TILTED NONPARAMETRIC ESTIMATION OF VOLATILITY FUNCTIONS

By

Peter C.B. Phillips and Ke-Li Xu

June 2007

COWLES FOUNDATION DISCUSSION PAPER NO. 1612

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Tilted Nonparametric Estimation of Volatility Functions*

PETER C. B. PHILLIPS† AND KE-LI XU‡

May 22, 2007

Abstract

This paper proposes a novel positive nonparametric estimator of the conditional variance function without relying on a logarithmic transformation. The basic idea is to apply the reweighted Nadaraya-Watson regression estimator of Hall and Presnell (1999, *Journal of the Royal Statistical Society B,* 61, 143–158) to squared residuals. The new conditional variance estimator is asymptotically equivalent to the local linear estimator and is restricted to be positive in finite samples. A small simulation is performed to compare the new methodology with Ziegelmann's (2002) local exponential and Yu and Jones's (2004) local likelihood-based estimators of the conditional variance.

Keywords: Conditional variance function; Empirical likelihood; Heteroskedasticity; Local linear estimator; Nadaraya-Watson estimator; Nonlinear time series; Nonparametric regression; Volatility.

JEL Classification: C22

*The authors thank Donald Andrews, Zongwu Cai, Yuichi Kitamura, and Taisuke Otsu for helpful comments. Phillips acknowledges partial research support from a Kelly Fellowship and the NSF under Grant Nos. SES 04-142254 and 06-47086.

†Department of Economics, Cowles Foundation for Research in Economics, Yale University, P. O. Box 208281, New Haven, CT 06520, USA. *E-mail address:* `peter.phillips@yale.edu`.

‡Program in Applied Mathematics, Department of Mathematics, Yale University, Room 111, 51 Prospect Street, New Haven, CT 06520, USA. *E-mail address:* `keli.xu@yale.edu`.

1 Introduction

Nonparametric approaches provide flexible alternatives to traditional parametric modeling methods. Consider the following nonparametric heteroskedastic regression model

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

where $\{X_i, Y_i, i = 1, \cdots, n\}$ are observations on two random variables $\{X, Y\}$, and $\{\varepsilon_i\}$ are innovations satisfying $E(\varepsilon_i|X_i) = 0$, $Var(\varepsilon_i|X_i) = 1$. The conditional mean function $m(x) = E(Y|X = x)$ and the conditional variance function $\sigma^2(x) = Var(Y|X = x) > 0$ (or volatility function $\sigma(x)$) are left unspecified and are the main interests of statistical investigation. The model (1) is of fundamental importance in financial econometrics due to its ability to allow nonlinearity and conditional heteroskedasticity in financial time series (Engle, 1982, Tong, 1990, Bossaerts *et al*, 1996). It can also be regarded as the discretized version of the continuous-time stochastic diffusion model which is commonly used in financial derivative pricing (Stanton, 1997, Bandi and Phillips, 2003).

This paper focuses on estimation of the conditional variance function $\sigma^2(\cdot)$, which is crucial in inference for the conditional mean function $m(\cdot)$, constructing confidence intervals and selecting data-driven bandwidths. It is also of great importance in practical applications, e.g., volatility or risk measurement in finance (Shephard, 2005). See Martins-Filho and Yao (2006a,b) for recent applications of conditional variance estimation in the estimation of value-at-risk and expected shortfall functions for a financial asset and production frontiers. Other applications of variance estimation are discussed in Carroll and Ruppert (1988). Earlier contributions on the estimation of $\sigma^2(\cdot)$ in nonparametric contexts include Carroll (1982), Müller and Stadtmüller (1987) and Hall and Carroll (1989) among others, which are mainly concerned with *iid* data applications.

Recently, assuming $X_i = Y_{i-1}$, Härdle and Tsybakov (1997) proposed a local polynomial estimation procedure of $\sigma^2(\cdot)$ based on a variance decomposition. Their method was criticized by Fan and Yao (1998) who pointed out that it is not fully adaptive to the unknown mean function, with the bias term depending on the derivative of $m(\cdot)$. To tackle this difficulty, Fan and Yao (1998) proposed a residual-based fully adaptive conditional variance estimator (see also Ruppert, Wand, Holst and Hössjer, 1997).[1] Applying the local linear technique to squared residuals of a first-stage nonparametric mean function modeling, they showed that the resulting variance estimator is asymptotically as efficient as the oracle estimator, which assumes knowledge of the mean function $m(\cdot)$. Although fitting a linear function locally seems appealing when compared with conventional local level (Nadaraya-Watson) estimation, as is well demonstrated in Fan and Gijbels (1996), it is not guaranteed to give positive values in finite samples when estimating a variance. The tendency to produce negative variance estimates is especially the case when large bandwidths are used. The negativity problem can lead to many practical difficulties. In consequence, it is commonly recommended in applications to use the (theoretically inferior) Nadaraya-Watson estimator when fitting conditional variance functions (Chen and Qin, 2002), especially at design points where the local linear estimators give negative results.

The dilemma has been widely recognized among practitioners, and some efforts suggesting alternatives to local linear volatility estimators have been made. Ziegelmann (2002) proposed to fit an exponential function locally (rather than a linear function in the local linear estimator) within the general locally parametric nonparametric framework of Hjort and Jones (1996). More recently, assuming *iid* Gaussian errors, Yu and Jones (2004) maximized the localized likelihood where the mean and variance function are parameterized locally as a linear function and an exponential function respectively. In both methods of Ziegelmann and Yu and Jones, the logarithm of the variance, rather than the variance itself, is estimated so that the resulting estimator is always positive without any restrictions. However, there seems little intuitive justification for fitting exponential forms without knowledge of the underlying model. Furthermore, the introduction of a logarithmic transformation produces one more term in the bias expression, which will increase the squared bias and subsequently the mean square error, *e.g.* when the second derivative of the conditional variance function is negative.

This paper proposes a novel conditional variance function estimator that preserves the appealing properties of local linear estimators while being always positive. It is based on the

[1] See also Dahl and Levine (2006), who dealt with nonparametric volatility estimation with serially dependent innovations ε_i.

intentionally biased bootstrap method due to Hall and Presnell (1999). The idea is to adjust the conventional Nadaraya-Watson estimator by minimally tilting the empirical distribution subject to a discrete bias reducing moment condition satisfied by the local linear estimator. The new estimator, which is called the *re-weighted Nadaraya-Watson* or *tilted estimator* here, inherits the non-negativity restriction of the variance function from the usual Nadaraya-Watson estimator, while possessing the superior properties of bias, boundary correction and minimax efficiency of the local linear estimator. Unlike the logarithmic transformation-based variance estimators mentioned above, the proposed estimator has a closed form and no multivariate optimization is needed. So the estimator is very easy to use. Furthermore, the estimator is constructed without knowledge of the error distribution and is robust to possible mis-specification.

