TILTED NONPARAMETRIC ESTIMATION OF VOLATILITY FUNCTIONS WITH EMPIRICAL APPLICATIONS

BY

Ke-Li Xu and Peter C. B. Phillips

COWLES FOUNDATION PAPER NO. 1337

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

2011

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

Tilted Nonparametric Estimation of Volatility Functions With Empirical Applications

Ke-Li Xu
Department of Economics, Texas A&M University, College Station, TX 77843-4228 (*keli.xu@econmail.tamu.edu*)

Peter C. B. Phillips
Department of Economics, Cowles Foundation for Research in Economics, Yale University, P. O. Box 208281, New Haven, CT 06520; The University of Auckland Business School, Auckland 1142, New Zealand; School of Social Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom; School of Economics, Singapore Management University, Singapore 178903 (*peter.phillips@yale.edu*)

This article 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 residuals of the mean regression. 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 nonnegative 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-sectional data and studies the relationship between occupational prestige and income, and the other uses time series data on Treasury bill rates to fit the total volatility function in a continuous-time jump diffusion model.

KEY WORDS: Conditional heteroscedasticity; Conditional variance function; Empirical likelihood; Heteroscedastic nonparametric regression; Jump diffusion; Local linear estimator.

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, Todd, and Van der Klaauw 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 generalized least squares-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 devolatized residuals.

When the conditional variance is modeled nonparametrically, as in the applications mentioned earlier, the estimation methods that are commonly recommended are based on local polynomial estimation, among which local linear estimation is especially popular because of its attractive properties. The theoretical foundation for this approach has been developed by Ruppert et al. (1997) and Fan and Yao (1998), among others. However, one drawback of the local linear variance estimator, which does not apply to 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. Consequently, it is commonly recommended in applications to use the theoretically less satisfactory local constant estimator (also known as the Nadaraya–Watson estimator) when fitting the variance function (Chen and Qin 2002; Porter 2003).

In this article we propose a new volatility function estimator that is almost asymptotically equivalent to the local linear estimator but is guaranteed to be nonnegative. Our estimator has the same asymptotic bias and variance as the local linear estimator when the explanatory variable has unbounded support. Such equivalence is important, because it allows extension of efficiency arguments along the lines of those of Fan (1992) for the local linear estimator to our new procedure. It also is 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 *reweighted local constant estimator*, or *tilted estimator*, inherits the nonnegativity restriction of the variance function from the usual local constant estimator while preserving the superior bias, boundary correction, and

Journal of Business & Economic Statistics
October 2011, Vol. 29, No. 4
DOI: 10.1198/jbes.2011.09012

minimax efficiency properties of the local linear estimator. We also show the adaptiveness of this procedure to the unknown mean function; that is, it estimates the volatility function as efficiently as if the true mean function were known.

Ziegelmann (2002) recently obtained a nonnegative 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, thereby leading to an additional bias term.

The remainder of the article is organized as follows. Section 2.1 describes the nonparametric heteroscedastic regression model, the framework within which the reweighted local constant volatility estimator is introduced in Section 2.2. Section 2.3 presents the asymptotic distributional theory for stationary and mixing time series for both interior and boundary points, and suggests a consistent estimator of the asymptotic variance. Section 3 evaluates the finite-sample performance of the proposed estimator via simulations. Section 4 reports two empirical applications, a study of the volatility of the relationship between income and occupational prestige in Canada using cross-sectional data and an estimation of the total volatility of 90-day Treasury bill yields in the context of a continuous-time jump diffusion model. Section 5 concludes and presents some extensions. Proofs are collected in the Appendix.

2. MAIN RESULTS

2.1 The Heteroscedastic Regression Model

We focus on the following nonparametric heteroscedastic regression model:

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

where $\{X_t, Y_t, t = 1, \ldots, n\}$ 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 our proposed volatility estimator applies straightforwardly to the mean-0 case, for example, 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). Martins-Filho and Yao (2007) presented a recent application in stochastic frontier analysis, and Hahn, Todd, and Van der Klaauw (2001), Porter (2003), and Imbens and Lemieux (2008) addressed the analysis of causal treatment effects. The model (1) is also of fundamental importance in financial econometrics because of its ability to allow for nonlinearity and conditional heteroscedasticity in financial time series modeling. It also can be considered the discretized version of the nonparametric continuous-time diffusion model that 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 that determine model complexity. Following recommendations in the theoretical and empirical literature, we can fit $m(\cdot)$ using the local linear method that 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; Yu and Jones 2004). In what follows, we use h' for mean regression estimation and h for variance estimation.

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 was done by Ruppert et al. (1997) and Fan and Yao (1998), we consider the following reweighted 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

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

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

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

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 $w_t^{\mathrm{LL}}(x) = \Gamma_{n,2} - (X_t - x)\Gamma_{n,1}$ with $\Gamma_{n,j} = \sum_{t=1}^{n} (X_t - x)^j K_h(X_t - x)$, $j = 1, 2$, and is considered 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 $w_t^{\mathrm{UNIF}}(x) = 1/n$ for all t, which reduces (3) to the usual local constant estimator (or the Nadaraya–Watson estimator). Thus the reweighted 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 be used (Cressie and Read 1984) and has an important connection to the empirical likelihood approach of Owen (2001).

