Nonstationary Nonparametric Volatility Model[1]

Heejoon Han[2] and Shen Zhang[3]

Abstract

We investigate a new nonstationary nonparametric volatility model, in which the conditional variance of time series is modeled as a nonparametric function of an integrated or near-integrated covariate. The important features of the model are that it can generate the long memory property in volatility and it allows that the unconditional variance of time series is time-varying. These properties are not allowed in most existing nonparametric or semiparametric volatility models. We establish the asymptotic distribution theory for the model. We show that the kernel estimate of the model is consistent and its asymptotic distribution is mixed normal. For an empirical application of the model, we consider the daily S&P 500 index return series and we use the VIX index as the covariate. It is shown that our model performs reasonably well both in within-sample and out-of-sample forecasts.

This version: May 2011
JEL classification: C14, C22, G12.

Keywords: nonparametric ARCH, Kernel estimation, nonparametric cointegrating regression, nonstationarity, volatility persistence, long memory property.

1 Introduction

ARCH type models have been widely used to model the volatility of economic and financial time series since the seminal work by Engle (1982) and the extension made by Bollerslev

[1] We are very grateful to the Co-Editor, Oliver Linton, and two anonymous referees for many useful comments, which improved the quality of the paper. This paper is based on parts of a working paper circulated under the title "Nonstationary semiparametric ARCH model and nonparametric cointegrating volatility model". We would like to thank Chang Sik Kim, Tatsushi Oka, Joon Park, Peter Phillips, Jun Yu and Yoon-Jae Whang for helpful discussions and suggestions. We also thank seminar participants in the Center for Financial Econometrics at the Singapore Management University, Sungkyunkwan University, Korea University, Seoul National University, 2009 Korean Econometric Society Summer Meeting, 2009 International Symposium on Econometric Theory and Applications (Kyoto), 20th EC^2 conference on Real Time Econometrics (Aarhus) for helpful comments. Han gratefully acknowledges research support from NUS Risk Management Institute. Address correspondence to: Heejoon Han, 1 Arts Link, Department of Economics, National University of Singapore, Singapore 117570. Tel: (+65)6516-6258, Fax: (+65)6775-2646, E-mail: ecshhj@nus.edu.sg.

[2] Department of Economics and Risk Management Institute, National University of Singapore.

[3] Department of Economics, National University of Singapore.

(1986). Recently there has been active research on nonparametric or semiparametric volatility models. See Linton (2009) for an excellent review. The nonparametric ARCH literature begins with Pagan and Schwert (1990a) and Pagan and Hong (1991). In the *nonparametric ARCH model* they considered, the conditional variance $\left(\sigma_t^2\right)$ of a martingale difference sequence (y_t) is given as

$$\sigma_t^2 = m\left(y_{t-1}\right) \tag{1}$$

where $m(\cdot)$ is a smooth but unknown function, and the multilag version is

$$\sigma_t^2 = m\left(y_{t-1}, y_{t-2}, \cdots, y_{t-d}\right).$$

They proposed these models to allow for a general shape to the news impact curve and their models can nest all the parametric ARCH processes. However, their models cannot capture adequately the time series properties of many financial time series, in particular volatility persistence, and the statistical properties of the estimators can be poor, due to curse of dimensionality. See Masry and Tjϕstheim (1995), Härdle and Tsybakov (1997) for related literature.

To overcome these problems, additive models have been proposed as a flexible but parsimonious alternative to nonparametric models. See Engle and Ng (1993), Yang, Härdle and Nielsen (1999), Kim and Linton (2004), Linton and Mammen (2005) and Yang (2006) for the related literature. Among many nonparametric or semiparametric ARCH models, only the models proposed by Audrino and Bühlmann (2001), Linton and Mammen (2005) and Yang (2006) can nest the GARCH(1,1) model.

However, it is well known that even the GARCH(1,1) model is inadequate to capture volatility persistence observed in many financial time series. While the autocorrelation of squared series of the GARCH(1,1) process decays exponentially and converges to zero very quickly, stock return or exchange rate return series commonly exhibit the *long memory property in volatility*; the autocorrelation of squared return series decays very slowly. Ding *et al.* (1993) found earlier that it is possible to characterize the power transformation of stock return series to be long memory.

In the literature of parametric ARCH type models, there has been active research on this issue and several models have been proposed to capture this long memory property in volatility.[4] These models accommodate fractional integration, structural changes or a persistent covariate in ARCH type models. For the related literature on the long memory property in volatility, see Baillie *et al.* (1996), Ding and Granger (1996), Bollerslev and Mikkelsen (1996) (fractionality of the order of integration), Engle and Lee (1999) (two components), Diebold and Inoue (2001) (switching regime), Mikosch and Starica (2004) (structural change), Granger and Hyung (2004) (occasional break) and Park (2002) and Han and Park (2008) (persistent covariate).

On the other hand, there has been less attention on the long memory property in volatility in the literature of nonparametric or semiparametric ARCH models. Even if it has been an important issue for nonparametric or semiparametric ARCH models to capture

[4] This is also an important issue in the literature of stochastic volatility models. See Hurvich and Soulier (2009) for stochastic volatility models with long memory property. But we do not consider stochastic volatility models. We focus only on ARCH type models that are parametric, nonparametric or semiparametric.

adequately volatility persistence, there has been no attempt to explain the long memory property in volatility in the framework of nonparametric or semiparametric ARCH models. This is the first limitation of existing nonparametric or semiparametric ARCH models that we focus on.

Moreover, most nonparametric or semiparametric ARCH models assume the covariance stationarity of (y_t). Hence, these models are valid only for stationary time series, which is the second limitation of existing models that we focus on. Among nonparametric or semiparametric ARCH models, the only exception without this limitation is the spline-GARCH model proposed by Engle and Rangel (2008) that allows the unconditional variance of (y_t) to be time-varying. If we model the volatility of financial return series, it is quite restrictive to assume that the unconditional variance of financial return series is constant for a long time span, in particular, considering that fundamental features of the financial markets are continuously and significantly changing.[5]

The aim of this paper is to develop and investigate a new nonparametric volatility model that could overcome the current limitations of most nonparametric or semiparametric ARCH models. We consider the following nonparametric volatility model, defined as

$$\sigma_t^2 = m(x_{t-1}) \tag{2}$$

where $m(\cdot)$ is a smooth but unknown function and (x_t) is an integrated or near-integrated covariate. We observe $\{y_t, x_t\}$ at time t. We refer to this model as the *nonstationary nonparametric volatility model*. The model can generate the long memory property in volatility if the unknown function belongs to the function classes considered by Park (2002), and moreover the model allows that the unconditional variance of (y_t) is time-varying.

