TILTED NONPARAMETRIC ESTIMATION OF VOLATILITY FUNCTIONS WITH EMPIRICAL APPLICATIONS

By

Peter C.B. Phillips and Ke-Li Xu

June 2007
Revised July 2010

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 with Empirical Applications

Ke-Li Xu*and Peter C. B. Phillips†

July 17, 2010

Abstract

This paper proposes a novel positive nonparametric estimator of the conditional variance function without reliance on logarithmic or other transformations. The estimator is based on an empirical likelihood modification of conventional local level nonparametric regression applied to squared mean regression residuals. The estimator is shown to be asymptotically equivalent to the local linear estimator in the case of unbounded support but, unlike that estimator, is restricted to be non-negative in finite samples. It is fully adaptive to the unknown conditional mean function. Simulations are conducted to evaluate the finite sample performance of the estimator. Two empirical applications are reported. One uses cross section data and studies the relationship between occupational prestige and income. The other uses time series data on Treasury bill rates to fit the total volatility function in a continuous-time jump diffusion model.

Keywords: Conditional variance function; Empirical likelihood; Conditional heteroskedasticity; Jump diffusion; Local linear estimator; Heteroskedastic nonparametric regression; Volatility. (*JEL Classification:* C13; C14; C22.)

* *Corresponding author.* University of Alberta School of Business and Department of Economics, Texas A&M University, College Station. Address: Business Building 3-40N, University of Alberta School of Business, Edmonton, Alberta, T6G 2R6, Canada. E-mail: keli.xu@ualberta.ca.

† Yale University, University of Auckland, University of Southampton, and Singapore Management University. Address: Department of Economics, Cowles Foundation for Research in Economics, Yale University, P. O. Box 208281, New Haven, CT 06520, USA. E-mail: peter.phillips@yale.edu.

1 Introduction

Conditional variance estimation is important in many applications. It is crucial in inference for the parameters in the conditional mean function. For example, to test for the causal treatment effect in a regression discontinuity design (Hahn et al., 2001, Porter, 2003, Imbens and Lemieux, 2008), the conditional variances of the outcome variable on the running variable at the threshold have to be estimated. In a time series context, Hansen (1995) obtained GLS-type efficient estimators of parameters in the mean function by incorporating nonparametric conditional variance estimates; see also Xu and Phillips (2008). Conditional variance estimation is also a key intermediate step in estimating some economic or financial quantities of practical importance. In a recent study, Martins-Filho and Yao (2007) proposed a nonparametric method to estimate a production frontier function starting from estimation of the conditional variance of the output given the input. Shang (2008) provided a two-stage value-at-risk forecasting procedure in a nonparametric ARCH framework based on preliminary estimation of the volatility function (viz. the conditional standard deviation) and then quantile estimation using the de-volatized residuals.

When the conditional variance is modeled nonparametrically, as in the applications mentioned above, the estimation methods usually recommended are based on local polynomial estimation, among which local linear estimation is especially popular due to its attractive properties. The theoretical foundation for this approach has been developed by Ruppert et al. (1997) and Fan and Yao (1998), inter alia. However, one drawback of the local linear variance estimator, which does not arise for the local linear mean function estimator, is that it may give negative values in finite samples which makes volatility estimation impossible. Negative variance estimates may occur for large or small smoothing bandwidths and are frequently observed at design points around which observations are relatively sparse. In consequence, it is commonly recommended in applications to use the theoretically less satisfactory local constant estimator (also known as Nadaraya-Watson estimator) when fitting the variance function (Chen and Qin, 2002, Porter, 2003).

This paper proposes a new volatility function estimator that is almost asymptotically equivalent to the local linear estimator but is guaranteed to be non-negative. It has the same asymptotic bias and variance as those of the local linear estimator when the explanatory variable has unbounded support. Such equivalence is important since it renders efficiency arguments along the lines of of Fan (1992) for the local linear estimator extendable to this new procedure. It is also convenient in that the mean squared error (MSE) or integrated MSE based selection criteria for a global or local variable smoothing bandwidth for the local linear estimator continue to apply. The new volatility function estimator is based on the idea of adjusting the conventional local constant estimator by minimally tilting the empirical distribution subject to a discrete bias-reducing moment condition satisfied by the local linear estimator (Hall and Presnell, 1999). The resultant *re-weighted local constant estimator,* or *tilted estimator*, inherits the non-negativity restriction of the variance function from the usual local constant estimator, while preserving the superior properties of bias, boundary correction and minimax efficiency of the local linear estimator. We also show adaptiveness of this procedure to the unknown mean function, i.e. it estimates the volatility function as efficiently as if the true mean function were known.

Ziegelmann (2002) recently obtained a non-negative nonparametric volatility estimator by fitting an exponential function locally (rather than a linear function as in the local linear estimator) within the general locally parametric nonparametric framework of Hjort and Jones (1996); see also Yu and Jones (2004) in a Gaussian iid setting. This estimator is not equivalent to the local linear estimator and it essentially estimates the logarithm of the variance rather than the variance itself, thus leading to an additional bias term.

The remainder of the paper is organized as follows. Section 2.1 describes the nonparametric heteroskedastic regression model, the framework within which the re-weighted local constant volatility estimator is introduced in Section 2.2. The asymptotic distributional theory is developed for stationary and mixing time series in Section 2.3 for both interior and boundary points, and a consistent estimator of the asymptotic variance is suggested. In Section 3 the finite sample performance of the proposed estimator is evaluated via simulations. Section 4 reports two empirical applications. One studies the volatility of the relationship between income and occupational prestige in Canada using cross section data. The other estimates the total volatility of 90-day Treasury bill yields in the context of a continuous time jump diffusion model. Section 5 concludes and discusses some extensions. Proofs are collected in the appendix.

2 Main Results

2.1 The heteroskedastic regression model

We focus on the following nonparametric heteroskedastic regression model

$$Y_t = m(X_t) + \sigma(X_t)\varepsilon_t, \tag{1}$$

where $\{X_t, Y_t, t = 1, \cdots, n\}$ are two stationary random processes, and $\{\varepsilon_t\}$ are innovations satisfying $\mathrm{E}(\varepsilon_t|X_t) = 0$, $\mathrm{Var}(\varepsilon_t|X_t) = 1$. The conditional mean function $m(x) = \mathrm{E}(Y_t|X_t = x)$ and the conditional variance function $\sigma^2(x) = \mathrm{Var}(Y_t|X_t = x) > 0$ are left unspecified and are the focus of statistical investigation. The reader should keep in mind that the volatility estimator proposed below applies straightforwardly to the zero mean case, e.g. the nonparametric ARCH model when $X_t = Y_{t-1}$ (Pagan and Schwert, 1990, Pagan and Hong, 1991). Many nonparametric economic models can be cast within the framework (1); e.g. see Martins-Filho and Yao (2007) for a recent application in stochastic frontier analysis and Hahn et al. (2001), Porter (2003) and Imbens and Lemieux (2008) in the analysis of causal treatment effects. As is well known, the model (1) is also of fundamental importance in financial econometrics due to its ability to allow for nonlinearity and conditional heteroskedasticity in financial time series modeling. It can further be regarded as the discretized version of the nonparametric continuous-time diffusion model which is commonly used in financial derivative pricing (Ait-Sahalia, 1996, Stanton, 1997, Bandi and Phillips, 2003).

2.2 The conditional variance estimator

Our nonparametric estimator of the conditional variance function $\sigma^2(\cdot)$ is residual-based, which relies on first-stage nonparametric estimation of the conditional mean function $m(\cdot)$. Let $W(\cdot)$ and $K(\cdot)$ be kernel functions and $h' = h'(n), h = h(n) > 0$ be smoothing bandwidths which determine model complexity. As is widely recommended in both the theoretical and empirical literatures, we can fit $m(\cdot)$ using the local linear method which solves

$$(\widehat{\gamma}_1, \widehat{\gamma}_2) = \arg\min_{(\gamma_1,\gamma_2)} \sum_{t=1}^{n} [Y_t - \gamma_1 - \gamma_2(X_t - x)]^2 W((X_t - x)/h') \tag{2}$$

leading to the estimate $\widehat{m}(x) = \widehat{\gamma}_1$ of $m(x)$ at the spatial point x. Application 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' for mean regression estimation and h for variance estimation in what follows.

To estimate the conditional variance function $\sigma^2(x)$, instead of fitting the squared residuals $\widehat{r}_t^2 = [Y_t - \widehat{m}(X_t)]^2$ to X_t using a second-stage local linear smoother as in Ruppert et al. (1997) and Fan and Yao (1998), we consider the following re-weighted local constant estimator

$$\widehat{\sigma}^2(x) = \frac{\sum_{t=1}^n \widehat{w}_t(x) K((X_t - x)/h) \widehat{r}_t^2}{\sum_{t=1}^n \widehat{w}_t(x) K((X_t - x)/h)}, \tag{3}$$

where $\widehat{w}_t(x)$ solves the constrained optimization problem

$$\{\widehat{w}_1(x), \cdots, \widehat{w}_n(x)\} = \arg\min_{\{w_1(x),\cdots,w_n(x)\}} l_n(w_1(x), \cdots, w_n(x)), \tag{4}$$

with $l_n(w_1(x), \cdots, w_n(x)) = -2\sum_{t=1}^n \log(nw_t(x))$, subject to restrictions

$$w_t(x) \geq 0, \ \sum_{t=1}^{n} w_t(x) = 1, \tag{5}$$

and

$$\sum_{t=1}^{n} w_t(x)(X_t - x) K_h(X_t - x) = 0, \tag{6}$$

where $K_h(\cdot) = K(\cdot/h)/h$. The discrete moment condition (6) is satisfied by the local linear weights $w_t^{LL}(x) = \Gamma_{n,2} - (X_t - x)\Gamma_{n,1}$ with $\Gamma_{n,j} = \sum_{t=1}^n (X_t - x)^j K_h(X_t - x)$, $j = 1, 2$, and is regarded as the key condition for local linear estimation to achieve bias reduction; see Fan and Gijbels (1996). Without (6), the optimization problem (4)-(5) is solved by the uniform weights $w_t^{UNIF}(x) = 1/n$

for all t which reduces (3) to the usual local constant estimator (or Nadaraya-Watson estimator). So the re-weighted local constant estimator (3) effectively minimizes the distance to the local constant estimator while preserving the bias-reducing condition of the local linear estimator. The distance used here is Kullback–Leibler divergence, although other distance measures can also be used (Cressie and Read, 1984), and has important connection to the empirical likelihood approach of Owen (2001).