Computationally, the reweighted estimator is very easy to use in practice, because (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, that is,

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

where the Lagrange multiplier λ satisfies

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

The reweighting idea is due to the intentionally biased bootstrap of Hall and Presnell (1999). It is especially powerful for conditional variance estimation, because 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 nonnegative results. The restriction in (6) is used, so that the original estimator (i.e., the local constant estimator) is modified to the smallest extent necessary 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 also 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 reweighting idea has been used fruitfully in other contexts, for example, by Hall, Wolff, and Yao (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 nonnegative diffusion functional estimation in a continuous-time nonstationary diffusion model.

2.3 Limit Theory

The asymptotic distribution of the reweighted 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. (a) Suppose that x is such that $x \pm h$ is in the support of $f(x)$. Under the assumptions stated in the Appendix, as $n \to \infty$,

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

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

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

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

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

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

and

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

where $\underline{K}_0 = \int_{c}^{1}[1 - \underline{\lambda}_c uK(u)]^{-1}K(u)\,\mathrm{d}u$, $\underline{K}_1 = \int_{c}^{1}[1 - \underline{\lambda}_c \times uK(u)]^{-1}u^2K(u)\,\mathrm{d}u$, $\underline{K}_2 = \int_{c}^{1}[K(u)/(1 - \underline{\lambda}_c uK(u))]^2\,\mathrm{d}u$ 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 (a) is concerned with interior points when f has bounded support or the case where f has unbounded support, and part (b) is concerned with boundary points. The theorem shows that the reweighted local constant variance estimator is asymptotically equivalent to the local linear variance estimator (cf. 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 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 $p_t(x) = [\sum_{t=1}^{n}\widehat{w}_t(x)K((X_t - x)/h)]^{-1}\widehat{w}_t(x)$; see 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),

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

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 0 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 reweighted local constant conditional variance estimator follows from Cai (2001) with the outcome variable $[Y_t - m(X_t)]^2$, because $\sigma^2(x) = \mathrm{E}[(Y_t - m(X_t))^2 | X_t = x]$. 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 shared by other residual-based variance estimators (see Fan and Yao 1998; Ziegelmann 2002).

Remark 4. Implementation of the reweighted volatility estimator involves determination of the amount of smoothing, that is, selection of the smoothing bandwidth h. Theorem 1 shows that minimization of the asymptotic mean squared error (MSE) or integrated MSE (IMSE) 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 by, for example, parametric fitting (the rule of thumb), iterations (the plug-in method) or cross-validation. An attractive feature of the reweighted 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 apply to the reweighted estimator as well.

Remark 5. Härdle and Tsybakov (1997) studied a volatility estimator for the model (1) assuming that $X_t = Y_{t-1}$ based on differencing the local polynomial estimators of the second conditional moment and the squared first conditional moment. Their estimator is not nonnegative and, as noted by Fan and Yao (1998), is not fully adaptive to the mean function. Ziegelmann's (2002) nonnegative residual-based local exponential (LE) variance estimator is obtained as $\widehat{\sigma}^2_{\mathrm{LE}} = \exp(\widehat{\psi}_1)$, 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 large class of local nonlinear estimators (Hjort and Jones 1996; Gozalo and Linton 2000). To ensure nonnegativity 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 pointwise confidence intervals. Let $\widehat{\Omega}(x) = \widehat{f}^{-2}(x)\widehat{V}(x)$ where $\widehat{V}(x) = nh^{-1} \times \sum_{t=1}^n K^2((X_t - x)/h)[\widehat{r}_t^2 - \widehat{\sigma}^2(x)]^2$ and $\widehat{f}(x) = h^{-1}\sum_{t=1}^n K((X_t - x)/h)$.

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

(b) Under the conditions of Theorem 1(b), 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.

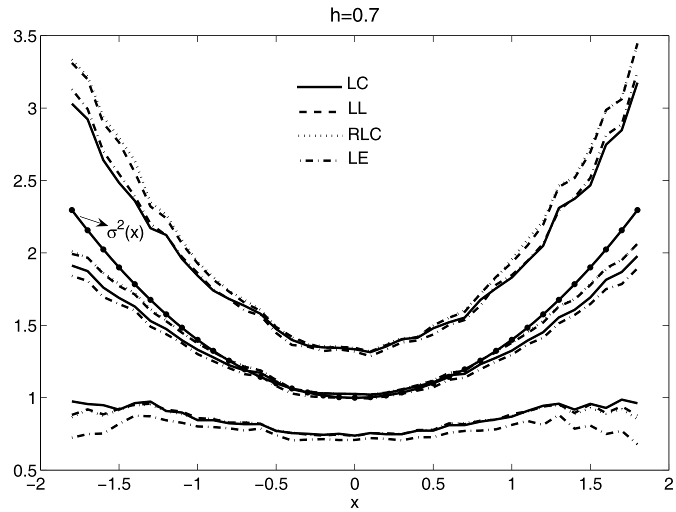

Figure 1. The means, 10% quantiles and 90% quantiles of the LC, LL, RLC, and LE estimates of the volatility function $\sigma^2(x) = 1 + 0.4x^2$ in the AR-ARCH model (11) when $\phi = 0$ over 1000 replications, using the smoothing bandwidth $h = 0.7$.