We derive the asymptotic distribution theory of the kernel estimate of our model. We show that the kernel estimate of the model is consistent and the limit distribution is mixed normal, giving straightforward asymptotics that are usable in practical work. Our volatility model is an application of the nonparametric cointegrating model by Wang and Phillips (2009a, 2009b). Their theoretic results are useful in the analysis of our model.

For an empirical application of the model, we consider the return series of the daily S&P 500 index for the 3 January 1996 to 27 February 2009 period (3260 trading days). Several tests of covariance stationarity by Loretan and Phillips (1994) indicate that the stock return series is not covariance stationary for the period, which is mainly due to the recent financial crisis. As the covariate (x_t), we use the VIX index, which can be modeled as a near-integrated process. We investigate the within-sample and out-of-sample predictive power of our model. The forecast evaluations are based on the QLIKE loss function that is not only robust to noise in the volatility proxy, but also has the highest power amongst the loss functions that are robust to noise in the proxy according to the study by Patton and Sheppard (2009). We use the realized kernel as the proxy for actual volatility because it has some robustness to the effect of market microstructure effects. Our model performs reasonably well exhibiting the lowest QLIKE loss both in within-sample and out-of-sample forecasts.

[5] Starica and Granger (2005) investigated a nonstationary unconditional variance model of stock return series. They discovered that most of the dynamics of stock return series are concentrated in shifts of the unconditional variance.

The rest of the paper is organized as follows. Section 2 introduces the model with required assumptions. Section 3 provides the asymptotic distribution theory of the kernel estimate of the model, and a simulation experiment is conducted in Section 4. Section 5 provides an empirical application of the model, which includes data description, evaluation criterion, and within-sample and out-of sample forecast evaluation results of the model. Section 6 concludes the paper, and Appendix contains mathematical proof for the technical result in the paper.

2 The Model

Before introducing our volatility model, we briefly review some closely related models. Park (2002) introduced the *nonstationary nonlinear heteroskedasticity* (NNH) model given as

$$\sigma_t^2 = f(x_{t-1}), \tag{3}$$

where $f(\cdot)$ is a parametric nonlinear function and (x_t) is a unit root process. As an extension of the NNH model, Han and Park (2008) investigated the model given as

$$\sigma_t^2 = \alpha y_{t-1}^2 + f(x_{t-1}), \tag{4}$$

where (x_t) is an integrated or near-integrated process. The parametric nonlinear function $f(\cdot)$ can be either *integrable* ($f \in \mathbb{I}$) or *asymptotically homogeneous* ($f \in \mathbb{H}$). The reader is referred to Park and Phillips (1999, 2001) for more details on these function classes. The classes $\mathbb{I}$ and $\mathbb{H}$ include a wide class, if not all, of transformations defined on $\mathbb{R}$. The bounded functions with compact supports and more generally all bounded integrable functions with fast enough decaying rates, for instance, belong to the class $\mathbb{I}$. On the other hand, power functions $a|x|^b$ with $b \geq 0$ belong to the class $\mathbb{H}$ having asymptotic order $a\lambda^b$ and $|x_t|^b$ as limit homogeneous functions. Moreover, logistic function $e^x/(1+e^x)$ and all the other distribution function-like functions are also the elements of the class $\mathbb{H}$ with asymptotic order 1 and limit homogeneous function $1\{x \geq 0\}$.

Recently, Wang and Phillips (2009a, 2009b) investigated *nonparametric cointegrating regression*

$$y_t = m(x_t) + u_t,$$

where $m(\cdot)$ is a smooth but unknown function and (x_t) is an integrated or fractionally integrated process. We apply their nonparametric estimation method of a structural cointegrating regression model in the framework of a nonparametric volatility model. Our new nonparametric volatility model is introduced in the following assumptions.

We write the time series (y_t) to be modeled as

$$y_t = \sigma_t \varepsilon_t$$

and let $(\mathcal{F}_t)$ be a filtration with $\mathcal{F}_t$ for each t denoting information available at time t.

Assumption 2.1 Assume that
(a) (ε_t) is iid (0,1) and adapted to $(\mathcal{F}_t)$,
(b)

$$\sigma_t^2 = m(x_t) \tag{5}$$

for a smooth but unknown function $m(\cdot)$ such that $m(x_t) > 0$ for all t, and
(c) (x_t) is adapted to $(\mathcal{F}_{t-1})$.

Under Assumption 2.1, we have

$$\mathbb{E}(y_t|\mathcal{F}_{t-1}) = 0 \quad \text{and} \quad \mathbb{E}(y_t^2|\mathcal{F}_{t-1}) = \sigma_t^2.$$

The time series (y_t) has conditional mean zero with respect to the filtration $(\mathcal{F}_t)$, and therefore, $(y_t, \mathcal{F}_t)$ is a martingale difference sequence. However, it is conditionally heteroskedastic with conditional variance (σ_t^2). We assume the part (c), instead of using (x_{t-1}) as in (2), for notational convenience in the proof.

Assumption 2.2 Assume that
(a)

$$x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + v_t,$$

where $c \geq 0$,
(b) (v_t) is generated by

$$v_t = \varphi(L)\eta_t = \sum_{k=0}^{\infty} \varphi_k \eta_{t-k},$$

where $\varphi_0 = 1$, $\varphi(1) \neq 0$ with $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$, and (η_t) are iid random variables with mean zero and $\mathbb{E}|\eta_t|^p < \infty$ for some $p > 2$.

Assumption 2.2 defines (x_t) as an integrated or near-integrated process driven by a general linear process. Throughout the paper, we set the long-run variance of (v_t) to be unity because it has only an unimportant scaling effect on our analysis. Note that we do not assume that (v_t) is independent of (ε_t). As explained in the next section, it is unnecessary to assume that (x_t) is independent of (ε_t) for the kernel estimate of our model.

Assumptions 2.1 and 2.2 define the nonstationary nonparametric volatility model. The parametric counterpart to this model is the NNH model by Park (2002). This volatility model is an application of the nonparametric cointegrating model by Wang and Phillips (2009a, 2009b). To the best of our knowledge, this model is the first empirical application of the nonparametric estimation method of a structural cointegrating regression model by Wang and Phillips (2009a, 2009b).

