Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Econometrics 130 (2006) 365–384
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

A semiparametric GARCH model for foreign exchange volatility

Lijian Yang*

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

Accepted 21 March 2005
Available online 10 May 2005

Abstract

A semiparametric extension of the GJR model (Glosten et al., 1993. Journal of Finance 48, 1779–1801) is proposed for the volatility of foreign exchange returns. Under reasonable assumptions, asymptotic normal distributions are established for the estimators of the model, corroborated by simulation results. When applied to the Deutsche Mark/US Dollar and the Deutsche Mark/British Pound daily returns data, the semiparametric volatility model outperforms the GJR model as well as the more commonly used GARCH(1, 1) model in terms of goodness-of-fit, and forecasting, by correcting overgrowth in volatility.

JEL classification: C13; C14; C22; C32; C53

Keywords: Equivalent kernel; Geometric mixing; Goodness-of-fit; Local polynomial

1. Introduction

In the study of foreign exchange returns, it has been a known fact that the return itself cannot be predicted. It is the forecasting of the returns' volatility that is of special interests. As a time series with zero conditional mean, the foreign exchange

*Tel.: +1 517 353 6369; fax: +1 517 432 1405.
E-mail address: yang@stt.msu.edu.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.03.006

returns are conveniently modelled as a process $\{Y_t\}_{t=0}^{\infty}$ of the form

$$Y_t = \sigma_t \xi_t, \quad t = 1, 2, \ldots,$$

where the $\{\xi_t\}_{t=1}^{\infty}$'s are i.i.d. random variables independent of Y_0 and satisfying $E(\xi_t) = 0, E(\xi_t^2) = 1, E(\xi_t^4) = m_4 < +\infty$, while $\{\sigma_t^2\}_{t=0}^{\infty}$ denotes the conditional volatility series $\sigma_t^2 = \mathrm{var}(Y_t | Y_{t-1}, Y_{t-2}, \ldots)$.

Empirical evidences had led to the understanding that σ_t^2 depends on infinitely many past returns Y_{t-j}, $j = 1, 2, \ldots$, with diminishing weights. The GARCH(p, q) model of Bollerslev (1986), for example, allows the volatility function to depend on all past observations, with geometrically decaying rate. Under the most commonly used GARCH(1, 1) model, the σ_t^2 is expressed recursively

$$\sigma_t^2 = w + \beta Y_{t-1}^2 + \alpha \sigma_{t-1}^2, \quad t = 1, 2, \ldots, 0 < \alpha, \beta < \alpha + \beta < 1,$$

which is equivalent to

$$\sigma_t^2 = v(Y_{t-1}) + \alpha v(Y_{t-2}) + \alpha^2 v(Y_{t-3}) + \cdots + \alpha^{t-1} v(Y_0), \tag{1.1}$$

where $v(y) \equiv \beta y^2 + w$, is the "news impact curve" according to Engle and Ng (1993), up to an additive constant.

The symmetric dependence of σ_t^2 on $Y_{t-1}, Y_{t-2}, \ldots$ has been questioned by many authors. In particular, the GARCH(1, 1) model (1.1) had been extended in many interesting directions in Engle and Ng (1993). Consider, for example, the following form of volatility:

$$\sigma_t^2 = w + \beta(Y_{t-1}^2 + \eta Y_{t-1}^2 1_{(Y_{t-1}<0)}) + \alpha \sigma_{t-1}^2, \quad t = 1, 2, \ldots, 0 < \alpha, \beta < \alpha + \beta < 1 \tag{1.2}$$

with an additional parameter $\eta \in R$. This was referred to as the GJR model in Engle and Ng (1993), named after and discussed in details by Glosten et al. (1993). Clearly, the case $\eta = 0$ corresponds to the GARCH(1, 1) model. In the GJR model, however, the news impact curve, $v(y) \equiv \beta(y^2 + \eta y^2 1_{(y<0)}) + w$, could be asymmetric, due to the inclusion of the parameter η. This allows one to study the different "leverage" of good news and bad news. Other parametric GARCH type models had been proposed by Hentschel (1995) and Duan (1997).

Engle and Ng (1993) also proposed a partially nonparametric ARCH news impact model, which essentially gives one the flexibility of having a smooth news impact function v, not necessarily of the form in either (1.1) or (1.2) or any specific parametric form. The proposed estimation method relied on linear spline and no consistency or rates of convergence results were presented. Along the same line of thoughts, there have been recent efforts to model the persistence of time series volatility nonparametrically. For instance, Yang et al. (1999) analyzed a multiplicative form of volatility using nonparametric smoothing. Hafner (1998) proposed to estimate both the smooth function v and the parameter α in the nonparametric GARCH model by a combination of recursive kernel smoothing and backfitting. The proposed method, however, lacks in consistency properties due to the presence of infinitely many variables in the nonparametric estimation of function v. Carroll et al. (2002) and Yang (2000, 2002), proposed a truncated version of the nonparametric

GARCH model, but pays the price of restricting the dependence of σ_t^2 on $Y_{t-1}, Y_{t-2}, \ldots$ to be finite. As a result, these truncated models do not have satisfactory prediction performance.

In this paper, I extend the GARCH(1, 1) model (1.1) in a different direction. Building on the "leverage" idea of the GJR model, one can introduce nonparametric flexibility by letting the volatility be given by

$$\sigma_t^2 = g\left\{\sum_{j=1}^{t} \alpha^{j-1} v(Y_{t-j};\eta)\right\}, \quad t = 1, 2, \ldots \tag{1.3}$$

for some smooth nonnegative link function g defined on $R_+ = [0, +\infty]$, constant $\alpha \in (0, 1)$ and where $\{v(y;\eta)\}_{\eta \in H}$ is a known family of nonnegative functions, continuous in y and twice continuously differentiable in the parameter $\eta \in H$, where $H = [\eta_1, \eta_2]$ is a compact interval with $\eta_1 < \eta_2$. The GJR model (1.2) then corresponds to the special case of $g(x) = \beta x + w/(1-\alpha), v(y;\eta) = y^2 + \eta y^2 1_{(y<0)}$. If, in addition, $\eta = 0$, one has the GARCH(1, 1) model (1.1). Hence, this semiparametric GARCH model contains both the GARCH(1, 1) and GJR as submodels, while at the same time, the nonparametric link function g calibrates the volatility's relationship with U_t, the cumulative sum of exponentially weighted past returns, as defined in (2.1). In this paper, model (1.3) is studied in full generality. In the two real data applications, it is found that the relationship of process σ_t^2 to U_t is not a simple linear relationship. Rather, it takes a sharp downward turn for larger values of U_t. This correction has given the semiparametric model a clear advantage in goodness-of-fit as well as prediction power.

The paper is organized as follows. Semiparametric estimation of both the unknown link function g and the unknown parameter vector $\gamma = (\alpha, \eta)$ are developed in Sections 2 and 3, respectively. The performance of the estimators is illustrated by analyzing a simulated example in Section 4 and two sets of daily foreign exchange return data in Section 5. All technical proofs are contained in the appendix.

2. Estimation when parameters are known

Suppose for now that the parameters α, η are known. The link function g and its derivatives can be estimated via local polynomial of degree $p \geqslant 1$. Specifics of such estimation are discussed in this section.

For convenience, define

$$U_t = \sum_{j=0}^{t} \alpha^j v(Y_{t-j};\eta), \quad t = 1, 2, \ldots, \tag{2.1}$$

which simplifies model (1.3) to

$$Y_t = g^{1/2}(U_{t-1})\xi_t, \quad \sigma_t^2 = g(U_{t-1}), \quad t = 1, 2, \ldots, \tag{2.2}$$

while the process $\{U_t\}_{t=0}^{\infty}$ satisfies the Markovian equation

$$U_t = \alpha U_{t-1} + v(g^{1/2}(U_{t-1})\xi_t; \eta), \quad t \geqslant 1. \tag{2.3}$$

The next lemma establishes the basic property of the process $\{U_t\}_{t=0}^{\infty}$.

