COMMUN. STATIST.—THEORY METH., 29(5&6), 1347–1365 (2000)

FINITE NONPARAMETRIC GARCH MODEL FOR FOREIGN EXCHANGE VOLATILITY

Lijian Yang

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

Key Words: additive model; coefficient parameter; geometric decay; local polynomial; out-of-sample prediction.

ABSTRACT

GARCH model has been commonly used to describe the volatility of foreign exchange returns, which typically depends on returns many lags before. While the GARCH model provides a simple geometric decaying structure for persistence in time, it restricts the impact of variables to quadratic functions. A finite nonparametric GARCH model is proposed that allows the variables' impact to be a smooth function of any form. A direct local polynomial estimation method for this finite GARCH model is proposed based on results on proportional additive model, and is applied to the German Mark (DEM)/US Dollar (USD) daily returns data. Estimators of both the decaying rate and the impact function are obtained. Diagnostics show satisfactory out-of-sample prediction based on the proposed model, which helps to better understand the dynamics of foreign exchange volatility.

1. INTRODUCTION

Daily returns of foreign exchange rate have been commonly modeled as having zero conditional mean and GARCH conditional variance. According to Bollerslev (1986), an GARCH(p,q) model describes a time series $\{Y_t\}_{t=0}^{\infty}$ as

$$Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_P Y_{t-P} + \varepsilon_t$$

where $\varepsilon_t = \sigma_t \xi_t$, the ξ_t's are i.i.d. white noise, and the conditional volatility $\sigma_t^2 = \mathrm{var}(Y_t|Y_{t-1}, Y_{t-2}, ...)$ satisfies the recursive equations

$$\sigma_t^2 = w + \beta_1 \varepsilon_{t-1}^2 + \cdots + \beta_q \varepsilon_{t-q}^2 + \alpha_1 \sigma_{t-1}^2 + \cdots + \alpha_p \sigma_{t-p}^2$$

with $w > 0, \beta_i \geq 0, i = 1, ..., q, \alpha_j \geq 0, j = 1, ..., p$. For the foreign exchange returns, the conditional mean is reasonably assumed to be zero, so $\mu = \phi_1 = \cdots = \phi_P = 0$. This implies that $Y_t = \varepsilon_t$, and the volatility then becomes

$$\sigma_t^2 = w + \beta_1 Y_{t-1}^2 + \cdots + \beta_q Y_{t-q}^2 + \alpha_1 \sigma_{t-1}^2 + \cdots + \alpha_p \sigma_{t-p}^2$$

and, in particular

$$\sigma_t^2 = w + \beta Y_{t-1}^2 + \alpha \sigma_{t-1}^2 \tag{1}$$

if $p = q = 1$. Here, one assumes that $0 < \alpha, \beta < \alpha + \beta < 1$.

There is, however, no strong reason to believe that the contribution of Y_{t-1} to σ_t^2 is of the quadratic form in (1). Yang, Härdle and Nielsen (1999), for instance, showed that when modeling the daily German Mark/US Dollar exchange rate returns volatility multiplicatively, the simple quadratic form of (1) is inadequate. Hafner (1998) proposed the following nonparametric GARCH model as an alternative

$$\sigma_t^2 = w + g(Y_{t-1}) + \alpha \sigma_{t-1}^2$$

in order to remove the restrictive quadratic form while keeping the exponential decaying feature of (1). Equivalently, the new model is

$$\sigma_t^2 = w + g(Y_{t-1}) + \alpha g(Y_{t-2}) + \alpha^2 g(Y_{t-3}) + \cdots \tag{2}$$

where g is an arbitrary smooth function. Model (2) attempts to strike a balance between the flexibility of nonparametric function and the simplicity of

GARCH model. Analysis of classic GARCH model relies on the simple polynomial structure, which is not available for model (2). The usual smoothing approach can not be directly applied either as there are infinitely many explanatory variables to smooth over. A more manageable form is the following finite truncation

$$\sigma_t^2 = w + g(X_{t1}) + \alpha g(X_{t2}) + \cdots + \alpha^{d-1} g(X_{td})$$

for some sufficiently large integer d, where one denotes $X_{ti} = Y_{t-i}, i = 1, ..., d$. Equivalently, one can write

$$\sigma_t^2 = g(X_{t1}) + \alpha g(X_{t2}) + \cdots + \alpha^{d-1} g(X_{td}) \tag{3}$$

by replacing function g with $g + w(1-\alpha)/(1-\alpha^d)$.

At least formally, the finite nonparametric GARCH model in (3) is a special case of proportional additive model (PAM) given in (4) below. Consider a regression model

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\epsilon$$

in which $\mathbf{X} = (X_1, X_2, ..., X_d)$ are explanatory variables, ϵ is independent of $\mathbf{X}$ with $E\epsilon = 0, Var(\epsilon) = 1$, Y is a scalar dependent variable, and the regression function is

$$m(\mathbf{X}) = g(X_1) + c_2(\gamma) g(X_2) + \cdots + c_d(\gamma) g(X_d). \tag{4}$$

Here $c_2, ..., c_d$ are known coefficients functions, and $\gamma = (\gamma_1, ..., \gamma_r)$ is a vector of parameters, which may or may not be known. A detailed study of (4) can be found in Yang (1999).

To estimate both α and g in (3), Carroll, Härdle and Mammen (1999) proposed to estimate function $g(\cdot)$ by either marginal integration or backfitting procedures, and then estimate α by an ordinary least squares procedure. Although theoretical properties of the backfitting method have been obtained by Opsomer and Ruppert (1997) and Opsomer (1999), automatic bandwidth selection based on closed bias and variance formulae has only been implemented for $d = 2$ by Opsomer and Ruppert (1998). Their Assumption A2, however, requires that the dependence among the covariates be very weak.