If we consider time series properties of our model, it is interesting to note that our model, depending on the unknown function $m(\cdot)$, could overcome some limitations of most nonparametric or semiparametric ARCH models that were described in the introduction. First, our model can generate the long memory property in volatility as long as the unknown function $m(\cdot)$ belongs to the function classes of $f(\cdot)$ considered by Park (2002) in (3). Park

(2002) shows that the autocorrelation of the squared process of the NNH model vanishes only very slowly, or do not even vanish at all, in the limit. This means that the NNH model can explain the long memory property in volatility. Since the function classes $\mathbb{I}$ and $\mathbb{H}$ considered by Park (2002) include a wide class of transformations defined on $\mathbb{R}$, it is possible that the unknown function $m(\cdot)$ in our model belongs to these function classes. And in this case our model also generates the long memory property in volatility. For example if $m(x) = a\,|x|^b$ for some $b > 0$ in (5), our model belongs to the NNH model with an asymptotically homogeneous function ($f \in \mathbb{H}$), which implies that the long memory property in volatility can be generated as shown in Park (2002).

Second, the nonstationarity of (y_t) is allowed in our model. The unconditional variance of (y_t) could be time-varying due to the nonstationary covariate (x_t), depending on the unknown function $m(\cdot)$.

It is important to note that these properties of our model are allowed because the covariate (x_t) is nonstationary. If (x_t) is stationary, the long memory property in volatility and the nonstationary of (y_t) will not be allowed. It is already noted by Park (2002) that the nonstationary covariate (x_t) plays a crucial role in generating volatility persistence. He showed that a nonlinear function of a stationary process, on the other hand, cannot generate the long memory property in volatility.

3 Asymptotic Distribution Theory

We establish the asymptotic distribution theory for the kernel estimate of our model. The nonstationary nonparametric volatility model in (5) can be rearranged as

$$y_t^2 = m(x_t) + u_t \tag{6}$$

where $u_t = m(x_t)\left(\varepsilon_t^2 - 1\right)$. The error term (u_t) in this model is a martingale difference sequence because (x_t) is adapted to $(\mathcal{F}_{t-1})$. Since the conditional variance of the error term is

$$\mathbb{E}\left(m^2(x_t)\left(\varepsilon_t^2 - 1\right)^2 |\mathcal{F}_{t-1}\right) = m^2(x_t)\left(\mathbb{E}\left(\varepsilon_t^4\right) - 1\right),$$

the above model is an extended case of the nonparametric cointegrating regression considered by Wang and Phillips (2009a, 2009b). The conventional kernel estimate of $m(x)$ in the above model[6] is given by

$$\hat{m}(x) = \frac{\sum_{t=1}^n y_t^2 K_h\left(x_t - x\right)}{\sum_{t=1}^n K_h\left(x_t - x\right)}$$

where $K_h(s) = h^{-1}K(s/h)$. This section investigates the limit behavior of $\hat{m}(x)$.

Assumption 3.1 The kernel K satisfies that $\int_{-\infty}^{\infty} K(s)\,ds = 1$ and $\sup_s K(s) < \infty$.

[6] As an alternative method, one can consider the local maximum likelihood estimation as in Avramidis (2002). See also Fan and Yao (1998).

Assumption 3.2 (a) For given x, there exists a real function $m_1(s,x)$ and is $0 < \gamma \le 1$ such that, when h sufficiently small, $|m(hy+x) - m(x)| \le h^\gamma m_1(y,x)$ for all $y \in \mathbb{R}$ and $\int_{-\infty}^{\infty} K(s)\, m_1(s,x)\, ds < \infty$.
(b) $\int_{-\infty}^{\infty} K^2(s)\, m_1(s,x)\, ds < \infty$ and $\int_{-\infty}^{\infty} K^2(s)\, m_1^2(s,x)\, ds < \infty$.

Assumption 3.3 $\sup_{1\le t\le n} E\left(|\varepsilon_t|^q \,|\mathcal{F}_{t-1}\right) < \infty$ a.s. for some $q > 4$.

Assumptions 3.1 and 3.2(a) are the same as Assumptions 3.1 and 3.2 in Wang and Phillips (2009a). As mentioned in Wang and Phillips (2009a), the conditions in Assumption 3.1 and 3.2(a) are quite weak and simply verified for various kernels $K(x)$ and functions $m(x)$. Assumption 3.2(b) is additional, but its marginal restriction is not substantial. For instance, if $K(x)$ is a standard normal kernel or has a compact support as in Karlsen *et al.* (2007), commonly occurring functions such as $m(x) = |x|^\beta$ and $m(x) = 1/\left(1 + |x|^\beta\right)$ for some $\beta > 0$ satisfy Assumption 3.2 (a) and (b) with $\gamma = \min\{\beta, 1\}$. Refer to Wang and Phillips (2009a) for detailed remarks on these assumptions. Regarding the value of γ in Assumption 3.2(a), $\gamma = 1$ is the most common case according to Wang and Phillips (2009a, 2009b). Assumption 3.3 is corresponding to Assumption 3.3 in Wang and Phillips (2009a). For $\sup_{1\le t\le n} E(|u_t|^{q_1} \,|\mathcal{F}_{t-1}) < \infty$ a.s. for some $q_1 > 2$ (Assumption 3.3 in Wang and Phillips (2009a)), we need $\sup_{1\le t\le n} E\left(|\varepsilon_t|^{2q_1} \,|\mathcal{F}_{t-1}\right) < \infty$ a.s. because $u_t = m(x_t)\left(\varepsilon_t^2 - 1\right)$ in our case.

Under Assumptions 2.1 and 2.2 in the previous section, Assumptions 3.4 and 3.5 in Wang and Phillips (2009a) are simply verified for our model (6) if we let $d_n = \sqrt{n}$. We can easily see that a continuous Gaussian process $G(t)$ in Wang and Phillips (2009a) is an Ornstein-Uhlenbeck process in our case. Under the conditions imposed on (v_t) in Assumption 2.2(b), the time series (x_t) included in the model becomes an integrated or near-integrated process satisfying the usual invariance principle. For $r \in [0,1]$,

$$n^{-1/2} x_{[nr]} \to_d V_c = \int_0^r \exp\left(-c(r-s)\right) dV_0(s)$$

where $[z]$ denotes the integer part of z and V_0 is the standard Brownian motion. Note that V_c is an Ornstein-Uhlenbeck process and its local time is defined as

$$L_c(t,s) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_0^t 1\left\{|V_c(r) - s| < \varepsilon\right\} dr.$$

Roughly speaking, 2ε times $L_c(t,s)$ measures the actual time spent by V_c in the ε-neighborhood of s up to time t. Hence, a continuous Gaussian process G and its local time L_G in Wang and Phillips (2009a) are the Ornstein-Uhlenbeck process V_c and its local time L_c in our case.

The limit theory for the kernel estimate of the nonstationary nonparametric volatility model is as follows.