The re-weighting idea used here is also useful in other contexts, *e.g.* in estimating the regression function (Hall and Presnell 1999, and Cai, 2001), the conditional distribution function (Hall, Wolff and Yao, 1999), quantiles (Cai, 2002), the conditional density function (De Gooijer and Zerom, 2003), and in continuous-time functional diffusion estimation (Xu, 2006). The same methodology is also used in regression function estimation subject to monotonicity restrictions (Hall and Huang, 2001).

The remainder of the paper is organized as follows. Section 2.1 describes the residual-based re-weighted Nadaraya-Watson estimator of the conditional variance. Section 2.2 develops the asymptotic distribution theory for the proposed estimator at both interior and boundary points and suggests a consistent estimator of the asymptotic variance. We compare the new estimator with the competitors suggested by Ziegelmann (2002) and Yu and Jones (2004) in Section 3 and a simulation experiment is reported in Section 4. Section 5 concludes and all proofs are collected in the appendix.

2 Main Results

2.1 The estimator

The residual-based nonparametric estimator of the conditional variance function $\sigma^2(\cdot)$ is built on a first-stage kernel-weighted least squares estimate of the conditional mean function $m(\cdot)$. Let $W(\cdot)$ and $K(\cdot)$ be kernel functions and $h_1 = h_1(n), h_2 = h_2(n) > 0$ be bandwidth parameters determining the complexity of the model. The local linear method solves

$$(\widehat{\gamma}_1, \widehat{\gamma}_2) = \arg\min_{(\gamma_1,\gamma_2)} \sum_{i=1}^{n} \Big(Y_i - \gamma_1 - \gamma_2(X_i - x)\Big)^2 W\left(\frac{X_i - x}{h_2}\right) \tag{2}$$

and then estimates $m(x)$ by $\widehat{m}(x) = \widehat{\gamma}_1$, for a design point x. The use of different bandwidths in mean and variance estimation has been stressed by several authors (Ruppert *et al,* 1997, and Yu and Jones, 2004), and we use h_2 in mean regression (2) and h_1 in variance estimation in what follows.

Instead of fitting the squared residuals $\widehat{r}_i^2 = \Big(Y_i - \widehat{m}(X_i)\Big)^2$ to X_i using a second-stage local linear smoother as in Ruppert *et al* (1997) and Fan and Yao (1998), we consider the following re-weighted Nadaraya-Watson estimator of $\sigma^2(x)$:

$$\widehat{\sigma}^2(x) = \frac{\sum_{i=1}^n \widehat{w}_i(x) K\left(\frac{X_i - x}{h_1}\right) \widehat{r}_i^2}{\sum_{i=1}^n \widehat{w}_i(x) K\left(\frac{X_i - x}{h_1}\right)}, \tag{3}$$

where $\widehat{w}_i(x)$ solves

$$\{\widehat{w}_i(x)\} = \arg\max_{\{w_i(x)\}} l_n(w_1(x), \cdots, w_n(x)), \tag{4}$$

with $l_n(w_1(x), \cdots, w_n(x)) = \sum_{i=1}^n \log w_i(x)$, subject to the restrictions $w_i(x) \geq 0$, $\sum_{i=1}^n w_i(x) = 1$ and

$$\sum_{i=1}^{n} w_i(x)(X_i - x)K_{h_1}(X_i - x) = 0, \tag{5}$$

where $K_h(\cdot) = K(\cdot/h)/h$. See Hall and Presnell (1999) and Cai (2001) for the motivation of such re-weighted Nadaraya-Watson (NW) estimators. One may interpret $\widehat{w}(x) = (\widehat{w}_1(x), \cdots, \widehat{w}_n(x))$ as revised estimates of the probability masses placed on observations $(X_1, \cdots, X_n)$ and $l_n(w_1(x), \cdots, w_n(x))$

as the log empirical likelihood (Owen, 1988, 1990, 2001).

The estimator defined jointly by (3), (4) and (5) belongs to a wide class of re-weighted estimators of the form

$$\widehat{g}(x) = \frac{\sum_{i=1}^{n} \widehat{w}_i(x) A_i(x) Y_i}{\sum_{i=1}^{n} \widehat{w}_i(x) A_i(x)}, \tag{6}$$

where $A_i(x)$ is the original weighting function and Y_i is the response variable (Hall and Huang, 2001). The probability vector $\widehat{w}(x)$ is chosen to minimize the distance $D(w(x))$ from the uniform distribution, $w_{unif}(x) = (1/n, \cdots, 1/n)$ subject to desirable constraints, thereby assuring that the original estimator $\sum_{i=1}^{n} A_i(x)Y_i / \sum_{i=1}^{n} A_i(x)$ is modified to the least extent needed to satisfy the constraints. Several distance measures are discussed in Hall and Huang (2001). Here we adapt the re-weighted estimator (6) in the conditional variance estimation context with $A_i(x) = K\left(\frac{X_i - x}{h_1}\right)$ and $Y_i = \widehat{r}_i^2$, and choose the distance $D(w(x)) = -2\sum_{i=1}^{n} \log(nw_i(x))$ which is the empirical log-likelihood ratio.

The constraint defined in (5) is a discrete bias-reducing moment condition (5) satisfied by the local linear smoothing weights[2] in which case $w_i(x) = S_{n,2} - (X_i - x)S_{n,1}$, where $S_{n,j} = \sum_{i=1}^{n} (X_i - x)^j K_h(X_i - x)$, $j = 1, 2$. Thus, we expect the constructed estimator to behave like the local linear estimator while preserving the non-negative weights of the NW estimator. Without the constraint (5), the re-weighted NW estimator reduces to the usual NW estimator since $l_n(w_1(x), \cdots, w_n(x))$ is maximized at $w_{unif}(x) = (1/n, \cdots, 1/n)$ for probability weights $w_1(x), \cdots, w_n(x)$. We can also choose weights $\widehat{w}_i(x)$ subject to the constraint $d\left(\widehat{\sigma}^2(x)\right)/dx \geq 0$ or $d\left(\widehat{\sigma}^2(x)\right)^2/dx^2 \geq 0$ to ensure monotonicity or convexity of the variance function as needed.