Lemma 1. *Under Assumptions* (A1)–(A4), *both* $\{U_t\}_{t=0}^{\infty}$ *and* $\{Y_t\}_{t=0}^{\infty}$ *are geometrically ergodic processes. Furthermore, if the initial distribution of either* $\{U_t\}_{t=0}^{\infty}$ *or* $\{Y_t\}_{t=0}^{\infty}$ *is stationary, then the processes are also geometrically* β*-mixing.*

Proof. Eq. (2.3) entails that

$$E(U_t | U_{t-1} = u) = \alpha u + Ev(g^{1/2}(u)\xi_t; \eta).$$

In order to prove geometric ergodicity of $\{U_t\}_{t=0}^{\infty}$, one can apply the Theorem 3 of Doukhan (1994) Section 2.4, p. 91. Note first that conditions (H_1) and (H_2) of Doukhan's Theorem 3 are trivially verified by Assumptions (A1) and (A2). It remains to establish the condition (H_3) of Doukhan's. Note next that

$$\sup_{0 \leqslant u \leqslant a} |\alpha u + Ev(g^{1/2}(u)\xi_t; \eta)| = A(a) < +\infty$$

for any $a \in (0, +\infty)$ which verifies the second condition of (H_3). Hence one only needs to verify the first condition of (H_3), i.e., that there exists an $a \in (0, +\infty)$ and $\rho \in (0, 1), \varepsilon > 0$ such that

$$\alpha u + Ev(g^{1/2}(u)\xi_t; \eta) < \rho|u| - \varepsilon \tag{2.4}$$

for all $u > a$. From (A.2) in Assumption (A3), one has

$$\begin{aligned} \alpha u + Ev(g^{1/2}(u)\xi_t; \eta) &\leqslant \alpha u + E|g^{1/2}(u)\xi_t|^{\delta}(c_1 + c_2|\eta|) \\ &= \alpha u + E|\xi_t|^{\delta} g^{\delta/2}(u)(c_1 + c_2|\eta|), \end{aligned}$$

which is in turn bounded by $\{\alpha + m_\delta(c_1 + c_2|\eta|)g_0\}u$ using (A.1) in Assumption (A3) that $g^{\delta/2}(u) \leqslant g_0 u$ for some g_0 such that $g_0 \in (0, (1-\alpha)m_\delta^{-1}(c_1 + c_2|\eta|)^{-1})$ and u sufficiently large. Thus (2.4) is established and $\{U_t\}_{t=0}^{\infty}$ is geometrically ergodic. Now $Y_t = g^{1/2}(U_{t-1})\xi_t$ and the fact that ξ_t is independent of U_{t-1} implies the geometric ergodicity of $\{Y_t\}_{t=0}^{\infty}$. The conclusions on mixing follows from the same Theorem 3 of Doukhan's. $\square$

Based on this lemma, kernel type smoothing can be carried out with processes $\{U_t\}_{t=0}^{\infty}$ and $\{Y_t\}_{t=0}^{\infty}$, as in Härdle et al. (1998).

Now observe that

$$Y_t^2 = g(U_{t-1}) + g(U_{t-1})(\xi_t^2 - 1)$$

and so

$$E(Y_t^2 | U_{t-1} = u) = g(u), \quad \operatorname{var}(Y_t^2 | U_{t-1} = u) = g^2(u)(m_4 - 1).$$

These are the basis for the estimation procedure proposed below. Intuitively, one has the following Taylor expansion

$$g(U_{t-1}) \approx g(u) + \sum_{\lambda=1}^{p} \hat{g}^{(\lambda)}(u)(U_{t-1}-u)^{\lambda}/\lambda! = g(u) + \sum_{\lambda=1}^{p} \frac{h^{\lambda}}{\lambda!} g^{(\lambda)}(u)\left(\frac{U_{t-1}-u}{h}\right)^{\lambda}$$

for every $t = 2, \ldots, n$, where $h > 0$ is a constant varying with the sample size n called the bandwidth. Hence the set of unknown function values $\{g(u), g^{(\lambda)}(u)\}_{\lambda=1,\ldots,p}$ at a given point u can be estimated via linear least squares regression with local weights as

$$\begin{aligned} &\{\hat{g}(u), \hat{g}^{(\lambda)}(u)\}_{\lambda=1,\ldots,p} \\ &\quad = \underset{\{g_0, g_\lambda\}_{\lambda=1,\ldots,p}}{\operatorname{argmin}} \sum_{t=2}^{n} \left\{ g(U_{t-1}) - g_0 - \sum_{\lambda=1}^{p} \frac{h^{\lambda}}{\lambda!} g_{\lambda} \left(\frac{U_{t-1}-u}{h}\right)^{\lambda} \right\}^2 \\ &\qquad \times \frac{1}{h} K\{(U_{t-1}-u)/h\} \end{aligned} \tag{2.5}$$

in which K is a compactly supported and symmetric probability density function called the kernel. Obviously, those U_{t-1} that are farther from u will have less contribution to the estimation of function values at u since the weight values $K\{(U_{t-1}-u)/h\}$ they receive are smaller. The resulted set of estimates $\{\hat{g}(u), \hat{g}^{(\lambda)}(u)\}_{\lambda=1,\ldots,p}$ are called local polynomial estimates of degree p, see, for example, Fan and Gijbels (1996).

To set up proper notations, for any fixed $u \in A$, where set A is defined in Assumption (A4), define estimators

$$\hat{g}(u) = E_0'(Z'WZ)^{-1}Z'W\mathbf{V}, \tag{2.6}$$

$$\hat{g}^{(\lambda)}(u) = \lambda! h^{-\lambda} E_{\lambda}'(Z'WZ)^{-1}Z'W\mathbf{V}, \tag{2.7}$$

where

$$Z = \left\{ \left(\frac{U_i - u}{h}\right)^{\lambda} \right\}_{1 \leqslant i \leqslant n-1,\, 0 \leqslant \lambda \leqslant p}, \quad W = \operatorname{diag}\left\{\frac{1}{n} K_h(U_i - u)\right\}_{i=1}^{n-1},$$

$$\mathbf{V} = (V_i)_{2 \leqslant i \leqslant n}, \quad V_i = Y_i^2, \ i = 2, \ldots, n,$$

E_λ is a $(p+1)$ vector of zeros whose $(\lambda+1)$-element is 1, $p > 0$ is an odd integer. In the following, I denote $K_h(u) = K(u/h)/h$ and $\|K\|_2^2 = \int K^2(u)\,du$ for any function K and let $K_\lambda^*(u)$ be defined as in (A.7).

The following theorem shows that $\hat{g}(u)$ behaves like a standard univariate local polynomial estimator.

Theorem 1. *Under Assumptions* (A1)–(A4), *for any fixed* $u \in A$ *and odd* p, *as* $nh \to \infty, nh^{2p+3} = O(1)$, *the estimator* $\hat{g}(u)$ *defined by* (2.6) *satisfies*

$$\sqrt{nh}\{\hat{g}(u) - g(u) - h^{p+1}b(u)\} \xrightarrow{D} N\{0, v(u)\},$$

where

$$b(u) = \Lambda_{0,p+1} g^{(p+1)}(u)/(p+1)!,$$

$$v(u) = \|K_0^*\|_2^2 (m_4 - 1) g^2(u) \varphi^{-1}(u), \tag{2.8}$$

$\varphi(\cdot)$ is the design density of U, and $\Lambda_{0,p+1}$ is defined in (A.8).

The performance of estimator $\hat{g}(u)$ is measured by its discrepancy over a compact set A, so one needs to minimize $E \int_A \{\hat{g}(u) - g(u)\}^2 \varphi(u) \, \mathrm{d}u$, where $\varphi(\cdot)$ is the stationary density of U_t. The next corollary follows directly from Theorem 1.

Corollary 1. *The global optimal bandwidth for estimating $g(u)$ is*

$$h_{\text{opt}} = \left[\frac{\{(p+1)!\}^2 \|K_0^*\|_2^2 \int_A (m_4 - 1) g^2(u) \, \mathrm{d}u}{2n(p+1)(\Lambda_{0,p+1})^2 \int_A \{g^{(p+1)}(u)\}^2 \varphi(u) \, \mathrm{d}u} \right]^{1/(2p+3)}. \tag{2.9}$$

According to Corollary 1, if the optimal bandwidth is used, the mean squared error of $\hat{g}(u)$ is of the optimal $n^{-2(p+1)/(2p+3)}$.

For the derivatives, one has the following theorem similar to Theorem 1.