Computationally the re-weighted estimator is very easy to use in practice as (4) can be solved by any empirical likelihood maximization program. To be specific, the weights $\widehat{w}_t(x)$ in (3) can be obtained via the Lagrange multiplier method, viz.

$$\widehat{w}_t(x) = (n[1 + \lambda(X_t - x)K_h(X_t - x)])^{-1}, \tag{7}$$

where the Lagrange multiplier λ satisfies

$$\sum_{t=1}^{n}[1 + \lambda(X_t - x)K_h(X_t - x)]^{-1}(X_t - x)K_h(X_t - x) = 0. \tag{8}$$

The re-weighting idea is due to the intentionally biased bootstrap of Hall and Presnell (1999). It is especially powerful for conditional variance estimation since the associated estimates always fall within the range $[\min_{1\le t\le n} \widehat{r}_t^2, \max_{1\le t\le n} \widehat{r}_t^2]$, thereby ensuring non-negative results. The restriction in (6) is used so that the original estimator (viz. the local constant estimator) is modified to the least extent needed to maintain the attractive properties of the local linear estimator. We can expand (6) so that the resulting variance estimator satisfies other desirable properties. For example, we can additionally impose the constraint $\mathrm{d}[\widehat{\sigma}^2(x)]/\mathrm{d}x \ge 0$ or $\mathrm{d}^2[\widehat{\sigma}^2(x)]/\mathrm{d}x^2 \ge 0$ to ensure monotonicity (Hall and Huang, 2001) or convexity of the estimated variance function as may be needed.

The re-weighting idea has been fruitfully used in other contexts, e.g. by Hall et al. (1999) for monotone estimation of the conditional distribution function that is within the range $[0, 1]$, by Cai (2002) for monotone conditional quantile estimation, and by Xu (2010) for non-negative diffusion functional estimation in a continuous-time nonstationary diffusion model.

2.3 Limit theory

The asymptotic distribution of the re-weighted local constant estimator of the conditional variance function is given in the following theorem for both interior and boundary spatial points. Let $f(\cdot)$ be the stationary density function of X_t and $\ddot{\sigma}^2(z) = \mathrm{d}^2[\sigma^2(z)]/\mathrm{d}z^2$. Assume that the kernel functions $W(\cdot)$ and $K(\cdot)$ are symmetric density functions each with bounded support $[-1, 1]$.

Theorem 1. *(i) Suppose that x is such that $x \pm h$ is in the support of $f(x)$. Under the assumptions stated in the appendix, as $n \to \infty$,*

$$\sqrt{nh}[\widehat{\sigma}^2(x) - \sigma^2(x) - h^2 K_1 \ddot{\sigma}^2(x)/2] \xrightarrow{d} \mathcal{N}\left(0, K_2\sigma^4(x)\xi^2(x)/f(x)\right), \tag{9}$$

where $K_1 = \int_{-1}^{1} u^2 K(u)\mathrm{d}u$, $K_2 = \int_{-1}^{1} K^2(u)\mathrm{d}u$, $\xi^2(x) = \mathrm{E}[(\varepsilon_t^2 - 1)^2 | X = x]$ with $\varepsilon_t = \sigma^{-1}(X_t)[Y_t - m(X_t)]$.

(ii) Suppose that $f(x)$ has bounded support $[a, b]$ and c is a constant such that $0 < c < 1$. Under the assumptions stated in the appendix, as $n \to \infty$,

$$\sqrt{nh}\left(\widehat{\sigma}^2(a+ch) - \sigma^2(a+ch) - h^2\overline{K}_1\ddot{\sigma}^2(a+ch)/[2\overline{K}_0]\right) \xrightarrow{d} \mathcal{N}\left(0, \overline{K}_2\sigma^4(a)\xi^2(a)/[\overline{K}_0^2 f(a)]\right), \tag{10}$$

where $\overline{K}_0 = \int_{-1}^{c}[1 - \overline{\lambda}_c u K(u)]^{-1} K(u)\mathrm{d}u$, $\overline{K}_1 = \int_{-1}^{c}[1 - \overline{\lambda}_c u K(u)]^{-1} u^2 K(u)\mathrm{d}u$, $\overline{K}_2 = \int_{-1}^{c}[K(u)/(1 - \overline{\lambda}_c u K(u))]^2\mathrm{d}u$ *and* $\overline{\lambda}_c$ *satisfies* $\overline{L}_c(\overline{\lambda}_c) = 0$ *with*

$$\overline{L}_c(\lambda) = \int_{-1}^{c} uK(u)/[1 - \lambda u K(u)]\mathrm{d}u,$$

and

$$\sqrt{nh}\left(\widehat{\sigma}^2(b-ch) - \sigma^2(b-ch) - h^2\underline{K}_1\ddot{\sigma}^2(b-ch)/[2\underline{K}_0]\right) \xrightarrow{d} \mathcal{N}\left(0, \underline{K}_2\sigma^4(b)\xi^2(b)/[\underline{K}_0^2 f(b)]\right),$$

where $\underline{K}_0 = \int_{c}^{1}[1 - \underline{\lambda}_c u K(u)]^{-1} K(u)\mathrm{d}u$, $\underline{K}_1 = \int_{c}^{1}[1 - \underline{\lambda}_c u K(u)]^{-1} u^2 K(u)\mathrm{d}u$, $\underline{K}_2 = \int_{c}^{1}[K(u)/(1 -$

$\underline{\lambda}_c uK(u))]^2\mathrm{d}u$ *and* $\underline{\lambda}_c$ *satisfies* $\underline{L}_c(\underline{\lambda}_c) = 0$ *with*

$$\underline{L}_c(\lambda) = \int_c^1 uK(u)/[1 - \lambda uK(u)]\mathrm{d}u.$$

Remark 1. In Theorem 1, part (i) is concerned with interior points when f has bounded support or the case where f has unbounded support, and part (ii) is concerned with boundary points. The theorem shows that the re-weighted local constant variance estimator is asymptotically equivalent to the local linear variance estimator (c.f. Ruppert et al., 1997, Fan and Yao, 1998) except for different scale constants for the bias and the variance at boundary points. The condition (6) is effective in removing a bias term of order $O_p(h^2)$ in the interior and a bias term of order $O_p(h)$ on the boundary of the local constant estimator. Thus, no additional boundary correction is needed. The following heuristic argument helps to elucidate this feature. The bias of $\widehat{\sigma}^2(x)$ is approximately accounted for by the term $(nh)^{-1}\sum_{t=1}^n p_t(x)K((X_t - x)/h)[\sigma^2(X_t) - \sigma^2(x)]$, where $p_t(x) = [\sum_{t=1}^n \widehat{w}_t(x)K((X_t - x)/h)]^{-1}\widehat{w}_t(x)$; c.f. the proof of Theorem 1 in the appendix. By a second-order Taylor expansion of $\sigma^2(X_t)$ at x and the discrete moment condition (6),

$$\begin{aligned}
& (nh)^{-1}\sum_{t=1}^n p_t(x)K((X_t - x)/h)[\sigma^2(X_t) - \sigma^2(x)] \\
= \;& (nh)^{-1}\sum_{t=1}^n p_t(x)K((X_t - x)/h)[\ddot{\sigma}^2(x)(X_t - x)^2/2] + \text{smaller order terms} \\
= \;& \begin{cases} h^2 f(x)K_1\ddot{\sigma}^2(x)/2 + o_p(h^2), & \text{if } x \text{ is in the interior;} \\ h^2 f(a)\overline{K}_1\ddot{\sigma}^2(a + ch)/2 + o_p(h^2), & \text{if } x \text{ is on the left boundary;} \\ h^2 f(b)\underline{K}_1\ddot{\sigma}^2(b - ch)/2 + o_p(h^2), & \text{if } x \text{ is on the right boundary.} \end{cases}
\end{aligned}$$

The bias term of order $O_p(h)$ is removed by the condition (6) for any n both at interior and boundary points just as for the local linear smoother. It is essentially different from the conventional local constant estimator for which the bias term of order $O_p(h)$ is eliminated in the limit via symmetry of the kernel function for interior points, but does not vanish for boundary points.

Remark 2. The constants $\overline{\lambda}_c$ and $\underline{\lambda}_c$ decrease with c and approach zero when $c \to 1$.

Theorem 1 (ii) also holds for an interior point x by noting that $\overline{K}_0 = \underline{K}_0 = 1$, $\overline{K}_1 = \underline{K}_1 = K_1$ and $\overline{K}_2 = \underline{K}_2 = K_2$ when $c \in [1, (b-a)/2h]$.

Remark 3. When the true mean function $m(\cdot)$ is known, the re-weighted local constant conditional variance estimator follows from Cai (2001) with the outcome variable $[Y_t - m(X_t)]^2$ since $\sigma^2(x) = \mathrm{E}[(Y_t - m(X_t))^2 | X_t = x]$. Theorem 1 shows that the residual-based estimator $\widehat{\sigma}^2(\cdot)$ which does not require $m(\cdot)$ to be known is asymptotically as efficient as the oracle estimator, which assumes knowledge of $m(\cdot)$. This adaptiveness property to the unknown conditional mean function is also shared by other residual-based variance estimators (see Fan and Yao, 1998, Ziegelmann, 2002).

Remark 4. Implementation of the re-weighted volatility estimator involves determination of the amount of smoothing, i.e. selection of the smoothing bandwidth h. Theorem 1 shows that minimization of the asymptotic MSE (mean squared error) or IMSE (integrated MSE) leads to an optimal local bandwidth or global bandwidth of the form $h = \varsigma n^{-1/5}$, where ς involves nuisance parameters $f(x)$, $\sigma^2(x)$, $\ddot{\sigma}^2(x)$, $\xi^2(x)$ and constants related to the kernel function. A feasible bandwidth is usually obtained by estimating ς, e.g. via parametric fitting (the rule of thumb), iterations (the plug-in method) or cross validation. An attractive feature of the re-weighted estimator is that given its asymptotic equivalence to the local linear estimator as implied by Theorem 1, the asymptotic MSE or IMSE based bandwidth selection criteria for the local linear estimator (see Fan and Yao, 1996) generally also apply to the re-weighted estimator.