Theorem 1 *Suppose Assumptions 2.1-2.2 and 3.1-3.3 hold. Then, for any h satisfying $nh^2 \to \infty$ and $h \to 0$,*

$$\hat{m}(x) \to_p m(x). \tag{7}$$

Furthermore, for any h satisfying $nh^2 \to \infty$ and $nh^{2(1+2\gamma)} \to 0$,

$$\left(h \sum_{t=1}^{n} K_h(x_t - x) \right)^{1/2} (\hat{m}(x) - m(x)) \to_d N\left(0, \sigma_1^2\right), \tag{8}$$

where $\sigma_1^2 = \left(E(\varepsilon_t^4) - 1\right) m^2(x) \int_{-\infty}^{\infty} K^2(s)\, ds$.

The result (7) implies that $\hat{m}(x)$ is a consistent estimate of $m(x)$. As shown in the proof of Theorem 1 in Appendix, we may obtain

$$\hat{m}(x) - m(x) = o_p \left\{ a_n \left[h^\gamma + \sqrt{1/(\sqrt{n}h)} \right] \right\}, \tag{9}$$

where γ is defined in Assumption 3.2 and a_n diverges to infinity as slowly as required. This leads to the following argument on bandwidth. In the most common case where $\gamma = 1$, a possible optimal bandwidth is suggested to be $h^* \sim an^{-1/6}$, so that $h = o(n^{-1/6})$ ensures undersmoothing. See Wang and Phillips (2009a, 2009b) for detailed remarks.

The result (8) shows that the asymptotic distribution of $\hat{m}(x)$ is mixed normal. The mixing variate in the limit distribution depends on the local time $L_c(1,0)$. Explicitly, in the most common case where $\gamma = 1$,

$$\left(nh^2\right)^{1/4} (\hat{m}(x) - m(x)) \to_d L_c^{-1/2}(1,0)\, N\left(0, \sigma_1^2\right)$$

by (15) in Appendix. The convergence rate is $\left(nh^2\right)^{1/4}$, which requires that $nh^2 \to \infty$. Wang and Phillips (2009a, 2009b) provide detailed explanations on this issue. The limiting variance of the (randomly normalized) kernel estimator in (8) contains the square of the volatility function $m^2(x)$ as it is expected in the remarks (e) in Wang and Phillips (2009b). This is similar to the result in Yang (2006). In his semiparametric GARCH model, the limiting variance of the estimator also contains the square of the volatility function.

Note that it is unnecessary to assume that (x_t) is independent of (ε_t) for the asymptotic theory. Our asymptotic theory holds regardless that (x_t) and (ε_t) are dependent or independent. A detailed explanation is given in the proof of Theorem 1 (below (19)) in the Appendix.

4 Simulation

This section reports the result of a simulation experiment investigating the finite sample performance of the kernel estimator of the model. The generating mechanism is

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= m(x_{t-1}) \end{aligned}$$

with

$$x_t = \left(1 - \frac{1}{n}\right) x_{t-1} + v_t,$$

and we consider the following function;

$$m(x) = 0.01 + 0.1x^2.$$

Note that the specified $m(x)$ is an asymptotically homogeneous function ($f \in \mathbb{H}$) and the model belongs to the NNH model in Park (2002). Park (2002) showed that this model can generate the long memory property in volatility. Our estimate explained in the previous section provides a nonparametric estimate of the NNH model.

We let the innovation process (ε_t) be iid $N(0,1)$. The initial values are set $x_1 = 0$ and $\sigma_1^2 = 0.01$. We consider both cases where (x_t) and (ε_t) are dependent and independent. For the former case, we let $v_{t+1} = 0.1\varepsilon_t$. For the latter case, we let (v_t) be independent of (ε_t) and iid $N(0, 0.01)$. According to the theorem in the previous section, we expect that there is not much difference between these two cases

The dependent and independent cases are described in Figure 1 and Figure 2, respectively. Each figure graphs the function $m(x)$ (solid line), the mean simulated kernel estimates (broken line) and 95% estimation bands (dotted line) over the intervals $[-1, 1]$. The bands contain 95% of the 10,000 simulated values of $\hat{m}(x)$ for a given x. Bias, variance and mean squared error for the estimates are computed on the grid of values $\{x = -1 + 0.02k;\ k = 0, 1, \cdots, 100\}$ based on 10,000 replications, which are reported in Table 1. We use the Gaussian kernel and the Silverman's bandwidth $\hat{\sigma}_z n^{-1/5}$ where $\hat{\sigma}_z$ is the sample standard deviation of (x_t). We use the cross validation bandwidth for the empirical application in the next section, but it is shown that, for our data, the result using the Silverman's bandwidth is very similar to the one using the cross validation bandwidth. We also tried $\hat{\sigma}_z n^{-1/6}$ that is a possible optimal bandwidth suggested in the previous section, and the simulation results still hold.

Figure 1 shows the results for the Monte Carlo approximations to $E(\hat{m}(x))$ with 95% estimation bands for sample sizes $n = 500$ and $n = 2000$ in the dependent case. The plots show that the estimated $\hat{m}(x)$ converges to the true function $m(x)$ as the sample size increases. Figure 2 also shows very similar results for the independent case. Moreover, Table 1 shows that, in both dependent and independent cases, variance as well as bias become smaller when the sample size is larger. Consequently, mean squared errors are much smaller when the sample size is larger. These simulation results confirm what our asymptotic theory implies. The estimated $\hat{m}(x)$ converges to the true function $m(x)$ as the sample size increases, and there exists no substantial difference between the dependent and independent cases.

5 Empirical Application

5.1 The Data, Models and Estimation Methods

We consider the daily S&P 500 index returns from 3 January 1996 to 27 February 2009 (3260 trading days). We demean the return series by subtracting its sample mean which is close to zero. We use the demeaned return series as (y_t). We conducted formal tests by Pagan and Schwert (1990b) and Loretan and Phillips (1994) for the covariance stationarity of the series (y_t). In general the null hypothesis of covariance stationarity is rejected for the series[7]. The unconditional variance of the series seems to be time-varying in particular due to the recent financial crisis. Since most nonparametric or semiparametric ARCH models assume the covariance stationarity of (y_t), these models are not suitable for the stock return series we consider. However, our nonstationary nonparametric volatility model allows the unconditional variance of (y_t) to be time-varying and, therefore, it could be better to use our model for the stock return series.

As the persistent covariate (x_t) for our nonstationary nonparametric volatility model, we use the VIX index by the Chicago Board Options Exchange. The VIX index is the implied volatility calculated from options on the S&P 500 index.[8] It is not a new idea to use implied volatilities from options to forecast volatility. See Latane and Rendleman (1976), Chiras and Manaster (1978), Christensen and Prabhala (1998), Fleming (1998), Blair *et al.* (2001) and Giot (2003). It is shown that the models based on implied volatilities provide better volatility forecasts of returns on stock indices, which motivates us to use the VIX index as our covariate (x_t).