A closed-form expression of the weights $\widehat{w}_i(x)$ in (3) can be obtained via the Lagrange multiplier method, *viz.*

$$\widehat{w}_i(x) = \frac{1}{n\left(1 + \lambda(X_i - x)K_{h_1}(X_i - x)\right)}, \tag{7}$$

[2] See section 3.3.2 in Fan and Gijbels (1996).

where the Lagrange multiplier λ satisfies

$$\sum_{i=1}^{n} \frac{(X_i - x)K_{h_1}(X_i - x)}{1 + \lambda(X_i - x)K_{h_1}(X_i - x)} = 0. \tag{8}$$

2.2 Limit theory

To derive the asymptotic distribution of $\widehat{\sigma}^2(x)$, we make the following assumptions. Let $f(\cdot)$ be the density function of X.

Assumption

(i) For a given design point x, the functions $f(x) > 0$, $\sigma^2(x) > 0$, $E(Y^3|X = x)$ and $E(Y^4|X = x)$ are continuous at x, and $\ddot{m}(z) = d^2m(z)/dz^2$ and $\ddot{\sigma}^2(z) = d^2(\sigma^2(z))/dz^2$ are uniformly continuous on an open set containing x;

(ii) For some $\delta \geq 0$ such that $EY^{4(1+\delta)} < \infty$;

(iii) There exists a constant $M < \infty$ such that $|g_{1,t}(y_1, y_2|x_1, x_2)| \leq M$ for all $t \geq 2$, where $g_{1,t}(y_1, y_2|x_1, x_2)$ is the conditional density of Y_1 and Y_t given $X_1 = x_1$ and $X_t = x_2$;

(iv) The kernel functions $W(\cdot)$ and $K(\cdot)$ are symmetric density functions each with a bounded support $[-1, 1]$. We also assume a Lipschitz condition is satisfied by each of functions $f(\cdot)$, $W(\cdot)$ and $K(\cdot)$;

(v) The process $\{(X_i, Y_i)\}$ are strictly stationary and absolutely regular[3] with mixing coefficients $\beta(j)$ satisfying $\sum_{j=1}^{\infty} j^2\beta^{\delta/(1+\delta)}(j) < \infty$, where δ is the same as in (ii);

(vi). As $n \to \infty$, $h_i \to 0$ and $\liminf_{n\to\infty} nh_i^4 > 0$ for $i = 1, 2$.

The asymptotic distribution of the re-weighted NW estimator of the conditional variance is given in the following theorem both at interior and boundary points.

Theorem 1. *(i) Suppose that x is such that $x \pm h_1$ is in the support of $f(x)$. Under the*

[3] See, e.g., Davidson (1994) (page 209) for the definition of an absolutely regular process.

Assumption stated, as $n \to \infty$,

$$\sqrt{nh_1}\left(\widehat{\sigma}^2(x) - \sigma^2(x) - \frac{h_1^2}{2}K_1\ddot{\sigma}^2(x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{K_2\sigma^4(x)\xi^2(x)}{f(x)}\right), \tag{9}$$

where $K_1 = \int_{-1}^1 u^2K(u)du$, $K_2 = \int_{-1}^1 K^2(u)du$, $\xi^2(x) = E\left((\varepsilon^2-1)^2|X=x\right)$, $\varepsilon = \frac{Y-m(X)}{\sigma(X)}$.

(ii) Suppose that $f(x)$ *has bounded support* $[a,b]$ *and* $x = a + ch_1$ $(0 < c < 1)$*. Under the Assumption stated, as* $n \to \infty$,

$$\sqrt{nh_1}\left(\widehat{\sigma}^2(a+ch_1) - \sigma^2(a+ch_1) - \frac{h_1^2\overline{K}_1}{2\overline{K}_0}\ddot{\sigma}^2(a+ch_1)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\overline{K}_2\sigma^4(a)\xi^2(a)}{\overline{K}_0^2 f(a)}\right), \tag{10}$$

where $\overline{K}_0 = \int_{-1}^c \frac{K(u)du}{1-\lambda_c uK(u)}, \overline{K}_1 = \int_{-1}^c \frac{u^2K(u)du}{1-\lambda_c uK(u)}, \overline{K}_2 = \int_{-1}^c \left(\frac{K(u)}{1-\lambda_c uK(u)}\right)^2 du$ *and* λ_c *satisfies* $L_c(\lambda_c) = 0$ *with*

$$L_c(\lambda) = \int_{-1}^c \frac{uK(u)}{1-\lambda uK(u)}du.$$

Remarks. Theorem 1 shows that the re-weighted NW conditional variance estimator is asymptotically fully adaptive to the unknown conditional mean function, a property that is shared by other residual-based conditional variance estimators (see Ruppert *et al*, 1997, Fan and Yao, 1998, Ziegelmann, 2002 and Yu and Jones, 2004). Theorem 1 also shows that the bias of the re-weighted NW estimator on the boundary is of the same order as the bias in the interior and thus no boundary correction is needed.

This feature can be better appreciated through the following heuristic argument. From the proof of Theorem 1 in the appendix, the bias of $\widehat{\sigma}^2(x)$ is approximately accounted for by the term $(nh_1)^{-1}\sum_{i=1}^n p_i(x)K\left(\frac{X_i-x}{h_1}\right)\left(\sigma^2(X_i) - \sigma^2(x)\right)$, neglecting terms of smaller order, where $p_i(x) = \left(\sum_{i=1}^n \widehat{w}_i(x)K\left(\frac{X_i-x}{h_1}\right)\right)^{-1}\sum_{i=1}^n \widehat{w}_i(x)$. By a second-order Taylor expansion of $\sigma^2(X_i)$ at x and the discrete moment condition (5),

$$
\begin{aligned}
&\frac{1}{nh_1}\sum_{i=1}^n p_i(x)K\left(\tfrac{X_i-x}{h_1}\right)\left(\sigma^2(X_i)-\sigma^2(x)\right)\\
=\ &\frac{1}{nh_1}\sum_{i=1}^n p_i(x)K\left(\tfrac{X_i-x}{h_1}\right)\left(\tfrac{1}{2}\ddot{\sigma}^2(x)(x-X_i)^2\right)+\text{higher order terms}\\
=\ &\begin{cases}\frac{h_1^2}{2}f(x)K_1\ddot{\sigma}^2(x)+o_p(h_1^2), & \text{if } x \text{ is in the interior;}\\ \frac{h_1^2\overline{K}_1 f(a)\ddot{\sigma}^2(a+ch_1)}{2}+o_p(h_1^2), & \text{if } x \text{ is on the boundary.}\end{cases}
\end{aligned}
$$

The bias term of order h_1 is eliminated by the condition (5) for any n both at interior and boundary points just as the local linear smoother. It is different from the conventional NW estimator which eliminates the bias term of order h_1 in the limit by the symmetry of the kernel function for interior points, but this term does not vanish for boundary points. The constant λ_c is decreasing in c and approaches zero when c goes to 1. Theorem 1 (ii) also holds for $c \geq 1$, *viz.* x is in the interior, by noting $\overline{K}_0$ and $\overline{K}_i$ $(i = 1, 2)$ reduce to 1 and K_i $(i = 1, 2)$, respectively. For the right boundary point $b - ch$, a similar result in (10) holds.