For most time series applications, this is infeasible, hence strictly speaking, there does not yet exist a backfitting method with automatic bandwidths for highly dependent time series data.

The marginal integration method, introduced by Tjøstheim and Auestad (1994), Linton and Nielsen (1995), achieves the optimal convergence rate and has closed formulae of bias and variance. From a practical point of view, however, the integration method suffers substantial errors due to the reuse of estimators. Such errors have been shown to be of asymptotically negligible higher orders, but the asymptotics do not take effect until the sample size becomes very large. Another well-known weakness of the integration method is that when there are too few observations due to high dimensionality, the method needs to use a very large bandwidth and therefore has unacceptably large bias. Some of the implementation issues are detailed in Sperlich, Linton and Härdle (1999). The standard integration method, therefore, is not suitable for model (3) either since the dimension d is inevitably quite large. It is also not clear if the modified integration method of Hengartner (1996) can be effectively used for model (3).

In view of the disadvantages of these well-known methods for model (3), a direct local polynomial estimator of the function $g(\cdot)$ in the more general model (4) is presented in Yang (1999) when γ is known, which makes use of the proportional structure. If γ were unknown, it can be estimated by minimizing the mean squared error of the proposed function estimator. The method requires at least $O(n)$ less computing than the integration method, and evidence from Monte Carlo study shows that its practical performance is immensely superior to the integration method, as discussed in the previous paragraph, despite the fact that in theory, its rate of convergence is slower than the integration method. It enjoys the relative advantage of not having to substitute functions with estimators. The direct estimator is adapted here to analyze model (3).

The paper is organized as follows. Section 2 defines estimators of the function $g(\cdot)$, its derivatives $g^{(\lambda)}(\cdot)$, and the parameter α if it is unknown. Asymptotics results are presented for all the estimators. Section 3 applies the method to estimate the volatility function of the DEM/USD daily exchange returns, while Section 4 discusses issues related to the usefulness of the model, possible extensions, and comparison to other models. A few technical proofs are contained in the Appendix.

2. ESTIMATION PROCEDURES

Estimation results are presented in this section. To begin with, suppose for now that the parameter α is known and let $\{Y_t\}_{t=0}^n$ be a finite realization of a time series satisfying (3). Note that $Y_t = \varepsilon_t = \sigma_t\xi_t$ and hence

$$Y_t^2 = \sigma_t^2\xi_t^2 = \sigma_t^2 + \sigma_t^2(\xi_t^2 - 1)$$

which entails, as $E(\xi_t^2 - 1) = 0$, that the conditional mean and volatility of Y_t^2 given $\{Y_i\}_{i=0}^{t-1}$ are

$$\begin{aligned} E(Y_t^2|Y_{t-1}, Y_{t-2}, ...Y_0) &= \sigma_t^2 = \textstyle\sum_{\nu=1}^d \alpha^{\nu-1} g(Y_{t-\nu}) = \sum_{\nu=1}^d \alpha^{\nu-1} g(X_{t\nu}) \\ var(Y_t^2|Y_{t-1}, Y_{t-2}, ...Y_0) &= \sigma_t^4(m_4 - 1) = \sigma^2(\mathbf{X}_t) \end{aligned}$$

where $m_4 = E\xi_t^4$ denotes the 4-th moment of the white noise ξ_t, $\mathbf{X}_t = (X_{t1}, ..., X_{td})$ denotes the i-th design variable, and the function $\sigma(x_1, ..., x_d)$ is defined as

$$\sigma(x_1, ..., x_d) = (m_4 - 1)^{1/2} \sum_{\nu=1}^{d} \alpha^{\nu-1} g(x_\nu). \quad (5)$$

Throughout the rest of the paper, one denotes by $p > 0$ an integer which is the smoothness of the function $g(\cdot)$, K a kernel function, $h > 0$ a smoothing parameter called the bandwidth, and $K_h(u) = K(u/h)/h$, while for a vector $\mathbf{u} = (u_1, ..., u_d) \in R^d$, $K(\mathbf{u}) = \prod_{\nu=1}^d K(u_\nu)$. Define the equivalent kernel $K_{\lambda,\alpha}^*(\mathbf{u}) = \sum_{\nu=1}^d \sum_{\lambda'=0}^p \alpha^{\nu-1} s_{\lambda\lambda'}(\alpha) u_\nu^{\lambda'} K(\mathbf{u})$ for any $\lambda = 0, 1, ..., p$, and $\alpha \in (0,1)$, with $\{s_{st}(\alpha)\}_{0\le s,t\le p} = S_\alpha^{-1}$ while S_α is defined in (11) in the Appendix.

Denote now $S_t = Y_t^2$, and $\mathbf{S} = (S_i)_{i=d}^n$, then one can regard $\sum_{\nu=1}^d \alpha^{\nu-1} g(x_\nu)$ as the regression function of $\mathbf{S}$ on $(X_{i\nu})_{d\le i\le n, 1\le\nu\le d} = (\mathbf{X}_i)_{i=d}^n$. Following the procedures of Yang (1999), for any $x \in R$, the function $g(x)$ and its derivatives $g^{(\lambda)}(x)$ are estimated by

$$\widehat{g^{(\lambda)}}(x) = \lambda! h^{-\lambda} E_\lambda' \left(Z_\alpha' W Z_\alpha\right)^{-1} Z_\alpha' W \mathbf{S}, \lambda = 0, 1, ..., \widehat{g}(x) = \widehat{g^{(0)}}(x) \quad (6)$$

in which E_λ is a $(p+1)$ vector of zeros whose $(\lambda + 1)$-element is 1

$$Z_\alpha = \left[\sum_{\nu=1}^{d} \alpha^{(\nu-1)} \left\{(X_{i\nu} - x)/h\right\}^\lambda\right]_{d\le i\le n, 0\le\lambda\le p}, \quad W = \mathrm{diag}\left\{\frac{1}{n}K_h(\mathbf{X}_i - \mathbf{x})\right\}_{i=d}^{n}$$

and $\mathbf{x} = (x, ..., x) \in R^d$.