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 $\varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. The first 200 observations are dropped to eliminate initialization effects, so the sample size is n. The heteroscedastic regression model (1) is then estimated with the generated data. Note that (11) is different from the ARCH(1) model regardless of the true value of ϕ , because it allows for uncertainty in the mean function. Figures 1 and 2 focus on the case where $\phi = 0$.

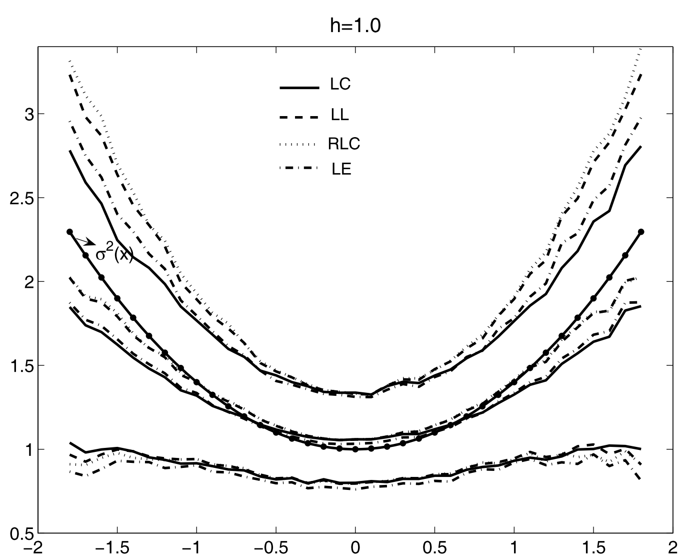

Figure 2. The means, 10% quantiles and 90% quantiles of the LC, LL, RLC, and LE estimates of the volatility function $\sigma^2(x) = 1 + 0.4x^2$ in the AR-ARCH model (11) when $\phi = 0$ over 1000 replications, using the smoothing bandwidth $h = 1.0$.

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

We plot the averages, 10% quantiles, and 90% quantiles (over 1000 replications) of the reweighted 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, we also plot the corresponding results for the local constant (LC), local linear (LL) and Ziegelmann's (2002) local exponential (LE) estimators along 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 an 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 case 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 with relatively fewer observations in their neighborhoods, for example, x with $|x| \geq 1$.

We also consider the case with serial correlation in Y_t, (i.e., $\phi = 0.4$), and we find that the results reported earlier are quite robust to weak serial correlation. Table 2 reports the MSEs of the RLC volatility estimates when the data-dependent bandwidths are used, that is, $h = \alpha\widehat{s}n^{-1/5}$, 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 visits sparingly, than 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 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 0, and the effects of estimating the unknown mean function on volatility estimation disappear asymptotically, thereby confirming the adaptiveness property suggested by the limit theory.

4. EMPIRICAL APPLICATIONS

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

4.1 Occupational Prestige versus 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) designated `Prestige`. It consists of cross-sectional observations for 102 occupations. Prestige for each occupation is measured by the Pineo–Porter prestige score from a social survey. Figure 3(a) shows a 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 also might be useful to provide variance estimates, for

Table 2. 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. represents the deviation of the MSE of the RLC volatility estimate from that of the estimate based on the true mean function)

		$x = 0$					$x = 1.5$				
	$\alpha\backslash n$	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					

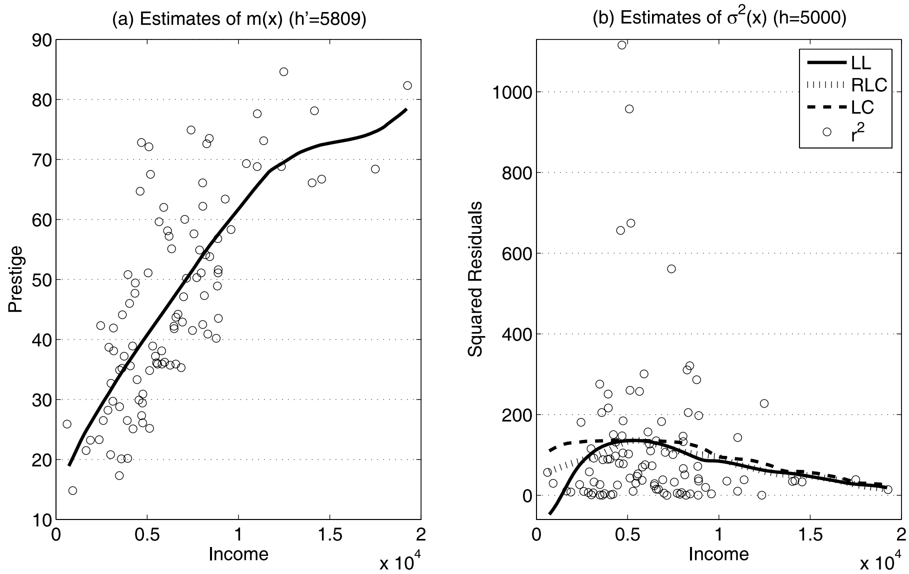

Figure 3. Prestige versus income. (a) Local linear estimation of the conditional mean function using the bandwidth $h' = 5809$; (b) Estimates of the conditional variance function based on the squared residuals using the LL, RLC, and conventional LC methods with the bandwidth $h = 5000$.

example, 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 19,211. For illustration, we use the bandwidth $h = 5000$. Clearly 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. Our proposed RLC estimates appear to provide a good compromise between those 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 income, for which the data are relatively rich, the RLC variance estimates are very close to the local linear estimates, not surprising given their first-order asymptotic similarity.