The following theorem gives a consistent estimator of the asymptotic variance of $\widehat{\sigma}^2(x)$ both at interior and boundary points, thereby allowing construction of consistent point-wise confidence intervals. Let

$$\widehat{\Omega}(x) = \widehat{f}^{-2}(x)\widehat{V}(x),$$

where

$$\widehat{V}(x) = \frac{n}{h_1}\sum_{i=1}^n \widehat{w}_i^2(x)K^2\left(\tfrac{X_i-x}{h_1}\right)(\widehat{r}_i^2-\widehat{\sigma}^2(x))^2,\ \widehat{f}(x) = \frac{1}{h_1}\sum_{i=1}^n \widehat{w}_i(x)K\left(\tfrac{X_i-x}{h_1}\right),$$

with $\widehat{w}_i(x)$ defined in (7).

Theorem 2. *Assume $EY^{8(1+\delta)} < \infty$ for some $\delta \geq 0$.*

(i) Under the conditions of Theorem 1 (i), as $n \to \infty$, $\widehat{\Omega}(x) \xrightarrow{p} \frac{K_2\sigma^4(x)\xi^2(x)}{f(x)}$;

(ii) Under the conditions of Theorem 1 (ii), as $n \to \infty$, $\widehat{\Omega}(x) \xrightarrow{p} \frac{\overline{K}_2 \sigma^4(a)\xi^2(a)}{\overline{K}_0^2 f(a)}$.

3 Other Estimators

Several positive nonparametric conditional variance estimators have been proposed as alternatives to the local linear estimator. Ziegelmann (2002)'s residual-based local exponential (LE) conditional variance estimator, denoted by $\widehat{\sigma}^2_{LE}$, belongs to a wide class of local nonlinear estimators (Hjort and Jones, 1996, Gozalo and Linton, 2000). Generally, it has the form $\vartheta(z, \widehat{\varphi})$ where ϑ is a known function and $\widehat{\varphi}$ minimizes the following sum of weighted squares, viz.

$$\widehat{\varphi} = \min_{\varphi} \sum_{i=1}^{n} \left(\widehat{r}_i^2 - \vartheta(X_i, \varphi) \right)^2 K\left(\tfrac{X_i - x}{h_1} \right),$$

where the $\widehat{r}_i^2$'s are the squared residuals from the nonparametric regression. To ensure positivity of the resultant conditional variance estimator, Ziegelmann (2002) proposed to use the exponential function $\vartheta(z, \varphi) = \exp(\varphi_1 + \varphi_2 z)$ rather than the linear function $\vartheta(z, \varphi) = \varphi_1 + \varphi_2 z$ as in local linear smoothers. With this re-parameterization,

$$\widehat{\sigma}^2_{LE} = \exp(\widehat{\psi}_1), \tag{11}$$

where $(\widehat{\psi}_1, \widehat{\psi}_2)$ solves

$$\arg\min_{(\psi_1, \psi_2)} \sum_{i=1}^{n} \left(\widehat{r}_i^2 - \exp(\psi_1 + \psi_2(X_i - x)) \right)^2 K\left(\tfrac{X_i - x}{h_1} \right).$$

In a recent paper, Yu and Jones (2004) adopted a local maximum likelihood (LML) framework (Staniswalis, 1989, Fan, Farmen, and Gijbels, 1998, and Fan and Chen, 1999) and proposed a slightly different estimator, denoted by $\widehat{\sigma}^2_{LML}$. Assuming *iid* Gaussian errors, given $X = x$, a localized normal log-likelihood for estimating the conditional mean and variance functions is

$$-\frac{1}{h} \sum_{i=1}^{n} \left(\tfrac{(Y_i - c(X_i))^2}{d(X_i)} + \log(d(X_i)) \right) K\left(\tfrac{X_i - x}{h} \right), \tag{12}$$

where $c(\cdot)$ and $d(\cdot)$ are functions to be fitted locally. The local maximum likelihood estimator amounts to maximizing (12) after an appropriate parameterization of the functions $c(\cdot)$ and $d(\cdot)$. Yu and Jones (2004) used a shortcut version by replacing $(Y_i - c(X_i))^2$ as $\widehat{r}_i^2$ and applied a linear form for the logarithm of $d(\cdot)$, viz.,

$$\widehat{\sigma}^2_{LML} = \exp(\widehat{d}_1), \tag{13}$$

where $(\widehat{d}_1, \widehat{d}_2)$ solves

$$\arg\min_{(d_1,d_2)} = \sum_{i=1}^{n} \left(\widehat{r}_i^2 \exp(-d_1 - d_2(X_i - x)) + d_1 + d_2(X_i - x) \right) K\left(\tfrac{X_i - x}{h_1} \right).$$

The local exponential estimator $\widehat{\sigma}^2_{LE}$ and local maximum likelihood estimator $\widehat{\sigma}^2_{LML}$ have the same asymptotic variance as that of the local linear estimator but with one extra term in the bias. As mentioned before, both estimators essentially estimate the logarithm of the variance, rather than the variance itself to ensure positivity. This logarithmic transformation complicates the bias term, which may have a negative effect depending on the nature of the true variance function (e.g., see the discussion in Yu and Jones, 2004, Table 1, page 141).