Table 2 shows the results of unit root tests for the VIX index, which indicate that the VIX index is persistent enough for the use in our model. We consider two alternative autoregressive specifications for the series: with and without a linear deterministic trend. In both cases, the estimated autoregressive coefficient is very close to unity (0.984). While the ADF (Augmented Dickey-Fuller) test rejects the null hypothesis of a unit root in the case without a linear deterministic trend, it cannot reject the null hypothesis in the case with a linear deterministic trend. And the KPSS test rejects the null hypothesis of stationarity in both cases, which suggests that there exists strong evidence in favor of the nonstationary alternative. Considering the results of the KPSS test and the fact that the estimated autoregressive coefficients are close to unity, we conclude that there exists at least a near unit root for the VIX index.

For the empirical application of our model, we estimate the following models and compare their within-sample and out-of-sample predictive ability;

$$\begin{aligned} &\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 && \text{GARCH(1,1) model} \\ &\sigma_t^2 = m(y_{t-1}) && \text{nonparametric ARCH model} \\ &\sigma_t^2 = m(x_{t-1}) && \text{nonstationary nonparametric volatility model} \end{aligned}$$

where (y_t) and (x_t) are the demeaned stock return series and the VIX index, respectively. The first two benchmark models are the GARCH(1,1) model and a nonparametric ARCH model by Pagan and Schwert (1990a). We also considered another nonparametric ARCH model $\left(\sigma_t^2 = m(y_{t-1}, y_{t-2})\right)$ by Pagan and Schwert (1990a). However, we decide not to report the result for this model because it performs very poorly in both within-sample and out-of-sample forecasts.

[7]The test results are not given to save the space. They will be available from the authors upon request.

[8]See www.cboe.com/VIX for more details of the VIX index. The VIX index is also available at the website.

For the GARCH(1,1) model, we use the quasi-maximum likelihood estimation method, which is the standard estimation method for parametric ARCH type models.[9] For the nonstationary nonparametric volatility model, we use the Nadaraya-Watson kernel estimation method. We also use the same method for the nonparametric ARCH model. These nonparametric models can be arranged as

$$y_t^2 = m(z_t) + u_t$$

where $u_t = \sigma_t^2 \left(\varepsilon_t^2 - 1\right)$ and (z_t) is either (y_{t-1}) or (x_{t-1}). Note that (u_t) is a martingale difference sequence. The kernel estimate of $m(z) = \mathbb{E}\left(y_t^2 | z_t = z\right)$ is

$$\hat{m}(z) = \sum_{j=1, j\neq t}^{n} \left(K_h\left(z - z_j\right) \Bigg/ \sum_{k=1}^{n} K_h\left(z_k - z\right) \right) y_j^2, \tag{10}$$

We adopt the 'leave-one-out' estimator as in Pagan and Schwert (1990a). We let $K(s)$ be a nonnegative real function and set $K_h(s) = h^{-1}K(s/h)$. The Gaussian kernel is used throughout the paper. We also tried other kernels but estimation results are affected only negligibly, which is common in the literature of nonparametric econometrics.

For the nonparametric models, we use the cross-validation bandwidth selection method that is designed to minimize the QLIKE loss function. For the nonparametric model $\sigma_t^2 = m(z_t)$ where (z_t) is either (y_{t-1}) or (x_{t-1}), we choose the bandwidth to minimize the following QLIKE loss function;

$$h_{CV} = \arg\min_h \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\sigma_t^2}{\hat{m}(z_t)} - \log \frac{\sigma_t^2}{\hat{m}(z_t)} - 1 \right\}$$

where $\hat{m}(z_t)$ is the 'leave-one-out' estimator given in (10). The realized kernel is used as the proxy for actual volatility σ_t^2. The descriptions of the QLIKE and the realized kernel are given in the next subsection.

5.2 Evaluation Criterion

To evaluate the performance of nonparametric ARCH models, Pagan and Schwert (1990a) compared the within-sample and out-of-sample predictive power of volatility models using R^2 of the Mincer-Zarnowitz regression

$$\sigma_t^2 = \alpha + \beta\, \hat{\sigma}_t^2 + u_t, \tag{11}$$

where $\left(\sigma_t^2\right)$ is the proxy for actual volatility and $\left(\hat{\sigma}_t^2\right)$ is the within-sample or out-of-sample forecast. Since actual volatility is unobservable, we need to use a proxy for actual volatility. Pagan and Schwert (1990a) used squared return series $\left(y_t^2\right)$ as the volatility proxy.

Following Pagan and Schwert (1990a), we will also evaluate the performance of our model by comparing predictive power of volatility models. However, since there have been

[9] For the consistency and asymptotic distribution of the quasi-maximum likelihood estimator (QMLE) of the GARCH(1,1), see Jensen and Rahbek (2004) and reference therein.

recent developments in the literature of volatility forecast evaluation, we will consider these developments.

First, we use a realized measure of volatility based on high frequency data as the proxy for actual volatility instead of squared return series. It is well known that squared return series is very noisy and realized measures are better estimates of actual volatility. See Barndorff-Nielsen and Shephard (2002) and Andersen *et al.* (2003). Moreover, Hansen and Lunde (2006) showed in an empirical application that using realized volatility leads to a more informative comparison with a tighter confidence intervals than using squared return. These support the use of realized measure as the proxy for actual volatility instead of squared return, and even imply that the evaluation based on realized measure could be more reliable.

More specifically, as the proxy for actual volatility, we will use the realized kernel, introduced by Barndorff-Nielsen *et al.* (2008), because it has some robustness to the effect of market microstructure effects. The realized kernel for the daily S&P 500 index return series is available at the database 'Oxford-Man Institute's realised library' produced by Heber *et al.* (2009)[10], to which we refer the reader for detailed explanations.

Second, we use the QLIKE loss function described below, instead of R^2 of the Mincer-Zarnowitz regression. Even if realized measures are known to be better measures, they are imperfect and noisy proxies for actual volatility. Therefore, it is possible, due to noisy proxies, that the evaluation based on some loss functions may identify an inferior volatility model as the 'best' and the inferior model may spuriously be found to be 'significantly' better than all other models. Hence, there has been research on loss functions that are robust to the use of a noisy volatility proxy. See Hansen and Lunde (2006), Patton (2010) and Patton and Sheppard (2009).