Theorem 2. *Under Assumptions* (A1)–(A4), *for any fixed $u \in A$ and $\lambda \geqslant 1$ such that $p - \lambda$ is odd, as $nh^{2\lambda+1} \to \infty, nh^{2p+3} = O(1)$, the estimator $\hat{g}^{(\lambda)}(u)$ defined by* (2.7) *satisfies*

$$\sqrt{nh^{2\lambda+1}} \{\hat{g}^{(\lambda)}(u) - g^{(\lambda)}(u) - h^{p+1-\lambda} b_\lambda(u)\} \xrightarrow{D} N\{0, v_\lambda(u)\},$$

where

$$b_\lambda(u) = \lambda! \Lambda_{\lambda,p+1} g^{(p+1)}(u)/(p+1)!,$$

$$v_\lambda(u) = (\lambda!)^2 \|K_\lambda^*\|_2^2 (m_4 - 1) g^2(u) \varphi^{-1}(u). \tag{2.10}$$

Note here that the variance terms $v(u)$ in (2.8) and $v_\lambda(u)$ in (2.10) contain the square of the mean function $g^2(u)$ instead of a general conditional variance function, as is the case in Härdle et al. (1998). This is the main reason that the set of estimators $\{\hat{g}(u), \hat{g}^{(\lambda)}(u)\}_{\lambda=1,\dots,p}$ is treated as a new set of estimators. A lesser reason being that the degree of local polynomial p is allowed to be higher than 1.

3. Estimating the parameters

Suppose now that the parameter vector $\gamma = (\alpha, \eta)$ is unknown. A nonlinear least squares procedure is shown to yield estimate of γ at the usual $\sqrt{n}$-rate. Without loss of generality, suppose that γ lies in the interior of $\boldsymbol{\Gamma} = [\alpha_1, \alpha_2] \times [\eta_1, \eta_2]$, where $0 < \alpha_1 < \alpha_2 < 1, -\infty < \eta_1 < \eta_2 < +\infty$ are boundary values known a priori.

Our approach is to go back to the estimation of function $g(\cdot)$ when γ is known and examine what occurs when one replaces γ with any unknown vector $\gamma' \in \boldsymbol{\Gamma}$. Consider therefore regressing the series $V_t = Y_t^2$ on $U_{\gamma',t-1}$, where $U_{\gamma',t}$ is a series analogous to

U_t defined by

$$U_{\gamma',t} = \sum_{j=0}^{t} \alpha'^j v(Y_{t-j};\eta') = \sum_{j=0}^{t} \alpha'^j v(g^{1/2}(U_{t-j-1})\xi_{t-j};\eta'),$$
$$t = 1,2,\ldots, \gamma' = (\alpha',\eta') \in \mathbf{\Gamma}. \tag{3.1}$$

For any $\gamma' \in \mathbf{\Gamma}$ define the predictor of V_t based on $U_{\gamma',t-1}$

$$g_{\gamma'}(u) = E(V_t | U_{\gamma',t-1} = u) \tag{3.2}$$

and the weighted mean square prediction error

$$L(\gamma') = \lim_{t\to\infty} E\{V_t - g_{\gamma'}(U_{\gamma',t-1})\}^2 \pi(\widetilde{U}_{t-1}), \tag{3.3}$$

where $\pi(\cdot)$ is a nonnegative and continuous weight function whose compact support is contained in A. The series $\widetilde{U}_t$ dominates all the possible explanatory variables $U_{\gamma',t}$: $|U_{\gamma',t}| \leqslant \widetilde{U}_t, t = 1,2,\ldots, \gamma' \in \mathbf{\Gamma}$ and it is defined as

$$\widetilde{U}_t = \sum_{j=0}^{t} \alpha_2^j v(g^{1/2}(U_{t-j-1})\xi_{t-j};\widetilde{\eta}), \quad t = 1,2,\ldots, \tag{3.4}$$

where $\widetilde{\eta}$ is defined in Assumption (A5), (A.3). Apparently $L(\gamma')$ allows the usual bias-variance decomposition

$$L(\gamma') = \lim_{t\to\infty} E\{g(U_{t-1}) - g_{\gamma'}(U_{\gamma',t-1})\}^2 \pi(\widetilde{U}_{t-1}) + (m_4 - 1)$$
$$\times \lim_{t\to\infty} Eg^2(U_{t-1})\pi(\widetilde{U}_{t-1}), \tag{3.5}$$

which equals

$$\lim_{t\to\infty} E\{g(U_{t-1}) - g_{\gamma'}(U_{\gamma',t-1})\}^2 \pi(\widetilde{U}_{t-1}) + L(\gamma).$$

Under Assumption (A8) in the appendix, $L(\gamma')$ has a unique minimum point at γ and is locally convex. Thus by minimizing the prediction error of $V_t = Y_t^2$ on $U_{\gamma',t-1}$, one should be able to locate the true parameter γ consistently.

For each $u \in A, \gamma' \in \mathbf{\Gamma}$, define now the following estimator of $g_{\gamma'}(u)$:

$$\hat{g}_{\gamma'}(u) = E_0'(Z_{\gamma'}' W_{\gamma'} Z_{\gamma'})^{-1} Z_{\gamma'}' W_{\gamma'} \mathbf{V}, \tag{3.6}$$

where

$$Z_{\gamma'} = \left\{ \left(\frac{U_{\gamma',i} - u}{h} \right)^{\lambda} \right\}_{1\leqslant i\leqslant n-1,\, 0\leqslant\lambda\leqslant p}, \quad W_{\gamma'} = \mathrm{diag}\left\{ \frac{1}{n} K_h(U_{\gamma',i} - u) \right\}_{i=1}^{n-1}.$$

Define next the estimated function

$$\hat{L}(\gamma') = \frac{1}{n} \sum_{i=2}^{n} \{V_i - \hat{g}_{\gamma'}(U_{\gamma',i-1})\}^2 \pi(\widetilde{U}_{i-1}) \tag{3.7}$$

and let $\hat{\gamma}$ be the minimizer of the function $\hat{L}(\gamma)$, i.e.

$$\hat{\gamma} = \arg\min_{\gamma'\in\mathbf{\Gamma}} \hat{L}(\gamma'). \tag{3.8}$$

Theorem 3. *Under Assumptions* (A1)–(A8), *if* $h \sim n^{-r}$ *for some* $r \in (1/2(p+1), 1/5)$, *then as* $n \to \infty$, *the* $\hat{\gamma}$ *defined by* (3.8) *satisfies*

$$\sqrt{n}(\hat{\gamma} - \gamma) \to N(\mathbf{0}, \{\nabla^2 L(\gamma)\}^{-1} \Sigma \{\nabla^2 L(\gamma)\}^{-1}), \tag{3.9}$$

where

$$\Sigma = 4(m_4 - 1)E[g^2(U_1)\pi^2(\widetilde{U}_1)\{\nabla g_{\gamma'}(U_{\gamma',1})\}\{\nabla g_{\gamma'}(U_{\gamma',1})\}^{\mathrm{T}}|_{\gamma'=\gamma}]. \tag{3.10}$$

Thus, the true parameter vector $\gamma = (\alpha, \eta)$ can be estimated by $\hat{\gamma}$ at $\sqrt{n}$-rate. One can then use the estimate $\hat{\gamma}$ in place of the unknown γ for the estimation of function g. In the next two sections, I present some numerical evidence of how the proposed procedures work for both simulated and real time series.

4. Simulation

To investigate the finite-sample precision of the proposed estimator, I have applied the procedure to time series data generated according to (1.3) with $\alpha = 0.5, \eta = 0.1, \mathbf{\Gamma} = [0.4, 0.6] \times [0, 0.2]$, and functions

$$g(u) = 0.1(2u + 1)/(1 - \alpha), \tag{4.1}$$

$$v(y; \eta) = y^2 + \eta y^2 1_{(y<0)}. \tag{4.2}$$

Notice that the data generating process actually follows the GJR model (1.2), hence possesses all the known theoretical properties of GJR model presented in Glosten et al. (1993). In particular, it is trivial to verify all the assumptions listed in the appendix.

For sample sizes $n = 400, 800, 1600$, a total of 100 realizations of length $n + 400$ are generated according to model (1.3), with functions $g(u)$ as in (4.1) and $v(y; \eta)$ as in (4.2). For each realization, the last n observations are kept as our data for inference. Truncating the first 400 observations off the series ensures that the remaining series behaves like a stationary one. Estimation of the function g is carried out according to the setups described in Sections 2 and 3, using local linear estimation (i.e., setting $p = 1$) and a rule-of-thumb plug-in bandwidth as described in Yang and Tschernig (1999). The sample sizes used in this simulation may seem rather large, but I point to the fact that the two real data sets used in the next section are both larger than the data sets for simulation.