4 Simulations

This section reports a brief simulation experiment comparing the small-sample performance of the following three positive estimators of the conditional variance function: the re-weighted NW (RNW) proposed here, Ziegelmann's local exponential (LE) and Yu and Jones's local maximum likelihood (LML) estimators, given by (3), (11) and (13) respectively, along with the local linear estimator as a benchmark. The following model is used for the simulation design:

$$Y_i = 0.5(X_i + 2\exp(-16X_i^2)) + (0.4\exp(-2X_i^2) + 0.2)\varepsilon_i, \tag{14}$$

where $X_i \overset{iid}{\sim} Unif(-2, 2)$ and $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$. The Epanechnikov kernel function is used for both W and K :

$$K(u) = \begin{cases} 0.75(1-u^2), & -1 < u < 1 \\ 0, \text{ otherwise} \end{cases} .$$

The bandwidth parameter for mean estimation, h_2, is selected by least squares cross-validation.

This design is the same as that used in Example 2 of Fan and Yao (1998). Each estimator is evaluated at 37 equally spaced grid points on $[-1.8, 1.8]$ with 500 replications. Table 1 reports the boxplots of the mean absolute deviation errors (MAD), $MAD = \frac{1}{37}\sum_{i=1}^{37} |\widehat{\sigma}^2(x_i) - \sigma^2(x_i)|$, where x_i is the $i-$th grid point, for the four estimators with 4 fixed bandwidths $h = 0.4, 0.5, 0.6, 0.7$ when the sample size is 100. We can see that the RNW method has very similar performance to that of its asymptotic analog local linear estimator. Compared with the LE and LML estimators, the RNW estimator has less bias over all bandwidths considered. These outcomes are not surprising since a logarithmic transformation leads to an adverse effect on the quality of estimation. We also consider the case with heavy-tailed errors (Table 2) when ε_i has a common t distribution with 5 degrees of freedom (the variance is normalized as unity). The advantage of the RNW estimator over its two competitors is similar in this case.

5 Summary

This paper provides a new nonparametric approach to estimating the conditional variance function based on the maximization of the empirical likelihood subject to a bias-reducing moment restriction. It is fully adaptive to the unknown mean function. Its construction does not depend on the error distribution, and it is applicable in quite general time series settings. The new estimator preserves the appealing design adaptive, bias and automatic boundary correction properties of the local linear estimator, while it is guaranteed to be non-negative in small samples. Moreover, compared with other positive variance estimators previously proposed in the literature, it has a simpler form and is easier to use. Simulation results suggest that the new estimator has good performance characteristics in finite samples and is a promising competitor in estimating conditional variance functions.

6 Appendix

This section proves Theorems 1 and 2. For simplicity, we write the weights in the re-weighted NW estimator $\widehat{w}_i(x)$ as w_i.

Proof of Theorem 1. Recall that $\widehat{r}_i = Y_i - \widehat{m}(X_i) = m(X_i) - \widehat{m}(X_i) + \sigma(X_i)\varepsilon_i$ and so

$$\widehat{r}_i^2 = \sigma^2(X_i)\varepsilon_i^2 + 2\sigma(X_i)\varepsilon_i\Big(m(X_i) - \widehat{m}(X_i)\Big) + \Big(m(X_i) - \widehat{m}(X_i)\Big)^2. \tag{15}$$

Thus, by (3)

$$\widehat{\sigma}^2(x) - \sigma^2(x) = \sum_{j=1}^{4} T_j, \tag{16}$$

where

$$T_1 = \frac{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)\sigma^2(X_i)(\varepsilon_i^2 - 1)}{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)},\ T_2 = \frac{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)\Big(\sigma^2(X_i) - \sigma^2(x)\Big)}{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)},$$

$$T_3 = \frac{2\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)\sigma(X_i)\varepsilon_i\Big(m(X_i) - \widehat{m}(X_i)\Big)}{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)},$$

and

$$T_4 = \frac{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)\Big(m(X_i) - \widehat{m}(X_i)\Big)^2}{\sum_{i=1}^n w_i K\left(\frac{X_i - x}{h_1}\right)}.$$

(i). Suppose that x is such that $x \pm h$ is in the support of $f(x)$. Since an absolutely regular time series is $\alpha-$mixing, Lemma A2 in Cai (2001) holds under our assumptions, i.e. $\lambda = -\frac{h_1 K_1 f'(x)}{\upsilon_2 f(x)} + O_{a.s.}(h_1^3)$, where $\upsilon_2 = \int u^2 K^2(u)du$, and

$$w_i(x) = n^{-1}\left(1 - \frac{h_1 K_1 f'(x)}{\upsilon_2 f(x)}(X_i - x)K_{h_1}(X_i - x)\right)^{-1}(1 + o_p(1)), \tag{17}$$

Consider the term T_2 first. The denominator of T_2 times $1/h_1$ is

$$\frac{1}{h_1}\sum_{i=1}^{n} w_i K\left(\tfrac{X_i-x}{h_1}\right) = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\tfrac{X_i-x}{h_1}\right) + o_p(1) \xrightarrow{p} f(x), \tag{18}$$

by (17) and an application of Birkhoff's ergodic theorem (*e.g.* Shiryaev, 1995) provided that $E\left(\frac{1}{h_1}K\left(\frac{X_i-x}{h_1}\right)\right) = \frac{1}{h_1}\int K\left(\frac{u-x}{h_1}\right) f(u)du \to f(x)$ as $h_1 \to 0$ after a change of variables. By a Taylor expansion of $\sigma^2(X_i)$ at x and the discrete moment condition (5), the numerator of T_2 times $1/h_1$ is

$$\begin{aligned}
&\frac{1}{h_1}\sum_{i=1}^{n} w_i K\left(\tfrac{X_i-x}{h_1}\right)\left(\sigma^2(X_i) - \sigma^2(x)\right) \\
=\;& \frac{1}{h_1}\sum_{i=1}^{n} w_i K\left(\tfrac{X_i-x}{h_1}\right)\left(\frac{1}{2}\ddot{\sigma}^2(x)(X_i-x)^2 + o((X_i-x)^2)\right) \\
=\;& \frac{h_1^2}{2} f(x) K_1 \ddot{\sigma}^2(x) + o_p(h_1^2),
\end{aligned} \tag{19}$$

by (17) and the ergodic theorem. Combining (18) and (19) gives $T_2 = \frac{h_1^2}{2}K_1\ddot{\sigma}^2(x) + o_p(h_1^2)$. Noting (17) and (18), it follows from (b)-(d) in the proof of Theorem 1 in Fan and Yao (1998) that $\sqrt{nh_1}T_1 \xrightarrow{d} \mathcal{N}\left(0, \frac{K_2\sigma^4(x)\xi^2(x)}{f(x)}\right)$, and $T_3, T_4 = o_p(h_1^2 + h_2^2)$. Hence, by (16) the proof of (i) is complete.