Patton (2010) provides necessary and sufficient conditions on the functional form of the loss function to ensure the ranking of various forecasts is preserved when using a noisy volatility proxy, and he shows that the MSE and QLIKE are robust. In particular, Patton and Sheppard (2009) shows in their simulation study that the QLIKE loss function has the highest power. The QLIKE loss function is defined as

$$L(\hat{\sigma}_t^2, \sigma_t^2) = \frac{\sigma_t^2}{\hat{\sigma}_t^2} - \log \frac{\sigma_t^2}{\hat{\sigma}_t^2} - 1. \tag{12}$$

As the simulation results by Patton and Sheppard (2009) points to this QLIKE as the preferred choice amongst the loss functions that are robust to noise in the proxy, we use the QLIKE as the loss function.

The significance of any difference in the QLIKE loss is tested via a Diebold-Marinao and West (henceforth DMW) test (see Diebold-Marinao(1995) and West (1996). A DMW statistic is computed using the difference in the losses of two models

$$\begin{aligned} d_t &= L(\hat{\sigma}_{t,1}^2, \sigma_t^2) - L(\hat{\sigma}_{t,2}^2, \sigma_t^2) \\ DMW_T &= \frac{\sqrt{T}\bar{d}_T}{\sqrt{\widehat{avar}\left(\sqrt{T}\bar{d}_T\right)}} \end{aligned} \tag{13}$$

[10]See http://realized.oxford-man.ox.ac.uk/.

where $\bar{d}_T$ is the sample mean of d_t and T is the number of forecasts. The asymptotic variance of the average is computed using a Newey-West variance estimator with the number of lags set to $\left[T^{1/3}\right]$.

5.3 Estimation and Forecast Evaluation Results

We estimate three volatility models and evaluate their within-sample and out-of-sample predictive abilities. The estimation result of the GARCH(1,1) model is following (standard errors are in parentheses):

$$\begin{array}{c} \hat{\sigma}_t^2 = 0.0000 + 0.0782 y_{t-1}^2 + 0.9156 \sigma_{t-1}^2 \\ \quad\quad (0.0000) \quad (0.0064) \quad\quad\;\; (0.0069) \end{array}$$

The ARCH effects are close to unity ($\hat{\alpha}+\hat{\beta} = 0.994$), which is a typical estimation result for the GARCH(1,1) model. This is why Engle and Bollerslev (1986) introduced the IGARCH model, where $\alpha + \beta = 1$.

The estimated volatility function for our nonparametric model is plotted in Figure 3. Figure 3 displays the mapping of $\hat{\sigma}_t^2 = \hat{m}(x_{-1})$ into the grid of values $\{x_{t-1} = 10 + k;\ k = 0, 1, \cdots, 70\}$. The VIX index, used as (x_t), is ranged from 9.98 to 80.86 in our sample. For smaller values of the VIX index, the shape of $\hat{m}(x_{t-1})$ is somewhat linear. However, for larger values of the VIX index, the shape of $\hat{m}(x_{-1})$ is clearly nonlinear. Moreover, it is not a monotonic increasing function. The volatility reaches its first peak when the VIX index is between 50 and 60 and takes a dip when the VIX index is between 60 and 70. After that, the volatility increases again and becomes much higher than the first peak when the VIX index is more than 70.

Within-sample forecast comparison

Table 3 contains the within-sample forecast evaluation result based on the QLIKE loss function. See Figure 4 for plots of the fitted values of the volatility models for the entire sample period. In this case, $\left(\hat{\sigma}_t^2\right)$ in (12) denotes the fitted values of the three volatility models for the entire sample period. Our nonparametric model shows the lowest QLIKE of 0.2981 while the QLIKE of the GARCH(1,1) model is 0.3316. The nonparametric ARCH model performs very poorly with the highest QLIKE of 0.6185. We test the null hypothesis of equal loss by the DMW test procedure, and the test results show that the null hypotheses of equal loss between our model and the rest models are all rejected at 1% significance level. This means that, in terms of the within-sample fitting, our nonparametric model provides a better explanation of the stock return volatility than the rest models.[11]

Out-of-sample forecast comparison

[11] We also estimated our model using the Silverman's bandwidth $\hat{\sigma}_z n^{-1/5}$ where $\hat{\sigma}_z$ is the sample standard deviation of (x_t). For our data, this Silverman's bandwidth is almost two times of the cross validation bandwidth. Using the Silverman's bandwidth, the QLIKE of our model is 0.3023, which is similar to the result obtained by the cross validation bandwidth.

To check the possibility of over-fitting, we follow Pagan and Schwert (1990a) and evaluate the out-of-sample forecasts. If over-fitting is a serious problem, the QLIKE statistics for the out-of-sample forecasts should be much higher than the QLIKE's for the within-sample forecasts. We adopt the rolling window forecast procedure with moving windows of four years (1008 trading days). This means that we obtain one-step-ahead forecasts of the models for the period from 10 February 2000 to 27 February 2009 (2251 trading days). For our model, we use the cross validation bandwidth chosen in the within-sample case. $\left(\hat{\sigma}_t^2\right)$ in (12) now denotes one-period ahead volatility forecasts at time $t-1$. Table 4 reports QLIKE's of the models and the DMW test statistics. Figure 5 provides plots of the out-of-sample forecasts of the models.

Similarly as the previous within-sample case, our nonstationary nonparametric volatility model shows the lowest QLIKE of 0.2694 while the QLIKE of the GARCH(1,1) model is 0.2819. The nonparametric ARCH model still has a poor performance with the QLIKE of 0.6316. According to the DMW test, the null hypothesis of equal loss between our model and the nonparametric ARCH model is rejected at 1% significance level, but we cannot reject the null hypothesis of equal loss between our model and the GARCH(1,1) model.

Table 4 shows that the out–of-sample QLIKE's are similar to the within-sample counterparts. Hence it seems that over-fitting is not a serious problem for our model. Moreover, both within-sample QLIKE's and out-of-sample QLIKE's exhibit the same rank order. It is shown that our model performs reasonably well both in within-sample fitting and out-of-sample forecast.

6 Conclusion

In the paper we propose and investigate a new nonstationary nonparametric volatility model that could overcome the current limitations of most nonparametric or semiparametric ARCH models. The model can generate the long memory property in volatility and allow that the unconditional variance of (y_t) is time-varying. We establish the asymptotic distribution theory of the kernel estimate of our model, which shows that the kernel estimate is consistent and the limit distribution is mixed normal. We also provide a simulation study to demonstrate the relevance of our asymptotic theory.

For the daily return series of the S&P 500 index of the 3 January 1996 to 27 February 2009 period (3260 trading days), we evaluate the within-sample fitting and the out-of-sample forecast of the model. We use the VIX index as the covariate. Considering unit root test results and the fact that the estimated autoregressive coefficients are very close to unity, we can conclude that there exists at least a near unit root for the VIX index. It is shown that our model performs well both in within-sample fitting and out-of-sample forecast.