Remark 5. Härdle and Tsybakov (1997) studied a volatility estimator for the model (1) assuming $X_t = Y_{t-1}$ based on differencing the local polynomial estimators of the second conditional moment and the squared first conditional moment. Their estimator is not non-negative and, as noted by Fan and Yao (1998), is not fully adaptive to the mean function. Ziegelmann's (2002) non-negative residual-based local exponential (LE) variance estimator is obtained as $\widehat{\sigma}^2_{LE} = \exp(\widehat{\psi}_1)$, where $(\widehat{\psi}_1, \widehat{\psi}_2) = \arg\min_{(\psi_1, \psi_2)} \sum_{t=1}^{n} [\widehat{r}_t^2 - \exp(\psi_1 + \psi_2(X_t - x))]^2 K((X_t - x)/h)$. It belongs to a wide class of local nonlinear estimators (Hjort and Jones, 1996, Gozalo and Linton, 2000). To ensure non-negativity of the resultant variance estimator, the procedure effectively approximates the logarithm of the variance (instead of the variance itself) locally by a linear function, thereby introducing an extra bias term.

Remark 6. The asymptotic variance of $\widehat{\sigma}^2(x)$ can be consistently estimated 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 $\widehat{V}(x) = nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)[\widehat{r}_t^2 - \widehat{\sigma}^2(x)]^2$ and $\widehat{f}(x) = h^{-1}\sum_{t=1}^{n} K((X_t - x)/h)$.

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

(ii) Under the conditions of Theorem 1 (ii), as $n \to \infty$, $\widehat{\Omega}(a+ch) \xrightarrow{p} \overline{K}_2\sigma^4(a)\xi^2(a)/[\overline{K}_0^2 f(a)]$ *and* $\widehat{\Omega}(b - ch) \xrightarrow{p} \underline{K}_2\sigma^4(b)\xi^2(b)/[\underline{K}_0^2 f(b)]$.

The following two sections provide several numerical examples illustrating the use of the new volatility estimator with simulated and real data. In all applications, the Epanechnikov function $K(u) = 0.75(1 - u^2)I_{(-1,1)}$ is used for both kernels W and K, and the bandwidth parameter in mean estimation h' is selected by least squares cross-validation.

3 Simulations

The finite-sample performance of the proposed estimator is assessed in the following simple time series setting. We generate $n + 201$ observations from the AR-ARCH model:

$$Y_t = \phi Y_{t-1} + \sqrt{\rho_0 + \rho_1 Y_{t-1}^2}\varepsilon_t \tag{11}$$

with $(\rho_0, \rho_1) = (1, 0.4)$, $Y_1 = 0$, $\phi \in \{0, 0.4\}$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, 1)$. The first 200 observations are dropped to eliminate initialization effects, so the sample size is n. The heteroskedastic regression model (1) is then estimated with the generated data. Note that (11) is different from the ARCH(1) model no matter what the true value of ϕ is since it allows for uncertainty in the mean function. Figures 1 and 2 focus on the case when $\phi = 0$. We plot the averages, 10% quantiles and 90% quantiles (over 1000 replications) of the re-weighted local constant (RLC) conditional variance estimates (when $n = 100$) at 37 equally spaced spatial points from $x = -1.8$ to $x = 1.8$, a range that is wide enough to cover most spatial points the time series visits. For comparison, the corresponding results for the local constant (LC), local linear (LL) and Ziegelmann's (2002) local exponential (LE) estimators are also plotted together with the true conditional variance function. In the two figures the smoothing bandwidths $h = 0.7$ and 1.0 are chosen to illustrate the bandwidth effects. The common bandwidth effects are observed; a larger bandwidth generally reduces the variability but increases the bias of the estimate.

A striking finding is that the RLC estimator has overall performance very close to that of the LL estimator for all spatial points considered in terms of both bias and variability. This is not surprising given the asymptotic similarity (and equivalence for unbounded support) of the two methods. But in particular samples, negative LL variance estimates are found (with frequencies listed in Table 1) mainly at spatial points with sparse neighborhoods or when a small bandwidth is used in which cases the estimates fluctuate widely. In such cases, of course, the volatility estimates are effectively useless. On the other hand, the LC and LE estimators generally suffer from large biases, especially at spatial points in whose neighborhoods there are relatively fewer observations, e.g. x with $|x| \geq 1$.

We also consider the case when there is serial correlation in Y_t, i.e. $\phi = 0.4$, and we find the results reported above are quite robust to weak serial correlation. Table 2 reports the mean squared errors (MSEs) of the RLC volatility estimates when the data-dependent bandwidths are used, i.e. $h = \alpha \widehat{s} n^{-1/5}$, where $\widehat{s}$ is the standard deviation of the sample and $\alpha \in \{1, 2, 3\}$. The MSEs decrease when the sample size increases, and they are larger for the design point $x = 1.5$ where the process sparsely visits than those for $x = 0$ where the process visits more frequently. The bandwidth with $\alpha = 2$ appears to work best in this setting and generally gives the smallest MSEs compared with the other two bandwidths. The distribution of the values of the data dependent bandwidths is also described in Table 2. For example, the median of the bandwidths (over 1000 replications) when $n = 100$ and $\alpha = 2$ is $0.559 \times 2 = 1.118$. Table 2 also reports the deviation of the MSE of the RLC volatility estimate from that of the estimate based on the true mean function $m(x) = 0.4x$. As the sample size increases, the deviation approaches zero and the effects of estimating the unknown mean function on volatility estimation disappear asymptotically, thereby confirming the adaptiveness property suggested by the limit theory.

Figures 1-2 and Table 1 about here

4 Empirical Applications

This Section provides two empirical examples to illustrate the usefulness of the proposed methodology. The first is a cross-section data application and the second involves financial time series.

4.1 Occupational Prestige vs. Income

Fox (2002) studied the relationship between occupational prestige and the average income of Canadian occupations. The dataset is available in the `car` package of `R` (`R` Development Core Team, 2010) named as `Prestige.` It consists of cross section observations for 102 occupations. Prestige for each occupation is measured by the Pineo-Porter prestige score from a social survey. Figure 3 (a) shows the scatterplot and a local linear mean fit with the bandwidth $h' = 5809$ chosen via cross validation (Li and Racine, 2004, see also Li and Racine, 2007, p.93). It might also be useful to provide variance estimates, e.g. for the construction of pointwise confidence intervals for the mean function or some automatic bandwidth selection criteria.

Figure 3 (b) plots the squared mean regression residuals against the explanatory variable (average income) and the fitted curves that give the functional conditional variance estimates by the LC, LL and RLC methods. The fitted curves are calculated over 186 levels of average incomes equally spaced from $x = 711$ to 19211. For illustration, we use the bandwidth $h = 5000$. It is clear that the LL variance estimates are negative at small values of average incomes, and the conventional LC estimates are always positive but suffer from large biases. The RLC estimates proposed in this paper appear to provide a good compromise between these two estimates, and evidently capture the declining variances in a reasonable way (being always positive) when the level of average income is low. At moderate and high levels of average incomes where the data are relatively rich, the RLC variance estimates are very close to the local linear estimates, which is not surprising given their first-order asymptotic similarity.

This example shows that bandwidth should be carefully selected to avoid the negativity problem when the LL estimator is used to estimate variance. We also consider the estimated integrated-MSE-based optimal bandwidth via rule of thumb (Fan and Gijbels, 1996, p.111) for the LL and RLC variance estimators. It has value $\widehat{h}_{op} = 1871$. We find that this bandwidth is too small and it gives wiggly estimated curves, which necessitates intervention on bandwidth selection. Figure 4 shows the estimated curves when $h = 2\widehat{h}_{op}$. It poses no problem for the LL estimator since the estimated curve is still above the zero line. Our empirical results show that further increasing the bandwidth would induce negative variance estimates.

To study the sensitivity of various functional variance estimates to the smoothing parameter, we estimate the conditional variance $\sigma^2(x)$ at two levels of average incomes $x = 1000$ and 6000 using 91 bandwidths equally spaced from $h = 1000$ to 10000 and the results are shown in Figure 5. At the boundary point $x = 1000$, negative estimates arising from the local linear fit occur within the bandwidth range approximately (4000, 6000), which might reasonably be chosen by empirical researchers. The RLC estimates generally lie between the LL and the conventional LC estimates, and are apparently quite stable over various bandwidths. At the interior point $x = 6000$, the three fitted values are much closer to each other, and the RLC and LL curves are almost indistinguishable.

Figures 3-5 about here

4.2 Jump Diffusion Volatilities

The re-weighting idea developed in this paper can be also used for functional estimation of continuous-time jump diffusions. Jump diffusion models are widely used in finance to account for discontinuities in the sample path, and are more flexible than the single-factor or multi-factor pure diffusion models in generating higher moments which match those typically observed in financial time series (see, e.g. Bakshi et al., 1997, Pan, 2002, Johannes, 2004).

