicting $X_i$ given $\mathcal{F}_0$. Following Wu (2005a), we view (4) as a physical system with $\mathcal{F}_i$, $X_i$ and $G$ being the input, output and transform, respectively. So (9) means that the cumulative contribution of $\eta_0$ in predicting future values $(X_i)_{i\geq 1}$ is finite, hence suggesting short-range dependence. The other condition (10) delivers a similar message.

**Remark 1.** In some applications it is not easy to deal with the $i$-th step ahead predictive density $f_i(x|\mathcal{F}_0)$. A simple sufficient condition of (9) which only involves one-step ahead predictive density is

$$\sum_{i=1}^{\infty} \sup_{x\in\mathbb{R}} \|f_1(x|\mathcal{F}_i) - f_1(x|\mathcal{F}_i^*)\|_\nu < \infty, \tag{11}$$

To see this, since $f_{1+i}(x|\mathcal{F}_0) = \mathbb{E}[f_1(x|\mathcal{F}_i)|\mathcal{F}_0]$ and $\mathbb{E}[f_1(x|\mathcal{F}_i^*)|\mathcal{F}_0] = \mathbb{E}[f_1(x|\mathcal{F}_i)|\mathcal{F}_0^*]$, by Jensen's inequality, $\|\mathbb{E}[f_1(x|\mathcal{F}_i)|\mathcal{F}_0] - \mathbb{E}[f_1(x|\mathcal{F}_i^*)|\mathcal{F}_0^*]\|_\nu \leq 2\|f_1(x|\mathcal{F}_i) - f_1(x|\mathcal{F}_i^*)\|_\nu$. So (11) implies (9). Similarly, (10) holds if

$$\sum_{i=1}^{\infty} \left[ \sup_{x\in\mathbb{R}} \|f_1(x|\mathcal{F}_i) - f_1(x|\mathcal{F}_i^*)\| + \sup_{x\in\mathbb{R}} \|f_1'(x|\mathcal{F}_i) - f_1'(x|\mathcal{F}_i^*)\| \right] < \infty. \tag{12}$$

In Section 4 we shall verify (11) and (12) for linear and nonlinear processes. ◇

**Condition 2** ( Regularity conditions )**.** Let $\mathcal{T} \subset \mathbb{R}$. We say that $(\mu, \sigma, f_1(\cdot|\mathcal{F}_0), f_X)$ satisfies regularity condition $\mathcal{R}(\mathcal{T})$ if there exists some $c < \infty$ such that

$$\mu(\cdot), \sigma(\cdot), f_X(\cdot) \in \mathcal{C}^4(\mathcal{T}), \quad \inf_{y\in\mathcal{T}} \sigma(y) > 0, \quad \sup_{y\in\mathcal{T}} f_1(y|\mathcal{F}_0) < c \text{ and } \inf_{y\in\mathcal{T}} f_X(y) > 0.$$

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

### 3.1 Point-wise Bahadur representation for $\hat{\mu}_{b_n}$

Let $x \in \mathbb{R}$ be fixed. Theorem 1 provides an asymptotic Bahadur representation for $\hat{\mu}_{b_n}(x)$. Namely, we approximate $\hat{\mu}_{b_n}(x) - \mu(x)$ by a linear form (cf. (13)), which is usually easier to deal with. Corollary 1 gives a CLT for $\hat{\mu}_{b_n}(x) - \mu(x)$.

**Theorem 1.** *Let $K \in \mathcal{K}$. Assume that $f_\varepsilon \in \mathcal{C}^3(\mathbb{R})$, $f_\varepsilon(0) > 0$, $b_n + (\log n)^2/(nb_n) \to 0$ and that (9) holds with some $\nu \in (1, 2]$. Further assume that, for some $\epsilon > 0$, $(\mu, \sigma, f_1(\cdot|\mathcal{F}_0), f_X)$ satisfies the regularity condition $\mathcal{R}(x_\epsilon)$. Let*

$$\begin{aligned}
\rho(x) &= \mu''(x) + 2\mu'(x)\frac{f'_X(x)}{f_X(x)} - \frac{\mu'(x)}{\sigma(x)}\Big[\frac{f'_\varepsilon(0)}{f_\varepsilon(0)}\mu'(x) + 2\sigma'(x)\Big] \quad \textit{and} \\
Q_{b_n}(x) &= \sum_{i=1}^{n}\Big\{(1/2 - \mathbf{1}_{Y_i \le \mu(x)})K_{b_n}(x - X_i) - \mathbb{E}[(1/2 - \mathbf{1}_{Y_i \le \mu(x)})K_{b_n}(x - X_i)]\Big\}.
\end{aligned}$$

*Then we have the Bahadur representation*

$$\hat{\mu}_{b_n}(x) - \mu(x) = \frac{\sigma(x)}{f_X(x) f_\varepsilon(0)}\frac{Q_{b_n}(x)}{nb_n} + \psi_K \rho(x) b_n^2 + O_{\mathrm{p}}(r_n), \tag{13}$$

*where*

$$r_n = \Big[\frac{\delta_n \log n}{nb_n}\Big]^{1/2} + \delta_n^2 \quad \textit{and} \quad \delta_n = b_n^2 + (nb_n)^{-1/2} + n^{1/\nu - 1}.$$

**Corollary 1.** *Assume that the conditions in Theorem 1 are satisfied and*

$$nb_n^9 + n^{2/\nu - 1} b_n^3 + n^{4/\nu - 3} b_n \to 0. \tag{14}$$

*Then we have*

$$(nb_n)^{1/2}[\hat{\mu}_{b_n}(x) - \mu(x) - \psi_K \rho(x) b_n^2] \Rightarrow N\Big(0, \frac{\varphi_K \sigma^2(x)}{4 f_X(x) f_\varepsilon^2(0)}\Big). \tag{15}$$

**Remark 2.** If $b_n \asymp n^{-\beta}$, then (14) holds if $\beta > \max\{1/9, 4/\nu - 3, (2 - \nu)/(3\nu)\}$. In particular, if (9) is satisfied with some $\nu \in [3/2, 2]$ and $\beta > 1/9$, then (14) holds. $\diamondsuit$

By Corollary 1, the asymptotic optimal mean squares error (MSE) bandwidth is

$$b_n = \Big[\frac{\varphi_K \sigma^2(x)}{16\psi_K^2 \rho^2(x) f_X(x) f_\varepsilon^2(0)}\Big]^{1/5} n^{-1/5} \asymp n^{-1/5}. \tag{16}$$

By the Bahadur representation in Theorem 1, the bias term $\psi_K \rho(x) b_n^2$ can be corrected by the simple jackknife estimator

$$\tilde{\mu}_{b_n}(x) = 2\hat{\mu}_{b_n}(x) - \hat{\mu}_{\sqrt{2}b_n}(x).$$

There certainly exists other forms, for example, $(\lambda\hat{\mu}_{b_n} - \hat{\mu}_{\sqrt{\lambda}b_n})/(\lambda - 1), \lambda \neq 1$; see Wu and Zhao (2005). In practice the choice $\lambda = 2$ has a good performance. By (13), following the proof of Corollary 1 in Section 6, we have under the conditions in Corollary 1 that

$$(nb_n)^{1/2}[\tilde{\mu}_{b_n}(x) - \mu(x)] \Rightarrow N\Big(0, \frac{\varphi_{K^*}\sigma^2(x)}{4 f_X(x) f_\varepsilon^2(0)}\Big),$$

where $K^*(u) = 2K(u) - 2^{-1/2}K(u/\sqrt{2}) \in \mathcal{K}$.

## 3.2 Point-wise Bahadur representation for $\hat{\sigma}_{h_n}$

As in Section 2, let $e_i = |Y_i - \tilde{\mu}_{b_n}(x)|$ be the estimated absolute conditional residuals given $X_i = x$. Then $\sigma(x)$ can be estimated by $\hat{\sigma}_{h_n}(x)$, the solution to the minimization problem (7). Theorem 2 below provides an asymptotic Bahadur representation for $\hat{\sigma}_{h_n}(x)$. To this end, let $Q_{b_n}(x)$ be as in Theorem 1, $T_{b_n}(x) = 2Q_{b_n}(x) - 2^{-1/2}Q_{\sqrt{2}b_n}(x)$ and

$$\begin{aligned} W_{h_n}(x) &= \sum_{i=1}^n \Big\{(1/2 - \mathbf{1}_{|Y_i - \mu(x)| \le \sigma(x)})\tilde{K}_{h_n}(x - X_i) \\ &\qquad - \mathbb{E}[(1/2 - \mathbf{1}_{|Y_i - \mu(x)| \le \sigma(x)})\tilde{K}_{h_n}(x - X_i)]\Big\}. \end{aligned}$$

**Theorem 2.** *Let $\tilde{K} \in \mathcal{K}$. Assume that $f_\varepsilon \in \mathcal{C}^3(\mathbb{R})$ and $\kappa_+ := f_\varepsilon(-1) + f_\varepsilon(1) > 0$, $h_n + (\log n)^2/(nh_n) \to 0$ and that the conditions in Theorem 1 are satisfied. Let*

$$\begin{aligned} \tilde{\rho}(x) &= \sigma''(x) + \kappa\mu''(x) + 2[\sigma'(x) + \kappa\mu'(x)][f_X'(x)/f_X(x) - \sigma'(x)/\sigma(x)] \\ &\quad - \{f_\varepsilon'(1)[\sigma'(x) + \mu'(x)]^2 - f_\varepsilon'(-1)[\sigma'(x) - \mu'(x)]^2\}/[\kappa_+\sigma(x)], \end{aligned}$$

*where $\kappa = [f_\varepsilon(1) - f_\varepsilon(-1)]/\kappa_+$. Then the Bahadur representation holds:*

$$\hat{\sigma}_{h_n}(x) - \sigma(x) = \frac{\sigma(x)}{f_X(x)}\Big[\frac{W_{h_n}(x)}{nh_n\kappa_+} - \frac{\kappa T_{b_n}(x)}{nb_n f_\varepsilon(0)}\Big] + \psi_{\tilde{K}}\tilde{\rho}(x)h_n^2 + O_{\mathrm{p}}(\tilde{r}_n), \qquad (17)$$

*where*

$$\tilde{r}_n = \Big[\frac{\tilde{\delta}_n \log n}{nh_n}\Big]^{1/2} + \tilde{\delta}_n^2 \quad \textit{and} \quad \tilde{\delta}_n = b_n^2 + (nb_n)^{-1/2} + h_n^2 + (nh_n)^{-1/2} + n^{1/\nu - 1}.$$

As in Corollary 1, we can derive the asymptotic distribution of $\hat{\sigma}_{h_n}(x)$ from the Bahadur representation in Theorem 2. The convergence rate, however, depends on the relative magnitudes of $b_n$ and $h_n$. Define

$$D_\alpha = \{(b_n, h_n) : \lim_{n\to\infty} h_n/b_n = \alpha\}, \quad 0 \le \alpha \le \infty.$$

We now consider the first term in the right hand side of (17). Assume that $(b_n, h_n) \in D_\alpha$ for some $\alpha \in [0, \infty]$. If $\alpha = 0$, then $W_{h_n}(x)/(nh_n)$ dominates; if $\alpha = \infty$, then $T_{b_n}(x)/(nb_n)$ dominates; if $0 < \alpha < \infty$, then both terms are of the same order of magnitude.