The NNH model by Park (2002) is an application of the parametric nonlinear nonstationary time series method by Park and Phillips (1999, 2001). Our model is the nonparametric counterpart of the NNH model, and we apply the nonparametric estimation method of a structural cointegrating regression model by Wang and Phillips (2009a, 2009b). The nonstationary covariate (x_t) plays a crucial role in generating important features of our model. The long memory property in volatility and the nonstationarity of time series (y_t) can be allowed in our model due to the nonstationary covariate (x_t), as noted by Park (2002).

Appendix

Proof of Theorem 1 The proof follows the same structure as the proof of Theorem 3.1 in Wang and Phillips (2009a, henceforth WP). We let $d_n = \sqrt{n}$ and $x_{t,n} = x_t/\sqrt{n}$. As explained in Section 3, Assumptions 3.4 and 3.5 in WP are simply verified for our case because $d_n = \sqrt{n}$ and (x_t) is integrated or near-integrated. Note that the term A denote constants that may be different at each appearance. We first prove (9). The consistency result (7) will then follow by choosing $a_n = \min\left\{h^\gamma, (nh/d_n)^{1/2}\right\}$. To prove (9), we split $\hat{m}(x) - m(x)$ as

$$\hat{m}(x) - m(x) = \frac{\sum_{t=1}^n m(x_t)\left(\varepsilon_t^2 - 1\right) K_h\left(x_t - x\right)}{\sum_{t=1}^n K_h\left(x_t - x\right)} + \frac{\sum_{t=1}^n \left[m(x_t) - m(x)\right] K_h\left(x_t - x\right)}{\sum_{t=1}^n K_h\left(x_t - x\right)}. \quad (14)$$

As in WP, we have

$$\frac{d_n}{nh}\sum_{t=1}^n K^\lambda\left(\frac{d_n}{h}x_{t,n} - \frac{x}{h}\right) \to_p L_c(1,0)\int_{-\infty}^{\infty} K^\lambda(s)\,ds, \quad (15)$$

for any $\lambda \geq 1$ and h satisfying $h \to 0$ and $nh/d_n \to \infty$. Recall that a continuous Gaussian process G and its local time L_G in WP are the Ornstein-Uhlenbeck process V_c and its local time L_c in our case. Since the result (15) implies that, for any a_n diverging to infinity as slowly as required, $1/\sum_{t=1}^n K_h\left(x_t - x\right) = o_p\left\{a_n d_n/n\right\}$, the result (9) will follow if

$$\Theta_{1n} := \sum_{t=1}^n m(x_t)\left(\varepsilon_t^2 - 1\right) K_h\left(x_t - x\right) = O_p\left\{\sqrt{n/(d_n h)}\right\} \quad (16)$$

and

$$\Theta_{2n} := \sum_{t=1}^n \left[m(x_t) - m(x)\right] K_h\left(x_t - x\right) = O_p\left(nh^\gamma/d_n\right). \quad (17)$$

We need to prove only (16) because (17) is already proven in WP.

For (16), we consider

$$\begin{aligned} E\Theta_{1n}^2 &= \left(E(\varepsilon_t^4) - 1\right) h^{-2}\sum_{t=1}^n E\left[m^2(x_t)K^2\left(\frac{d_n}{h}x_{t,n} - \frac{x}{h}\right)\right] \\ &= \left(E(\varepsilon_t^4) - 1\right) h^{-2}\sum_{t=1}^n E\left[m^2(x)K^2\left(\frac{d_n}{h}x_{t,n} - \frac{x}{h}\right)\right] \\ &\quad + \left(E(\varepsilon_t^4) - 1\right) h^{-2}\sum_{t=1}^n E\left[\left(m^2(x_t) - m^2(x)\right)K^2\left(\frac{d_n}{h}x_{t,n} - \frac{x}{h}\right)\right]. \end{aligned}$$

Since

$$h^{-2}\sum_{t=1}^{n} E\left[m^2(x)K^2\left(\frac{d_n}{h}x_{t,n}-\frac{x}{h}\right)\right] \le An/(d_n h)$$

as shown in WP, the result (16) will follow if we prove

$$\Theta_{1n,a}^2 := \sum_{t=1}^{n} E\left[\left(m^2(x_t)-m^2(x)\right)K^2\left(\frac{d_n}{h}x_{t,n}-\frac{x}{h}\right)\right] = O_p\left(nh^{1+\gamma}/d_n\right). \qquad (18)$$

Since $x_{t,n}$ satisfies Assumption 2.3 in WP, $x_{t,n}/d_{t,0,n}$ has a density $h_{t,0,n}$ following the notation of Assumption 2.3 in WP. Following the way to prove (17) in WP, we have

$$\begin{aligned}
E\left|\Theta_{1n,a}^2\right| &\le \sum_{t=1}^{n} E\left[\left|m^2(d_n x_{t,n})-m^2(x)\right|K^2\left(\frac{d_n}{h}x_{t,n}-\frac{x}{h}\right)\right] \\
&= \sum_{t=1}^{n}\int_{-\infty}^{\infty}\left\{\left|m^2(d_n d_{t,0,n}y)-m^2(x)\right|K^2\left(\frac{d_n d_{t,0,n}}{h}y-\frac{x}{h}\right)\right\}h_{t,0,n}(y)dy \\
&\le \frac{h}{d_n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}\left\{\left|m^2(hy+x)-m^2(x)\right|K^2(y)\right\}dy \\
&\le \frac{h}{d_n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}\left\{\left|m(hy+x)-m(x)\right|h^\gamma m_1(y,x)K^2(y)\right\}dy \\
&\quad +2\frac{h}{d_n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}\left\{\left|m(hy+x)-m(x)\right|m(x)K^2(y)\right\}dy \\
&\le \frac{nh^{1+2\gamma}}{d_n}\frac{1}{n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}m_1^2(s,x)K^2(s)\,ds \\
&\quad +2m(x)\frac{nh^{1+\gamma}}{d_n}\frac{1}{n}\sum_{t=1}^{n}(d_{t,0,n})^{-1}\int_{-\infty}^{\infty}m_1(s,x)K^2(s)\,ds \\
&\le A\frac{nh^{1+\gamma}}{d_n}
\end{aligned}$$

by Assumption 3.2. Since $m(\cdot)$ is positive, $m(hy+x) \le m(x)+h^\gamma m_1(y,x)$. The fourth line follows from this. This completes the proof of (9).