Our empirical application uses $T = 54$ years of daily secondary market quotes for 3-month

T-bills from January 4, 1954 to March 13, 2008, containing $n = 13538$ observations, which are plotted in Figure 6 (a). The dataset is available from Board of Governors of the Federal Reserve System (http://research.stlouisfed.org/fred2). The spot rate r_t is assumed to follow the jump diffusion process

$$\mathrm{d}\log(r_t) = \mu(r_t)\mathrm{d}t + \sigma(r_{t-})\mathrm{d}W_t + \mathrm{d}(\Sigma_{i=1}^{I_t} Z_i),$$

where $r_{t-} = \lim_{s\uparrow t} r_s$, W_t is a standard Brownian motion, I_t is a doubly stochastic point process with stochastic intensity $\lambda(r_t)$ and $Z_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_z^2)$. We have assumed that the mean jump size is zero without loss of generality. The four quantities of interest in estimation (i.e. the drift function $\mu(r)$, the diffusion function $\sigma^2(r)$, the jump intensity $\lambda(r)$, for interest rate level r, and the jump variance σ_z^2) can be identified for a sufficiently small sampling interval Δ by the moments $M_j(r) = \mathrm{E}(\log(r_{t+\Delta}/r_t)^j | r_t = r)/\Delta$ for $j = 1, 2, 4, 6$ using the following approximate moment conditions:

$$\begin{aligned} M_1(r) &\simeq \mu(r),\ M_2(r) \simeq \sigma^2(r) + \lambda(r)\sigma_z^2, \\ M_4(r) &\simeq 3\lambda(r)\sigma_z^4,\ M_6(r) \simeq 15\lambda(r)\sigma_z^6. \end{aligned}$$

We use local linear fitting to estimate $M_1(r)$, and apply the re-weighted local constant method proposed in this paper to estimate the even-order moments $M_2(r)$, $M_4(r)$ and $M_6(r)$ to avoid the occasional but unreasonable negative estimates that result from local linear fitting. The estimates are denoted as $\widehat{M_j}(r)$, $j = 1, 2, 4, 6$. Based on the daily data $\{r_{i\Delta},\ i = 1, \cdots, n\}$, following Johannes (2004) we obtain the estimates step by step:

$$\widehat{\sigma}_z^2 = n^{-1} \sum_{i=1}^{n} \widehat{M_6}(r_{i\Delta})/[5\widehat{M_4}(r_{i\Delta})],\ \widehat{\lambda}(r) = \widehat{M_4}(r)/(3\widehat{\sigma}_z^4),$$

$$\widehat{\sigma}^2(r) = \widehat{M_2}(r) - \widehat{\lambda}(r)\widehat{\sigma}_z^2,\ \widehat{\mu}(r) = \widehat{M_1}(r).$$

The jump variance σ_z^2 is first estimated by integrating the ratio of sixth-to-fourth moments over the stationary density with the same bandwidth for the fourth and sixth moments $h_4 = 1.7\widehat{s}T^{-1/5} = 2.1\%$, where $\widehat{s}$ is the standard deviation of the sample. The estimate $\widehat{\sigma}_z^2$ is 2.39×10^{-3}.

Then, to estimate $\lambda(r)$ we consider bandwidths $h_4^{(j)} = 1.2^j \cdot h_4$ ($j = 0, 1, 2$) in $\widehat{M}_4(r)$. To estimate $\sigma^2(r)$ we use the bandwidth h_4 in computing $\widehat{M}_4(r)$ (and therefore $\widehat{\lambda}(r)$) and bandwidths $h_2^{(j)} = 1.2^j h_2$ ($j = 0, 1, 2$) in $\widehat{M}_2(r)$, where $h_2 = 1.3\widehat{s}T^{-1/5} = 1.7\%$. Lastly, $\mu(r)$ is estimated by $\widehat{M}_1(r)$ using the bandwidth $h_1^{(j)} = 1.2^j h_1$, $j = 0, 1, 2$, where $h_1 = 2.8\widehat{s}T^{-1/5} = 3.5\%$. We characterize the bandwidths used in terms of the time span T (instead of the sample size n) since the convergence rates of the $\widehat{M}_j(r)$ depend on T (or, more generally, the local time process), as shown by Bandi and Nguyen (2003). The scale constants chosen above are such that the resulting bandwidths are close to the ones reported in empirical studies of US short rates dynamics.

The estimated curves $\widehat{\mu}(r)$, $\widehat{\lambda}(r)$, $\widehat{\sigma}^2(r)$ are plotted in Figure 6 (b), Figure 7 (a) and (b), respectively. They are expected to have smaller biases than the estimates of Johannes (2004) and Bandi and Nguyen (2003), which are based on local constant estimation of the four moments. Figure 7 (b) also contains the estimates (given in the higher three lines) of the total volatility function $\sigma^2(r)+\lambda(r)\sigma_z^2$. The implication is that for most short rate levels the diffusion components explain about two thirds of the total volatility and the jump components account for about a third. This can be compared with Johannes (2004) who used a subset of our data and found that jumps typically generate more than half the volatility of interest rate changes and Eraker et al. (2003) who found that jumps in equity indices explain 10-15 percent of return volatility.

It is noteworthy that limit theories for the local linear and the re-weighted local constant estimators of the four moments in the jump diffusion model have not yet become available in the literature. We conjecture that they can be studied along the lines of Bandi and Nguyen (2003). For the pure diffusion models (where $\sigma_z^2 = 0$), the asymptotic theories for these two methods were studied by Moloche (2001), Fan and Zhang (2003) and Xu (2010).

Figures 6-7 about here

5 Concluding Remarks

This paper provides a new nonparametric approach to estimating the conditional variance function based on maximization of the empirical likelihood subject to a bias-reducing moment restriction. The method is fully adaptive for the unknown mean function. The construction of the estimator does not depend on the error distribution, and it is applicable in quite general time series and cross section settings. The new estimator preserves the appealing design adaptive, bias and automatic boundary correction properties of the local linear estimator, and it is guaranteed to be non-negative in finite samples. Numerical examples suggest that the new estimator performs well in finite samples and is a promising competitor in estimating conditional variance functions.

The proposed method can be extended to the case when X_t is multivariate, e.g. in the nonparametric AR-ARCH(p) model, $Y_t = m(Y_{t-1}, \cdots, Y_{t-p}) + \sigma(Y_{t-1}, \cdots, Y_{t-p})\varepsilon_t$ with $X_t = (Y_{t-1}, \cdots, Y_{t-p})'$. In such cases, the constrained optimization (4) is conducted under multiple restrictions. To be specific, suppose we have p covariates, and $X_t = (X_{1,t}, \cdots, X_{p,t})'$, $x = (x_1, \cdots, x_p)'$ are $p\times 1$ vectors. The RLC variance estimator is defined as $\widehat{\sigma}^2(x) = [\sum_{t=1}^n \widehat{w}_t(x)\mathbf{K}_h(X_t - x)]^{-1} \sum_{t=1}^n \widehat{w}_t(x)\mathbf{K}_h(X_t - x)\widehat{r}_t^2$ where $\widehat{r}_t$ are residuals of a $p-$dimensional nonparametric mean fit (e.g. a local linear fit) and $\mathbf{K}_h(X_t - x) = h^{-p}\Pi_{i=1}^p K((X_{i,t} - x_i)/h)$ are product kernel weights. Different bandwidths and kernels could be used for each covariate but we assume they are the same for expositional simplicity. The weights $\widehat{w}_t(x)$ are such that (4) is solved subject to (5) and the p-dimensional restrictions

$$\sum_{t=1}^n w_t(x)(X_t - x)\mathbf{K}_h(X_t - x) = 0. \tag{12}$$

The local linear weights satisfy (12) and they take the form, e.g. when $p = 2$, $w_t^{LL}(x) = \widetilde{\Gamma}_1 - \widetilde{\Gamma}_2(X_{1,t} - x_1) + \widetilde{\Gamma}_3(X_{2,t} - x_2)$ with

$$\widetilde{\Gamma}_1 = \det\begin{pmatrix} \widetilde{\Gamma}_{(2,0)} & \widetilde{\Gamma}_{(1,1)} \\ \widetilde{\Gamma}_{(1,1)} & \widetilde{\Gamma}_{(0,2)} \end{pmatrix}, \widetilde{\Gamma}_2 = \det\begin{pmatrix} \widetilde{\Gamma}_{(1,0)} & \widetilde{\Gamma}_{(1,1)} \\ \widetilde{\Gamma}_{(0,1)} & \widetilde{\Gamma}_{(0,2)} \end{pmatrix}, \widetilde{\Gamma}_3 = \det\begin{pmatrix} \widetilde{\Gamma}_{(1,0)} & \widetilde{\Gamma}_{(2,0)} \\ \widetilde{\Gamma}_{(0,1)} & \widetilde{\Gamma}_{(1,1)} \end{pmatrix},$$

where $\det(A)$ denotes the determinant of the matrix A and $\widetilde{\Gamma}_{(i,j)} = \Sigma_{t=1}^{n}(X_{1,t} - x_1)^j (X_{2,t} - x_2)^k \mathbf{K}_h(X_t - x)$ for $j, k = 0, 1, 2$. Just as in the univariate case, the re-weighted estimator selects the weights such that the good bias properties of the local linear estimator are preserved and the resulting variance estimate is always non-negative.

However, the fully nonparametric volatility estimators above suffer from slow convergence rates when p is large and difficulties of interpretation. A popular alternative that can achieve the one-dimensional convergence rate and imposes reasonably weak assumptions on the specification of the volatility function is the additive model, e.g. the additive ARCH model considered by Kim and Linton (2004), where $\sigma(Y_{t-1}, \cdots, Y_{t-p}) = \sqrt{\theta + \sigma_1^2(Y_{t-1}) + \cdots + \sigma_p^2(Y_{t-p})}$.The functions $\sigma_1^2(\cdot), \cdots, \sigma_p^2(\cdot)$ can be estimated, e.g. by the method of marginal integration or backfitting, which essentially involves iterative univariate smoothing. Again, the re-weighted local constant method proposed here is expected to be a promising alternative to the local linear estimator which is commonly recommended.

Acknowledgments

The authors thank the Editors (Arthur Lewbel and Keisuke Hirano), Associate Editor, two anonymous referees, Donald Andrews and Taisuke Otsu for helpful comments. Xu acknowledges partial research support from University of Alberta School of Business under H. E. Pearson fellowship and J. D. Muir grant. Phillips acknowledges partial research support from a Kelly Fellowship and the NSF under Grant Nos. SES 04-142254, 06-47086 and 09-56687.

Appendix

This section provides proofs of Theorems 1 and 2. To derive the asymptotic distribution of $\widehat{\sigma}^2(x)$, we make the following assumptions.