**Corollary 2.** *Let $\kappa, \kappa_+, \tilde{\rho}(x)$ be as in Theorem 2 and $(b_n, h_n) \in D_\alpha$ for some $\alpha \in [0, \infty]$. Assume the conditions in Theorem 2 are fulfilled and*

$$n(b_n + h_n)^9 + n^{2/\nu-1}(b_n + h_n)^3 + n^{4/\nu-3}(b_n + h_n) \to 0. \tag{18}$$

*(i) If $\kappa \neq 0, \alpha = \infty$, then*

$$(nb_n)^{1/2}[\hat{\sigma}_{h_n}(x) - \sigma(x) - \psi_{\tilde{K}}\tilde{\rho}(x)h_n^2] \Rightarrow N\Big(0, \frac{\kappa^2\varphi_{K^*}\sigma^2(x)}{4f_X(x)f_\varepsilon^2(0)}\Big).$$

*(ii) If $\kappa \neq 0$ and $\alpha \in [0, \infty)$, then*

$$(nh_n)^{1/2}[\hat{\sigma}_{h_n}(x) - \sigma(x) - \psi_{\tilde{K}}\tilde{\rho}(x)h_n^2] \Rightarrow N(0, \varsigma_\alpha^2), \quad \textit{where}$$
$$\varsigma_\alpha^2 = \frac{\sigma^2(x)}{4f_X(x)}\Big\{\frac{\varphi_{\tilde{K}}}{\kappa_+^2} + \frac{\alpha^2\kappa^2\varphi_{K^*}}{f_\varepsilon^2(0)} - \frac{2\alpha\kappa[1-4F_\varepsilon(-1)]}{\kappa_+ f_\varepsilon(0)}\int_{\mathbb{R}} \tilde{K}(u)K^*(\alpha u)du\Big\}. \tag{19}$$

*(iii) If $\kappa = 0$ and $h_n/(nb_n^2) \to 0$, then (19) holds with $\varsigma_\alpha^2 = \varphi_{\tilde{K}}\sigma^2(x)/[4\kappa_+^2 f_X(x)], \alpha \in [0, \infty]$.*

**Remark 3.** If $\varepsilon_0$ is symmetric, then $f_\varepsilon(-1) = f_\varepsilon(1)$ and $\kappa = 0$. In this case, $\varsigma_\alpha^2 = \varphi_{\tilde{K}}\sigma^2(x)/[16f_X(x)f_\varepsilon^2(1)], \alpha \in [0, \infty)$. If $0 < \alpha < \infty$, condition (18) reduces to (14). $\diamondsuit$

If $\mu$ were known and we use $e_i = |Y_i - \mu(X_i)|$ in (7) to estimate $\sigma$, then (19) holds with $\kappa = 0$. Corollary 2 delivers the message that the convergence rate of $\hat{\sigma}_{h_n}$ depends on whether $\mu$ is known or not. If $\kappa \neq 0, \alpha \in [0, \infty)$ or $\kappa = 0$, then one has the interesting oracle property that the convergence rate of $\tilde{\sigma}_{h_n}$ is the same as the one as if $\mu$ were known (cf. Section 8.7 in Fan and Yao (2003)). On the other hand, if $\kappa \neq 0$ and $\alpha = \infty$, then $\tilde{\sigma}_{h_n}$ with estimated $\mu$ has a slower convergence rate. Assuming that $(X_i, Y_i)$ are independent and $\varepsilon_0$ has zero mean and unit variance, Hall and Carroll (1989) investigated the effect of estimating $\mu$ on the estimation of the conditional variance $\sigma^2(\cdot)$.

From Corollary 2, we can derive that, the optimal MSE bandwidth $h_n \asymp n^{-1/5}$, regardless of the choice of $b_n$ as long as $n^{-1/5} = O(b_n)$ and the conditions in Corollary 2 (ii) and (iii) hold. As in the case of $\hat{\mu}_{b_n}$, we can also use the bias-corrected estimator $\tilde{\sigma}_{h_n}(x) = 2\hat{\sigma}_{h_n}(x) - \hat{\sigma}_{\sqrt{2}h_n}(x)$ and a similar central limit theorem holds. For example, when $\kappa \neq 0$ and $\alpha \in [0, \infty)$, $(nh_n)^{1/2}[\tilde{\sigma}_{b_n}(x) - \sigma(x)] \Rightarrow N(0, \varsigma_\alpha^{*2})$, where $\varsigma_\alpha^{*2}$ is obtained by replacing $\tilde{K}$ with $\tilde{K}^*(u) = 2\tilde{K}(u) - 2^{-1/2}\tilde{K}(u/\sqrt{2})$ in Corollary 2 (ii).

## 3.3 Uniform Bahadur representations

Let $T > 0$ be fixed. Theorem 3 provides uniform Bahadur representations of $\hat{\mu}_{b_n}$ and $\hat{\sigma}_{h_n}$ over the interval $\mathcal{T} := [-T, T]$. Such results are useful in developing a limit theory for refined estimates based on Fan and Zhang's (2000) two-step procedure (cf. Section 3.4).

**Theorem 3.** *Assume (10), $f_\varepsilon \in \mathcal{C}^3(\mathbb{R}), f_\varepsilon(0) > 0, f_\varepsilon(-1) + f_\varepsilon(1) > 0$, $K, \tilde{K} \in \mathcal{K}$ and*

$$b_n + h_n + (\log n)^3/(nb_n) + (\log n)^3/(nh_n) \to 0. \tag{20}$$

*Further assume that there exists an $\epsilon > 0$ such that $(\mu, \sigma, f_1(\cdot|\mathcal{F}_0), f_X)$ satisfies the regularity condition $\mathcal{R}(\mathcal{T}_\epsilon)$. Then the Bahadur representations in Theorems 1 and 2 hold uniformly over $x \in \mathcal{T}$ with the uniform error bounds $r_n^{\mathrm{unif}}$ and $\tilde{r}_n^{\mathrm{unif}}$, respectively, given by*

$$\begin{aligned} r_n^{\mathrm{unif}} &= b_n^4 + \Big[\frac{b_n \log n}{n}\Big]^{1/2} + \Big[\frac{\log n}{nb_n}\Big]^{3/4}, \\ \tilde{r}_n^{\mathrm{unif}} &= b_n^4 + h_n^4 + \Big[\frac{h_n \log n}{n}\Big]^{1/2} + \Big[\frac{b_n^2 \log n}{nh_n}\Big]^{1/2} + \Big[\frac{(\log n)^3}{n^3 b_n h_n^2}\Big]^{1/4} + \Big[\frac{\log n}{nh_n}\Big]^{3/4}. \end{aligned}$$

## 3.4 Refining the raw estimates $\tilde{\mu}_{b_n}$ and $\tilde{\sigma}_{h_n}$

The raw estimates $\tilde{\mu}_{b_n}$ and $\tilde{\sigma}_{h_n}$ are not smooth and improved estimates can be obtained by smoothing. Suppose we have the values of the estimates at the evenly distributed grid points $x_i = iT/N$, $-N \le i \le N, N = N_n \to \infty$. Then we can smooth the estimates $\tilde{\mu}_{b_n}(x_i)$ and $\tilde{\sigma}_{h_n}(x_i)$ by locally averaging:

$$\breve{\mu}_{b_n}(x) = \sum_{i=-N}^{N} \pi(x, x_i)\tilde{\mu}_{b_n}(x_i) \quad \text{and} \quad \breve{\sigma}_{h_n}(x) = \sum_{i=-N}^{N} \pi(x, x_i)\tilde{\sigma}_{h_n}(x_i),$$

where $\pi(x, x_i)$ are non-negative weights satisfying $\sum_{i=-N}^{N} \pi(x, x_i) = 1$. Our approach can be viewed as a combination of the bias-reduction jackknife method and Fan and Zhang's (2000) two-step smoothing procedure since our "raw" estimates are jackknifed.

There are various smoothing techniques to construct the weights $\pi(x, x_i)$, including kernel, local polynomial, wavelets among others. Here we use the kernel method:

$$\pi(x, x_i) = \frac{\breve{K}_{\tau_N}(x - x_i)}{\sum_{i=-N}^{N} \breve{K}_{\tau_N}(x - x_i)},$$

where $\breve{K}_{\tau_N}(u) = \breve{K}(u/\tau_N)$ for some kernel $\breve{K}$ and the bandwidth $\tau_N > 0$.