(ii). Suppose that $f(x)$ has a bounded support $[a, b]$ and $x = a + ch_1$ $(0 < c < 1)$. By Lemma A.3 in Cai (2001),

$$w_i = \frac{1}{n(1 - \lambda_c(X_i - a - ch_1)K_h(X_i - a - ch_1))}(1 + o_p(1)).$$

Consider the term T_2 in (16). Note that

$$\frac{1}{h_1}\sum_{i=1}^{n} w_i K\left(\tfrac{X_i-a-ch_1}{h_1}\right) = \frac{1}{nh_1}\sum_{i=1}^{n} \frac{K\left(\frac{X_i-a-ch_1}{h_1}\right)}{1 - \lambda_c(X_i - a - ch_1)K_{h_1}(X_i - a - ch_1)} + o_p(1) \xrightarrow{p} \overline{K}_0 f(a), \tag{20}$$

by the ergodic theorem provided that

$$
\begin{aligned}
E\left(\frac{1}{h_1}\frac{K\left(\frac{X_i-a-ch_1}{h_1}\right)}{1-\lambda_c(X_i-a-ch_1)K_{h_1}(X_i-a-ch_1)}\right) &= \int_a^b \frac{1}{h_1}\frac{K\left(\frac{z-a-ch_1}{h_1}\right)}{1-\lambda_c(z-a-ch_1)K_{h_1}(z-a-ch_1)}f(z)dz \\
&\to \int_{-1}^{c}\frac{K(u)du}{1-\lambda_c uK(u)}f(a) = \overline{K}_0 f(a),
\end{aligned}
$$

as $h_1 \to 0$ after a change of variables. By a Taylor expansion of $\sigma^2(X_i)$ at $a + ch_1$ and the discrete moment condition (5),

$$
\begin{aligned}
&\frac{1}{h_1}\sum_{i=1}^{n} w_i K\left(\frac{X_i-a-ch_1}{h_1}\right)\left(\sigma^2(X_i)-\sigma^2(a+ch_1)\right) \\
= &\frac{1}{h_1}\sum_{i=1}^{n} w_i K\left(\frac{X_i-a-ch_1}{h_1}\right)\left(\frac{1}{2}\ddot{\sigma}^2(a+ch)(X_i-a-ch_1)^2 + o((X_i-a-ch_1)^2\right) \\
= &\frac{h_1^2\overline{K}_1 f(a)\ddot{\sigma}^2(a+ch_1)}{2} + o_p(h_1^2),
\end{aligned}
$$

again by the ergodic theorem. Thus, by (20), $T_2 = \frac{h_1^2\overline{K}_1}{2\overline{K}_0}\ddot{\sigma}^2(a+ch_1) + o_p(h_1^2)$. Following the proof of Theorem 1 in Fan and Yao (1998), it may be proven that $T_3, T_4 = o_p(h_1^2 + h_2^2)$ and T_1 is asymptotically normal with mean zero and variance $1/nh_1$ times (noting (20))

$$
\begin{aligned}
&\frac{1}{h_1\overline{K}_0^2 f^2(a)}E\left(nw_iK\left(\frac{X_i-a-ch_1}{h_1}\right)\sigma^2(X_i)(\varepsilon_i^2-1)\right)^2 \\
= &\frac{1}{h_1\overline{K}_0^2 f^2(a)}E\left(\frac{1}{(1-\lambda_c(X_i-a-ch_1)K_{h_1}(X_i-a-ch_1))}K\left(\frac{X_i-a-ch_1}{h_1}\right)\sigma^2(X_i)(\varepsilon_i^2-1)\right)^2 + o_p(1) \\
\to &\frac{1}{\overline{K}_0^2 f^2(a)}\int_{-1}^{c}\left(\frac{K(u)}{1-\lambda_c uK(u)}\right)^2 du\sigma^4(a)\xi^2(a)f(a) = \frac{\overline{K}_2\sigma^4(a)\xi^2(a)}{\overline{K}_0^2 f(a)}.
\end{aligned}
$$

Thus, by (16) the proof of (ii) is complete.

Proof of Theorem 2. (i). We write $\widehat{V}(x) = \widehat{V}_1(x) + \widehat{V}_2(x) + \widehat{V}_3(x)$, where

$$\begin{aligned}
\widehat{V}_1(x) &= \frac{n}{h_1}\sum_{i=1}^{n}\widehat{w}_i^2(x)K^2\left(\tfrac{X_i-x}{h_1}\right)\widehat{r}_i^4,\ \widehat{V}_2(x) = -\frac{2n\widehat{\sigma}^2(x)}{h_1}\sum_{i=1}^{n}\widehat{w}_i^2(x)K^2\left(\tfrac{X_i-x}{h_1}\right)\widehat{r}_i^2,\\
\widehat{V}_3(x) &= \frac{n\widehat{\sigma}^4(x)}{h_1}\sum_{i=1}^{n}\widehat{w}_i^2(x)K^2\left(\tfrac{X_i-x}{h_1}\right).
\end{aligned}$$

By (15), we have

$$\begin{aligned}
\widehat{r}_i^4 &= \sigma^4(X_i)\varepsilon_i^4 + 4\sigma^2(X_i)\varepsilon_i^2\left(m(X_i)-\widehat{m}(X_i)\right)^2 + \left(m(X_i)-\widehat{m}(X_i)\right)^4 + 4\sigma^3(X_i)\varepsilon_i^3\cdot\\
&\quad \left(m(X_i)-\widehat{m}(X_i)\right) + 2\sigma^2(X_i)\varepsilon_i^2\left(m(X_i)-\widehat{m}(X_i)\right)^2 + 4\sigma(X_i)\varepsilon_i\left(m(X_i)-\widehat{m}(X_i)\right)^3,
\end{aligned}$$

and denote $\widehat{V}_1(x) = \sum_{j=1}^{6} S_j$, where

$$\begin{aligned}
S_1 &= \frac{n}{h_1}\sum_{i=1}^{n} w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\sigma^4(X_i)\varepsilon_i^4,\\
S_2 &= \frac{4n}{h_1}\sum_{i=1}^{n} w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\sigma^2(X_i)\varepsilon_i^2\left(m(X_i)-\widehat{m}(X_i)\right)^2\\
S_3 &= \frac{n}{h_1}\sum_{i=1}^{n} w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\left(m(X_i)-\widehat{m}(X_i)\right)^4,\\
S_4 &= \frac{4n}{h_1}\sum_{i=1}^{n} w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\sigma^3(X_i)\varepsilon_i^3\left(m(X_i)-\widehat{m}(X_i)\right),\\
S_5 &= \frac{2n}{h_1}\sum_{i=1}^{n} w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\sigma^2(X_i)\varepsilon_i^2\left(m(X_i)-\widehat{m}(X_i)\right)^2,\\
\text{and } S_6 &= \frac{4n}{h_1}\sum_{i=1}^{n} w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\sigma(X_i)\varepsilon_i\left(m(X_i)-\widehat{m}(X_i)\right)^3.
\end{aligned}$$