The following theorem shows that $\widehat{g}(x)$ and $\widehat{g^{(\lambda)}}(x)$ behave like a univariate local polynomial estimator in terms of bias, but a multivariate one in terms of variance.

Theorem 1 Under assumptions A1-A5, for any fixed x and odd p, as $nh^d \to \infty, h \to 0$, the estimator $\widehat{g}(x)$ defined by (6) satisfies

$$\sqrt{nh^d}\left\{\widehat{g}(x) - g(x) - h^{p+1}b(x)\right\} \xrightarrow{D} N\left\{0, v(x)\right\}$$

where

$$\begin{aligned} b(x) &= \Lambda_{0,p+1,\alpha} g^{(p+1)}(x)/(p+1)! \\ v(x) &= \left\|K^*_{0,\alpha}\right\|_2^2 \sigma^2(\mathbf{x})\varphi^{-1}(\mathbf{x}) \end{aligned}$$

$\varphi(\cdot)$ is the design density of $\mathbf{X}_i$, the noise function $\sigma(\cdot)$ is defined in (5), and the constant $\Lambda_{0,p+1,\alpha}$ is defined in (12). Furthermore, for any fixed x and $\lambda \ge 1$ such that $p - \lambda$ is odd, as $nh^{2\lambda+d} \to \infty, h \to 0$, the estimator $\widehat{g^{(\lambda)}}(x)$ defined by (6) satisfies

$$\sqrt{nh^{2\lambda+d}}\left\{\widehat{g^{(\lambda)}}(x) - g^{(\lambda)}(x) - h^{p+1-\lambda}b_\lambda(x)\right\} \xrightarrow{D} N\left\{0, v_\lambda(x)\right\}$$

where

$$\begin{aligned} b_\lambda(x) &= \lambda!\Lambda_{\lambda,p+1,\alpha} g^{(p+1)}(x)/(p+1)! \\ v_\lambda(x) &= (\lambda!)^2 \left\|K^*_{\lambda,\alpha}\right\|_2^2 \sigma^2(\mathbf{x})\varphi^{-d}(\mathbf{x}). \end{aligned}$$

If some form of mean squared error is used as measure of discrepancy, the optimal bandwidth rate for the estimation of function $g(\cdot)$ is then $n^{-1/(2p+2+d)}$. An optimal bandwidth was derived for the proportional additive model in Yang (1999), which can be adapted to the situation here as well. The rate of convergence of $\widehat{g}(x)$ is therefore $n^{-2(p+1)/(2p+2+d)}$.

If the parameter α is unknown, but is known to be contained in a compact interval $\mathbf{A} = [a, b]$, with $0 < a < b < 1$, then the procedure in Yang (1999) can be adapted here to yield an estimator $\widehat{\alpha}$ of α that achieves the $\sqrt{n}$-rate. Define for each $\alpha' \in \mathbf{A}$

$$\widehat{g}_{\alpha'}(x) = E_0' \left(Z_{\alpha'}' W Z_{\alpha'} \right)^{-1} Z_{\alpha'}' W \mathbf{S} \tag{7}$$

where

$$Z_{\alpha'} = \left[\sum_{\nu=1}^{d} \alpha'^{(\nu-1)} \left\{ (X_{i\nu} - x)/h \right\}^{\lambda} \right]_{d \le i \le n, 0 \le \lambda \le p}$$

and define the following function of $\alpha' \in \mathbf{A}$

$$L(\alpha') = \frac{1}{n} \sum_{i=d}^{n} \left\{ Y_i^2 - \sum_{\nu=1}^{d} \alpha'^{(\nu-1)} \widehat{g}_{\alpha'}(X_{i\nu}) \right\}^2 \pi(\mathbf{X}_i) \tag{8}$$

where $\pi(\cdot)$ is a compact supported weight function. Let $\widehat{\alpha}$ be the minimizer of the function $L(\alpha')$, i.e.

$$\widehat{\alpha} = \arg \min_{\alpha' \in \mathbf{A}} L(\alpha'). \tag{9}$$

The minimizer $\widehat{\alpha}$ of L is obtained via a grid search cross-validation over the interval $\mathbf{A}$. Similar to Yang (1999), one has

Theorem 2 Under assumptions A1-A5, if $h^{p+1} + nh^d = o(1/\sqrt{n})$, as $n \to \infty$, the $\widehat{\alpha}$ defined by equations (7), (8) and (9) satisfies

$$\sqrt{n} \left(\widehat{\alpha} - \alpha \right) \to N(0, \Sigma_\alpha)$$

for some positive definite matrix Σ_α.

All the estimators are implemented with some rule-of-thumb (ROT) bandwidths. These ROT bandwidths are defined in a fashion similar to Fan and Gijbels [1996, equation (4.3), p.111], or Yang and Tschernig (1999), i.e., to use a polynomial regression as a pilot estimator of the regression function and its derivatives, in order to estimate functionals needed for the asymptotically optimal bandwidth. Details of such implementation in the case of the proportional additive model can be found in Yang (1999), together with some discussion about the plug-in bandwidths.

3. AN APPLICATION

In this section, the above estimation scheme is applied to the daily returns of Deutche Mark/US Dollar (DEM/USD) from January 2, 1980 to October

30, 1992. The number of observations is 3212, but the asymptotic results can be expected to kick in very slowly as the data is highly dependent. For our analysis, the data is trimmed at the 1.5 percentile and the 98.5 percentile to avoid the extreme outliers.

Step 1. To first estimate the parameter α, the data is divided into 6 consecutive subsamples, each consists of $n = 500$ observations. For each subsample, model (3) is fitted with $d = 5$ or 10, respectively. The interval used is always $[a, b] = [0.70, 0.98]$. The use of subsamples is for the purpose of getting more than one estimate of α, so that a more representative value may be obtained.