This example demonstrates the need to carefully select the bandwidth to avoid the negativity problem when the LL estimator is used to estimate variance. We also consider the estimated IMSE-based optimal bandwidth via rule of thumb (Fan and Gijbels 1996, p. 111) for the LL and RLC variance estimators. This has value $\widehat{h}_{op} = 1871$. We find that this bandwidth is too small and it gives wiggly estimated curves, necessitating intervention for bandwidth selection. Figure 4 shows the estimated curves when $h = 2\widehat{h}_{op}$. This poses no problem for the LL estimator, because the estimated curve is still above the zero line. Our empirical results indicate 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 10,000. 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 of 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 apparently are quite stable over various bandwidths. At the interior point $x = 6000$, the three fitted values are much closer to one another, and the RLC and LL curves are almost indistinguishable.

4.2 Jump Diffusion Volatilities

The reweighting idea developed in this article also can be used for functional estimation of continuous-time jump diffu-

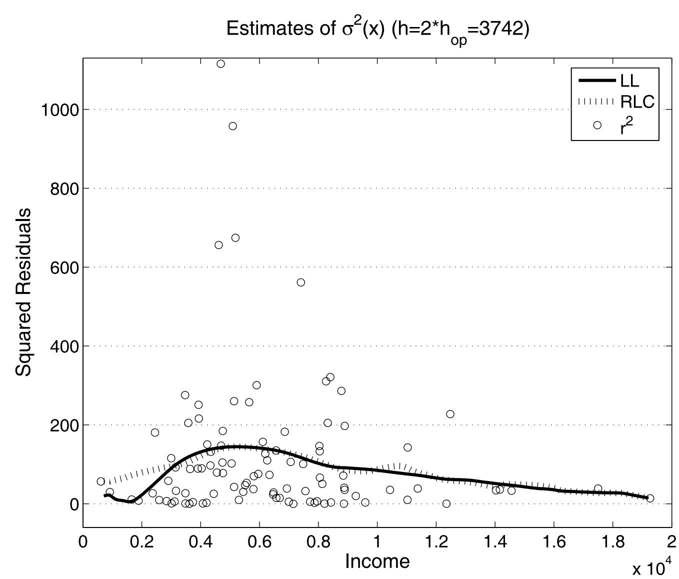

Figure 4. Prestige versus income. Estimates of the conditional variance function based on the squared residuals using the LL and RLC methods with the bandwidth $h = 2\widehat{h}_{op} = 3742$.

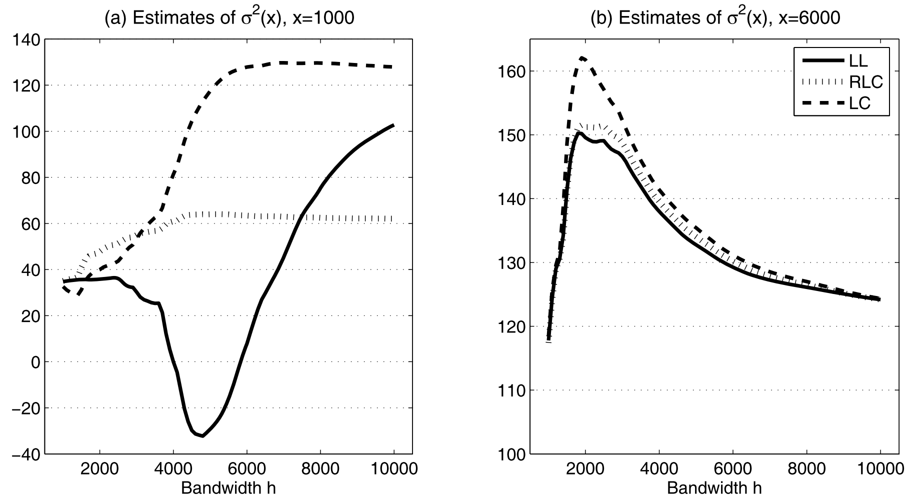

Figure 5. Prestige versus income. Estimates of the conditional variance function over 91 bandwidths using LL, RLC and LC methods with design points (a) $x = 1000$ and (b) $x = 6000$.

sions. Jump diffusion models are widely used in finance to account for discontinuities in the sample path. They are more flexible than single-factor or multifactor pure diffusion models in generating higher moments that match those typically observed in financial time series (see, e.g., Bakshi, Cao, and Chen 1997; Pan 2002; Johannes 2004).