Assumptions

(i) For a given design point x*, the functions* $f(x) > 0$, $\sigma^2(x) > 0$, $\mathrm{E}(Y^3|X = x)$ *and*

$E(Y^4|X = x)$ *are continuous at* x, *and* $\ddot{m}(x) = \mathrm{d}^2 m(x)/\mathrm{d}x^2$ *and* $\ddot{\sigma}^2(x) = \mathrm{d}^2(\sigma^2(x))/\mathrm{d}x^2$ *are uniformly continuous on an open set containing* x;

(ii) $E|Y|^{4(1+\delta)} < \infty$ *for some* $\delta \geq 0$;

(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* $[-1, 1]$. *A Lipschitz condition is satisfied by each of functions* $f(\cdot)$, $W(\cdot)$ *and* $K(\cdot)$;

(v) The process (X_t, Y_t) *is strictly stationary and absolutely regular*[1] *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 $n \to \infty$, $h, h' \to 0$ *and* $\liminf_{n\to\infty} nh^4 > 0$, $\liminf_{n\to\infty} nh'^4 > 0$.

Proof of Theorem 1. Note that the weights $\widehat{w}_t(x)$ in the RLC estimator as in (3) has the computationally convenient representation in (7). For simplicity we write $\widehat{w}_t(x)$ as w_t in what follows. Note that $\widehat{r}_t = Y_t - \widehat{m}(X_t) = [m(X_t) - \widehat{m}(X_t)] + \sigma(X_t)\varepsilon_t$, so

$$\widehat{r}_t^2 = \sigma^2(X_t)\varepsilon_t^2 + 2\sigma(X_t)\varepsilon_t[m(X_t) - \widehat{m}(X_t)] + [m(X_t) - \widehat{m}(X_t)]^2. \tag{13}$$

Thus by (3)

$$\widehat{\sigma}^2(x) - \sigma^2(x) = \Sigma_{j=1}^4 N_j, \tag{14}$$

where

$$N_1 = \frac{\sum_{t=1}^n w_t K((X_t - x)/h)\sigma^2(X_t)(\varepsilon_t^2 - 1)}{\sum_{t=1}^n w_t K((X_t - x)/h)}, \quad N_2 = \frac{\sum_{t=1}^n w_t K((X_t - x)/h)[\sigma^2(X_t) - \sigma^2(x)]}{\sum_{t=1}^n w_t K((X_t - x)/h)},$$

$$N_3 = \frac{2\sum_{t=1}^n w_t K((X_t - x)/h)\sigma(X_t)\varepsilon_t[m(X_t) - \widehat{m}(X_t)]}{\sum_{t=1}^n w_t K((X_t - x)/h)},$$

and

$$N_4 = \frac{\sum_{t=1}^n w_t K((X_t - x)/h)[m(X_t) - \widehat{m}(X_t)]^2}{\sum_{t=1}^n w_t K((X_t - x)/h)}.$$

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

(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. $\lambda = -\frac{hK_1 f'(x)}{\upsilon_2 f(x)} + O_{a.s.}(h^3)$, where $\upsilon_2 = \int u^2 K^2(u)\mathrm{d}u$, and

$$w_t = n^{-1}\left(1 - \frac{hK_1 f'(x)}{\upsilon_2 f(x)}(X_t - x)K_h(X_t - x)\right)^{-1}(1 + o_p(1)), \tag{15}$$

Consider the term N_2 first. The denominator of N_2 times $1/h$ is

$$h^{-1}\sum_{t=1}^{n} w_t K((X_t - x)/h) = (nh)^{-1}\sum_{t=1}^{n} K((X_t - x)/h) + o_p(1) \xrightarrow{p} f(x), \tag{16}$$

by (15) and an application of Birkhoff's ergodic theorem (see, e.g., Shiryaev, 1996) since $\mathrm{E}[h^{-1}K((X_t - x)/h)] = h^{-1}\int K((u-x)/h)f(u)\mathrm{d}u \to f(x)$ as $h \to 0$ after a simple change of variables. By Taylor expansion of $\sigma^2(X_t)$ at x and the discrete moment condition (6), the numerator of N_2 times $1/h$ is

$$\begin{aligned}
& h^{-1}\sum_{t=1}^{n} w_t K((X_t - x)/h)[\sigma^2(X_t) - \sigma^2(x)] \\
= \; & h^{-1}\sum_{t=1}^{n} w_t K((X_t - x)/h)[\ddot{\sigma}^2(x)(X_t - x)^2/2 + o((X_t - x)^2)] \\
= \; & h^2 f(x) K_1 \ddot{\sigma}^2(x)/2 + o_p(h^2), \qquad (17)
\end{aligned}$$

by (15) and the ergodic theorem. Combining (16) and (17) gives $N_2 = h^2 K_1 \ddot{\sigma}^2(x)/2 + o_p(h^2)$. Noting (15) and (16), it follows from Fan and Yao (1998, the proof of Theorem 1, (b)-(d)) that $\sqrt{nh}N_1 \xrightarrow{d} \mathcal{N}\left(0, K_2\sigma^4(x)\xi^2(x)/f(x)\right)$, and $N_3, N_4 = o_p(h^2 + h'^2)$. Hence by (14) Theorem (i) holds.

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

$$w_t = \frac{1}{n(1 - \lambda_c(X_t - a - ch)K_h(X_t - a - ch))}(1 + o_p(1)).$$

Consider the term N_2 in (14) first. Note that

$$h^{-1}\sum_{t=1}^{n} w_t K((X_t-a-ch)/h) = (nh)^{-1}\sum_{t=1}^{n}\frac{K((X_t-a-ch)/h)}{1-\lambda_c(X_t-a-ch)K_h(X_t-a-ch)}+o_p(1) \xrightarrow{p} \overline{K}_0 f(a), \tag{18}$$

by the ergodic theorem since

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{h}\frac{K((X_t-a-ch)/h)}{1-\lambda_c(X_t-a-ch)K_h(X_t-a-ch)}\right) &= \int_a^b \frac{1}{h}\frac{K((z-a-ch)/h)}{1-\lambda_c(z-a-ch)K_h(z-a-ch)} f(z)\mathrm{d}z \\
&\to \int_{-1}^{c} \frac{K(u)\mathrm{d}u}{1-\lambda_c u K(u)} f(a) = \overline{K}_0 f(a),
\end{aligned}$$

as $h \to 0$ after a change of variables. By Taylor expansion of $\sigma^2(X_t)$ at $a+ch$ and the discrete moment condition (6),

$$\begin{aligned}
& h^{-1}\sum_{t=1}^{n} w_t K((X_t-a-ch)/h)[\sigma^2(X_t)-\sigma^2(a+ch)] \\
= \; & h^{-1}\sum_{t=1}^{n} w_t K((X_t-a-ch)/h)[\ddot{\sigma}^2(a+ch)(X_t-a-ch)^2/2 + o((X_t-a-ch)^2)] \\
= \; & h^2\overline{K}_1 f(a)\ddot{\sigma}^2(a+ch)/2 + o_p(h^2),
\end{aligned}$$

again by the ergodic theorem. Thus, by (18) $N_2 = [2\overline{K}_0]^{-1}h^2\overline{K}_1\ddot{\sigma}^2(a+ch)+o_p(h^2)$. Following the proof of Theorem 1 in Fan and Yao (1998), it can be shown that $N_3, N_4 = o_p(h^2+h'^2)$ and N_1 is asymptotically normal with mean zero and variance $1/nh$ times (noting (18))

$$\begin{aligned}
& \frac{1}{h\overline{K}_0^2 f^2(a)}\mathrm{E}\left(nw_t K((X_t-a-ch)/h)\sigma^2(X_t)(\varepsilon_t^2-1)\right)^2 \\
= \; & \frac{1}{h\overline{K}_0^2 f^2(a)}\mathrm{E}\left(\frac{1}{(1-\lambda_c(X_t-a-ch)K_h(X_t-a-ch))}K((X_t-a-ch)/h)\sigma^2(X_t)(\varepsilon_t^2-1)\right)^2 + o_p(1) \\
\to \; & \frac{1}{\overline{K}_0^2 f^2(a)}\int_{-1}^{c}\left(\frac{K(u)}{1-\lambda_c u K(u)}\right)^2 \mathrm{d}u \cdot \sigma^4(a)\xi^2(a)f(a) = \frac{\overline{K}_2\sigma^4(a)\xi^2(a)}{\overline{K}_0^2 f(a)}.
\end{aligned}$$

So the result desired follows by (14). The case when $x = b - ch$ can be proved similarly. 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

$$\begin{aligned} \widehat{V}_1(x) &= h^{-1}n\sum_{t=1}^{n} K^2((X_t - x)/h)\widehat{r}_t^4, \ \widehat{V}_2(x) = -2h^{-1}n\widehat{\sigma}^2(x)\sum_{t=1}^{n} K^2((X_t - x)/h)\widehat{r}_t^2, \\ \widehat{V}_3(x) &= h^{-1}n\widehat{\sigma}^4(x)\sum_{t=1}^{n} K^2((X_t - x)/h). \end{aligned}$$

Consider the term $\widehat{V}_1(x)$ first. By (13), we have

$$\begin{aligned} \widehat{r}_t^4 &= \sigma^4(X_t)\varepsilon_t^4 + 4\sigma^2(X_t)\varepsilon_t^2[m(X_t) - \widehat{m}(X_t)]^2 + [m(X_t) - \widehat{m}(X_t)]^4 + 4\sigma^3(X_t)\varepsilon_t^3 \cdot \\ &\quad [m(X_t) - \widehat{m}(X_t)] + 2\sigma^2(X_t)\varepsilon_t^2[m(X_t) - \widehat{m}(X_t)]^2 + 4\sigma(X_t)\varepsilon_t[m(X_t) - \widehat{m}(X_t)]^3, \end{aligned}$$

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

$$\begin{aligned} \widehat{V}_{11} &= nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma^4(X_t)\varepsilon_t^4, \\ \widehat{V}_{12} &= 4nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma^2(X_t)\varepsilon_t^2[m(X_t) - \widehat{m}(X_t)]^2 \\ \widehat{V}_{13} &= nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)[m(X_t) - \widehat{m}(X_t)]^4, \\ \widehat{V}_{14} &= 4nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma^3(X_t)\varepsilon_t^3[m(X_t) - \widehat{m}(X_t)], \\ \widehat{V}_{15} &= 2nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma^2(X_t)\varepsilon_t^2[m(X_t) - \widehat{m}(X_t)]^2 \ , \\ \text{and } \widehat{V}_{16} &= 4nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma(X_t)\varepsilon_t[m(X_t) - \widehat{m}(X_t)]^3. \end{aligned}$$