**Corollary 3.** *Let the conditions in Theorem 3 be fulfilled, $\breve{K} \in \mathcal{K}$ and $\tau_N \to 0, N\tau_N \to \infty$. Let $T_{b_n}(x)$ and $r_n^{\mathrm{unif}}$ be as in Theorems 2 and 3, respectively. Let $\delta \in (0, T)$ be fixed and $\breve{r}_n = r_n^{\mathrm{unif}} + N^{-1} + \tau_N^4$. Then*

$$\breve{\mu}_{b_n}(x) = \mu(x) + \sum_{i=-N}^{N} \frac{\pi(x, x_i)\sigma(x_i)}{f_X(x_i) f_\varepsilon(0)} \frac{T_{b_n}(x_i)}{n b_n} + \psi(\breve{K})\tau_N^2 \mu''(x) + O_{\mathrm{p}}(\breve{r}_n) \tag{21}$$

*holds uniformly over $x \in [-T + \delta, T - \delta]$. A similar representation holds for $\breve{\sigma}_{h_n}(x)$.*

*Proof.* Elementary calculations show that, for $j = 0, 1, 2, 3$,

$$\frac{T}{N\tau_N} \sum_{i=-N}^{N} \breve{K}(\frac{x - x_i}{\tau_N})(\frac{x - x_i}{\tau_N})^j = \int_{\mathbb{R}} \breve{K}(u) u^j du + O(\frac{1}{N\tau_N})$$

hold uniformly over $x \in [-T + \delta, T - \delta]$. By Theorem 3,

$$\breve{\mu}_{b_n}(x) \quad = \quad \sum_{i=-N}^{N} \pi(x, x_i)\mu(x_i) + \sum_{i=-N}^{N} \frac{\pi(x, x_i)\sigma(x_i)}{f_X(x_i) f_\varepsilon(0)} \frac{T_{b_n}(x_i)}{n b_n} + O_{\mathrm{p}}(r_n^{\mathrm{unif}}).$$

Since $\breve{K}$ has a bounded support, we only need to consider those $x_i$'s with $x_i - x = O(\tau_N)$. By Taylor's expansion $\mu(x_i) = \sum_{j=0}^{3} \mu^{(j)}(x)(x_i - x)^j / j! + O(\tau_N^4)$, (21) holds. $\diamondsuit$

Based on the representation (21), one can obtain a central limit theorem for $\breve{\mu}_{b_n}(x)$. For example, if $\tau_N = o(b_n)$, by the argument in Proposition 3, we have

$$\sum_{i=-N}^{N} \frac{\pi(x, x_i)\sigma(x_i)}{f_X(x_i)} T_{b_n}(x_i) = \frac{\sigma(x)}{f_X(x)} T_{b_n}(x) + O_{\mathrm{p}}[(n\tau_N \log n)^{1/2} + n^{1/2} b_n].$$

By the argument in Corollary 1, one can establish a CLT for $T_{b_n}(x)$. Techniques in Proposition 4 can be similarly applied to $\breve{\sigma}_{h_n}(x)$. The details are omitted.

# 4 Examples

In this section, we shall verify Condition 1 for some popular processes.

## 4.1 Nonlinear AR models

If $X_i = Y_{i-1}$ and $\eta_i = \varepsilon_{i-1}$, then (1) becomes the CHARN model (Bossaerts *et al.*, 1996)

$$Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\varepsilon_i. \tag{22}$$

Many popular nonlinear time series models are special cases of (22), including TAR model $Y_n = a(0 \vee Y_{n-1}) + b(0 \wedge Y_{n-1}) + \varepsilon_n$, ARCH model $Y_n = \varepsilon_n\sqrt{a^2 + b^2 Y_{n-1}^2}$, EAR model $Y_n = (a + be^{-cY_{n-1}^2})Y_{n-1} + \varepsilon_n$ and others.

**Proposition 1.** *Let $f_\varepsilon$ be the density of $\varepsilon_0$. Assume (i) $\varepsilon_0 \in \mathcal{L}^q$ for some $q > 0$ and $\lambda := \sup_{x\in\mathbb{R}} \|\mu'(x) + \sigma'(x)\varepsilon_0\|_q < 1$ and (ii) $\mu, \sigma \in \mathcal{C}^1(\mathbb{R})$, $\inf_{x\in\mathbb{R}} \sigma(x) > 0$ and $\sup_{x\in\mathbb{R}}(1 + |x|)[|f_\varepsilon'(x)| + |f_\varepsilon''(x)|] < \infty$. Then (10) holds.*

*Proof.* By Theorem 2 in Wu and Shao (2004), if (i) is satisfied, then $X_i$ has a unique stationary solution of the form $G(\mathcal{F}_i)$ with the property that

$$\|G(\mathcal{F}_i) - G(\mathcal{F}_i')\|_q = O(\lambda^i), \tag{23}$$

where $\mathcal{F}_i' = (\ldots, \eta_{-1}', \eta_0', \eta_1, \ldots, \eta_i)$ couples $\mathcal{F}_i$ with $\eta_j$ replaced by $\eta_j', j \le 0$. Let $X_i^* = G(\mathcal{F}_i^*)$. Since $|a + b|^q \le 2^q(|a|^q + |b|^q)$, by stationarity, (23) implies that

$$\|G(\mathcal{F}_i) - X_i^*\|_q^q \le 2^q[\|G(\mathcal{F}_i) - G(\mathcal{F}_i')\|_q^q + \|G(\mathcal{F}_{i+1}) - G(\mathcal{F}_{i+1}')\|_q^q] = O(\lambda^{iq}).$$

Note that $f_1(x|\mathcal{F}_i) = f_\varepsilon(\{x - \mu(X_i)\}/\sigma(X_i))$ and $f_1(x|\mathcal{F}_i^*) = f_\varepsilon(\{x - \mu(X_i^*)\}/\sigma(X_i^*))$. By (ii), elementary calculations shows that (12), and hence (10), follows. $\diamondsuit$

## 4.2 Linear processes

Let $\eta_i, i \in \mathbb{Z}$, be iid random variables with $\eta_0 \in \mathcal{L}^q, q > 0$; let $f_\eta$ be the density of $\eta_0$. Assume $\mathbb{E}(\eta_0) = 0$ if $q \ge 1$. Consider the linear process

$$X_i = \sum_{j=0}^{\infty} a_j \eta_{i-j}, \tag{24}$$

where the real sequence $(a_i)_{i\geq 0}$ satisfies $\sum_{i=0}^{\infty} |a_i|^{2\wedge q} < \infty$. The latter condition is needed to guarantee the existence of $X_n$. Special cases of the linear processes (24) include ARMA and fractional ARIMA models.

**Proposition 2.** *Let $\eta_0 \in \mathcal{L}^q, q > 0$, and $\mathbb{E}(\eta_0) = 0$ if $q \geq 1$. Assume $\sup_{x\in\mathbb{R}}[f_\eta(x) + |f'_\eta(x)|] < \infty$. (i) If $\sum_{i=0}^{\infty} |a_i|^{(q\wedge 2)/\nu} < \infty$ holds with $\nu \in (1, 2]$, then (9) is satisfied. (ii) If $\sup_{x\in\mathbb{R}} |f''_\eta(x)| < \infty$ and $\sum_{i=0}^{\infty} |a_i|^{(q\wedge 2)/2} < \infty$, then (10) holds.*

*Proof.* 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 $f_1(x|\mathcal{F}_i) = f_\eta(x - \bar{X}_{i+1})$. (i) We shall show that $\theta_i(\nu) = O[|a_i|^{(q\wedge 2)/\nu}]$. If $q \geq 2$, then

$$\|f_\eta(x - \bar{X}_{i+1}) - f_\eta(x - \bar{X}^*_{i+1})\|_2^2 \leq c^2\|\bar{X}_{i+1} - \bar{X}^*_{i+1}\|_2^2 = O(a_{i+1}^2),$$

where $c = \sup_{x\in\mathbb{R}}[f_\eta(x) + |f'_\eta(x)|] < \infty$. If $0 < q < 2$, then for any $\nu \geq q \vee 1$

$$\begin{aligned} \|f_\eta(x - \bar{X}_{i+1}) - f_\eta(x - \bar{X}^*_{i+1})\|_\nu^\nu &\leq c^\nu\|1 \wedge |a_{i+1}(\eta_0 - \eta'_0)|\|_\nu^\nu \\ &\leq c^\nu\|a_{i+1}(\eta_0 - \eta'_0)\|_q^q = O(|a_{i+1}|^q). \end{aligned}$$

Similarly, under (ii), $\|f'_\eta(x - \bar{X}_{i+1}) - f'_\eta(x - \bar{X}^*_{i+1})\|_2^2 = O(|a_{i+1}|^{q\wedge 2})$. So (10) follows. $\diamond$

If $q = \nu = 2$, then the summability conditions in (i) and (ii) of Proposition 2 become $\sum_{i=0}^{\infty} |a_i| < \infty$, a classical condition for short-range dependence. If the latter is violated, then one has long-range dependence and it is beyond the scope of the paper.

# 5 A simulation study and a real data example

In this section we shall first perform a simulation study and then apply our methods to the FX rates between U.K. Pound and U.S. Dollar. The two-step procedure is applied here. In the first step we use the Epanechnikov kernel $K(u) = 0.75(1 - u^2)\mathbf{1}_{|u|\leq 1}$ to obtain raw estimates, while in the second step, we apply the local linear smoothing procedure for refinements.

The bandwidth (16) is not immediately usable since it involves unknown quantities $\rho(x), f_X(x), f_\varepsilon(0)$ and $\sigma(x)$. It is a difficult problem to develop a good automatic bandwidth selector in the context of quantile regression under dependence. Here for the illustration purpose, we select bandwidth through visual inspection of the estimates in the first step and apply Ruppert *et al.*'s (1995) automatic bandwidth selection procedure in the local linear smoothing step. Throughout our numerical work, we have also tried different choices of bandwidths and obtained similar results.

## 5.1 A simulation study

Consider the continuous-time model

$$dX_t = (\alpha_0 + \alpha_1 X_t)dt + (\beta_0 + \beta_1 |X_t|^{\gamma_0})^{\gamma_1} dZ_t. \tag{25}$$

When $\{Z_t\} = \{W_t\}$ is a standard Brownian motion, many popular models are the special cases of (25) with $X_t$ being the logarithm of some stock price. For example, in the CIR model (Cox *et al.*, 1985), one has $\beta_0 = 0, \gamma_0 = 1/2$ and $\gamma_1 = 1$, while in the CKLS model (Chan *et al.*, 1992), $\beta_0 = 0$. Fan and Zhang (2003) considered nonparametric inference for such models based on higher order approximations.