The parameter estimates are

$d = 5$	0.90	0.76	0.90	0.86	0.84	0.98
$d = 10$	0.90	0.72	0.72	0.96	0.86	0.78

which are summarized as follows. When $d = 5$, one simply uses the mode as a reasonable estimate, so one takes $\alpha = 0.90$. When $d = 10$, the mode 0.72 is too extreme, one then uses the mean as a reasonable estimate, so one takes $\alpha = 0.82$. Apparently, when d, the number of lags used is increased from 5 to 10, more variables are contributing to the volatility, and hence the decaying coefficient α becomes smaller.

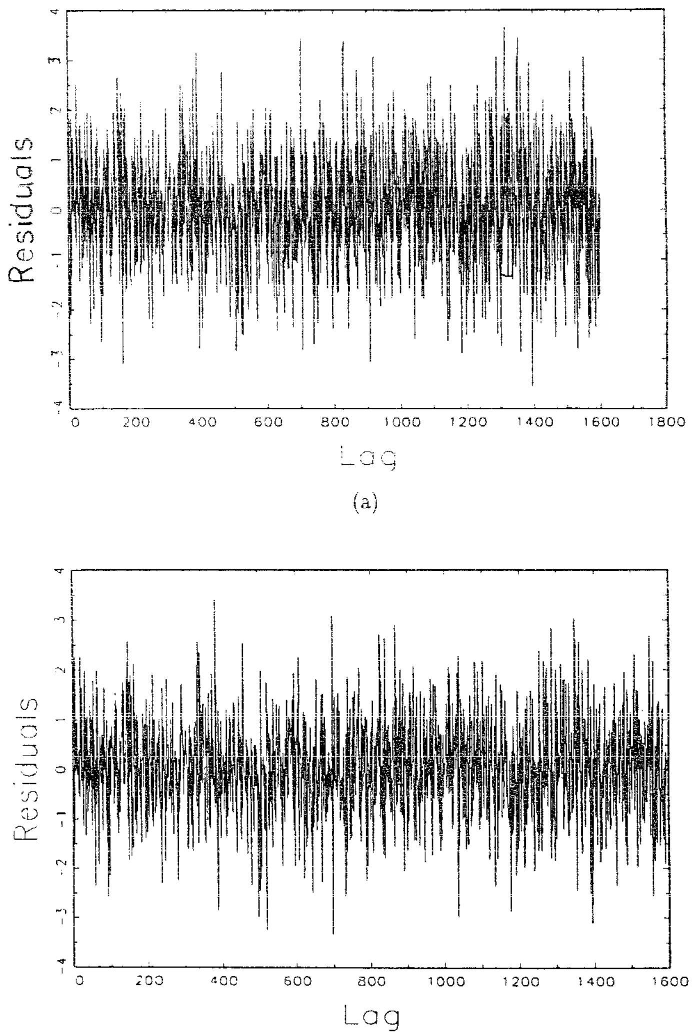
Step 2. To estimate the function $g(\cdot)$ based on the estimates of the parameter α obtained in Step 1, the data is divided into 2 consecutive subsamples, each consists of $n = 1606$ observations. The first sample $\{Y_t\}_{t=0}^{1605}$ is used to construct the estimated function $\widehat{g}(\cdot)$, and then one defines for the second subsample $\{Y_t\}_{t=1606}^{3211}$, the following estimated volatility series

$$\widehat{\sigma}_t^2 = \sum_{\nu=1}^{d} \alpha^{\nu-1} \widehat{g}(Y_{t-\nu}), t = 1606 + d, ..., 3211$$

and then the residuals

$$\widehat{\xi}_t = Y_t/\widehat{\sigma}_t = (\sigma_t/\widehat{\sigma}_t)\xi_t, t = 1606 + d, ..., 3211 \tag{10}$$

where $d = 5$ or 10 depending on which model is considered. The residuals series $\left\{\widehat{\xi}_t\right\}_{t=1606+d}^{3211}$ are shown in FIG.1, while the estimated functions $\widehat{g}(\cdot)$ (for $d = 5$ and 10 respectively) are shown in FIG.2.



(a)

(b)

FIG.1. Residuals when fitting finite nonparametric GARCH model to the DEM/USD daily returns: (a) residuals using $d = 5, p = 7$; (b) residuals using $d = 10, p = 11$.