In Fig. 1, I have overlaid the 100 function estimates $\hat{g}(u)$ with the true function $g(u)$ on the same scale, for all sample sizes. The plots show that the estimated $\hat{g}(u)$ always has an unmistakably linear shape and clearly converges to the true function $g(u)$ as the sample size increases, corroborating the asymptotics in Theorem 1. In Section 5, when the estimated function $\hat{g}(u)$ shows a clearly nonlinear shape for a real data set, the semiparametric GARCH model is decidedly preferred over the GJR model.

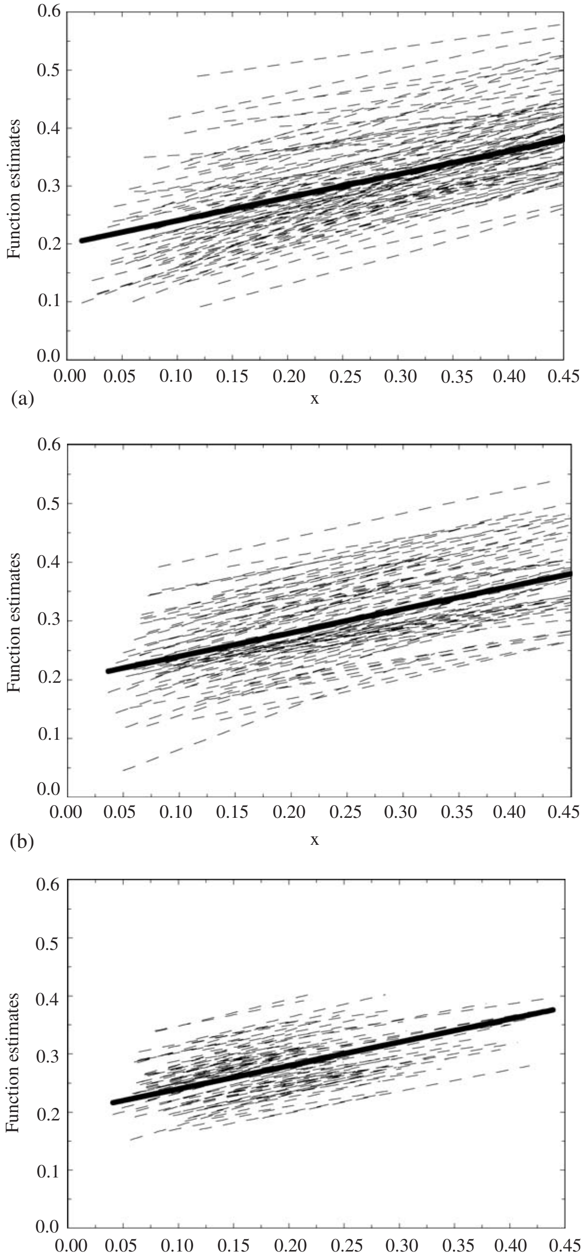

Fig. 1. Plots of the 100 Monte Carlo estimates $\hat{g}(u)$ of the function $g(u)$ in the semiparametric GARCH model: (a) sample size $n = 400$; (b) sample size $n = 800$; (c) sample size $n = 1600$. The true function $g(u)$ is plotted as the solid thick line while estimates are plotted as dashed lines.

5. Applications

In this section, I compare the goodness-of-fit of three models to the daily returns of Deutsche Mark against US Dollar (DEM/USD), and Deutsche Mark against British Pound (DEM/GBP) from January 2, 1980 to October 30, 1992. Both data sets consist of $n = 3212$ observations. The three models are: the semiparametric GARCH model (1.3); the GJR model obtained from (1.3) by setting the function g to be linear; the GARCH(1, 1) model obtained from (1.3) by setting the function g to be linear and the parameter η to be 0.

In analyzing the two data sets, a process $\{U_{\gamma',t}\}_{t=1}^{3212}$ is generated for every parameter value γ'. To have all such processes as close to strict stationarity as possible, I use only the last half for inference. Hence all estimation of parameters and function is done using $\{U_{\gamma',t}\}_{t=1607}^{3212}$ and $\{V_t\}_{t=1607}^{3212}$. The parameter estimate $\hat{\gamma}$ is first obtained according to Theorem 3 of Section 3. In the second step, following Theorem 1, local linear estimation is used, but with an undersmoothing bandwidth $h = h_{\mathrm{ROT}}/\sqrt{\ln(n/2)}$, where h_{ROT} is the rule-of-thumb optimal bandwidth, as described in Yang and Tschernig (1999). This bandwidth is called undersmoothing because $h = o(h_{\mathrm{opt}})$ with the optimal bandwidth h_{opt} as given in (2.9), yet it still satisfies the requirements of Theorem 1. Such undersmoothing technique is commonly used for obtaining nonparametric confidence intervals. It ensures that the bias term $h^{p+1}b(u)$ in Theorem 1 is negligible and one can construct pointwise 95% confidence interval of $g(u)$ as

$$\hat{g}(u) \pm z_{0.975}\sqrt{\frac{\|K_0^*\|_2^2(m_4-1)\hat{g}^2(u)}{\hat{\varphi}(u)nh}},$$

where $z_{0.975} = 1.96$, $\hat{g}(u)$ is based on local linear estimation described above and $\hat{\varphi}(u)$ is an ordinary kernel density estimator with the rule-of-thumb bandwidth defined by Silverman (1986), p. 86–87. The volatility forecasts are then $\hat{\sigma}_t^2 = \hat{g}_{\hat{\gamma}}(U_{\hat{\gamma},t})$, $t = 1607, \ldots, 3212$, while the residuals are calculated as $\hat{\xi}_t = Y_t/\hat{\sigma}_t, t = 1607, \ldots,$ 3212. For the two parametric models, the forecasts and residuals are computed similarly.

In Tables 1 and 2, the goodness-of-fit is compared for all three modelling methods, in terms of volatility prediction error and the log-likelihood, which are calculated respectively as $\sum_{t=1607}^{3212}(Y_t^2 - \hat{\sigma}_t^2)^2/1606$ and $-(1/2)\sum_{t=1607}^{3212}\{Y_t^2/\hat{\sigma}_t^2 + \ln(\hat{\sigma}_t^2)\}/1606$. Clearly, the semiparametric method has an edge over the two parametric models in

Table 1
Fitting the DEM/USD returns, all three models have $\hat{\alpha} = 0.88$

Fitted model	Log-likelihood	Volatility prediction error
GARCH(1, 1)	−0.15678287	0.66679410
GJR	−0.15663672	0.66615086
Semi. GARCH	−0.15084264	0.65293253

Table 2
Fitting the DEM/GBP returns, the GARCH(1, 1) model has $\hat{\alpha} = 0.80$, the GJR model has $\hat{\alpha} = 0.82$ while the semiparametric GARCH has $\hat{\alpha} = 0.83$

Fitted model	Log-likelihood	Volatility prediction error
GARCH(1, 1)	0.52314098	0.10452587
GJR	0.52335277	0.10397757
Semi. GARCH	0.53066411	0.099472033

Table 3
Fitting the DEM/USD returns, frequencies of autocorrelation functions (ACFs) of the absolute residuals $|\hat{\xi}_t|$ that exceed the 95% confidence bound of $1.96/\sqrt{1606}$ vs. the same for a random normal sample of the same size

ACF up to lag	$\lvert\hat{\xi}_t\rvert, Z_t$	$\lvert\hat{\xi}_t\rvert^2, Z_t^2$	$\lvert\hat{\xi}_t\rvert^3, Z_t^3$	$\lvert\hat{\xi}_t\rvert^4, Z_t^4$
100	0.04, 0.09	0.05, 0.06	0.08, 0.05	0.08, 0.05
200	0.05, 0.065	0.035, 0.04	0.06, 0.035	0.06, 0.035
300	0.04, 0.06	0.03, 0.037	0.04, 0.033	0.043, 0.047

Table 4
Fitting the DEM/GBP returns, frequencies of autocorrelation functions (ACFs) of the absolute residuals $|\hat{\xi}_t|$ that exceed the 95% confidence bound of $1.96/\sqrt{1606}$ vs. the same for a random normal sample of the same size