Here we assume that $\{Z_t\}$ is a stable Lévy process rather than a Brownian motion. In particular, we consider the standard symmetric $\alpha$-stable Lévy process with index $\alpha = 1.6$. Let $\mu(x) = \alpha_0 + \alpha_1 x$ and $\sigma(x) = (\beta_0 + \beta_1 |x|^{\gamma_0})^{\gamma_1}$, where the parameters $\alpha_0 = 0.05, \alpha_1 = -0.8, \beta_0 = 0.16, \beta_1 = 0.32, \gamma_0 = 2$ and $\gamma_1 = 0.5$. A discretized version of (25) is

$$Y_i := X_{i+1} - X_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{26}$$

where $\varepsilon_i$ are iid S$\alpha$S random variables with index 1.6. In this case, it is easy to see that conditions in (i) of Proposition 1 are satisfied with $q < \alpha = 1.6$. Thus (26) admits a stationary solution. We simulate a sample $(X_i), i = 1, 2, \ldots, n = 2500$, from model (26) and use bandwidths $b_n, h_n = 0.5, 0.6, 0.7$ to obtain the raw estimates. We perform the estimation procedure at 500 grid points, distributed evenly between 5% and 95% quantiles of $X_i$'s, $i = 1, 2, \ldots, n$, and then use linear interpolation to obtain the estimates at other points. The actual and estimated functions are plotted in Figure 1. The plots suggest that the estimates are reasonably good. The classical LS method is not applicable since $\varepsilon_0$ has infinite variance. Our quantile estimates provide a useful alternative.

**Insert Figure 1 about here**

## 5.2 An application: FX rates between Pound and USD

The dataset is obtained from *http://www.federalreserve.gov/releases/h10/hist/*, the website of Federal Reserve Bank of New York. It contains 8848 weekdays records of Pound/USD noon buying rates from January 4th 1971 to May 5th 2006. In model (1), let $X_i, i = 1, 2, \ldots, 8848$, be the exchange rates and $Y_i = \log(X_{i+1}/X_i)$ be the daily log-returns. We use bandwidths $b_n, h_n = 0.005, 0.010$ and $0.015$ to obtain the raw estimates. Estimated drift function $\tilde{\mu}_{b_n}(\cdot)$ and scale function $\tilde{\sigma}_{h_n}(\cdot)$ are plotted in Figure 2. As in the simulation study, estimation procedure is performed at 2000 grid points, distributed evenly on the range of $X_i$'s, and linear interpolation is used for other points.

Nolan (2001) considered the FX rates of USD/Pound (reciprocal of Pound/USD) from January 2nd 1980 to May 21st 1996. He fitted the log-returns with a stable distribution with stable index $\alpha = 1.530$, skewness $\beta = -0.088$, scale $\gamma = 0.00376$ and location $\delta = 0.00009$. Nolan also fitted stable distributions for FX rates for other currencies. Here we shall consider the innovations $\varepsilon_i$ instead of the log-returns. In model (1), after estimating $\mu(\cdot)$ and $\sigma(\cdot)$, the innovations $\varepsilon_i$ can be estimated by

$$\hat{\varepsilon}_i = \frac{Y_i - \tilde{\mu}_{b_n}(X_i)}{\tilde{\sigma}_{h_n}(X_i)}, \ i = 1, 2, \ldots, 8847.$$

We use bandwidths $b_n = 0.010$ and $h_n = 0.005$ to compute $\hat{\varepsilon}_i$. We fit the estimated innovations $\hat{\varepsilon}_i$ with a stable distribution using the program STABLE (available from J. P. Nolan's website: *http://academic2.american.edu/~jpnolan*). The estimated parameters are $\alpha = 1.62, \beta = -0.10, \gamma = 0.97, \delta = 0.03$ for the maximum likelihood method; $\alpha = 1.45, \beta = -0.05, \gamma = 0.89, \delta = 0.02$ for the quantile method and $\alpha = 1.74, \beta = -0.08, \gamma = 1.00, \delta = 0.03$ for the characteristic function method. All these 3 cases suggest that $\alpha < 2$ and $\varepsilon_i$ has a heavy tail. The plot of sample kurtosis of $\hat{\varepsilon}_i$ in Figure 3 gives further evidence of heavy tails.

**Insert Figures 2 and 3 about here**

# 6 Proofs

Throughout this section $c, c_1, c_2, \ldots$ stand for positive constants which may vary from line to line. Recall that $\mathcal{F}_i = (\ldots, \eta_{i-1}, \eta_i)$ and $\mathcal{F}_i^* = (\mathcal{F}_{-1}, \eta_0', \eta_1, \ldots, \eta_i)$. Let $\mathcal{G}_i =$

$(\ldots, \eta_i, \eta_{i+1}; \varepsilon_i, \varepsilon_{i-1}, \ldots)$. Let $x$ be fixed and define

$$
\begin{aligned}
&L_n(s) = \sum_{i=1}^n K_{b_n}(x - X_i)\mathbf{1}_{Y_i \le \mu(x)+s}, \quad L_n = \sum_{i=1}^n K_{b_n}(x - X_i), \\
&\tilde{L}_n(s,t) = \sum_{i=1}^n \tilde{K}_{h_n}(x - X_i)\mathbf{1}_{|Y_i - \mu(x) - s| \le \sigma(x)+t}, \quad \tilde{L}_n = \sum_{i=1}^n \tilde{K}_{h_n}(x - X_i), \\
&J_n(s) = \mathbb{E}L_n(s), \ J_n = \mathbb{E}L_n, \ \tilde{J}_n(s,t) = \mathbb{E}\tilde{L}_n(s,t) \ \text{ and } \ \tilde{J}_n = \mathbb{E}\tilde{L}_n.
\end{aligned}
$$

**Lemma 1.** *Let $x$ be fixed, $K, \tilde{K} \in \mathcal{K}$ and $b_n \to 0, h_n \to 0$. Assume $\sigma(x) > 0, f_\varepsilon \in \mathcal{C}^3(\mathbb{R})$ and that there exists an $\epsilon > 0$ such that $\mu, \sigma, f_X \in \mathcal{C}^4(x_\epsilon)$. Let $\rho(x), \tilde{\rho}(x), \kappa, \kappa_+$ be as in Theorems 1 and 2. Then*

$$
\begin{aligned}
&J_n = nb_n f_X(x) + \psi_K f_X''(x) nb_n^3 + O(nb_n^5), \\
&J_n(0) = J_n/2 - nb_n[b_n^2 \psi_K \rho(x) f_\varepsilon(0) f_X(x)/\sigma(x) + O(b_n^4)], \\
&J_n(s) - J_n(0) = nb_n s[f_\varepsilon(0) f_X(x)/\sigma(x) + O(b_n^2 + s)], \\
&\tilde{J}_n(s,0) = \tilde{J}_n/2 - nh_n\{[h_n^2 \psi_{\tilde{K}} \tilde{\rho}(x) - s\kappa]\kappa_+ f_X(x)/\sigma(x) + O(h_n^4 + s^2)\}, \\
&\tilde{J}_n(s,t) - \tilde{J}_n(s,0) = nh_n t[\kappa_+ f_X(x)/\sigma(x) + O(h_n^2 + s + t)].
\end{aligned}
$$

*Proof.* Let $h$ be a bounded function and $H_n(x) = \sum_{i=1}^n K_{b_n}(x - X_i) h(Y_i)$. Assume that $g(\cdot) := f_X(\cdot)\mathbb{E}[h(\mu(\cdot) + \sigma(\cdot)\varepsilon_0)] \in \mathcal{C}^4(x_\epsilon)$. Since $K \in \mathcal{K}$, by Taylor's expansion,

$$
\mathbb{E}[H_n(x)] = nb_n \int_{\mathbb{R}} K(u) g(x - ub_n) du = nb_n[g(x) + \psi_K g''(x) b_n^2 + O(b_n^4)].
$$

Lemma 1 follows by considering the specific forms of $h(\cdot)$ in the above expressions. ◇

Recall that $f_1(x|\mathcal{F}_{i-1})$ is the conditional density of $X_i$ at $x$ given $\mathcal{F}_{i-1}$. Define

$$
I_n(x) = \sum_{i=1}^n [f_1(x|\mathcal{F}_{i-1}) - \mathbb{E}f_1(x|\mathcal{F}_{i-1})], \quad x \in \mathbb{R}. \tag{27}
$$

**Lemma 2.** *(i) Suppose (9) holds with $\nu \in (1, 2]$. Then $\sup_{x\in\mathbb{R}} \|I_n(x)\|_\nu = O(n^{1/\nu})$. (ii) Let $T > 0$ be fixed. Then under (10), $\| \sup_{|x| \le T} |I_n(x)| \|_2 = O(n^{1/2})$.*

*Proof.* (i) For $i \ge 0, \mathbb{E}[f_1(x|\mathcal{F}_i)|\mathcal{F}_{-1}] = \mathbb{E}[f_1(x|\mathcal{F}_i^*)|\mathcal{F}_0]$. Thus, by Jensen's inequality

$$
\|\mathcal{P}_0 f_1(x|\mathcal{F}_i)\|_\nu \ = \ \|\mathbb{E}[f_1(x|\mathcal{F}_i) - f_1(x|\mathcal{F}_i^*)|\mathcal{F}_0]\|_\nu
$$

$$\leq \|f_1(x|\mathcal{F}_i) - f_1(x|\mathcal{F}_i^*)\|_\nu \leq \theta_i(\nu).$$

Since $\{\mathcal{P}_j I_n(x)\}_{j=-\infty}^{n-1}$ are martingale differences, by the von Bahr-Esseén inequality,

$$\begin{aligned}
\|I_n(x)\|_\nu^\nu &\leq 2\sum_{j=-\infty}^{n-1} \|\mathcal{P}_j I_n(x)\|_\nu^\nu \leq 2\sum_{j=-\infty}^{n-1} \Big[\sum_{i=0}^{n-1} \|\mathcal{P}_j f_1(x|\mathcal{F}_i)\|_\nu\Big]^\nu \\
&\leq 2\sum_{j=-\infty}^{n-1} \Big[\sum_{i=0\vee j}^{n-1} \theta_{i-j}(\nu)\Big]^\nu \leq 2\sum_{j=-\infty}^{n-1} \Big[\Theta^{\nu-1}(\nu)\sum_{i=0\vee j}^{n-1} \theta_{i-j}(\nu)\Big] = O(n).
\end{aligned}$$