Similar to the analysis of the term N_1 in the proof of Theorem 1 (i), we have

$$n\sqrt{n}h^{-1/2}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma^4(X_t)(\varepsilon_t^4 - (\xi^2(x) + 1)) = O_p(1)$$

provided that

$$\mathrm{E}[K^2((X_t - x)/h)\sigma^4(X_t)(\varepsilon_t^4 - (\xi^2(x) + 1))]^{2+\delta/2} < \infty,$$

which holds by assumption. Thus $\widehat{V}_{11} = \widetilde{V}_{11} + o_p(1)$, where

$$\widetilde{V}_{11} = (\xi^2(x)+1)nh^{-1}\sum_{t=1}^{n} K^2((X_t - x)/h)\sigma^4(X_t) \xrightarrow{p} (\xi^2(x)+1)K_2\sigma^4(x)f(x)$$

by the ergodic theorem. It follows from Fan and Yao (1998) and the proof of Theorem 1 (c)) that $\widehat{V}_{1j} = o_p(1)$ for $j = 2, \cdots, 6$. Thus, $\widehat{V}_1(x) \xrightarrow{p} (\xi^2(x)+1)K_2\sigma^4(x)f(x)$. Similarly using (13) we can show that $\widehat{V}_2(x) \xrightarrow{p} -2K_2\sigma^4(x)f(x)$. Lastly $\widehat{V}_3(x) \xrightarrow{p} K_2\sigma^4(x)f(x)$. So $\widehat{V}(x) \xrightarrow{p} \xi^2(x)K_2\sigma^4(x)f(x)$ and Theorem 2 (i) follows from (16).

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

References

Aït-Sahalia, Y., 1996, Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527-560.

Bakshi, G., C. Cao and Z. Chen, 1997, Empirical performance of alternative option pricing models. Journal of Finance 52, 2003-2049.

Bandi, F. and T. Nguyen, 2003, On the functional estimation of jump-diffusion processes. Journal of Econometrics 116, 293-328.

Bandi, F. and P.C.B. Phillips, 2003, Fully nonparametric estimation of scalar diffusion models. Econometrica 71, 241-283.

Cai, Z., 2001, Weighted Nadaraya-Watson regression estimation. Statistics and Probability Letters 51, 307-318.

Cai, Z., 2002, Regression quantiles for time series. Econometric Theory 18, 169-192.

Chen, S.X. and Y. Qin, 2002, Confidence interval based on a local linear smoother. Scandinavian Journal of Statistics 29, 89-99.

Cressie, N. and T. Read, 1984, Multinomial goodness-of-fit tests. Journal of the Royal Statistical Society B, 46, 440-464.

Davidson, J., 1994, Stochastic limit theory: An introduction for econometricians. Oxford University Press, Oxford.

Eraker, B., M. Johannes and N. Polson, 2003, The impact of jumps in volatility and returns. Journal of Finance 58, 1269–1300.

Fan, J., 1992, Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998-1004.

Fan, J. and I. Gijbels, 1996, Local polynomial modeling and its applications. Chapman and Hall, London.

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

Fan, J. and C. Zhang, 2003, A re-examination of Stanton's diffusion estimations with applications to financial model validation. Journal of American Statistical Association 98, 118-134.

Fox, J., 2002, An R and S-PLUS companion to applied regression. Sage, Thousand Oaks.

Gozalo, P. and O. Linton, 2000, Local nonlinear least squares: Using parametric information in nonparametric regression. Journal of Econometrics 99, 63-106.

Hahn, J., P. Todd, and W. Van der Klaauw, 2001, Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69, 201–209.

Hall, P. and L.-S. Huang, 2001, Nonparametric kernel regression subject to monotonicity constraints. Annals of Statistics 29, 624-647.

Hall, P. and B. Presnell, 1999, Intentionally biased bootstrap methods. Journal of the Royal Statistical Society B 61, 143-158.

Hall, P., R.C.L. Wolff and Q. Yao, 1999, Methods for estimating a conditional distribution function. Journal of the American Statistical Association 94, 154-163.

Hansen, B.E., 1995, Regression with nonstationary volatility. Econometrica 63, 1113-1132.

Härdle, W. and A.B. Tsybakov, 1997, Local polynomial estimators of the volatility function in nonparametric autoregression. Journal of Econometrics 81, 223-242.

Hjort, N.L. and M.C. Jones, 1996, Local parametric nonparametric density estimation. Annals of Statistics 24, 1619-1647.

Imbens, G. W. and T. Lemieux, 2008, Regression discontinuity designs: a guide to practice, Journal of Econometrics 142, 615-635.

Kim, W. and O. Linton, 2004, A local instrumental variable estimation method for generalized additive volatility models. Econometric Theory 20, 1094-1139.

Li, Q. and J.S. Racine, 2004, Cross-validated local linear nonparametric regression. Statistica Sinica 14, 485-512.

Li, Q. and J.S. Racine, 2007, Nonparametric Econometrics: Theory and Practice. Princeton University Press, New Jersey.

Johannes, M., 2004, The statistical and economic role of jumps in continuous-time interest rate models. Journal of Finance 59, 227-260.

Martins-Filho, C. and F. Yao, 2007, Nonparametric frontier estimation via local linear regression. Journal of Econometrics 141, 283-319.

Moloche, G., 2001, Local nonparametric estimation of scalar diffusions. Unpublished paper, MIT.

Owen, A., 2001, Empirical likelihood. Chapman and Hall/CRC.

Pagan, A. and G. Schwert, 1990, Alternative models for conditional stock volatility. Journal of Econometrics 45, 267-290.

Pagan, A.R., and Y.S. Hong (1991): .Nonparametric Estimation and the Risk Premium,. in W. Barnett, J. Powell, and G.E. Tauchen (eds.) Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press.

Pan, J., 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, Journal of Financial Economics 63, 3-50.

Porter, J., 2003, Estimation in the regression discontinuity model. Working paper, Department of Economics, University of Wisconsin.

R Development Core Team, 2010, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ruppert, D., M.P. Wand, U. Holst and O. Hössjer, 1997, Local polynomial variance function estimation. Technometrics 39, 262-273.

Shang, D., 2008, Robust interval forecasts of value-at-risk for nonparametric ARCH with heavytailed errors. Working paper, University of Wisconsin.

Shiryaev, A.N., 1996, Probability. Springer-Verlag, New York.

Stanton, R., 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk. Journal of Finance 52, 1973-2002.

Xu, K.-L., 2010, Re-weighted functional estimation of nonlinear diffusions. Econometric Theory 26, 541-563.

Xu, K.-L. and P.C.B. Phillips, 2008, Adaptive estimation of autoregressive models with time-varying variances, Journal of Econometrics 142, 265-280.

Yu, K. and M.C. Jones, 2004, Likelihood-based local linear estimation of the conditional variance function. Journal of the American Statistical Association 99, 139-144.

Ziegelmann, F.A., 2002, Nonparametric estimation of volatility functions: The local exponential estimator. Econometric Theory 18, 985-991.

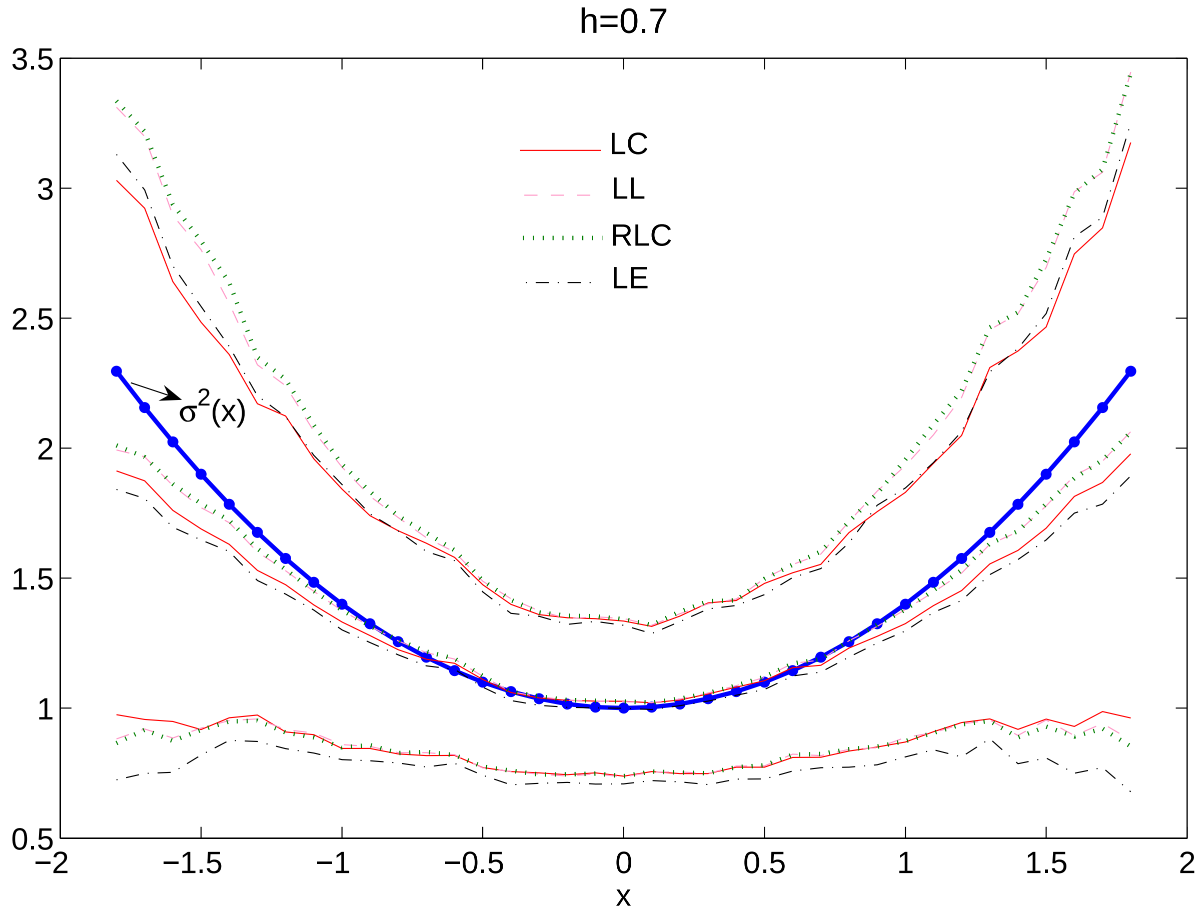

Figure 1: The means, 10% quantiles and 90% quantiles of the local constant (LC), local linear (LL), re-weighted local constant (RLC) and local exponential (LE) estimates of the volatility function $\sigma^2(x) = 1 + 0.4x^2$ in the AR-ARCH model (11) when $\phi = 0$ over 1000 replications, using the smoothing bandwidth $h = 0.7$.