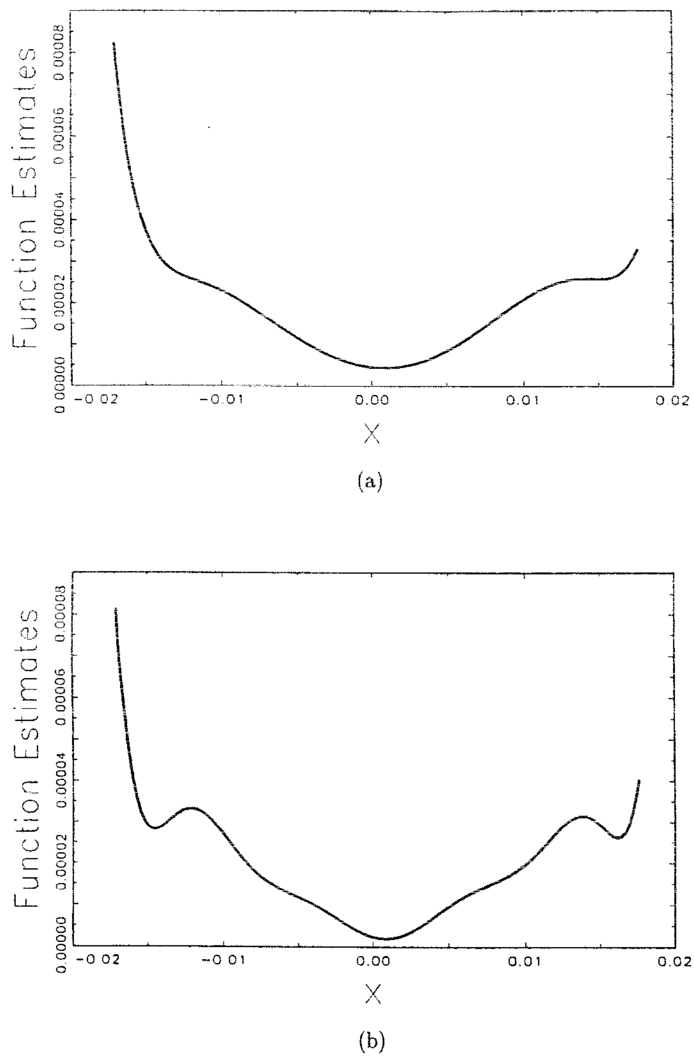
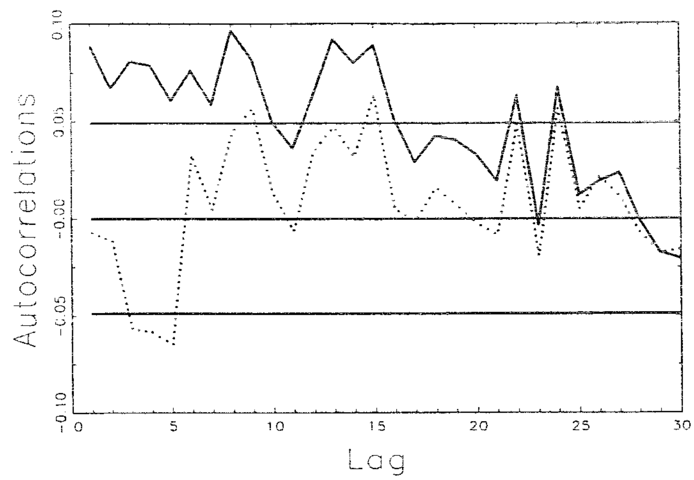


FIG.2. The estimated function g when fitting finite nonparametric GARCH model to the DEM/USD daily returns: (a) estimated function g, using $d = 5, p = 7$; (b) estimated function g, using $d = 10, p = 11$.

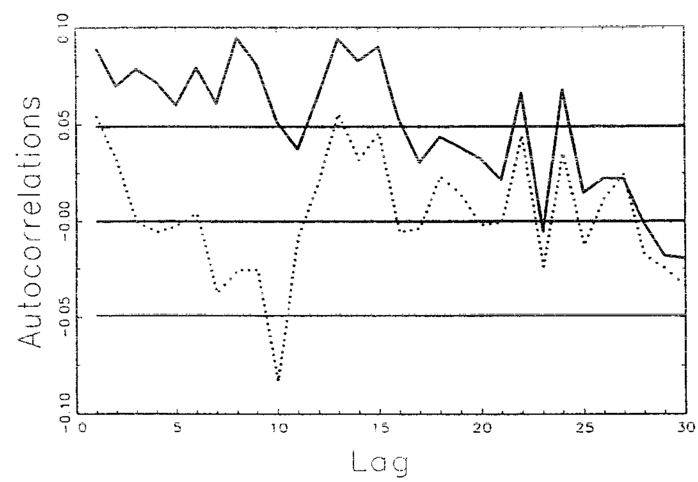
One sees from FIG.2 that functions $\widehat{g}(\cdot)$ are clearly not quadratic, and even unsymmetric. The shape near 0 is a dip, and the concavity changes at several places. If one zooms in on the middle portion of the curve, however, the function is of a roughly symmetric shape, but still not quadratic around 0. This suggests that standard GARCH model can describe fairly well the volatility if and only if all the relevant previous returns are close to neither 0 nor the boundary. In other words, if the absolute values of the previous returns are bounded from 0 and the upper bound. One can not expect this to be the case, however, when many lags are involved (say $d = 5$ or 10). This is the limitation of the standard GARCH model.

Note that the curves in FIG.2 are constructed using only data $\{Y_t\}_{t=0}^{1605}$, so it is interesting to check how these curves fit the dynamic structure of data $\{Y_t\}_{t=1606}^{3211}$. Note that the residuals $\left\{\widehat{\xi}_t\right\}_{t=1606+d}^{3211}$ in (10) are the results of fitting $\{Y_t\}_{t=1606}^{3211}$ with functions $\widehat{g}(\cdot)$, and hence should behave very much like a white noise series if the fit is good. FIG.3 shows the autocorrelation functions of both the absolute data $\{|Y_t|\}_{t=1606}^{3211}$ itself and of the absolute residuals $\left\{\left|\widehat{\xi}_t\right|\right\}_{t=1606+d}^{3211}$, the two solid horizontal lines represent the 95% upper and lower critical bounds for testing the significance of autocorrelation. One sees that the absolute data $\{|Y_t|\}_{t=1606}^{3211}$ is highly correlated and hence highly dependent, while the absolute residuals $\left\{\left|\widehat{\xi}_t\right|\right\}_{t=1606+d}^{3211}$ have much less autocorrelation. When d is increased from 5 to 10, there is further reduction of autocorrelation. For the absolute residuals, the significant autocorrelation occur roughly at lags d and its multiples (regardless of $d = 5$ or 10). This phenomenon is due to the fact that the finite model (3) truncates at lag d, and fails to take into account the contribution to the volatility by variables more than d lags before. Naturally, this implies that by using larger d, the fit can always be improved. This is not pursued further here as it does not give more insights about the problem.