(ii) By the argument in (i), $\sup_{x\in\mathbb{R}}[\|I_n(x)\|_2 + \|I_n'(x)\|_2] = O(n^{1/2})$. Since $|I_n(x) - I_n(-T)| \leq \int_{-T}^T |I_n'(u)|du$, by Jensen's inequality, we have $\mathbb{E}[\sup_{|x|\leq T} I_n^2(x)] = O(n)$. $\diamondsuit$

**Proposition 3.** *Let $x$ be a continuity point of $\mu$ and $\sigma$. Assume that $\sigma(x) > 0, F_\varepsilon \in \mathcal{C}^1(\mathbb{R})$, $b_n \vee h_n \vee l_n \to 0, \log n = O(n(b_n \wedge h_n)l_n)$ and that there exists an $\epsilon > 0$ such that $\sup_{y\in x_\epsilon} f_1(y|\mathcal{F}_0) < c_0$ for some constant $c_0 < \infty$. Further assume that (9) holds with some $\nu \in (1, 2]$. For $b = b_n, h_n$, define $\Omega_n(b) = (nbl_n \log n)^{1/2} + n^{1/\nu} bl_n$. Then*

$$\sup_{|s|\leq l_n} |[L_n(s) - J_n(s)] - [L_n(0) - J_n(0)]| = O_{\mathrm{p}}[\Omega_n(b_n)], \tag{28}$$

$$\sup_{|s|+|t|\leq l_n} |[\tilde{L}_n(s,t) - \tilde{J}_n(s,t)] - [\tilde{L}_n(0,0) - \tilde{J}_n(0,0)]| = O_{\mathrm{p}}[\Omega_n(h_n)]. \tag{29}$$

*Proof.* We only prove (29) since (28) similarly follows. Let $Z_i := Z_i(s,t) = \tilde{K}_{h_n}(x - X_i)[\mathbf{1}_{|Y_i-\mu(x)-s|\leq\sigma(x)+t} - \mathbf{1}_{|Y_i-\mu(x)|\leq\sigma(x)}]$. Then $[\tilde{L}_n(s,t) - \tilde{J}_n(s,t)] - [\tilde{L}_n(0,0) - \tilde{J}_n(0,0)] = \sum_{i=1}^n [Z_i - \mathbb{E}(Z_i)]$. Write

$$\begin{aligned}
\sum_{i=1}^n [Z_i - \mathbb{E}(Z_i)] &= M_n(s,t) + R_n(s,t), \text{ where} \\
M_n(s,t) &= \sum_{i=1}^n [Z_i - \mathbb{E}(Z_i|\mathcal{G}_{i-2})] \text{ and } R_n(s,t) = \sum_{i=1}^n [\mathbb{E}(Z_i|\mathcal{G}_{i-2}) - \mathbb{E}(Z_i)]. 
\end{aligned}\tag{30}$$

Hereafter we call (30) $M/R$-decomposition with $M_n(s,t)$ and $R_n(s,t)$ being the $M$ and $R$-parts, respectively. We shall only consider $0 \leq s \leq t$ and assume WLOG that $\tilde{K}$ has support $[-1, 1]$. Let $\underline{B}(y) = [\mu(x) - \mu(y) - \sigma(x)]/\sigma(y), \overline{B}(y) = [\mu(x) - \mu(y) + \sigma(x)]/\sigma(y)$ and

$$G(y; s,t) = F_\varepsilon\Big(\overline{B}(y) + \frac{s+t}{\sigma(y)}\Big) - F_\varepsilon(\overline{B}(y)) + F_\varepsilon(\underline{B}(y)) - F_\varepsilon\Big(\underline{B}(y) - \frac{t-s}{\sigma(y)}\Big).$$

Then $\sup_{|u|\le 1, |s|+|t|\le l_n} |G(x-uh_n;s,t)| = O(l_n)$. Since $\mathbb{E}(Z_i|\mathcal{G}_{i-2}) = \mathbb{E}[\mathbb{E}(Z_i|\mathcal{G}_{i-1})|\mathcal{G}_{i-2}] = \mathbb{E}[\tilde{K}_{h_n}(x-X_i)G(X_i;s,t)|\mathcal{F}_{i-1}]$, we have

$$R_n(s,t) = h_n \int_{\mathbb{R}} \tilde{K}(u)G(x-uh_n;s,t)I_n(x-uh_n)du = O_{\mathrm{p}}(n^{1/\nu}h_n l_n)$$

in view of Lemma 2(i). By Lemma 3 below, (29) follows. $\diamondsuit$

**Lemma 3.** *Recall (30) for $M_n(s,t)$. Then under conditions in Proposition 3, we have*

$$\sup_{|s|+|t|\le l_n} |M_n(s,t)| = O_{\mathrm{p}}[(nh_n l_n \log n)^{1/2}].$$

*Proof.* Let $Z_i = Z_i(s,t)$ be as in Proposition 3 and $M_n^\circ(s,t) = \sum_{i:\mathrm{even}}[Z_i - \mathbb{E}(Z_i|\mathcal{G}_{i-2})]$ the sum over even indices $i \le n$. Let $d_n = nh_n l_n$ and $\mathcal{S} = \{(s,t) : 0 \le s \le t, s+t \le l_n\}$. It suffices to show that $\sup_{(s,t)\in\mathcal{S}} |M_n^\circ(s,t)| = O_{\mathrm{p}}[(d_n \log n)^{1/2}]$ since other cases can be similarly treated. Let $N = \lceil d_n^{1/2} \rceil$, $\omega_n = l_n/N$, $t_i = i\omega_n$, $0 \le i \le N$. For any $(s,t) \in \mathcal{S}$, there exist $j$ and $k$ such that $s \in [t_j, t_{j+1}), t \in [t_k, t_{k+1})$. Define

$$\begin{aligned} \overline{A}_{ijk} &= \tilde{K}_{h_n}(x-X_i)[\mathbf{1}_{t_j-\sigma(x)-t_{k+1}\le Y_i-\mu(x)<-\sigma(x)} + \mathbf{1}_{\sigma(x)<Y_i-\mu(x)\le t_{j+1}+\sigma(x)+t_{k+1}}], \\ \underline{A}_{ijk} &= \tilde{K}_{h_n}(x-X_i)[\mathbf{1}_{t_{j+1}-\sigma(x)-t_k\le Y_i-\mu(x)<-\sigma(x)} + \mathbf{1}_{\sigma(x)<Y_i-\mu(x)\le t_j+\sigma(x)+t_k}]. \end{aligned}$$

Then $\underline{A}_{ijk} \le Z_i \le \overline{A}_{ijk}$. Since $F_\varepsilon \in \mathcal{C}^1(\mathbb{R})$ and $\sup_{y\in x_\epsilon} f_1(y|\mathcal{F}_0) < c_0$, there exists a constant $c_1 < \infty$ such that $\mathbb{E}[(\overline{A}_{ijk} - \underline{A}_{ijk})|\mathcal{G}_{i-2}] \le c_1 h_n \omega_n$. Therefore,

$$\underline{A}_{ijk} - \mathbb{E}(\underline{A}_{ijk}|\mathcal{G}_{i-2}) - c_1 h_n\omega_n \le Z_i - \mathbb{E}(Z_i|\mathcal{G}_{i-2}) \le \overline{A}_{ijk} - \mathbb{E}(\overline{A}_{ijk}|\mathcal{G}_{i-2}) + c_1 h_n \omega_n$$

and consequently

$$\sup_{(s,t)\in\mathcal{S}} |M_n^\circ(s,t)| \le \max_{0\le j,\,k\le N-1} (|\underline{H}_{jk}| + |\overline{H}_{jk}|) + c_1 d_n/N \tag{31}$$

where $\underline{H}_{jk} = \sum_{i:\mathrm{even}}[\underline{A}_{ijk} - \mathbb{E}(\underline{A}_{ijk}|\mathcal{G}_{i-2})]$ and $\overline{H}_{jk} = \sum_{i:\mathrm{even}}[\overline{A}_{ijk} - \mathbb{E}(\overline{A}_{ijk}|\mathcal{G}_{i-2})]$. Note that for fixed $j$ and $k$, $\{\underline{A}_{ijk} - \mathbb{E}(\underline{A}_{ijk}|\mathcal{G}_{i-2}) : i \text{ even}\}$ form martingale differences with respect to the sigma field generated by $\mathcal{G}_i$ and there exists some constant $c_2 < \infty$ such that

$$\sum_{i:\mathrm{even}} \mathbb{E}[\underline{A}_{ijk}^2|\mathcal{G}_{i-2}] = \sum_{i:\mathrm{even}} \mathbb{E}[\mathbb{E}(\underline{A}_{ijk}^2|\mathcal{G}_{i-1})|\mathcal{G}_{i-2}] \le c_2 d_n.$$

Assume WLOG that $\sup_u \tilde{K}(u) \leq 1$. Let $c_3 = \sup_n(\log n/d_n)^{1/2} < \infty$. By Freedman's inequality (Freedman, 1975) for bounded martingale differences,

$$\begin{aligned} \mathbb{P}\Big(\max_{0\leq j,k\leq N-1}|\underline{H}_{jk}| \geq c\sqrt{d_n \log n}\Big) &\leq \sum_{j,k=0}^{N-1}\mathbb{P}\Big(|\underline{H}_{jk}| \geq c\sqrt{d_n \log n}\Big) \\ &\leq 2N^2 \exp\Big[-\frac{c^2 d_n \log n}{2c\sqrt{d_n \log n} + 2c_2 d_n}\Big] = O(n^{-\lambda_c}), \end{aligned}$$

where $\lambda_c = c^2/(2cc_3 + 2c_2) - 1$. A similar expression holds for $\overline{H}_{jk}$. Choose sufficiently large $c$, then the desired result follows from (31) since $d_n/N = o[(d_n \log n)^{1/2}]$. $\diamondsuit$

Recall that $\mathcal{C}^0(\mathbb{R})$ is the set of bounded and measurable functions and that $f_\varepsilon$ is the density of $\varepsilon_0$. Define the operator $\mathcal{H}$ by

$$\mathcal{H}(g)(t) = \mathbb{E}[g(Y_0)|X_0 = t] = \int_{\mathbb{R}} g(\mu(t) + \sigma(t)v) f_\varepsilon(v)dv, \quad g \in \mathcal{C}^0(\mathbb{R}).$$