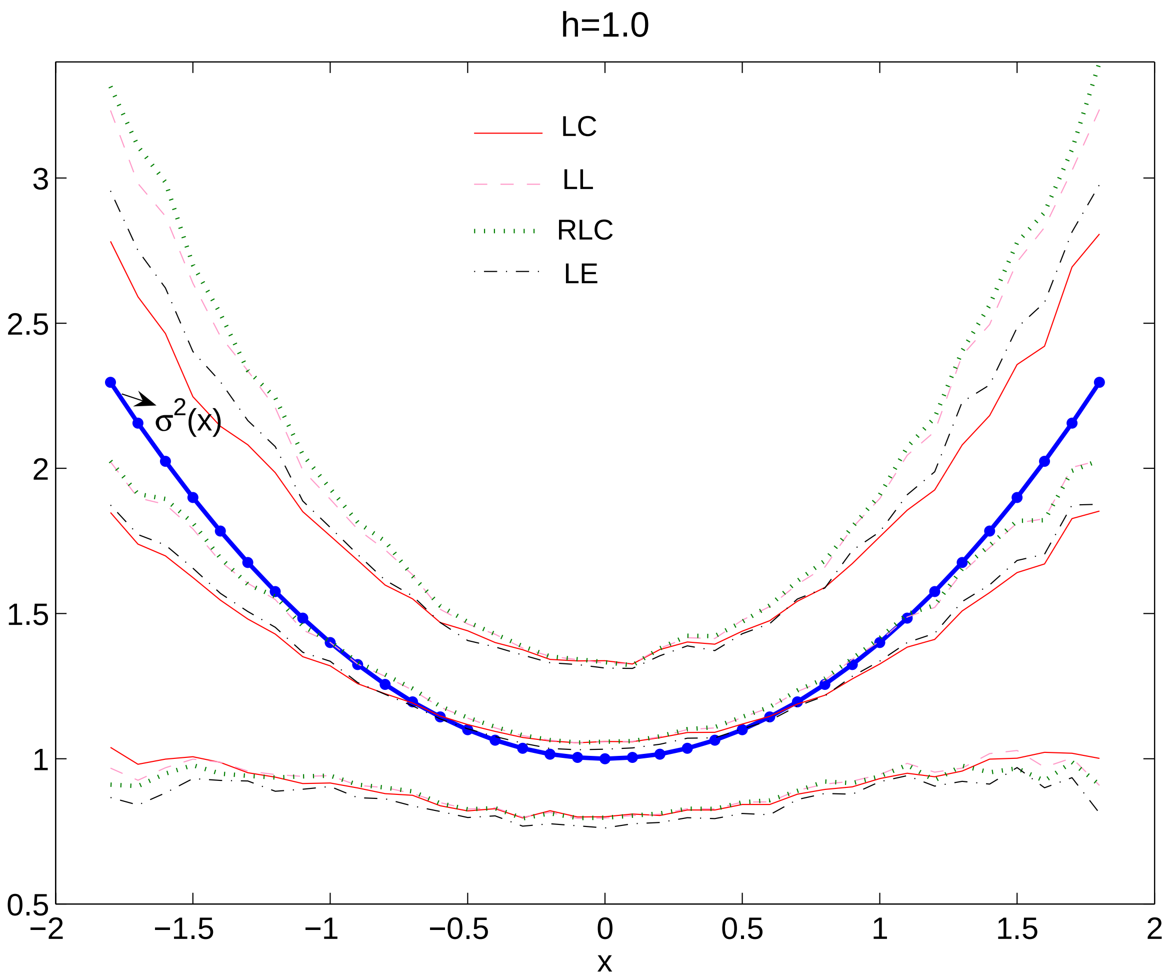

Figure 2: The means, 10% quantiles and 90% quantiles of the local constant (LC), local linear (LL), re-weighted local constant (RLC) and local exponential (LE) estimates of the volatility function $\sigma^2(x) = 1 + 0.4x^2$ in the AR-ARCH model (11) when $\phi = 0$ over 1000 replications, using the smoothing bandwidth $h = 1.0$.

Table 1: Frequencies of negative local linear conditional variance estimates in the AR-ARCH model (11) when $\phi = 0$ over 1000 replications (zeros for blank cells).

Bandwidth	$h = 0.7$	$h = 0.6$	$h = 0.5$	$h = 0.4$	$h = 0.3$	$h = 0.2$
$x = 1.8$	3	4	6	13	19	61
$x = 1.6$		2	3	3	16	39
$x = 1.4$				1	4	18
$x = 1.2$						6
$x = 1.1$					1	8
$x = 1.0$						8
$x = 0.9$						6
$x = 0.8$						2

Table 2: Mean squared errors (MSEs) of the RLC volatility estimates and the adaptiveness to the unknown mean function in the AR-ARCH model (11) when $\phi = 0.4$. [Dev. stands for the deviation of the MSE of the RLC volatility estimate from that of the estimate based on the true mean function]

		$x = 0$					$x = 1.5$				
	$\alpha \setminus n$	50	100	200	400	800	50	100	200	400	800
RLC	$\alpha = 1$	0.375	0.279	0.208	0.141	0.118	1.129	0.815	0.648	0.419	0.319
Dev.		0.039	0.012	0.005	0.002	0.001	0.122	0.102	0.017	0.021	0.001
RLC	$\alpha = 2$	0.317	0.230	0.172	0.133	0.093	1.020	0.758	0.563	0.369	0.254
Dev.		0.066	0.032	0.020	0.010	0.005	0.181	0.112	0.036	0.021	0.012
RLC	$\alpha = 3$	0.355	0.277	0.212	0.158	0.125	1.054	0.787	0.546	0.385	0.269
Dev.		0.119	0.059	0.031	0.017	0.009	0.379	0.286	0.164	0.045	0.021
Value of data-dependent h when $\alpha=1$											
Mean		0.661	0.587	0.514	0.449	0.394					
Std.		0.162	0.142	0.080	0.052	0.037					
Median		0.630	0.559	0.501	0.441	0.389					

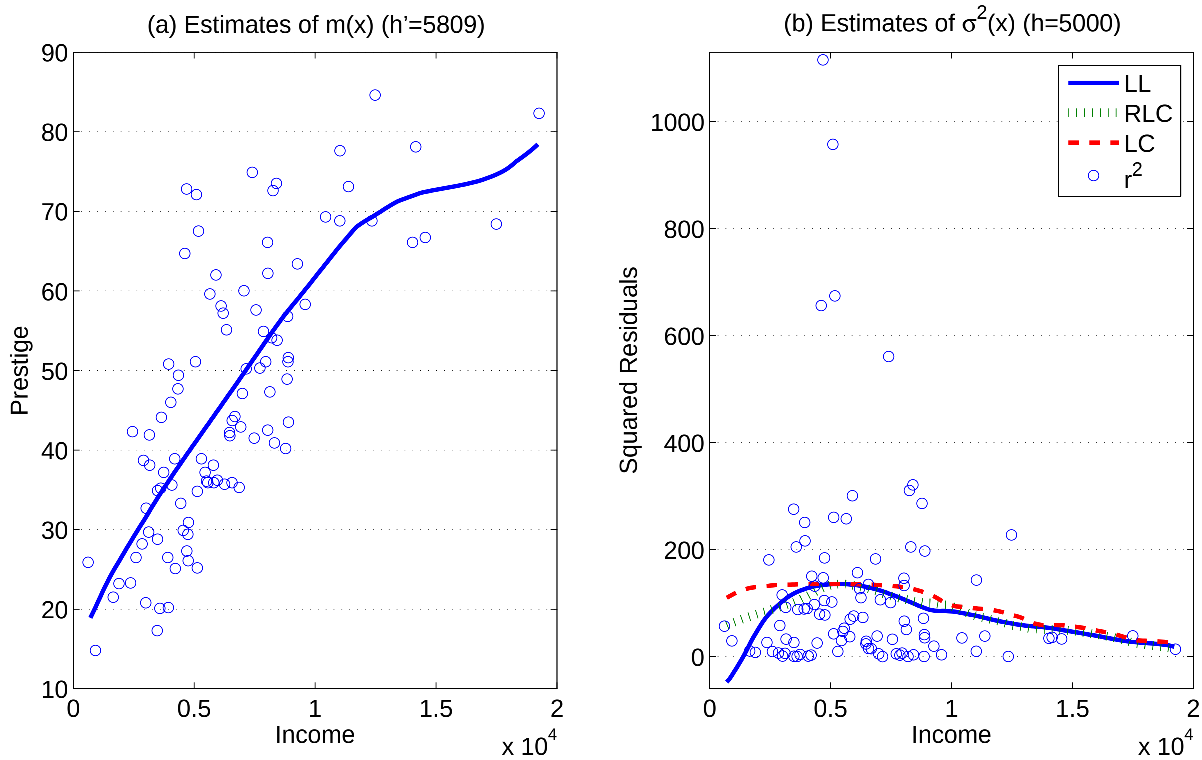

Figure 3: Prestige vs. Income: (a) Local linear estimation of the conditional mean function using the bandwidth $h' = 5809$; (b) Estimates of the conditional variance function based on the squared residuals using the local linear (LL), re-weighted local constant (RLC) and conventional local constant (LC) methods with the bandwidth $h = 5000$.