ACF up to lag	$\lvert\hat{\xi}_t\rvert, Z_t$	$\lvert\hat{\xi}_t\rvert^2, Z_t^2$	$\lvert\hat{\xi}_t\rvert^3, Z_t^3$	$\lvert\hat{\xi}_t\rvert^4, Z_t^4$
100	0.08, 0.09	0.05, 0.06	0.05, 0.05	0.01, 0.05
200	0.055, 0.065	0.04, 0.04	0.05, 0.035	0.025, 0.035
300	0.053, 0.06	0.037, 0.037	0.043, 0.033	0.027, 0.047

terms of prediction error and log-likelihood. One can see from these tables that the improvement of the semiparametric model over the GJR model in terms of prediction error is much greater than that of the GJR model over the GARCH(1, 1) model. This phenomenon suggests that the leverage effects of the GJR model can be further enhanced by a nonlinear link function g to yield a much better volatility fit. For diagnostic purpose, the autocorrelation functions are also calculated for the absolute powers of residuals based on the semiparametric model. In Tables 3 and 4, the frequencies of the ACF exceeding the significance limits are shown, and they are close enough for the residual absolute powers and for independent normal random samples, and hence one is reasonably sure that there is very little if any dependence left in the residuals. Further evidence of the residuals' randomness is provided in Table 5, where p-values are listed for the Ljung–Box and McLeod–Li tests of the semiparametric GARCH residuals. All p-values are larger than 0.1, and hence there is no evidence of any serial dependence in the residuals.

Table 5
Significance probabilities of Portmanteau tests on the residuals of the semiparametric GARCH model

Lag	LB for DEM/USD	ML for DEM/USD	LB for DEM/GBP	ML for DEM/GBP
20	0.684	0.673	0.109	0.714
30	0.472	0.665	0.229	0.881
40	0.262	0.252	0.272	0.978

LB: Ljung–Box tests, ML: McLeod–Li tests.

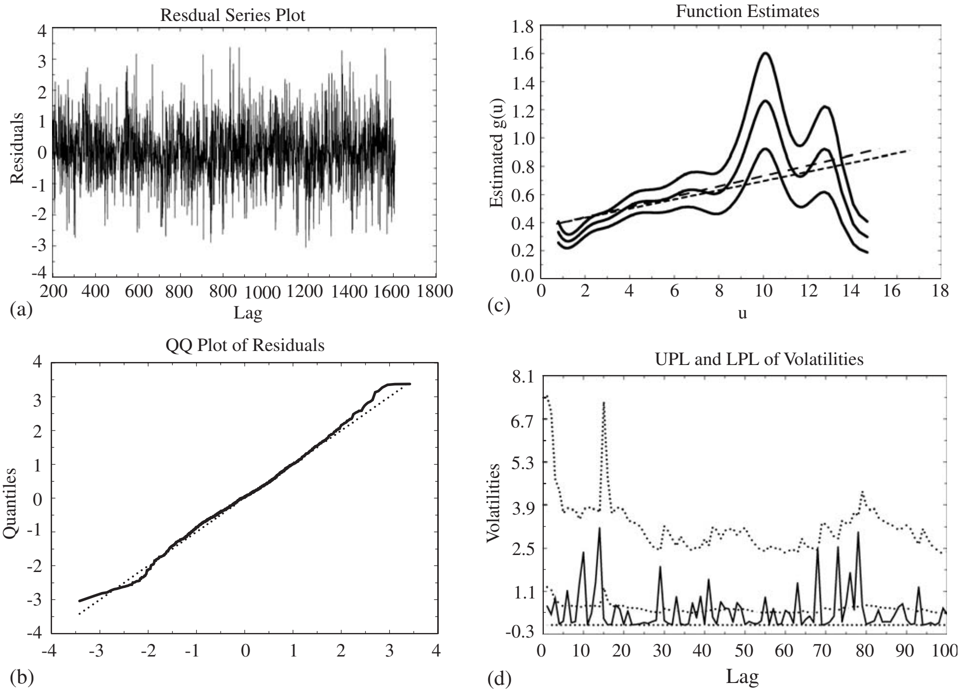

Fig. 2. Semiparametric GARCH modelling of DEM/USD daily returns: (a) residuals; (b) QQ plot of the residuals; (c) estimated function g for the semiparametric GARCH model with 95% confidence limits (solid curves) and the estimated linear functions g for the GARCH(1, 1) and GJR models (the dash lines); (d) the 95% prediction limits of the 100 squared daily returns, $V_t = Y_t^2, t = 1607, \ldots, 1706$ (solid line). The predicted volatility, lower and upper limits, are $\hat{g}_{\hat{\gamma}}(U_{\hat{\gamma},t})$, $\hat{g}_{\hat{\gamma}}(U_{\hat{\gamma},t})\hat{Q}_{0.025}$ and $\hat{g}_{\hat{\gamma}}(U_{\hat{\gamma},t})\hat{Q}_{0.975}, t = 1607, \ldots, 1706$ respectively (dotted lines).

Fig. 2 represents graphically the fit to DEM/USD, where (a) shows the standardized residuals, which seem to have a heavy-tailed distribution as one can observe in the normal quantile plot in (b). The estimated functions $\hat{g}_{\hat{\gamma}}$ with pointwise confidence intervals are overlaid in (c), showing that the semiparametric estimator behaves similarly to its parametric counterparts for smaller values of $U_{\hat{\gamma},t}$, but has a clear correction effect with higher values of $U_{\hat{\gamma},t}$, with the two straight lines outside the confidence limits. This implies that the maximum amount of volatility is reached not with the highest $U_{\hat{\gamma},t}$, as the two parametric models suggest. Rather, the volatility reaches its peak and then takes a straight dip. This gives a second reason why the semiparametric model is preferred, in addition to the optimal prediction power. Denote by $\hat{Q}_\alpha$ the αth quantile of all the residual squares $\hat{\xi}_t^2, t = 1607, \ldots, 3212$. Pointwise 95% prediction intervals are constructed for the 100 squared returns $\{V_t = Y_t^2\}_{t=1607}^{1706}$ as $[\hat{g}_{\hat{\gamma}}(U_{\hat{\gamma},t})\hat{Q}_{0.025}, \hat{g}_{\hat{\gamma}}(U_{\hat{\gamma},t})\hat{Q}_{0.975}]$, and one sees from (d) that all the 100 squared returns fall into their respective intervals. For the whole set of squared returns $\{V_t = Y_t^2\}_{t=1607}^{3212}$, the percentage of V_t's that fall outside their own prediction intervals is 0.051. Similar phenomenon is also observed for the DEM/GBP data.

In summary, the semiparametric model can fit the volatility dynamic of daily foreign exchange returns much better than the parametric models, by curtailing the overgrowth of volatility. One has good reasons to believe that this model superiority will also hold when modelling other types of volatility with geometric rate of decay and possible leverage effects.

Acknowledgements

Gratitude is due to Feike Drost for suggesting a preliminary version of the model during a visit of the author at CentER of Tilburg University in 1996. The helpful comments from two referees and the Associate Editor are very much appreciated. The research was supported in part by NSF Grants DMS 9971186, SES 0127722 and DMS 0405330.

Appendix A

The following assumptions are used:

A1: The random variable ξ_1 has a continuous density function which is positive everywhere.

A2: The link function $g(\cdot)$ is positive everywhere on R_+ and has Lipschitz continuous $(p+1)$th derivative.

A3: There exist constants $\delta, c_1, c_2 > 0$ such that

$$\limsup_{u \to +\infty} g^{\delta/2}(u)/u = g_0 \in \left(0, \frac{1-\alpha}{m_\delta(c_1 + c_2|\eta|)}\right) \tag{A.1}$$

where $m_\delta = E|\xi_1|^\delta < \infty$ and that for every $a > 0$

$$Ev(a\xi_1; \eta) \leqslant a^\delta m_\delta(c_1 + c_2|\eta|). \tag{A.2}$$

A4: The variable U_t has a stationary density $\varphi(\cdot)$ which is Lipschitz continuous and satisfy $\inf_{u \in A} \varphi(u) > 0$, where A is a compact subset of R with nonempty interior.