**Proposition 4.** *Let $g_k \in \mathcal{C}^0(\mathbb{R})$ and $q_k \in \mathcal{K}$, $k = 1, 2, \ldots, d$. For each $x \in \mathbb{R}$, define*

$$S_n(x) = \sum_{i=1}^{n}\sum_{k=1}^{d}\Big\{g_k(Y_i)q_{k,b_n}(x - X_i) - \mathbb{E}[g_k(Y_i)q_{k,b_n}(x - X_i)|\mathcal{G}_{i-2}]\Big\},$$

*where $q_{k,b_n}(u) = q_k(u/b_n)$. Assume that $x$ is a continuity point of $\mu, \sigma, f_X, f_1(\cdot|\mathcal{F}_0), \mathcal{H}(g_k)(\cdot)$ and $\mathcal{H}(g_k g_{k'})(\cdot), 1 \leq k, k' \leq d$. If $b_n \to 0$ and $nb_n \to \infty$, then we have*

$$\frac{S_n(x)}{\sqrt{nb_n}} \Rightarrow N(0, \sigma^2), \text{ where } \sigma^2 = f_X(x)\sum_{1\leq k,k'\leq d}\mathcal{H}(g_k g_{k'})(x)\int_{\mathbb{R}} q_k(u)q_{k'}(u)du.$$

*Proof.* For $1 \leq k \leq d$ let $\zeta_{k,i} = g_k(Y_i)q_{k,b_n}(x - X_i), \xi_{k,i} = \zeta_{k,i} - \mathbb{E}(\zeta_{k,i}|\mathcal{G}_{i-1})$ and $\gamma_{k,i} = \mathbb{E}(\zeta_{k,i+1}|\mathcal{G}_i) - \mathbb{E}(\zeta_{k,i+1}|\mathcal{G}_{i-1})$. Then we have

$$S_n(x) = \sum_{i=1}^{n}\sum_{k=1}^{d}(\xi_{k,i} + \gamma_{k,i}) + O(1). \tag{32}$$

Note that $(\xi_{1,i}, \gamma_{1,i}, \xi_{2,i}, \gamma_{2,i}, \ldots, \xi_{d,i}, \gamma_{d,i}), i = 1, 2, \ldots, n$, form triangular array martingale differences with respect to $\mathcal{G}_i$. By the martingale central limit theorem, it suffices to verify convergence of the conditional variance and the Lindeberg condition. For $1 \leq k, k' \leq d$, we write $g_{k,k'}(u) = g_k(u)g_{k'}(u), q_{k,k'}(u) = q_k(u)q_{k'}(u)$ and $q_{k,k',b_n}(u) = q_{k,k'}(u/b_n)$. Then

$$\mathbb{E}(\xi_{k,i}\xi_{k',i}|\mathcal{G}_{i-1}) = q_{k,k',b_n}(x - X_i)[\mathbb{E}(g_{k,k'}(Y_i)|\mathcal{G}_{i-1}) - \mathbb{E}(g_k(Y_i)|\mathcal{G}_{i-1})\mathbb{E}(g_{k'}(Y_i)|\mathcal{G}_{i-1})]$$

$$= \quad q_{k,k',b_n}(x-X_i)[\mathcal{H}(g_{k,k'})(x) - \mathcal{H}(g_k)(x)\mathcal{H}(g_{k'})(x) + o(1)]. \qquad (33)$$

Define $\alpha_i = q_{k,k',b_n}(x-X_i) - \mathbb{E}[q_{k,k',b_n}(x-X_i)|\mathcal{F}_{i-1}]$ and $\beta_i = \mathbb{E}[q_{k,k',b_n}(x-X_i)|\mathcal{F}_{i-1}] - \mathbb{E}[q_{k,k',b_n}(x-X_i)]$. Since $\alpha_i, 1 \le i \le n$, form stationary martingale differences with respect to $\mathcal{F}_i$, $\sum_{i=1}^n \alpha_i = O_{\mathrm{p}}[(nb_n)^{1/2}]$. By Lemma 2(i), $\sum_{i=1}^n \beta_i = b_n \int_{\mathbb{R}} q_{k,k'}(u) I_n(x-ub_n)du = O_{\mathrm{p}}(n^{1/\nu}b_n)$. So $\sum_{i=1}^n(\alpha_i+\beta_i) = o_{\mathrm{p}}(nb_n)$ and consequently

$$\frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}(\xi_{k,i}\xi_{k',i}|\mathcal{G}_{i-1}) \xrightarrow{p} \quad [\mathcal{H}(g_{k,k'})(x) - \mathcal{H}(g_k)(x)\mathcal{H}(g_{k'})(x)] f_X(x) \textstyle\int_{\mathbb{R}} q_{k,k'}(u)du.$$

Similarly, we can show that

$$\begin{aligned} \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}(\gamma_{k,i}\gamma_{k',i}|\mathcal{G}_{i-1}) &\xrightarrow{p} \mathcal{H}(g_k)(x)\mathcal{H}(g_{k'})(x) f_X(x)\int_{\mathbb{R}} q_{k,k'}(u)du, \\ \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}|\xi_{k,i}\gamma_{k',i}| &= O(b_n^{-1})\mathbb{E}[q_{k,b_n}(x-X_0)q_{k',b_n}(x-X_1)] = O(b_n). \end{aligned}$$

The Lindeberg condition is easily verified since $g_k$ and $q_k$ are bounded and $nb_n \to \infty$. $\diamondsuit$

*Proof of Theorem 1.* Recall $\delta_n = (nb_n)^{-1/2} + b_n^2 + n^{1/\nu - 1}$. Let $k_n \uparrow \infty$ be a positive sequence satisfying $\delta_n k_n \to 0$. We first show that $\Delta := \hat{\mu}_{b_n}(x) - \mu(x) = O_{\mathrm{p}}(k_n\delta_n)$. Since $\hat{\mu}_{b_n}(x)$ is a solution of (5), $L_n(\Delta) = L_n/2 + O_{\mathrm{p}}(1)$. Applying the $M/R$-decomposition technique in (30), we have $|L_n(0) - J_n(0)| + |L_n - J_n| = O_{\mathrm{p}}[(nb_n)^{1/2} + n^{1/\nu}b_n] = O_{\mathrm{p}}(nb_n\delta_n)$. Hence, by (28) in Proposition 3 and Lemma 1, after elementary calculations,

$$\begin{aligned} L_n(k_n\delta_n) &= J_n(k_n\delta_n) + L_n(0) - J_n(0) + O_{\mathrm{p}}[(nb_nk_n\delta_n \log n)^{1/2} + n^{1/\nu}b_nk_n\delta_n] \\ &= L_n/2 + [J_n(k_n\delta_n) - J_n(0)] + [J_n(0) - J_n/2] + [J_n - L_n]/2 \\ &\quad + [L_n(0) - J_n(0)] + O_{\mathrm{p}}[(nb_nk_n\delta_n \log n)^{1/2} + n^{1/\nu}b_nk_n\delta_n] \\ &= L_n/2 + nb_n\delta_nk_n f_\varepsilon(0) f_X(x)/\sigma(x)[1 + o_{\mathrm{p}}(1)] \end{aligned} \qquad (34)$$

Since $L_n(s)$ is non-decreasing and $nb_n\delta_nk_n \to \infty$, $\mathbb{P}(\Delta < k_n\delta_n) \to 1$. Likewise $\mathbb{P}(\Delta > -k_n\delta_n) \to 1$. Thus $\Delta = O_{\mathrm{p}}(k_n\delta_n)$. Since the rate of $k_n \uparrow \infty$ can be arbitrarily slow, $\Delta = O_{\mathrm{p}}(\delta_n)$. Again, by (28) in Proposition 3 and Lemma 1,

$$\begin{aligned} L_n(0) - J_n(0) &= L_n(\Delta) - J_n(\Delta) + O_{\mathrm{p}}[(nb_n\delta_n \log n)^{1/2} + n^{1/\nu}b_n\delta_n] \\ &= [L_n(\Delta) - L_n/2] + [L_n - J_n]/2 + [J_n/2 - J_n(0)] - [J_n(\Delta) - J_n(0)] \\ &\quad + O_{\mathrm{p}}(nb_nr_n) \end{aligned}$$

$$= [L_n - J_n]/2 + nb_n[b_n^2\psi_K\rho(x) - \Delta + O_{\rm p}(r_n)]f_\varepsilon(0)f_X(x)/\sigma(x).$$

The representation (13) then follows by solving the above equation. $\diamondsuit$

*Proof of Theorem 2.* The argument in the proof of Theorem 1 can be applied here. For example, since $s := \tilde{\mu}_{b_n}(x) - \mu(x) = O_{\rm p}(\tilde{\delta}_n)$ by Theorem 1, (34) now becomes

$$\begin{aligned}\tilde{L}_n(s, k_n\tilde{\delta}_n) &= \tilde{J}_n(s, k_n\tilde{\delta}_n) + \tilde{L}_n(0,0) - \tilde{J}_n(0,0) + O_{\rm p}[(nh_nk_n\tilde{\delta}_n\log n)^{1/2} + n^{1/\nu}h_nk_n\tilde{\delta}_n] \\ &= \tilde{L}_n/2 + nh_nk_n\tilde{\delta}_n\kappa_+ f_X(x)/\sigma(x)[1 + o_{\rm p}(1)]\end{aligned}$$

via (29) in Proposition 3 and Lemma 1. The rest of the proof follows the same lines as in the proof of Theorem 1 via Proposition 3 and Lemma 1. We omit the details. $\diamondsuit$