Q-Q plots are made for the residuals $\left\{\widehat{\xi}_t\right\}_{t=1606+d}^{3211}$ and presented in FIG.4. The dashed line has slope 1 and passes through the origin, and the Q-Q plots are clearly very close to this line. This suggests that the distribution of the white noise ξ_t is fairly close to being normal. This, of course, should be taken with caution as one notices that the residuals $\left\{\widehat{\xi}_t\right\}_{t=1606+d}^{3211}$ are still not totally independent.

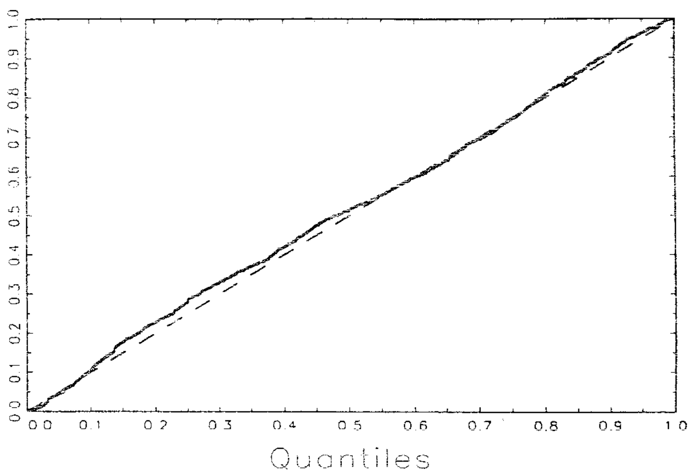


(a)

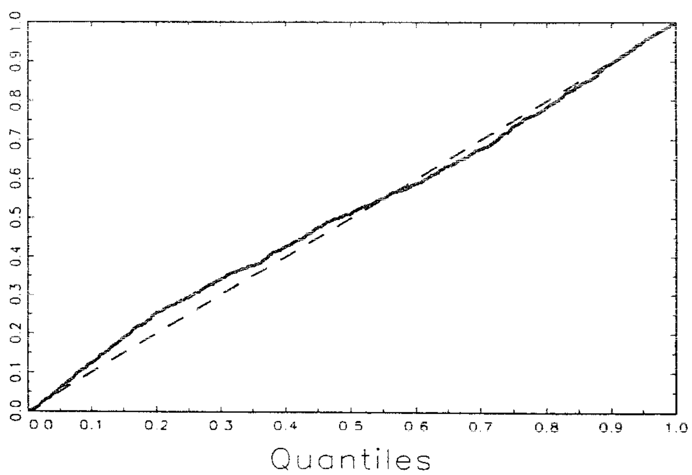


(b)

FIG.3. Autocorrelations of the absolute daily returns $|Y_t|$ (solid) and of the absolute residuals $\left|\hat{\xi}_t\right|$ (dotted) (a) residuals with $d = 5, p = 7$; (b) residuals with $d = 10, p = 11$.



(a)



(b)

FIG.4. Residuals' Q-Q plot when fitting finite nonparametric GARCH model to the DEM/USD daily returns: (a) residuals' Q-Q plot using $d = 5, p = 7$; (b) residuals' Q-Q plot using $d = 10, p = 11$.

4. DISCUSSION

In this paper we proposed a direct local polynomial estimation method for a finite nonparametric GARCH model (3). The method has good theoretical properties while its implementation advantages made it easy to analyze the DEM/USD daily return volatility. The same can be done for other foreign exchange returns as well, such as JPY (Japanese Yen) against USD, etc. The only drawback of model (3) is the finite lag restriction. Further research will extend the scheme to infinite number of lags, and one can expect much more accurate estimation of the function $g(\cdot)$ and the parameter α in model (2). Another possible improvement is to use some kind of smoothed quasi-likelihood estimation.

Model (3) can be extended to multivariate data. For example, let the bivariate time series $\{(Y_{t1}, Y_{t2})\}_{t=0}^{\infty}$ represent the daily returns of two related foreign exchange rates, such as German Mark/US Dollar and German Mark/British Pound. Then the conditional covariance matrix function

$$\Sigma_t = \text{cov}(Y_{t1}, Y_{t2} | Y_{t-1,1}, Y_{t-1,2} Y_{t-2,1}, Y_{t-2,2}, ..., Y_{01}, Y_{02})$$

could be expressed as

$$\Sigma_t = G(\mathbf{X}_{t1}) + \alpha G(\mathbf{X}_{t2}) + \cdots + \alpha^{d-1} G(\mathbf{X}_{td})$$

where $\mathbf{X}_{ti} = (Y_{t-i,1}, Y_{t-i,2}), i = 1, ..., d$ and $G(\cdot)$ is some unknown smooth symmetric 2×2 matrix function.

Further research could yield multi-step ahead nonparametric prediction of the volatility function in (3), and testing procedures to decide if some parametric form is suitable for the function $g(\cdot)$ in model (3). The latter is especially important, since one referee suggested that the asymmetric GARCH model of Engle and Ng (1993) might be sufficiently flexible to explain the asymmetry one observes in FIG.2. This has to be resolved through nonparametric hypotheses testing, and if it turned out that the referee is right, one would have a good example of nonparametric study leading to a more accurate parametric model. It does not seem, however, that this would be the case since FIG.2 shows not only asymmetry, but also significant variation in curvature, which cannot be accommodated by any second order polynomial.

Another issue raised by the referee is the comparison between model (3) and the multiplicative volatility model in Yang, Härdle and Nielsen (1999). Both attempt to measure the contribution of lagged variables to the volatility, and heuristically, they are related via the following observation. If the multiplicative model is restricted to a special form of

$$\sigma_t^2 = v(X_{t1})v^{\alpha}(X_{t2})\cdots v^{\alpha^{d-1}}(X_{td})$$

then for $g(x) = \ln v(x)$, one has

$$\sigma_t^2 = \exp\left\{g(X_{t1}) + \alpha g(X_{t2}) + \cdots + \alpha^{d-1}g(X_{td})\right\}$$

whose first order approximation is

$$1 + g(X_{t1}) + \alpha g(X_{t2}) + \cdots + \alpha^{d-1}g(X_{td}).$$

So the two models do not necessarily exclude one another.