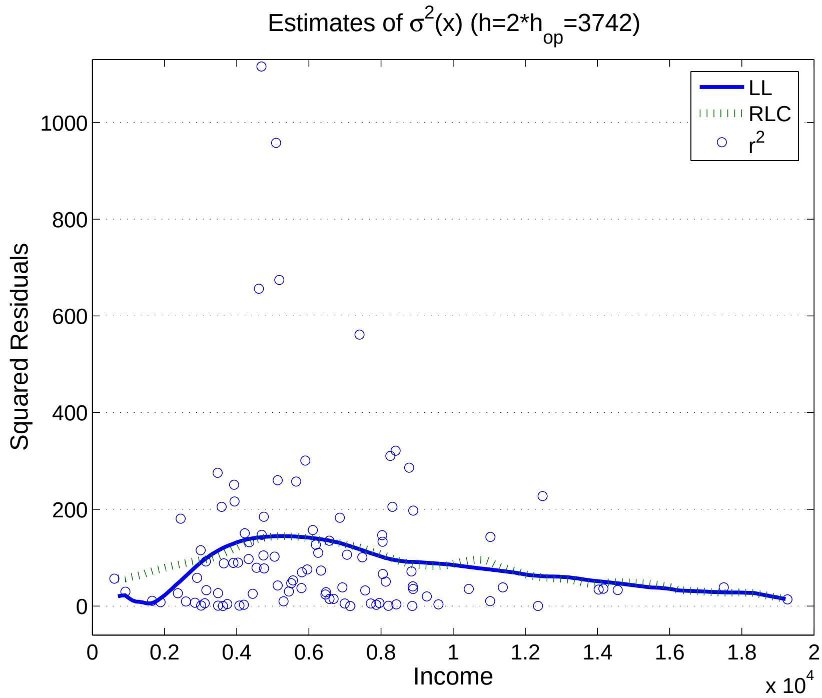

Figure 4: Prestige vs. Income: Estimates of the conditional variance function based on the squared residuals using the local linear (LL) and re-weighted local constant (RLC) methods with the bandwidth $h = 2\widehat{h}_{op} = 3742$.

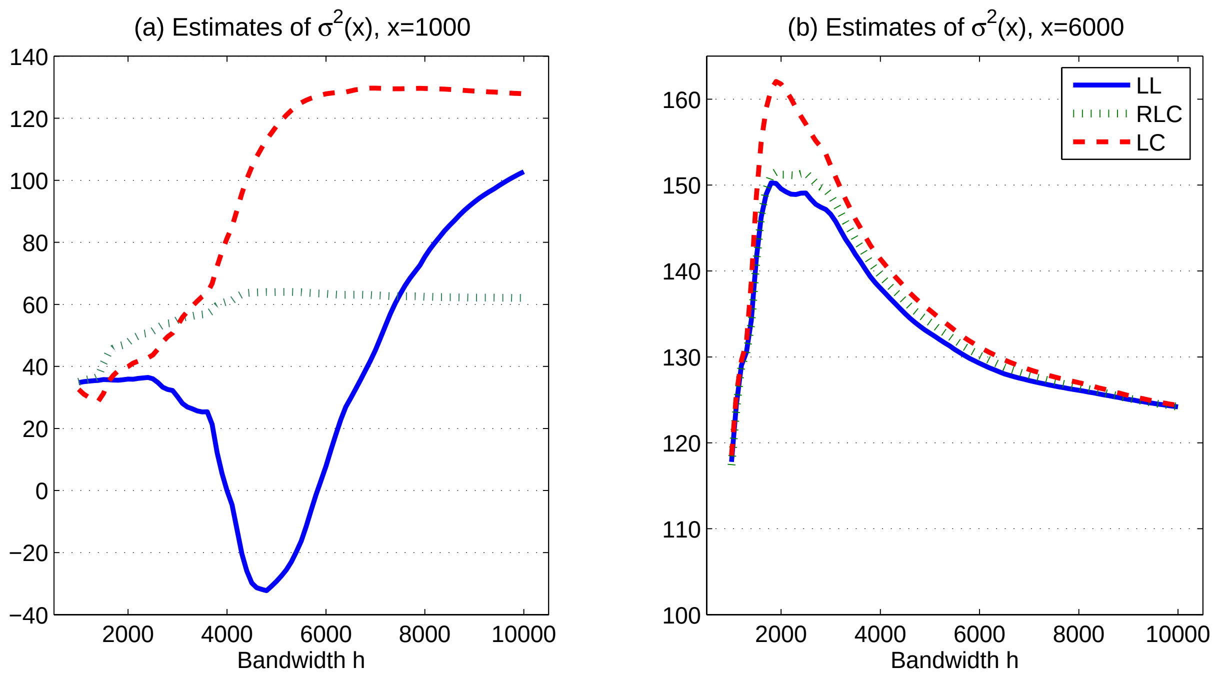

Figure 5: Prestige vs. Income: estimates of the conditional variance function over 91 bandwidths using LL, RLC and LC methods when the design point (a) $x = 1000$; (b) $x = 6000$.

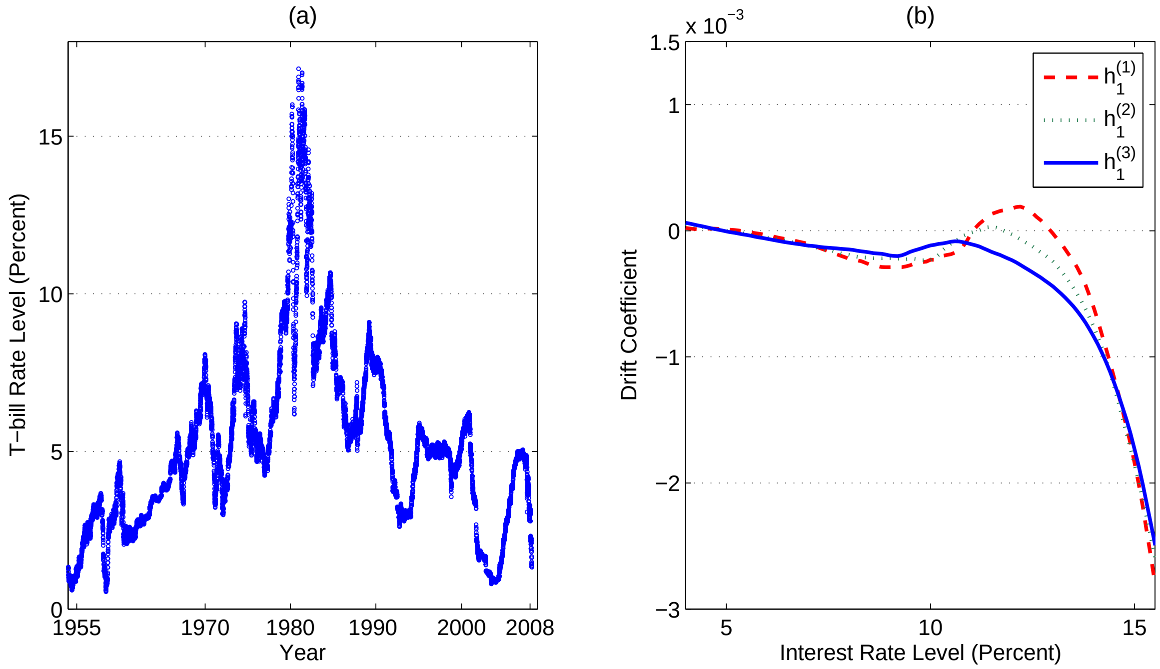

Figure 6: (a) The time series of daily 3-month Treasury bill rates (secondary market rates) from January 4, 1954 to March 13, 2008; (b) the local linear estimators of the drift function using three bandwidths 3.5%, 4.2% and 5.0%.

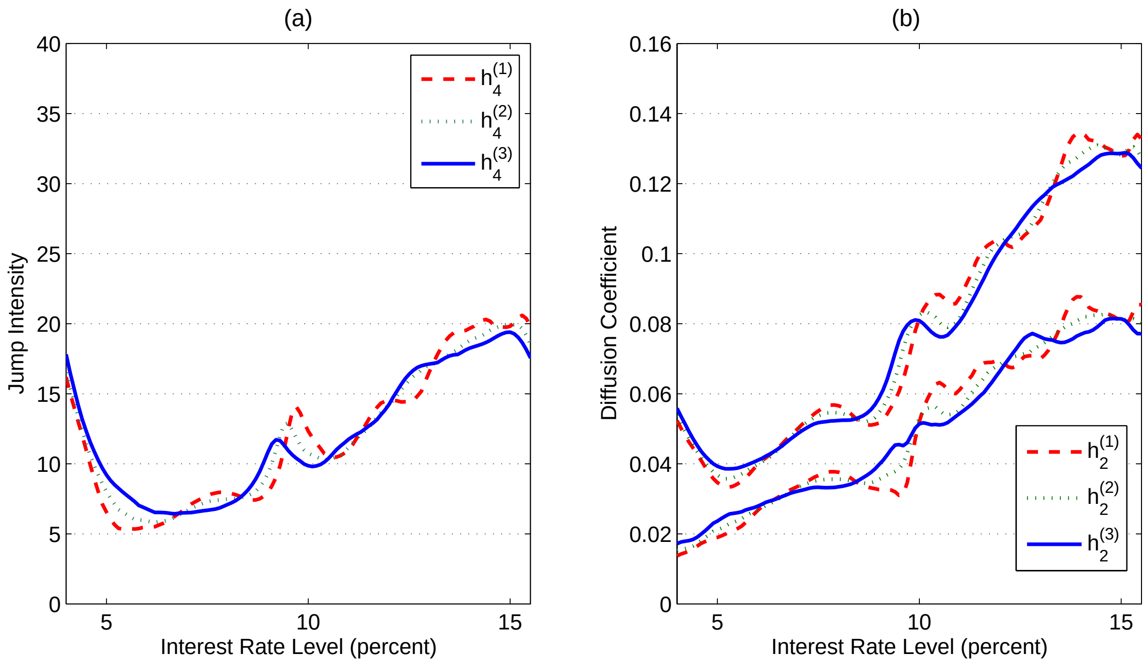

Figure 7: (a) The re-weighted local constant estimators of the jump intensity using three bandwidths; (b) the re-weighted local constant estimators of the second moment $\widehat{M}_2(r)$ (the higher three lines) and the diffusion coefficient over three bandwidths respectively.