*Proof of Theorem 3.* Since the essential idea of the proof is same as that of Theorem 1, we only sketch necessary changes. To prove Theorem 3, the key step is to show that (28) and (29) in Proposition 3 hold uniformly over $\mathcal{B}_1 := \{(s,x) : |s| \le l_n, x \in \mathcal{T}\}$ and $\mathcal{B}_2 := \{(s,t,x) : |s| + |t| \le l_n, x \in \mathcal{T}\}$, respectively. Here we shall only prove the first expression since the second one similarly follows. As in (30), we use the $M/R$-decomposition. By Lemma 2(ii) and the argument in Proposition 3, the $R$-part is of order $O_{\rm p}(n^{1/2}b_nl_n)$ uniformly over $\mathcal{B}_1$. By Lemma 4 below, the $M$-part is of the uniform order $O_{\rm p}[(nb_nl_n\log n)^{1/2}]$. By the same $M/R$-decomposition technique and the argument in Lemma 4, we can show that $\sup_{x\in\mathcal{T}}[|L_n(0) - J_n(0)| + |L_n - J_n|] = O_{\rm p}[(nb_n\log n)^{1/2} + n^{1/2}b_n] = O_{\rm p}[(nb_n\log n)^{1/2}]$. The rest of the proof is as same as that of Theorem 1 with $\delta_n$ therein replaced by $\delta_n^{\rm unif} := [\log n/(nb_n)]^{1/2} + b_n^2$. We can similarly deal with $\hat{\sigma}_{h_n}(x)$. Details are omitted. $\diamondsuit$

**Lemma 4.** *Let $K \in \mathcal{K}, l_n \to 0, b_n \to 0$. For $x \in \mathbb{R}$ define*

$$N_n(x,t) = \sum_{i=1}^n [U_i - \mathbb{E}(U_i|\mathcal{G}_{i-2})], \text{ where } U_i = K_{b_n}(x - X_i)\mathbf{1}_{\mu(x)<Y_i\le\mu(x)+t}.$$

*Assume $\log n = O(nb_nl_n)$ and that (10) holds. Further assume $F_\varepsilon \in \mathcal{C}^1(\mathbb{R})$ and that there exists an $\epsilon > 0$ such that $\mu, \sigma \in \mathcal{C}^1(\mathcal{T}_\epsilon), \inf_{y\in\mathcal{T}_\epsilon}\sigma(y) > 0$ and $\sup_{y\in\mathcal{T}_\epsilon} f_1(y|\mathcal{F}_0) < c_0$ for some constant $c_0 < \infty$. Then*

$$\sup_{x\in\mathcal{T}, 0\le t\le l_n} |N_n(x,t)| = O_{\rm p}[(nb_nl_n\log n)^{1/2}]. \tag{35}$$

*Proof.* We shall apply the argument in the proof of Lemma 3. It suffices to verify (35) with $N_n(x,t)$ being replaced by $N_n^\circ(x,t) := \sum_{i:\text{even}}[U_i - \mathbb{E}(U_i|\mathcal{G}_{i-2})]$. Let $N = \lceil (nb_n^{-3}l_n^{-1})^{1/2} \rceil$ and $x_i = -T + 2iT/N$, $t_i = il_n/N, 0 \le i \le N$. For any $x \in \mathcal{T}$ and $t \in [0, l_n]$, there exist $j$ and $k$ such that $x \in [x_j, x_{j+1})$ and $t \in [t_k, t_{k+1})$. Let $c_1 = \sup_{x \neq y} |K(x) - K(y)|/|x-y| + \sup_{x\in\mathcal{T}} |\mu'(x)| < \infty$ and

$$\begin{aligned}
\overline{Z}_{ijk} &= K_{b_n}(x_j - X_i)[\mathbf{1}_{Y_i \le \mu(x_j) + 2c_1T/N + t_{k+1}} - \mathbf{1}_{Y_i \le \mu(x_j) - 2c_1T/N}],\\
\underline{Z}_{ijk} &= K_{b_n}(x_j - X_i)[\mathbf{1}_{Y_i \le \mu(x_j) - 2c_1T/N + t_k} - \mathbf{1}_{Y_i \le \mu(x_j) + 2c_1T/N}].
\end{aligned}$$

Since $|K_{b_n}(x - X_i) - K_{b_n}(x_j - X_i)| \le 2c_1T/(Nb_n)$ and $|\mu(x) - \mu(x_j)| \le 2c_1T/N$,

$$\underline{Z}_{ijk} - \frac{2c_1T}{Nb_n} \le U_i \le \overline{Z}_{ijk} + \frac{2c_1T}{Nb_n}.$$

Since $F_\varepsilon(y) \in \mathcal{C}^1(\mathbb{R})$ and $\sup_{y\in\mathcal{T}_\epsilon} f_1(y|\mathcal{F}_0) < c_0$, we have $\mathbb{E}(\overline{Z}_{ijk} - \underline{Z}_{ijk}|\mathcal{G}_{i-2}) = O(b_n/N)$, and consequently there exists a constant $c_2 < \infty$ such that

$$\underline{Z}_{ijk} - \mathbb{E}(\underline{Z}_{ijk}|\mathcal{G}_{i-2}) - \frac{c_2}{Nb_n} \le U_i - \mathbb{E}(U_i|\mathcal{G}_{i-2}) \le \overline{Z}_{ijk} - \mathbb{E}(\overline{Z}_{ijk}|\mathcal{G}_{i-2}) + \frac{c_2}{Nb_n}.$$

For the rest of the proof, the argument for proving Lemma 3 applies here. Details are omitted. ♢

*Proof of Corollaries 1 and 2.* The results follow from the $M/R$-decomposition technique as in (30). The $R$-part is negligible under the conditions specified there. For Corollary 1, by Proposition 4, the $M$-part is asymptotically normally distributed; for Corollary 2, we can extend Proposition 4 to cases with two bandwidths $b_n$ and $h_n$ by following the argument therein. ♢

REFERENCES


Bhattacharya, P.K and Gangopadhyay, A.K. (1990) Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31** 307–327.

Borkovec, M. and Klüppelberg, C. (2001) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Ann. Appl. Probab.* **11** 1220–1241.

Bossaerts, P., Härdle, W. and Hafner, C. (1996) Foreign exchange-rates have surprising volatility. *In Athens Conference on Applied Probability and Time Series*, Vol. 2 (Eds. Robinson, P.M.). Lecture Notes in Statistics 115, New York: Springer, 55–72.

Cai, Z.W. (2002) Regression quantiles for time series. *Econometric Theory* **18** 169–192.

Chan, K.C., Karolyi, A.G., Longstaff, F.A. and Sanders, A.B. (1992) An empirical comparison of alternative models of the short-term interest rate. *J. Finance* **47** 1209–1227.

Chaudhuri, P. (1991a) Global nonparametric estimation of conditional quantile functions and their derivatives. *J. Multivariate Anal.* **39** 246–269.

Chaudhuri, P. (1991b) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) A theory of the term structure of interest rates. *Econometrica* **53** 385–403.

Chen, R. and Liu, L.M. (2001) Functional Coefficient Autoregressive Models: Estimation and Tests of Hypotheses. *J. Time Ser. Anal.* **22** 151–173.

Chen, R. and Tsay, R.S. (1993) Functional-coefficient autoregressive models. *J. Am. Statist. Assoc.* **88** 298–308.

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

Fan, J. (2005) A selective overview of nonparametric methods in financial econometrics. *Statist. Sci.* **20** 317–337.

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

Fan, J. and Yao, Q. (1998) Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85** 645–660.

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

Fan, J. and Zhang, C. (2003) A reexamination of diffusion estimators with applications to financial model validation. *J. Am. Statist. Assoc.* **98** 118–134.

Fan, J. and Zhang, J. (2000) Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Stat. Soc. Ser. B* **62** 303–322.

Franke, J. and Mwita, P. (2003) Nonparametric estimates for conditional quantiles of time series. *Tech. report*, # 87, Dept. of Math., Univ. of Kaiserslautern, Germany.

Freedman, D.A. (1975) On tail probabilities for martingales. *Ann. Probab.* **3** 100–118.

Gouriéroux, Ch. and Monfort, A. (1992) Qualitative threshold ARCH models. *J. Econometrics* **52** 159–199.

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

Hall, P., Kay, J.W. and Titterington, D. M. (1990) Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.

Hall, P. and Carroll, R.J. (1989) Variance function estimation in regression: the effect of estimating the mean. *J. R. Stat. Soc. Ser. B* **51** 3–14.

Härdle, W. and Tsybakov, A.B. (1997) Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics* **81** 223–242.

Honda, T. (2000) Nonparametric estimation of a conditional quantile for $\alpha$-mixing processes. *Ann. Inst. Statist. Math.* **52** 459–470.

Jiang, J. and Mack, Y. (2001) Robust local polynomial regression for dependent data. *Statist. Sinica* **11** 705–722.

Jones, M.C. and Hall, P. (1990) Mean squared error properties of kernel estimates of regression quantiles. *Statist. Probab. Lett.* **10** 283–289.

Koenker, R. (2005) *Quantile Regression.* Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978) Regression quantiles. *Econometrica* **46** 33–50.

Masry, E. and Fan, J. (1997) Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* **24** 165–179.

McKeague, I.W. and Zhang, M.J. (1994) Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.* **22** 495-514.

Müller, H.C. and Stadtmüller, U. (1987) Estimation of Heteroscedasticity in Regression Analysis. *Ann. Statist.* **15** 610–625.

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **59** 347–370.

Nolan, J.P. (2001) Maximum likelihood estimation and diagnostics for stable distributions. *In Lévy Processes: Theory and Applications*, (Eds. Barndorff-Nielsen, O.E., Mikosch, T. and Resnick, S.I.), Birkhäuser, 379–400.

Resnick, S.I. (1997) Heavy tail modeling and teletraffic data. *Ann. Statist.* **25** 1805–1849.

Ruppert, D., Sheather, S.J. and Wand, M.P. (1995) An effective bandwidth selector for local least squares regresion. *J. Am. Statist. Assoc.*, **90** 1257–1270.

Samanta, M. (1989) Nonparametric estimation of conditional quantiles. *Statist. Probab. Lett.* **7** 407–412.

Shi, P. (1995) Asymptotic behaviour of nonparametric conditional quantile estimates for time series. *Canad. J. Statist.* **23** 161–169.

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.

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

Truong, Y.K. and Stone, C.J. (1992) Nonparametric function estimation involving time series. *Ann. Statist.* **20** 77–97.

Tsay, R.S. (2005) *Analysis of financial time series*, 2nd edition, Wiley, New York.

Wu, W.B. (2005a) Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci.* **102** 14150–14154.