Our empirical application uses $T = 54$ years of daily secondary market quotes for 3-month Treasury bills from January 4, 1954, to March 13, 2008, containing $n = 13{,}538$ observations, plotted in Figure 6(a). The dataset is available from the 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,

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

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 $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_z^2)$. We assume that the mean jump size is 0 without loss of generality. The four values 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:

$$M_1(r) \simeq \mu(r), \qquad M_2(r) \simeq \sigma^2(r) + \lambda(r)\sigma_z^2,$$

$$M_4(r) \simeq 3\lambda(r)\sigma_z^4, \qquad M_6(r) \simeq 15\lambda(r)\sigma_z^6.$$

We use local linear fitting to estimate $M_1(r)$, and apply our proposed reweighted local constant method 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

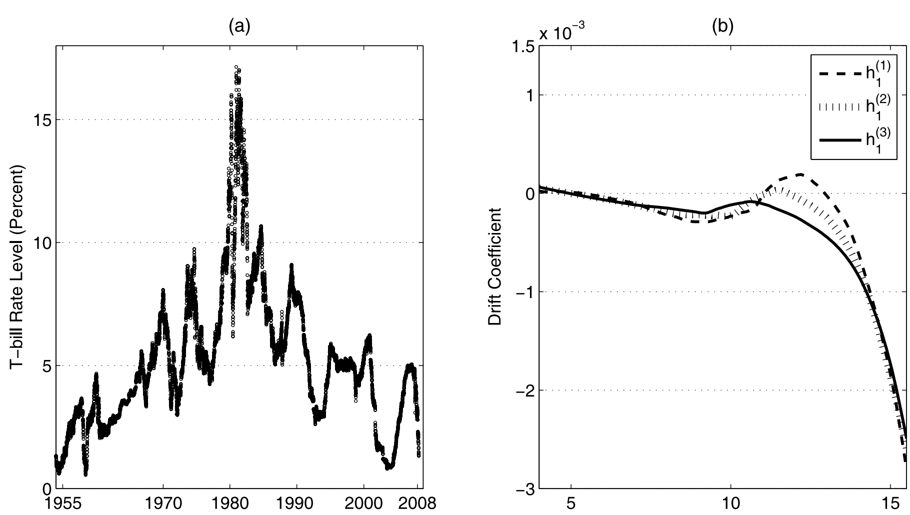

Figure 6. (a) The time series of daily 3-month Treasury bill rates (secondary market rates) from January 4, 1954, to March 13, 2008. (b) Local linear estimators of the drift function using three bandwidths, 3.5%, 4.2%, and 5.0%.

from local linear fitting. The estimates are denoted as $\widehat{M}_j(r)$, $j = 1, 2, 4, 6$. Based on the daily data, $\{r_{i\Delta},\ i = 1, \ldots, n\}$, following Johannes (2004), we obtain the estimates step by step:

$$\widehat{\sigma}_z^2 = n^{-1} \sum_{i=1}^{n} \widehat{M}_6(r_{i\Delta})/[5\widehat{M}_4(r_{i\Delta})],$$

$$\widehat{\lambda}(r) = \widehat{M}_4(r)/(3\widehat{\sigma}_z^4),$$

$$\widehat{\sigma}^2(r) = \widehat{M}_2(r) - \widehat{\lambda}(r)\widehat{\sigma}_z^2, \qquad \widehat{\mu}(r) = \widehat{M}_1(r).$$

The jump variance σ_z^2 is first estimated by integrating the ratio of sixth moments 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 thus $\widehat{\lambda}(r)$] and bandwidths $h_2^{(j)} = 1.2^j h_2$ $(j = 0, 1, 2)$ in $\widehat{M}_2(r)$, where $h_2 = 1.3\widehat{s}T^{-1/5} = 1.7\%$. Finally, $\mu(r)$ 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), because the convergence rates of the $\widehat{M}_j(r)$ depend on T (or, more generally, on the local time process), as shown by Bandi and Nguyen (2003). The scale constants that we chose are such that the resulting bandwidths are close to those reported in empirical studies of US short rate dynamics.

The estimated curves $\widehat{\mu}(r)$, $\widehat{\lambda}(r)$, $\widehat{\sigma}^2(r)$ are plotted in Figure 6(b) and Figure 7(a) and (b). 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 presents the estimates (given in the top 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 approximately two-thirds of the total volatility and the jump components account for the remaining one-third. This can be compared with the work of 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, Johannes, and Polson (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 reweighted local constant estimators of the four moments in the jump diffusion model have not yet become available in the literature. We conjecture that these theories can be studied along the lines of the approach of Bandi and Nguyen (2003). For the pure diffusion models (where $\sigma_z^2 = 0$), the asymptotic theories for these two methods have been studied by Moloche (2001), Fan and Zhang (2003), and Xu (2010).

5. CONCLUDING REMARKS

This article provides a new nonparametric approach to estimating the conditional variance function based on maximizing 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 the estimator is applicable in quite general time series and cross-sectional settings. The estimator preserves the appealing design adaptive, bias, and automatic boundary correction properties of the local linear estimator and is guaranteed to be nonnegative 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.

Our proposed method can be extended to the case where $\mathbf{X}_t$ is multivariate, for example, in the nonparametric AR-ARCH(p) model, $Y_t = m(Y_{t-1}, \ldots, Y_{t-p}) + \sigma(Y_{t-1}, \ldots, Y_{t-p})\varepsilon_t$ with $\mathbf{X}_t = (Y_{t-1}, \ldots, Y_{t-p})'$. In such cases, the constrained optimization (4) is conducted under multiple restrictions. To be specific, suppose that we have p covariates, and $\mathbf{X}_t = (X_{1,t}, \ldots, X_{p,t})'$, $\mathbf{x} = (x_1, \ldots, x_p)'$ are $p \times 1$ vectors. The RLC variance estimator is defined as $\widehat{\sigma}^2(\mathbf{x}) = [\sum_{t=1}^{n} \widehat{w}_t(\mathbf{x})\mathbf{K}_h(\mathbf{X}_t - \mathbf{x})]^{-1} \sum_{t=1}^{n} \widehat{w}_t(\mathbf{x})\mathbf{K}_h(\mathbf{X}_t - \mathbf{x})\widehat{r}_t^2$ where $\widehat{r}_t$ are residuals of a p-dimensional nonparametric mean fit (e.g., a local linear fit)