Similar to the analysis of T_1 in the proof of Theorem 1 (i), we have $\frac{n\sqrt{n}}{\sqrt{h_1}}\sum_{i=1}^{n} w_i^2K^2\left(\frac{X_i-x}{h_1}\right)\cdot$ $\sigma^4(X_i)(\varepsilon_i^4-(\xi^2(x)+1)) = O_p(1)$ provided that

$$E\left(w_i^2K^2\left(\tfrac{X_i-x}{h_1}\right)\sigma^4(X_i)(\varepsilon_i^4-(\xi^2(x)+1))\right)^{2+\delta/2} < \infty.$$

Thus, $S_1 = \widetilde{S}_1 + o_p(1)$, where $\widetilde{S}_1 = (\xi^2(x)+1)\frac{n}{h_1}\sum_{i=1}^n w_i^2 K^2\left(\frac{X_i-x}{h_1}\right)\sigma^4(X_i) \xrightarrow{p} (\xi^2(x)+1)K_2\sigma^4(x)f(x)$ by the ergodic theorem. Noting (17), it follows from (c) in the proof of Theorem 1 in Fan and Yao (1998) that $S_i = o_p(1)$ for $i = 2, \cdots, 6$. Thus, $\widehat{V}_1(x) \xrightarrow{p} (\xi^2(x)+1)K_2\sigma^4(x)f(x)$. Similarly, we can show $\widehat{V}_3(x) \xrightarrow{p} K_2\sigma^4(x)f(x)$, and noticing (15) we can show $\widehat{V}_2(x) \xrightarrow{p} -2K_2\sigma^4(x)f(x)$. Hence, $\widehat{V}(x) \xrightarrow{p} \xi^2(x)K_2\sigma^4(x)f(x)$ and, therefore, Theorem 2 (i) follows from (18).

(ii). This part can be proved as in (i) using the arguments in the proof of Theorem 1 (ii).

References

[1] Bandi, F., and Phillips, P. C. B. (2003), "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica,* 71, 241–283.

[2] Bossaerts, P., Hardle, 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.

[3] Cai, Z. (2001), "Weighted Nadaraya-Watson Regression Estimation," *Statistics and Probability Letters,* 51, 307–318.

[4] Cai, Z. (2002), "Regression Quantiles for Time Series," *Econometric Theory,* 18, 169–192.

[5] Carroll, R. J. (1982), "Adapting for Heteroscedasticity in Linear Models," *The Annals of Statistics*, 10, 1224-1233.

[6] Carroll, R., and Ruppert, D. (1988), *Transforming and Weighting in Regression.* London: Chapman and Hall.

[7] Chen, S. X., and Qin, Y. (2002), "Confidence Interval Based on a Local Linear Smoother," *Scandinavian Journal of Statistics*, 29, 89-99.

[8] Dahl, C. M., and Levine, M. (2006), "Nonparametric Estimation of Volatility Models with Serially Dependent Innovations", *Statistics and Probability Letters*, 76, 2007–2016.

[9] Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

[10] De Gooijer, J. G., and Zerom, D. (2003), "On Conditional Density Estimation," *Statistica Neerlandica,* 57, 159-176.

[11] Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica,* 50, 987-1008.

[12] Fan, J., and Chen, J. (1999), "One-step Local Quasi-Likelihood Estimation," *Journal of the Royal Statistical Society Series B,* 61, 927-943.

[13] Fan, J., Farmen, M., and Gijbels, I. (1998), "Local Maximum Likelihood Estimation and Inference," *Journal Royal Statistics Society Series B*, 60, 591-608.

[14] Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and its Applications.* London: Chapman and Hall.

[15] Fan, J., and Yao, Q. (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika,* 85, 645-660.

[16] Gozalo, P., and Linton, O. (2000), "Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression," *Journal of Econometrics*, 99, 63-106.

[17] Härdle, W. (1990), *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

[18] Härdle, W., and Tsybakov, A. B. (1997), "Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression," *Journal of Econometrics*, 81, 223–242.

[19] Hall, P., and Carroll, R. J. (1989), "Variance Estimation in Regression: the Effect of Estimating the Mean," *Journal of the Royal Statistical Society B,* 51, 3-14.

[20] Hall, P., and Huang, L.-S. (2001), "Nonparametric Kernel Regression Subject to Monotonicity Constraints," *The Annals of Statistics,* 29, 624–647.

[21] Hall, P., and Presnell, B. (1999), "Intentionally Biased Bootstrap Methods," *Journal of the Royal Statistical Society B,* 61, 143–158.

[22] Hall, P., Wolff, R. C. L. and Yao, Q. (1999), "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association,* 94, 154–163.

[23] Hjort, N. L., and Jones, M. C. (1996), "Local Parametric Nonparametric Density Estimation", *Annals of Statistics*, 24, 1619-1647.

[24] Martins-Filho, C., and Yao, F. (2006a), "Estimation of Value-At-Risk and Expected Shortfall Based on Nonlinear Models of Return Dynamics and Extreme Value Theory," *Studies in Nonlinear Dynamics and Econometrics*, 10 (2).

[25] Martins-Filho, C., and Yao, F. (2006b), "Nonparametric Frontier Estimation via Local Linear Regression," forthcoming in *Journal of Econometrics*.

[26] Müller, H. G., and Stadtmüller U. (1987), "Estimation of Heteroscedasticity in Regression Analysis," *The Annals of Statistics*, 15, 610-625.

[27] Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

[28] Owen, A. (1990), "Empirical Likelihood Ratio Confidence Regions", *The Annals of Statistics*, 18, 90–120.

[29] Owen, A. (2001), *Empirical Likelihood*. Chapman and Hall/CRC.

[30] Ruppert, D., Wand, M. P., Holst, U. and Hössjer, O. (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262-273.