Wu, W.B. (2005b) On the Bahadur representation of sample quantiles for stationary sequences. *Ann. Statist.* **33** 1934–1963.

Wu, W.B. and Shao, X. (2004) Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436.

Wu, W.B. and Yu, K. (2005) Kernel estimation of value-at-risk. Manuscript.

Wu, W.B. and Zhao, Z. (2005) Inference of trends in time series. Manuscript.

Yu, K. and Jones, M.C. (1998) Local linear quantile regression. *J. Am. Statist. Assoc.* **93** 228–237.

Yu, K., Lu, Z. and Stander, J. (2003) Quantile regression: applications and current research areas. *The Statistician* **52** 331–350.

Ziegelmann, F.A. (2002) Nonparametric estimation of volatility functions: the local exponential estimator. *Econometric Theory* **18** 985–991.

Ziegelmann, F.A. (2005) A nonparametric least-absolute-deviations estimator of volatility functions. In *XXVII Encontro Brasileiro de Econometria*, Natal.

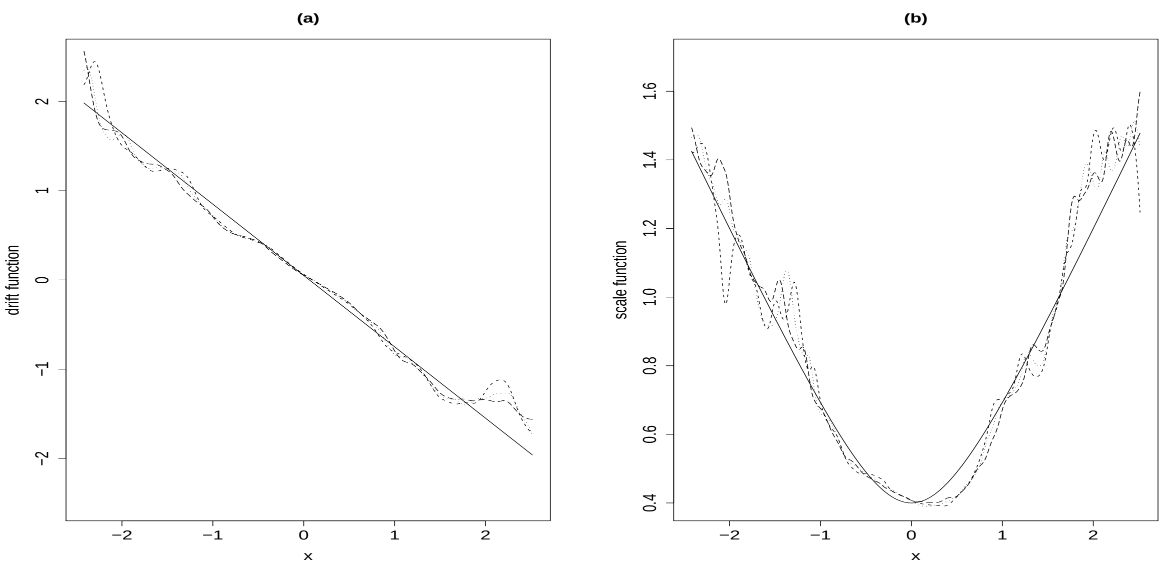


Figure 1: Estimated drift function $\mu(\cdot)$ (plot a) and scale function $\sigma(\cdot)$ (plot b) for the simulated model (26). In (a), solid line is the true drift $\mu(x) = 0.05 - 0.8x$, dashed, dotted and long-dashed lines correspond to quantile estimates of $\mu(\cdot)$ with bandwidth $b_n = 0.5, 0.6$ and $0.7$, respectively. In (b), fixed $b_n = 0.5$ is used for estimating $\mu(\cdot)$, solid line is the true scale $\sigma(x) = 0.4(1+2x^2)^{1/2}$, dashed, dotted and long-dashed lines correspond to quantile estimates of $\sigma(\cdot)$ with bandwidth $h_n = 0.5, 0.6$ and $0.7$, respectively.

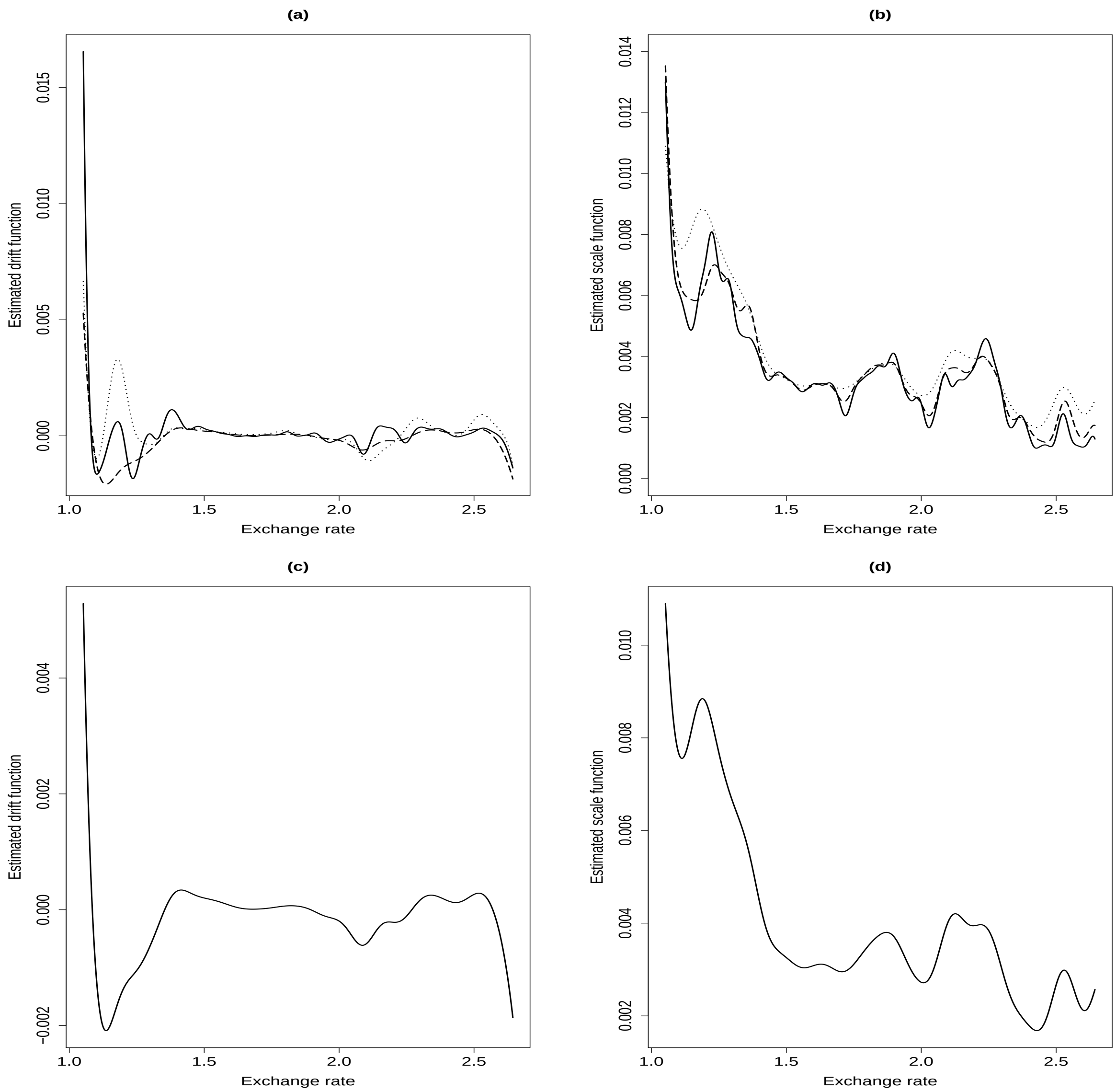


Figure 2: Quantile estimates of the drift function $\mu(\cdot)$ (plots a and c) and scale function $\sigma(\cdot)$ (plots b and d) for Pound/USD. In (a), dotted, long-dashed and solid lines correspond to bandwidth $b_n = 0.005, 0.010$ and $0.015$, respectively; in (b), fixed $b_n = 0.010$ is used for estimating $\mu(\cdot)$, dotted, long-dashed and solid lines correspond to bandwidth $h_n = 0.005, 0.010$ and $0.015$, respectively; in (c), $b_n = 0.010$; in (d), $b_n = 0.010$ and $h_n = 0.005$.

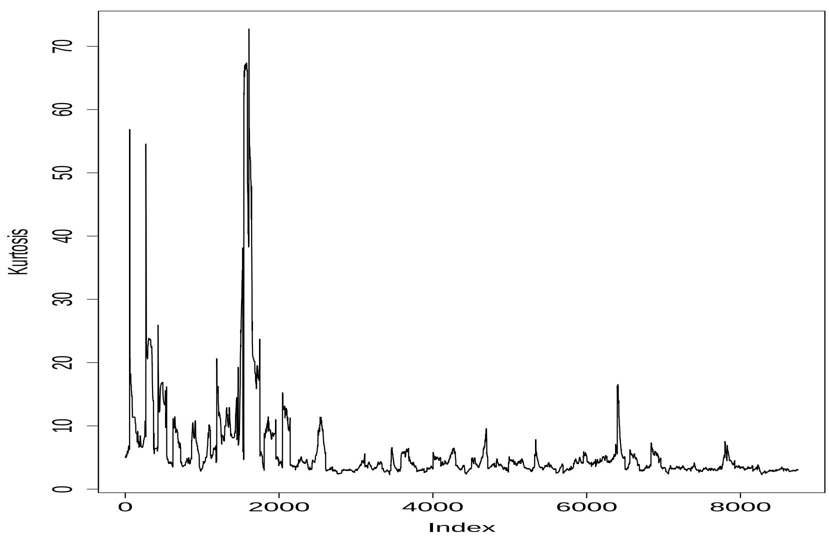


Figure 3: Sample kurtosis for the estimated innovations $\hat{\varepsilon}_i$. The innovations are estimated using bandwidths $b_n = 0.010$ and $h_n = 0.005$. The kurtosises are obtained by binning the data and computing sample kurtosis for each bin. Here we have used the window size 100.