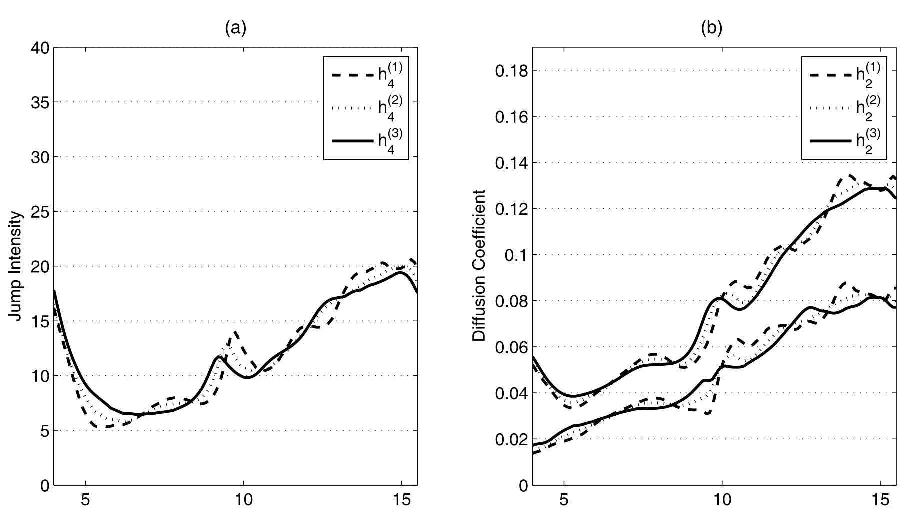

Figure 7. (a) Reweighted local constant estimators of the jump intensity using three bandwidths. (b) Reweighted local constant estimators of the second moment $\widehat{M}_2(r)$ (the top three lines) and the diffusion coefficient over three bandwidths, respectively.

and $\mathbf{K}_h(\mathbf{X}_t - \mathbf{x}) = h^{-p}\prod_{i=1}^{p} K((X_{i,t} - x_i)/h)$ are product kernel weights. Different bandwidths and kernels could be used for each covariate, but here we assume that they are the same for expositional simplicity. The weights $\widehat{w}_t(\mathbf{x})$ are such that (4) is solved subject to (5) and the p-dimensional restrictions,

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

The local linear weights satisfy (12) and take the form of, for example, when $p = 2$, $w_t^{\mathrm{LL}}(\mathbf{x}) = \widetilde{\boldsymbol{\Gamma}}_1 - \widetilde{\boldsymbol{\Gamma}}_2(X_{1,t} - x_1) + \widetilde{\boldsymbol{\Gamma}}_3(X_{2,t} - x_2)$ with

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

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

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

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

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 $\mathrm{E}(Y^4|X = x)$ are continuous at x, and $\ddot{m}(x) = \mathrm{d}^2m(x)/\mathrm{d}x^2$ 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) $\mathrm{E}|Y|^{4(1+\delta)} < \infty$ 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 (see, e.g., Davidson 1994, p. 209) 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) have a computationally convenient representation in (7). For simplicity, we write $\widehat{w}_t(x)$ as w_t in what follows. Note that $\widehat{r}_t = Y_t - \widehat{m}(X_t) = [m(X_t) - \widehat{m}(X_t)] + \sigma(X_t)\varepsilon_t$, and thus

$$\widehat{r}_t^2 = \sigma^2(X_t)\varepsilon_t^2 + 2\sigma(X_t)\varepsilon_t[m(X_t) - \widehat{m}(X_t)] + [m(X_t) - \widehat{m}(X_t)]^2. \tag{A.1}$$

Thus, by (3),

$$\widehat{\sigma}^2(x) - \sigma^2(x) = \sum_{j=1}^{4} N_j, \tag{A.2}$$

where

$$N_1 = \frac{\sum_{t=1}^{n} w_tK((X_t - x)/h)\sigma^2(X_t)(\varepsilon_t^2 - 1)}{\sum_{t=1}^{n} w_tK((X_t - x)/h)},$$

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

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

and

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

(a) Suppose that x is such that $x \pm h$ is in the support of $f(x)$. Because an absolutely regular time series is α-mixing, lemma A2 of Cai (2001) holds under our assumptions, that is, $\lambda = -\frac{hK_1f'(x)}{\upsilon_2f(x)} + O_{a.s.}(h^3)$, where $\upsilon_2 = \int u^2K^2(u)\,\mathrm{d}u$, and

$$w_t = n^{-1}\left(1 - \frac{hK_1f'(x)}{\upsilon_2f(x)}(X_t - x)K_h(X_t - x)\right)^{-1} \times (1 + o_p(1)), \tag{A.3}$$