[31] Shephard, N. (2005), *Stochastic Volatility: Selected Readings.* Oxford University Press, Oxford.

[32] Shiryaev, A. N. (1995), *Probability*. New York: Springer-Verlag.

[33] Staniswalis, J. G. (1989), "The Kernel Estimate of a Regression Function in Likelihood Based Models," *Journal of the American Statistical Association*, 84, 276-283.

[34] Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance,* 52, 1973-2002.

[35] Tong, H. (1990), *Nonlinear Time Series Analysis: A Dynamic Approach.* Oxford University Press, Oxford.

[36] Xu, K.-L. (2006), "Empirical Likelihood Re-weighted Functional Estimation of Nonlinear Diffusions," Working paper, Yale University.

[37] Yu, K., and Jones, M. C. (2004), "Likelihood-based Local Linear Estimation of the Conditional Variance Function," *Journal of the American Statistical Association*, 99, 139-144.

[38] Ziegelmann, F. A. (2002), "Nonparametric Estimation of Volatility Functions: the Local Exponential Estimator," *Econometric Theory*, 18, 985-991.

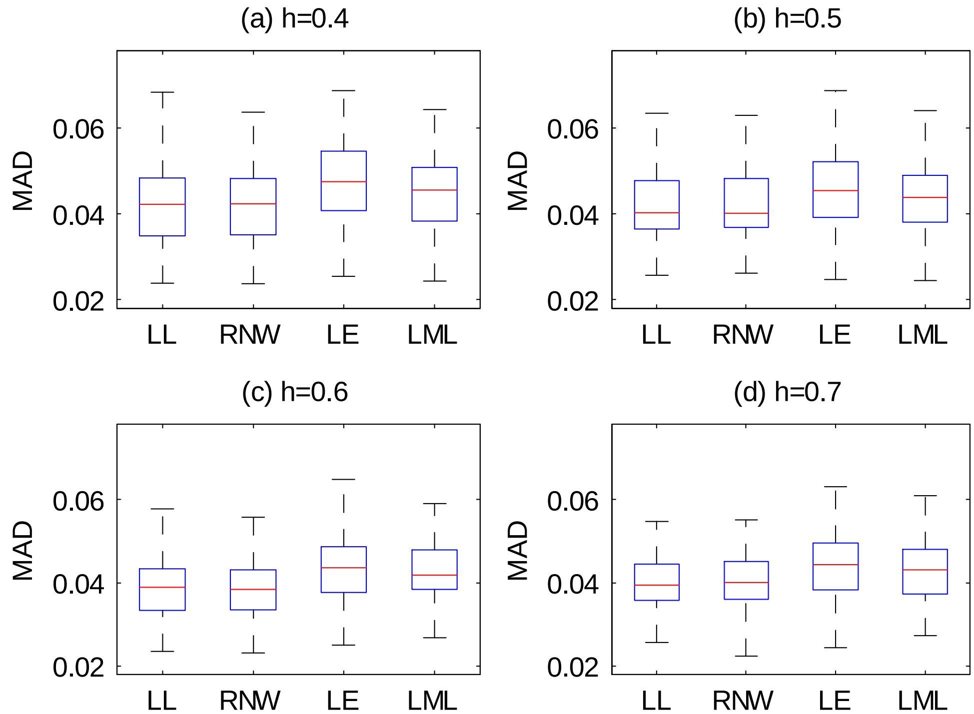

Figure 1: Boxplots of the mean absolute deviation errors (MAD) for the local linear (LL), reweighted NW (RNW), local exponential (LE) and local maximum likelihood (LML) estimators with 4 fixed bandwidths: (a) $h = 0.4$;(b) $h = 0.5$;(c) $h = 0.6$;(d) $h = 0.7$. The sample size $n = 100$ and the replications is 500. $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$.

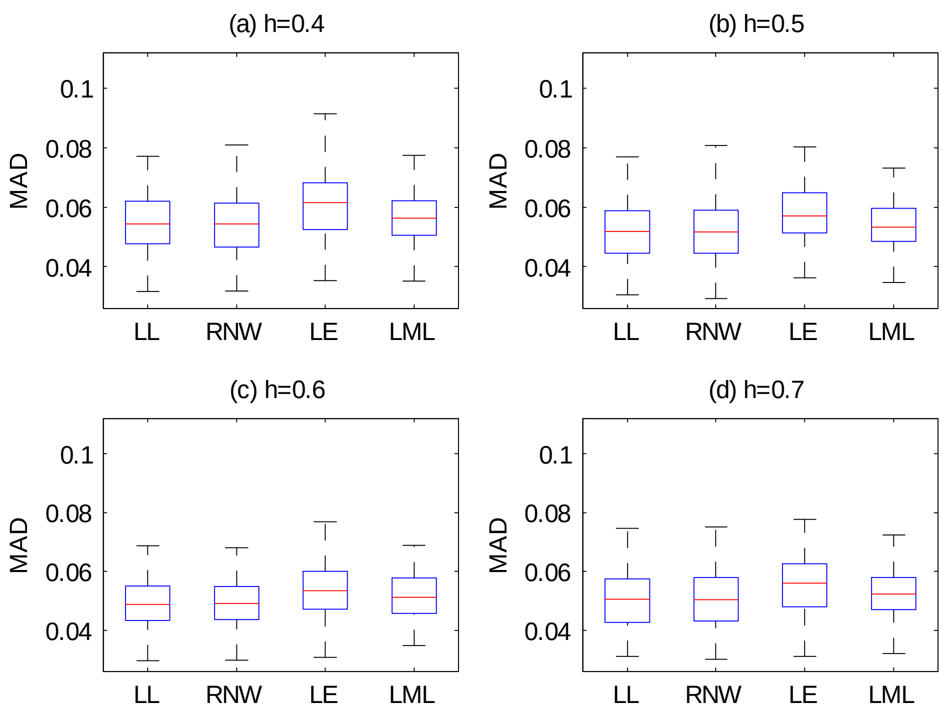

Figure 2: Boxplots of the mean absolute deviation errors (MAD) for the local linear (LL), reweighted NW (RNW), local exponential (LE) and local maximum likelihood (LML) estimators with 4 fixed bandwidths: (a) $h = 0.4$;(b) $h = 0.5$;(c) $h = 0.6$;(d) $h = 0.7$. The sample size $n = 100$ and the replications is 500. $\varepsilon_i \overset{iid}{\sim} (0.6)^{1/2} t_5$.