A5: There exists a $\widetilde{\eta} \in H$ such that for any $y \in R$

$$v(y;\widetilde{\eta}) = \max_{\eta \in H} v(y;\eta) \tag{A.3}$$

and that

$$\limsup_{u \to +\infty} g^{\delta/2}(u)/u = g_0 \in \left(0, \frac{1-\alpha_2}{m_\delta\{c_1 + c_2|\widetilde{\eta}|\}}\right). \tag{A.4}$$

The next lemma establishes the ergodicity and mixing properties of the family of processes $\{(U_t, U_{\gamma',t}, \widetilde{U}_t)\}_{t \geqslant 1}$:

Lemma A.1. *Under Assumptions* (A1)–(A5), *processes* $\{(U_t, U_{\gamma',t}, \widetilde{U}_t)\}_{t \geqslant 1}$, $\gamma' \in \mathbf{\Gamma}$ *are uniformly geometrically ergodic and* ϕ*-mixing*: *there exists a constant* $\rho \in (0,1)$ *such that*

$$\|P^k_{\gamma'}(.|u, u_{\gamma'}, \widetilde{u}) - P^k_{\gamma'}(.|u', u'_{\gamma'}, \widetilde{u}')\|_{\mathrm{Var}} \leqslant 2\rho^k, \|P^k_{\gamma'}(.|u, u_{\gamma'}, \widetilde{u}) - \pi_{\gamma'}(.)\|_{\mathrm{Var}} \leqslant 2\rho^k \tag{A.5}$$

for all $u, u_{\gamma'}, \widetilde{u}, u', u'_{\gamma'}, \widetilde{u}' \in R_+, k = 1, 2, \ldots, \gamma' \in \mathbf{\Gamma}$, *where* $P^k_{\gamma'}(.|u, u_{\gamma'}, \widetilde{u})$ *is the probability measure of* $(U_{k+1}, U_{\gamma',k+1}, \widetilde{U}_t)$ *conditional on* $U_1 = u, U_{\gamma',1} = u_{\gamma'}, \widetilde{U}_t = \widetilde{u}$, $\pi_{\gamma'}(.)$ *the stationary distribution of* $\{(U_t, U_{\gamma',t}, \widetilde{U}_t)\}_{t \geqslant 1}$, *and* $\|\cdot\|_{\mathrm{Var}}$ *denotes the total variation distance. Consequently*, *the conditional distribution of* $(U_t, U_{\gamma',t}, \widetilde{U}_t)$ *converges to* $\pi_{\gamma'}$ *in total variation at the rate of* $2\rho^t$ *and its* ϕ*-mixing coefficient* $\phi_t \leqslant 2\rho^t$, *regardless of its initial distribution and the parameter value* $\gamma' \in \mathbf{\Gamma}$.

Proof. See the downloadable manuscript Yang (2004), pp. 11–12 for complete proof. □

Eq. (A.5) is very useful as it allows one to obtain the asymptotics of function $\hat{L}(\gamma')$ uniformly for all $\gamma' \in \mathbf{\Gamma}$. This then allows the use of $\hat{L}(\gamma')$ as an uniform approximation of $L(\gamma')$ and establishes the consistency of $\hat{\gamma}$ as an estimator of γ. For this scheme to work, one needs the following assumptions as well:

A6: The processes $\{(U_t, U_{\gamma',t}, \widetilde{U}_t)\}_{t \geqslant 1}$ have stationary densities $\varphi(u, u_{\gamma'}, \widetilde{u})$, and there are constants m and M such that $0 < m \leqslant \varphi_{\gamma'}(u) \leqslant M < \infty, u \in A, \gamma' \in \mathbf{\Gamma}$ where $\varphi_{\gamma'}(\cdot)$ is the marginal stationary density of $U_{\gamma',t}$.

A7: The functions $g_{\gamma'}(u), \gamma' \in \mathbf{\Gamma}$ defined in (3.2) satisfy $\sup_{\gamma' \in \mathbf{\Gamma}} \sup_{u \in A} |g^{(p+1)}_{\gamma'}(u)| < +\infty$ while the process $\{Y_t\}_{t=0}^{\infty}$ satisfies $E \exp\{a|Y_t|^r\} < +\infty$ for some constants $a > 0$ and $r > 0$.

A8: The function $L(\gamma')$ has a positive definite Hessian matrix at its unique minimum γ and is locally convex: there is a constant $C > 0$ such that $L(\gamma') \geqslant L(\gamma) + C\|\gamma' - \gamma\|^2, \gamma' \in \mathbf{\Gamma}$ where $\|\cdot\|$ is the Euclidean norm.

Denote $\mu_r(K) = \int u^r K(u)\,\mathrm{d}u$ and let $(s_{\lambda\lambda'})_{0\leqslant \lambda,\,\lambda'\leqslant p} = S^{-1}$ where the matrix S is defined as

$$S = \{\mu_{\lambda+\lambda'}(K)\}_{0\leqslant \lambda,\,\lambda'\leqslant p} = \begin{Bmatrix} \mu_0(K) & 0 & \mu_2(K) & \cdots & 0 \\ 0 & \mu_2(K) & 0 & \cdots & \mu_{p+1}(K) \\ \mu_2(K) & 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \mu_{p+1}(K) & 0 & \cdots & \mu_{2p}(K) \end{Bmatrix}. \tag{A.6}$$

Define next the equivalent kernel

$$K_\lambda^*(u) = \sum_{\lambda'=0}^{p} s_{\lambda\lambda'} u^{\lambda'} K(u), \quad \lambda = 0, 1, \ldots, p. \tag{A.7}$$

Note here by the definition of matrix S, $K_\lambda^*(u)$ satisfies the following moment equation:

$$\int K_\lambda^*(u) u^{\lambda''}\,\mathrm{d}u = \begin{cases} 1 & \lambda'' = \lambda, \\ 0 & 0\leqslant \lambda'' \leqslant p, \lambda'' \neq \lambda, \\ \Lambda_{\lambda,p+1} & \lambda'' = p+1, \end{cases} \tag{A.8}$$

where $\Lambda_{\lambda,p+1}$ is a nonzero constant, see Fan and Gijbels (1996) p. 64. Note also by definition that

$$S^{-1} Z'_{\gamma'} W_{\gamma'} = \frac{1}{n\varphi_{\gamma'}(u)} \begin{Bmatrix} K_{0,h}^*(U_{\gamma',1} - u) & \cdots & K_{0,h}^*(U_{\gamma',n-1} - u) \\ \vdots & \ddots & \vdots \\ K_{p,h}^*(U_{\gamma',1} - u) & \cdots & K_{p,h}^*(U_{\gamma',n-1} - u) \end{Bmatrix}. \tag{A.9}$$

Lemma A.2. *Under Assumptions* (A1)–(A6), *as* $n \to \infty$

$$Z'_{\gamma'} W_{\gamma'} Z_{\gamma'} = \varphi_{\gamma'}(u) S\{I + o_p(1)\} \tag{A.10}$$

and therefore

$$(Z'_{\gamma'} W_{\gamma'} Z_{\gamma'})^{-1} = \varphi_{\gamma'}(u)^{-1} S^{-1}\{I + o_p(1)\} \tag{A.11}$$

uniformly for all $\gamma' \in \boldsymbol{\Gamma}$.

Proof. See Fan and Gijbels (1996) p. 64. In addition, for the uniformity of convergence, one uses Lemma A.1. □

Proof of Theorems 1 and 2. By definition

$$\hat{g}^{(\lambda)}(u) = \lambda! h^{-\lambda} E'_\lambda (Z'WZ)^{-1} Z'W\mathbf{V} = \lambda! h^{-\lambda} E'_\lambda (Z'_\gamma W Z_\gamma)^{-1} Z'_\gamma W \mathbf{V}$$

as the true parameter vector equals γ. By definition of the matrices, for a fixed λ

$$E'_\lambda (Z'_\gamma W Z_\gamma)^{-1} Z'_\gamma W Z_\gamma E_\lambda = 1, \quad E'_\lambda (Z'_\gamma W Z_\gamma)^{-1} Z'_\gamma W Z_\gamma E_{\lambda'} = 0,$$
$$0 \leqslant \lambda' \leqslant p, \quad \lambda' \neq \lambda$$

so

$$\hat{g}^{(\lambda)}(u) - g^{(\lambda)}(u) = \lambda! h^{-\lambda} E'_\lambda (Z'_\gamma W Z_\gamma)^{-1} Z'_\gamma W \mathbf{V} - g^{(\lambda)}(u) E'_\lambda (Z'_\gamma W Z_\gamma)^{-1} Z'_\gamma W Z_\gamma E_\lambda - \sum_{\lambda' \geqslant 0,\, \lambda' \neq \lambda} \frac{\lambda!}{\lambda'!} g^{(\lambda')}(u) h^{\lambda'} E'_\lambda (Z'_\gamma W Z_\gamma)^{-1} Z'_\gamma W Z_\gamma E_{\lambda'}.$$