Finally, one referee objected to the name "finite nonparametric GARCH", arguing that any GARCH model should always involve infinitely many variables. While this objection is justified, I decided after much thinking to keep the name for lack of a better one.

APPENDIX

The following assumptions are used

A1: The kernel $K(\cdot)$ is a symmetric, compactly supported, and Lipschitz continuous probability density;

A2: The function $g(\cdot)$ has bounded Lipschitz continuous $(p+1)$th derivative;

A3: The d-dimensional vector process $\mathbf{X}_t$ is strictly stationary and β-mixing with $\beta(n) \le c_0\rho^n$ for some $0 < \rho < 1$, $c_0 > 0$ where

$$\beta(n) = E\sup\left\{\left|P(A|\mathcal{F}_d^k) - P(A)\right| : A \in \mathcal{F}_{n+k}^{\infty}\right\}$$

in which $\mathcal{F}_t^{t'}$ is the σ-algebra generated by $\mathbf{X}_t, \mathbf{X}_{t+1}, ..., \mathbf{X}_{t'}$;

A4: The stationary distribution of the process $\mathbf{X}_t$ has a density $\varphi(\cdot)$ and marginal densities $\varphi_\nu(\cdot), \nu = 1, ..., d$, which has bounded Lipschitz continuous $(p+1)$-th derivatives and a positive lower bound on the support of $\pi(\cdot)$;

A5: The $\{\xi_t\}_{t \geq i_m}$ have a finite fourth moment m_4.

Denote $\mu_r = \mu_r(K) = \int u^r K(u) du$ and let $\{s_{st}(\alpha)\}_{0 \leq s,t \leq p} = S_\alpha^{-1}$ where the matrix S_α is defined as

$$S_\alpha = \sum_{\nu=1}^{d} \alpha^{2(\nu-1)} \begin{pmatrix} \mu_0 & 0 & \mu_2 & \cdots & 0 & \mu_{p-1} & 0 \\ 0 & \mu_2 & 0 & \cdots & \mu_{p-1} & 0 & \mu_{p+1} \\ \mu_2 & 0 & \mu_4 & \cdots & 0 & \mu_{p+1} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \mu_{p-1} & 0 & \cdots & \mu_{2p-4} & 0 & \mu_{2p-2} \\ \mu_{p-1} & 0 & \mu_{p+1} & \cdots & 0 & \mu_{2p-2} & 0 \\ 0 & \mu_{p+1} & 0 & \cdots & \mu_{2p-2} & 0 & \mu_{2p} \end{pmatrix}$$

$$+ \sum_{1 \leq \nu \neq \nu' \leq d} \alpha^{\nu+\nu'-2} \begin{pmatrix} \mu_0 \\ 0 \\ \mu_2 \\ \vdots \\ 0 \\ \mu_{p-1} \\ 0 \end{pmatrix} \begin{pmatrix} \mu_0 \\ 0 \\ \mu_2 \\ \vdots \\ 0 \\ \mu_{p-1} \\ 0 \end{pmatrix}^T \tag{11}$$

which is proportional to the dispersion matrix of the direct local polynomial regression, as seen in Lemma 1.

By the definition of matrix S_α and of the multivariate equivalent kernel $K^*_{\lambda,\alpha}(\mathbf{u})$, one derives immediately the following property

$$\sum_{\nu'=1}^{d} \int K^*_{\lambda,\alpha}(\mathbf{u}) \alpha^{\nu'-1} u_{\nu'}^{\lambda''} d\mathbf{u} = \begin{cases} 1 & \lambda'' = \lambda \\ 0 & 0 \leq \lambda'' \leq p, \lambda'' \neq \lambda \\ \Lambda_{0,p+1,\alpha} & \lambda'' = p+1 \end{cases} \tag{12}$$

where $\Lambda_{0,p+1,\alpha}$ is a mixture of $(p+1)$-th moments depending on the parameter α. It is also direct to verify that

$$S_\alpha^{-1} Z'_\alpha W = \frac{1}{n\varphi(\mathbf{x}})} \left\{ \begin{matrix} K^*_{0,\alpha,h}(\mathbf{X}_1 - \mathbf{x}) & \cdots & K^*_{0,\alpha,h}(\mathbf{X}_n - \mathbf{x}) \\ \vdots & \ddots & \vdots \\ K^*_{p,\alpha,h}(\mathbf{X}_1 - \mathbf{x}) & \cdots & K^*_{p,\alpha,h}(\mathbf{X}_n - \mathbf{x}) \end{matrix} \right\} \tag{13}$$

by multiplying the matrices.

Lemma 1 As $n \to \infty$

$$Z_\alpha' W Z_\alpha = \varphi(\mathbf{x}) S_\alpha \{I + o_p(1)\} \tag{14}$$

and therefore

$$\left(Z_\alpha' W Z_\alpha\right)^{-1} = \varphi(\mathbf{x})^{-1} S_\alpha^{-1} \{I + o_p(1)\}. \tag{15}$$

Proof. Equation (14) is proved by first showing, for any $1 \le \nu, \nu' \le d$ and $0 \le \lambda, \lambda' \le p$, that

$$\frac{1}{n}\sum_{i=1}^{n} K_h(\mathbf{X}_i - \mathbf{x}) \left(\frac{X_{i\nu} - x}{h}\right)^{\lambda} \left(\frac{X_{i\nu'} - x}{h}\right)^{\lambda'} =$$

$$\int K_h(\mathbf{w} - \mathbf{x}) \left(\frac{w_\nu - x}{h}\right)^{\lambda} \left(\frac{w_{\nu'} - x}{h}\right)^{\lambda'} \varphi(\mathbf{w}) d\mathbf{w} \{1 + o_p(1)\} \tag{16}$$

$$= \varphi(\mathbf{x}) \int K(\mathbf{u}) u_\nu^{\lambda} u_{\nu'}^{\lambda'} d\mathbf{u} \{1 + o_p(1)\}$$

and then using the same calculation of moments as in Yang (1999). Equation (16) is derived as a type of weak law of large numbers under geometrical β-mixing conditions, as in Härdle, Tsybakov and Yang (1998). Equation (15) follows directly from equation (14).