First, consider the term N_2. The denominator of N_2 times $1/h$ is

$$h^{-1}\sum_{t=1}^{n} w_tK((X_t - x)/h) = (nh)^{-1}\sum_{t=1}^{n} K((X_t - x)/h) + o_p(1) \xrightarrow{p} f(x), \tag{A.4}$$

by (A.3) and an application of Birkhoff's ergodic theorem (see, e.g., Shiryaev 1996), because $\mathrm{E}[h^{-1}K((X_t - x)/h)] = h^{-1}\int K((u - x)/h)f(u)\,\mathrm{d}u \to f(x)$ as $h \to 0$ after a simple change of variables. By a Taylor expansion of $\sigma^2(X_t)$ at x and the discrete moment condition (6), the numerator of N_2 times $1/h$ is

$$h^{-1}\sum_{t=1}^{n} w_tK((X_t - x)/h)[\sigma^2(X_t) - \sigma^2(x)]$$

$$= h^{-1}\sum_{t=1}^{n} w_tK((X_t - x)/h)$$

$$\times [\ddot{\sigma}^2(x)(X_t - x)^2/2 + o((X_t - x)^2)]$$
$$= h^2 f(x) K_1 \ddot{\sigma}^2(x)/2 + o_p(h^2), \quad \text{(A.5)}$$

by (A.3) and the ergodic theorem. Combining (A.4) and (A.5) gives $N_2 = h^2 K_1 \ddot{\sigma}^2(x)/2 + o_p(h^2)$. Based on (A.3) and (A.4), it follows from Fan and Yao [1998, proof of thm. 1(b)–(d)] that $\sqrt{nh} N_1 \xrightarrow{d} \mathcal{N}(0, K_2 \sigma^4(x) \xi^2(x)/f(x))$, and $N_3, N_4 = o_p(h^2 + h'^2)$. Thus, by (A.2), part (a) holds.

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

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

First, consider the term N_2 in (A.2). Note that

$$h^{-1} \sum_{t=1}^{n} w_t K((X_t - a - ch)/h)$$
$$= (nh)^{-1} \sum_{t=1}^{n} \frac{K((X_t - a - ch)/h)}{1 - \lambda_c(X_t - a - ch) K_h(X_t - a - ch)} + o_p(1)$$
$$\xrightarrow{p} \overline{K}_0 f(a), \quad \text{(A.6)}$$

by the ergodic theorem, because

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

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

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

again by the ergodic theorem. Thus, by (A.6), $N_2 = [2\overline{K}_0]^{-1} \times h^2 \overline{K}_1 \ddot{\sigma}^2(a + ch) + o_p(h^2)$. Following the proof of theorem 1 of 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 0 and variance $1/nh$ times [noting (A.6)],

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

Thus the desired result follows by (A.2). The case where $x = b - ch$ can be proved similarly. The proof of (b) is complete.

Proof of Theorem 2

(a) We write $\widehat{V}(x) = \widehat{V}_1(x) + \widehat{V}_2(x) + \widehat{V}_3(x)$, where

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

First, consider the term $\widehat{V}_1(x)$. By (A.1), we have

$$\widehat{r}_t^4 = \sigma^4(X_t)\varepsilon_t^4 + 4\sigma^2(X_t)\varepsilon_t^2[m(X_t) - \widehat{m}(X_t)]^2$$
$$+ [m(X_t) - \widehat{m}(X_t)]^4 + 4\sigma^3(X_t)\varepsilon_t^3[m(X_t) - \widehat{m}(X_t)]$$
$$+ 2\sigma^2(X_t)\varepsilon_t^2[m(X_t) - \widehat{m}(X_t)]^2$$
$$+ 4\sigma(X_t)\varepsilon_t[m(X_t) - \widehat{m}(X_t)]^3,$$

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

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

and

$$\widehat{V}_{16} = 4nh^{-1} \sum_{t=1}^{n} K^2((X_t - x)/h) \sigma(X_t) \varepsilon_t [m(X_t) - \widehat{m}(X_t)]^3.$$

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

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

provided that

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

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(a) that $\widehat{V}_{1j}=o_p(1)$ for $j=2,\dots,6$. Thus, $\widehat{V}_1(x)\xrightarrow{p}(\xi^2(x)+1)K_2\sigma^4(x)f(x)$. Similarly, using (A.1), we can show that $\widehat{V}_2(x)\xrightarrow{p}-2K_2\sigma^4(x)f(x)$. Finally, $\widehat{V}_3(x)\xrightarrow{p}K_2\sigma^4(x)f(x)$. Thus $\widehat{V}(x)\xrightarrow{p}\xi^2(x)K_2\sigma^4(x)f(x)$, and Theorem 2(a) follows from (A.4).

(b) This can be proved as in (a) using the arguments given in the proof of Theorem 1(b).

ACKNOWLEDGMENTS

The authors thank the editors, the 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 an H. E. Pearson fellowship and a J. D. Muir grant. Phillips acknowledges partial research support from a Kelly Fellowship and National Science Foundation grants SES 04-142254, 06-47086 and 09-56687.

[Received January 2009. Revised July 2010.]