Applying (A.11), the above equals

$$= I_1 + I_2,$$

where

$$I_1 = \frac{\lambda!}{n\varphi(u) h^\lambda} \sum_{i=1}^{n-1} K^*_{\lambda,h}(U_i - u) \left\{ g(U_i) - \sum_{\lambda'=0}^{p} \frac{1}{\lambda'!} g^{(\lambda')}(u)(U_i - u)^{\lambda'} \right\} \{1 + o_p(1)\},$$

$$I_2 = \frac{\lambda!}{n\varphi(u) h^\lambda} \sum_{i=1}^{n-1} K^*_{\lambda,h}(U_i - u) g(U_i)(\xi^2_{i+1} - 1)\{1 + o_p(1)\}$$

by (A.9). Note next that

$$\frac{\lambda!}{n\varphi(u) h^\lambda} \sum_{i=1}^{n-1} K^*_{\lambda,h}(U_i - u) \left\{ g(U_i) - \sum_{\lambda'=0}^{p} \frac{1}{\lambda'!} g^{(\lambda')}(u)(U_i - u)^{\lambda'} \right\}$$
$$= \frac{\lambda!}{\varphi(u) h^\lambda} \int K^*_{\lambda,h}(U - u) \sum_{\alpha=1}^{d} c_\alpha(\gamma) \left\{ g(U) - \sum_{\lambda'=0}^{p} \frac{1}{\lambda'!} g^{(\lambda')}(u)(U - u)^{\lambda'} \right\} \times \varphi(U)\, \mathrm{d}U \{1 + o_p(1)\},$$

which, by a change of variable $U = u + hv$, becomes

$$\frac{\lambda!}{\varphi(u) h^\lambda} \int K^*_\lambda(v) \left\{ g(u + hv) - \sum_{\lambda'=0}^{p} \frac{1}{\lambda'!} g^{(\lambda')}(u) h^{\lambda'} v^{\lambda'} \right\} \varphi(u + hv)\, \mathrm{d}v \{1 + o_p(1)\}$$
$$= \frac{\lambda!}{\varphi(u) h^\lambda} \varphi(u) \frac{h^{p+1}}{(p+1)!} g^{(p+1)}(u) \int K^*_\lambda(v) v^{p+1}\, \mathrm{d}v \{1 + o_p(1)\},$$

which yields

$$I_1 = \frac{\lambda! \Lambda_{\lambda,p+1} g^{(p+1)}(u)}{(p+1)!} h^{p+1-\lambda} + o_p(h^{p+1-\lambda}) = h^{p+1-\lambda} b_\lambda(u) + o_p(h^{p+1-\lambda}). \qquad \text{(A.12)}$$

On the other hand, using martingale central limit theorem as in Härdle et al. (1998), the term I_2 is asymptotically normal, with variance

$$\frac{(\lambda!)^2(m_4-1)}{n\varphi^2(u)h^{2\lambda}}\int \{K^*_{\lambda,h}(U-u)g(U)\}^2\varphi(U)\,\mathrm{d}U\{1+o_p(1)\},$$

which, by a change of variable $U=u+hv$, becomes

$$\frac{(\lambda!)^2(m_4-1)\|K^*_\lambda\|_2^2 g^2(u)}{nh^{2\lambda+1}\varphi(u)}\{1+o_p(1)\}=\frac{1}{nh^{2\lambda+1}}v_\lambda(u)\{1+o_p(1)\}. \tag{A.13}$$

Combining (A.12) and (A.13), we have finished the proof of (2.10). □

The proof of Theorem 3 makes use of the following technical lemma:

Lemma A.3. *Under Assumptions* (A1)–(A7), *for* $k=0,1,2$, *as* $n\to\infty$

$$\sup_{\gamma'\in\Gamma}|\nabla^{(k)}\hat{L}(\gamma')-\nabla^{(k)}L(\gamma')|=O\left(h^{p+1-k}+\left(\sqrt{nh}\right)^{-1}h^{-k}\log n\right)\ a.s. \tag{A.14}$$

Proof. I illustrate the case of $k=0$, the other cases involve more cumbersome notations but are essentially the same steps. For any $u\in A$

$$\hat{g}_{\gamma'}(u)-g_{\gamma'}(u)=I_1(u)\{1+a_1(u)\}+I_2(u)\{1+a_2(u)\},$$

where

$$I_1(u)=\frac{1}{n\varphi(u)}\sum_{i=1}^{n-1}K^*_{0,h}(U_{\gamma',i}-u)\left\{g_{\gamma'}(U_{\gamma',i})-g_{\gamma'}(u)-\sum_{\lambda'=1}^{p}\frac{1}{\lambda'!}g_{\gamma'}^{(\lambda')}(u)(U_{\gamma',i}-u)^{\lambda'}\right\},$$

$$I_2(u)=\frac{1}{n\varphi(u)}\sum_{i=1}^{n-1}K^*_{0,h}(U_{\gamma',i}-u)\{V_{i+1}-g_{\gamma'}(U_{\gamma',i})\}$$

by (A.9), where $\sup_{\gamma'\in\Gamma}\{\sup_{u\in A}|a_1(u)|+\sup_{u\in A}|a_2(u)|\}\to 0$ in probability, according to Lemma A.2. It is clear from the proof of (A.12) and the first half of Assumption (A7) that

$$\sup_{\gamma'\in\Gamma}\sup_{u\in A}|I_1(u)|\leqslant Ch^{p+1}\sup_{\gamma'\in\Gamma}\sup_{u\in A}|g_{\gamma'}^{(p+1)}(u)|=O(h^{p+1})\text{ a.s.}$$

Using Theorem 3.2, p. 73 of Bosq (1998), and second half of Assumption (A7), one has also

$$\sup_{\gamma'\in\Gamma}\sup_{u\in A}|I_2(u)|=O\left(\left(\sqrt{nh}\right)^{-1}\log n\right)\text{ a.s.}$$

Putting all these together, one has

$$\sup_{\gamma'\in\Gamma}\sup_{u\in A}|g_{\gamma'}(u)-\hat{g}_{\gamma'}(u)|=O\left(h^{p+1}+\left(\sqrt{nh}\right)^{-1}\log n\right)\text{ a.s.}$$

Next one notes that $\hat{L}(\gamma')$ apparently allows the following decomposition:

$$\hat{L}(\gamma') = \frac{1}{n}\sum_{i=1}^{n-1}\{V_{i+1} - g_{\gamma'}(U_{\gamma',i}) + g_{\gamma'}(U_{\gamma',i}) - \hat{g}_{\gamma'}(U_{\gamma',i})\}^2 \pi(\widetilde{U}_i)$$

and the fact that $\pi(\widetilde{U}_i) > 0$ implies that $U_{\gamma',i} \in A$, so

$$\sup_{\gamma' \in \boldsymbol{\Gamma}} \hat{L}(\gamma') - L(\gamma') = O\left(h^{p+1} + \left(\sqrt{nh}\right)^{-1} \log n\right) \text{ a.s.} \qquad \square$$

Proof of Theorem 3. The almost sure convergence of stochastic function $\hat{L}(\gamma')$ to the deterministic function $L(\gamma')$ uniformly for all $\gamma' \in \boldsymbol{\Gamma}$, with the usual application of the Borel–Cantelli Lemma, establishes that $\hat{\gamma} \to \gamma$ a.s.