Proof of Theorem 1. We only prove the results on the function estimator, since the proof for the derivative estimator is similar. Making use of equations (15), (12) and (13), the proof given in Yang (1999) of a parallel result for the PAM model can be adapted if one establishes for the following 2 terms

$$I_1 = \frac{1}{n\varphi(\mathbf{x})} \sum_{i=1}^{n} K_{0,\alpha,h}^*(\mathbf{X}_i - \mathbf{x}) \sum_{\nu=1}^{d} \alpha^{\nu-1} \left\{ g(X_{id}) - \sum_{\lambda=0}^{p} \frac{g^{(\lambda)}(x)}{\lambda!} (X_{id} - x)^{\lambda} \right\}$$

$$I_2 = \frac{1}{n\varphi(\mathbf{x})} \sum_{i=1}^{n} K_{0,\alpha,h}^*(\mathbf{X}_i - \mathbf{x}) \sigma(\mathbf{X}_i) \varepsilon_i$$

that

$$I_1 = \frac{\Lambda_{0,p+1,\alpha} g^{(p+1)}(x)}{(p+1)!} h^{p+1} + o_p(h^{p+1}) \tag{17}$$

and that

$$\sqrt{nh^d} I_2 \to N\left(0, \left\|K_{0,\alpha}^*\right\|_2^2 \sigma^2(\mathbf{x}) \varphi^{-1}(\mathbf{x})\right). \tag{18}$$

Equation (17) follows by using the mixing conditions and showing that

$$E\left\{I_1/h^{p+1} - \Lambda_{0,p+1,\alpha}g^{(p+1)}(x)/(p+1)!\right\}^2 \to 0$$

as in the proof of Lemma 1. To prove (18), use the martingale central limit theorem as in Härdle, Tsybakov and Yang (1998).

Proof of Theorem 2. To adapt the parallel proof of Yang (1999), one needs to establish that

$$\int\left\{\sum_{\nu=1}^{d}\frac{1}{n\varphi(\mathrm{w}_\nu)}\sum_{i=1}^{n}K^*_{0,\alpha,h}(\mathrm{X}_i - \mathrm{w}_\nu)\sigma(\mathrm{X}_i)\varepsilon_i\right\}^2\pi(\mathrm{w})d\mathrm{w}$$

$$=O(nh^d) = o_p(1/\sqrt{n})$$

which is accomplished by applying a standard U-statistic argument, using our β-mixing assumption (A3) and Lemma 1 of Yoshihara (1976).

ACKNOWLEDGEMENTS

The author's research was partially supported by NSF grant DMS 9971186. The author also gratefully acknowledges the thoughtful comments of the two anonymous referees and of the participants at the 1999 ICSA Applied Statistics Symposium.

BIBLIOGRAPHY

Bollerslev, T. P. (1986). "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics*, 31, 307-327.

Carroll, R., Härdle, W. and Mammen, E. (1999). "Estimation in an Additive Model when the Components Are Linked Parametrically." *Econometric Theory*, tentatively accepted.

Engle, R. F. and Ng, V. (1993). "Measuring and Testing the Impact of News on Volatility." *Journal of Finance*, 48, 1749-1778.

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

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

Härdle, W., Tsybakov, A. B. and Yang, L. (1998). "Nonparametric Vector Autoregression." *Journal of Statistical Planning and Inference*, 68, 221-245.

Hengartner, N. (1996). "Rate Optimal Estimation of Additive Regression via the Integration Method in the Presence of Many Covariates." *Annals of Statistics*, submitted.

Linton, O. B. and Nielsen, J. P. (1995). "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration." *Biometrika*, 82, 93-101.

Opsomer, J. D. and Ruppert, D. (1997). "Fitting a Bivariate Additive Model by Local Polynomial Regression." *Annals of Statistics*, 25, 186-211.

Opsomer, J. D. and Ruppert, D. (1998). "A Fully Automated Bandwidth Selection Method for Fitting Additive Models." *Journal of the American Statistical Association*, 93, 605-619.

Opsomer, J. D. (1999). "Asymptotic Properties of Backfitting Estimators." *Journal of Multivariate Analysis*, forthcoming.

Sperlich, S., Linton, O. B. and Härdle, W. (1999). "A Simulation Comparison between Integration and Backfitting Methods of Estimating Separable Nonparametric Regression Models." *Test*, forthcoming.

Tjøstheim, D. and Auestad, B. H. (1994). "Nonparametric Identification of Nonlinear Time Series: Projections." *Journal of the American Statistical Association*, 89, 398-1409.

Yang, L. (1999). Direct Estimation of Proportional Additive Model, submitted.

Yang, L., Härdle, W. and Nielsen, J. P. (1999). "Nonparametric Autoregression with Multiplicative Volatility and Additive Mean." *Journal of Time Series Analysis*, 20, 579-604.

Yang, L. and Tschernig, R. (1999). "Multivariate Bandwidth Selection for Local Linear Regression." *Journal of the Royal Statistical Society, Series B*, 61, 793-815.

Yoshihara, K. (1976). "Limiting Behavior of *U*-statistics for Stationary, Absolutely Regular Processes." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 237-252.

Received June, 1999; Revised January, 2000.