REFERENCES

Aït-Sahalia, Y. (1996), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560. [519]

Bakshi, G., Cao, C., and Chen, Z. (1997), "Empirical Performance of Alternative Option Pricing Models," *Journal of Finance*, 52, 2003–2049. [524]

Bandi, F., and Nguyen, T. (2003), "On the Functional Estimation of Jump-Diffusion Processes," *Journal of Econometrics*, 116, 293–328. [525]

Bandi, F., and Phillips, P. C. B. (2003), "Fully Nonparametric Estimation of Scalar Diffusion Models," *Econometrica*, 71, 241–283. [519]

Cai, Z. (2001), "Weighted Nadaraya–Watson Regression Estimation," *Statistics and Probability Letters*, 51, 307–318. [521,526,527]

——— (2002), "Regression Quantiles for Time Series," *Econometric Theory*, 18, 169–192. [520]

Chen, S. X., and Qin, Y. (2002), "Confidence Interval Based on a Local Linear Smoother," *Scandinavian Journal of Statistics*, 29, 89–99. [518]

Cressie, N., and Read, T. (1984), "Multinomial Goodness-of-Fit Tests," *Journal of the Royal Statistical Society, Ser. B*, 46, 440–464. [519]

Davidson, J. (1994), *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford: Oxford University Press. [526]

Eraker, B., Johannes, M., and Polson, N. (2003), "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58, 1269–1300. [525]

Fan, J. (1992), "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004. [518]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman & Hall. [519,521,523]

Fan, J., and Yao, Q. (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645–660. [518-521,527, 528]

Fan, J., and Zhang, C. (2003), "A Re-Examination of Stanton's Diffusion Estimations With Applications to Financial Model Validation," *Journal of American Statistical Association*, 98, 118–134. [525]

Fox, J. (2002), *An R and S-PLUS Companion to Applied Regression*, Thousand Oaks: Sage. [522]

Gozalo, P., and Linton, O. (2000), "Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression," *Journal of Econometrics*, 99, 63–106. [521]

Hahn, J., Todd, P., and Van der Klaauw, W. (2001), "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69, 201–209. [518,519]

Hall, P., and Huang, L.-S. (2001), "Nonparametric Kernel Regression Subject to Monotonicity Constraints," *The Annals of Statistics*, 29, 624–647. [520]

Hall, P., and Presnell, B. (1999), "Intentionally Biased Bootstrap Methods," *Journal of the Royal Statistical Society, Ser. B*, 61, 143–158. [518,520]

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. [520]

Hansen, B. E. (1995), "Regression With Nonstationary Volatility," *Econometrica*, 63, 1113–1132. [518]

Härdle, W., and Tsybakov, A. B. (1997), "Local Polynomial Estimators of the Volatility Function in Nonparametric Autoregression," *Journal of Econometrics*, 81, 223–242. [521]

Hjort, N. L., and Jones, M. C. (1996), "Local Parametric Nonparametric Density Estimation," *The Annals of Statistics*, 24, 1619–1647. [519,521]

Imbens, G. W., and Lemieux, T. (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635. [518,519]

Johannes, M. (2004), "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models," *Journal of Finance*, 59, 227–260. [524,525]

Kim, W., and Linton, O. (2004), "A Local Instrumental Variable Estimation Method for Generalized Additive Volatility Models," *Econometric Theory*, 20, 1094–1139. [526]

Li, Q., and Racine, J. S. (2004), "Cross-Validated Local Linear Nonparametric Regression," *Statistica Sinica*, 14, 485–512. [522]

——— (2007), *Nonparametric Econometrics: Theory and Practice*, New Jersey: Princeton University Press. [522]

Martins-Filho, C., and Yao, F. (2007), "Nonparametric Frontier Estimation via Local Linear Regression," *Journal of Econometrics*, 141, 283–319. [518, 519]

Moloche, G. (2001), "Local Nonparametric Estimation of Scalar Diffusions," unpublished manuscript, Boca Raton, FL: MIT. [525]

Owen, A. (2001), *Empirical Likelihood*, Chapman & Hall/CRC. [519]

Pagan, A., and Schwert, G. (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290. [519]

Pagan, A. R., and Hong, Y. S. (1991), "Nonparametric Estimation and the Risk Premium," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W. Barnett, J. Powell, and G. E. Tauchen, Cambridge, United Kingdom: Cambridge University Press. [519]

Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence From an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50. [524]

Porter, J. (2003), "Estimation in the Regression Discontinuity Model," working paper, Dept. of Economics, University of Wisconsin. [518,519]

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, available at *http://www.R-project.org*. [522]

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997), "Local Polynomial Variance Function Estimation," *Technometrics*, 39, 262–273. [518-520]

Shang, D. (2008), "Robust Interval Forecasts of Value-at-Risk for Nonparametric ARCH With Heavy-Tailed Errors," working paper, University of Wisconsin. [518]

Shiryaev, A. N. (1996), *Probability*, New York: Springer-Verlag. [526]

Stanton, R. (1997), "A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk," *Journal of Finance*, 52, 1973–2002. [519]

Xu, K.-L. (2010), "Re-Weighted Functional Estimation of Nonlinear Diffusions," *Econometric Theory*, 26, 541–563. [520,525]

Xu, K.-L., and Phillips, P. C. B. (2008), "Adaptive Estimation of Autoregressive Models With Time-Varying Variances," *Journal of Econometrics*, 142, 265–280. [518]

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. [519]

Ziegelmann, F. A. (2002), "Nonparametric Estimation of Volatility Functions: The Local Exponential Estimator," *Econometric Theory*, 18, 985–991. [519, 521,522]