Because $\hat{L}(\gamma')$ is a second order smooth function, one has $\nabla\hat{L}(\hat{\gamma}) = 0$ and hence for some $\widetilde{\gamma}_1, \widetilde{\gamma}_2$ between $\hat{\gamma}$ and γ

$$\begin{aligned}\nabla\hat{L}(\gamma) = \nabla\hat{L}(\gamma) - \nabla\hat{L}(\hat{\gamma}) &= \begin{pmatrix} (\partial^2/\partial\alpha^2)\hat{L}(\widetilde{\gamma}_1) & (\partial^2/\partial\alpha\partial\eta)\hat{L}(\widetilde{\gamma}_1) \\ (\partial^2/\partial\alpha\partial\eta)\hat{L}(\widetilde{\gamma}_2) & (\partial^2/\partial\eta^2)\hat{L}(\widetilde{\gamma}_2) \end{pmatrix}(\gamma - \hat{\gamma}) \\ &= A(\gamma - \hat{\gamma}),\end{aligned}$$

which means

$$\hat{\gamma} - \gamma = -A^{-1}\nabla\hat{L}(\gamma). \tag{A.15}$$

Since $\hat{\gamma}$ is strong consistent, $\widetilde{\gamma}_1, \widetilde{\gamma}_2$ are between $\hat{\gamma}$ and γ, and

$$\sup_{\gamma' \in \boldsymbol{\Gamma}} |\nabla^2\hat{L}(\gamma') - \nabla^2 L(\gamma')| = O\left(h^{p-1} + \left(\sqrt{nh}\right)^{-1} h^{-2} \log n\right) = o(1) \text{ a.s.}$$

According to (A.14), one concludes that $A^{-1} \to \{\nabla^2 L(\gamma)\}^{-1}$ almost surely. For any real numbers a and b, one has

$$a\frac{\partial}{\partial\alpha}\hat{L}(\gamma) + b\frac{\partial}{\partial\eta}\hat{L}(\gamma) = I_1 + I_2 + I_3 + I_4,$$

where

$$I_1 = \frac{2}{n}\sum_{i=1}^{n-1}\{g(U_i) - \hat{g}(U_i)\}\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)\{\hat{g}_{\gamma'}(U_{\gamma',i}) - g_{\gamma'}(U_{\gamma',i})\}|_{\gamma'=\gamma}\pi(\widetilde{U}_i),$$

$$I_2 = \frac{2}{n}\sum_{i=1}^{n-1} g(U_i)(\xi_{i+1}^2 - 1)\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)\{\hat{g}_{\gamma'}(U_{\gamma',i}) - g_{\gamma'}(U_{\gamma',i})\}|_{\gamma'=\gamma}\pi(\widetilde{U}_i),$$

$$I_3 = \frac{2}{n}\sum_{i=1}^{n-1}\{g(U_i) - \hat{g}(U_i)\}\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',i})|_{\gamma'=\gamma}\pi(\widetilde{U}_i),$$

$$I_4 = \frac{2}{n}\sum_{i=1}^{n-1} g(U_i)(\xi_{i+1}^2 - 1)\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',i})|_{\gamma'=\gamma}\pi(\widetilde{U}_i).$$

It remains to show that

$$|I_1| + |I_2| + |I_3| = o_p(n^{-1/2}), \quad \sqrt{n}I_4 \to N\left(0, (a \quad b)\Sigma\begin{pmatrix} a \\ b \end{pmatrix}\right), \tag{A.16}$$

where Σ is defined as in (3.10). Wald device and (A.16) entail that $\hat{L}(\gamma) \to N(0, \Sigma)$, which, together with (A.15) and the limit of A, establishes the limiting distribution of $\hat{\gamma} - \gamma$ as in Theorem 3.

Note first that the ith summand in I_4 is

$$\zeta_i = 2g(U_i)(\xi_{i+1}^2 - 1)\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',i})|_{\gamma'=\gamma}\pi(\widetilde{U}_i)$$

and $\{\zeta_i\}_{i=1}^{n-1}$ forms a martingale with respect to the σ-field sequence $\mathscr{F}_t = \sigma\{(U_i, U_{\gamma',i}, \widetilde{U}_i)\}_{i=1}^t, t = 1, \ldots, n-1$. By the martingale central limit theorem of Liptser and Shirjaev (1980), has asymptotic normal distribution with mean zero and variance as

$$\begin{aligned}
&\frac{4}{n}E\left\{g(U_1)(\xi_2^2 - 1)\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',1})|_{\gamma'=\gamma}\pi(\widetilde{U}_1)\right\}^2 \\
&\quad = \frac{4(m_4 - 1)}{n}E\left\{g(U_1)\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',1})|_{\gamma'=\gamma}\pi(\widetilde{U}_1)\right\}^2 = \frac{1}{n}(a \quad b)\Sigma\begin{pmatrix} a \\ b \end{pmatrix}.
\end{aligned}$$

According to (A.14), the term I_1 is bounded by

$$O_p\left(h^{p+1} + \left(\sqrt{nh}\right)^{-1}\log n\right)O_p\left(h^p + \left(\sqrt{nh}\right)^{-1}h^{-1}\log n\right) = o_p(n^{-1/2}).$$

By applying Lemmas 2 and 3 of Yoshihara (1976) for degenerate U-statistics of geometrically mixing series, the term I_2 is bounded by $O_p(h^p n^{-1/2} + n^{-1}h^{-1/2}h^{-1}\log n) = o_p(n^{-1/2})$, the verification is routine.

The term I_3 equals, up to a term of order $O_p(h^{p+1} + h\log n/\sqrt{n})$, the following:

$$\begin{aligned}
&\frac{2}{n}\sum_{i=1}^{n-1}\frac{1}{n\varphi(u)}\sum_{j=1}^{n-1}K_{0,h}^*(U_j - U_i)\left\{g(U_j) - g(U_i) - \sum_{\lambda=1}^{p}\frac{1}{\lambda!}g^{(\lambda)}(U_i)(U_j - U_i)^\lambda\right\} \\
&\quad \times\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',i})|_{\gamma'=\gamma}\pi(\widetilde{U}_i) \\
&\quad + \frac{2}{n}\sum_{i=1}^{n-1}\left\{\frac{1}{n\varphi(u)}\sum_{j=1}^{n-1}K_{0,h}^*(U_j - U_i)g(U_j)(\xi_{j+1}^2 - 1)\right\} \\
&\quad \times\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',i})|_{\gamma'=\gamma}\pi(\widetilde{U}_i) \\
&\quad = O_p(h^{p+1}) + Z,
\end{aligned}$$

where

$$Z = \frac{2}{n}\sum_{i=1}^{n-1}\left\{\frac{1}{n\varphi(u)}\sum_{j=1}^{n-1}K_{0,h}^{*}(U_j - U_i)g(U_j)(\xi_{j+1}^2 - 1)\right\}$$
$$\times\left(a\frac{\partial}{\partial\alpha} + b\frac{\partial}{\partial\eta}\right)g_{\gamma'}(U_{\gamma',i})|_{\gamma'=\gamma}\pi(\widetilde{U}_i).$$

Applying again the Lemmas 2 and 3 of Yoshihara (1976), one obtains that $Z = O_p(n^{-1}h^{-1/2}) = o_p(n^{-1/2})$. Hence one concludes that $I_3 = o_p(n^{-1/2})$. $\square$

References

Bollerslev, T.P., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes. Springer, New York.

Carroll, R., Härdle, W., Mammen, E., 2002. Estimation in an additive model when the components are linked parametrically. Econometric Theory 18, 886–912.

Doukhan, P., 1994. Mixing: Properties and Examples. Springer, New York.

Duan, J.C., 1997. Augmented GARCH(p, q) process and its diffusion limit. Journal of Econometrics 79, 97–127.

Engle, R.F., Ng, V., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman & Hall, London.

Glosten, L.R., Jaganathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Hafner, C., 1998. Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility. Physica-Verlag, Heidelberg.

Härdle, W., Tsybakov, A.B., Yang, L., 1998. Nonparametric vector autoregression. Journal of Statistical Planning and Inference 68, 221–245.

Hentschel, L., 1995. All in the family: nesting symmetric and asymmetric GARCH models. Journal of Financial Economics 39, 71–104.

Liptser, R.S., Shirjaev, A.N., 1980. A functional central limit theorem for martingales. Theory of Probability and its Applications 25, 667–688.

Silverman, B.W., 1986. Density Estimation. Chapman & Hall, London.

Yang, L., 2000. Finite nonparametric GARCH model for foreign exchange volatility. Communications in Statistics-Theory and Methods 5 and 6, 1347–1365.

Yang, L., 2002. Direct estimation in an additive model when the components are proportional. Statistica Sinica 12, 801–821.

Yang, L., 2004. A semiparametric GARCH model for foreign exchange volatility. Unpublished manuscript, downloadable at http://www.msu.edu/~yangli/smgarchfull.pdf.

Yang, L., Tschernig, R., 1999. Multivariate bandwidth selection for local linear regression. Journal of the Royal Statistical Society Series B 61, 793–815.

Yang, L., Härdle, W., Nielsen, J.P., 1999. Nonparametric autoregression with multiplicative volatility and additive mean. Journal of Time Series Analysis 20, 597–604.

Yoshihara, K., 1976. Limiting behavior of U-statistics for stationary, absolutely regular processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 35, 237–252.