We next prove (8). From (14), we have

$$\left(h\sum_{t=1}^{n}K_h(x_t-x)\right)^{1/2}(\hat{m}(x)-m(x)) = \sum_{t=1}^{n}\left(\varepsilon_t^2-1\right)Z_{nt}+\left(\frac{d_n h}{n}\right)^{1/2}\Theta_{2n}/\Theta_{3n},$$

where

$$Z_{nt} = \left(\frac{d_n}{nh}\right)^{1/2} m(x_t)K\left(\frac{d_n}{h}x_{t,n}-\frac{x}{h}\right)/\Theta_{3n}$$

with

$$\Theta_{3n}^2 = \frac{d_n}{nh} \sum_{t=1}^{n} K\left(\frac{d_n}{h} x_{t,n} - \frac{x}{h}\right).$$

Note that Z_{nt} includes $m(x_t)$, which is different from the case in WP. Hence, the limit of $\Lambda_n^2 = \sum_{t=1}^n Z_{nt}^2$ is also different and given by the following;

$$\begin{aligned}
\Lambda_n^2 &= \Theta_{3n}^{-2} \frac{d_n}{nh} \sum_{t=1}^{n} m^2(x_t) K^2\left(\frac{d_n}{h} x_{t,n} - \frac{x}{h}\right) \\
&= \Theta_{3n}^{-2} \frac{d_n}{nh} \sum_{t=1}^{n} m^2(x) K^2\left(\frac{d_n}{h} x_{t,n} - \frac{x}{h}\right) \\
&\quad + \Theta_{3n}^{-2} \frac{d_n}{nh} \sum_{t=1}^{n} \left(m^2(x_t) - m^2(x)\right) K^2\left(\frac{d_n}{h} x_{t,n} - \frac{x}{h}\right) \\
&\to_p m^2(x) \int_{-\infty}^{\infty} K^2(s)\, ds
\end{aligned}$$

by (14) and (18).

As in WP, $(d_n h/n)^{1/2}\, \Theta_{2n}/\Theta_{3n} = o_p(1)$ because $nh^{1+2\gamma}/d_n \to 0$. Note that $\left(\varepsilon_t^2 - 1\right) Z_{nt}$ is a martingale difference sequence because (x_t) is adapted to $(\mathcal{F}_{t-1})$. Hence, we can deduce

$$V_n \equiv \frac{1}{\Lambda_n} \sum_{t=1}^{n} \left(\varepsilon_t^2 - 1\right) Z_{nt} \to_d N(0, \left(E(\varepsilon_t^4) - 1\right)), \tag{19}$$

for any h satisfying $nh/d_n \to \infty$ and $nh^{1+2\gamma}/d_n \to 0$, as in WP. (19) corresponds to the equation (5,21) in WP. They assume that (x_t) is independent of the error term (u_t) so that $u_t Z_{nt}$ becomes a martingale difference sequence. However, note that we do not require the independence between (x_t) and (ε_t). Even if (x_t) and (ε_t) are dependent, $\left(\varepsilon_t^2 - 1\right) Z_{nt}$ is a martingale difference sequence because (x_t) is adapted to $(\mathcal{F}_{t-1})$ and, therefore, (19) holds. This completes the proof of (8). □.

Table 1. Simulation results

		Bias	Std	MSE
Dependent case	$n = 500$	-0.0038	0.0170	5.20×10^{-4}
	$n = 2000$	0.0033	0.0107	1.70×10^{-4}
Independent case	$n = 500$	-0.0048	0.0169	5.33×10^{-4}
	$n = 2000$	0.0029	0.0109	1.75×10^{-4}

Notes: Bias, standard deviation and mean squared error are computed on the grid of values $\{x = -1 + 0.02k;\ k = 0, 1, \cdots, 100\}$ based on $10,000$ replications.

Table 2. Unit root test results for the VIX index

	With intercept	With intercept and trend
AR Coefficient	0.984	0.984
ADF Test	-3.369^{*}	-3.365
KPSS Test	0.468^{*}	0.462^{**}

Notes: The critical values for the ADF (Augmented Dickey-Fuller) test are -2.862 (5%) and -3.432 (1%) with intercept, and -3.411 (5%) and -3.961 (1%) with intercept and linear time trend. The critical values for the KPSS test are 0.463 (5%) and 0.739 (1%) with intercept, and 0.146 (5%) and 0.216 (1%) with intercept and linear time trend. In the table * and ** signify that H_0 is rejected by 5% and 1% tests, respectively.

Table 3. Comparison of within-sample predictive power for the stock return volatility (1996.01.04-2009.02.27)

models		QLIKE	DMW
GARCH(1,1)	$\omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$	0.3316	3.2341^{**}
nonpara-ARCH	$m(y_{t-1})$	0.6185	9.4616^{**}
our model	$m(x_{t-1})$	**0.2981**	

Notes: The QLIKE loss is defined in (12) and the DMW test statistic is defined in (13). * and ** signify rejecting the null hypothesis of equal loss for 5% and 1% tests, respectively.

Table 4. Comparison of out-of-sample predictive power for the stock return volatility (2000.02.10-2009.02.27)

models		QLIKE	DMW
GARCH(1,1)	$\omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$	0.2819	0.4893
nonpara-ARCH	$m(y_{t-1})$	0.6316	6.6612^{**}
our model	$m(x_{t-1})$	**0.2694**	

Notes: The QLIKE loss is defined in (12) and the DMW test statistic is defined in (13). * and ** signify rejecting the null hypothesis of equal loss for 5% and 1% tests, respectively.

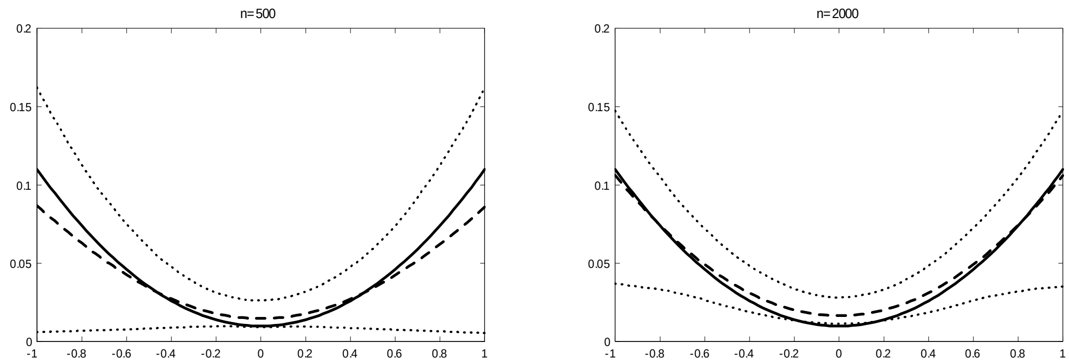

Figure 1. Graphs over the interval [-1,1] of $m(x)$ (solid line), the Monte Carlo estimates of $E(\hat{m}(x))$ (broken line) and 95% estimation bands (dotted line) for the dependent case.

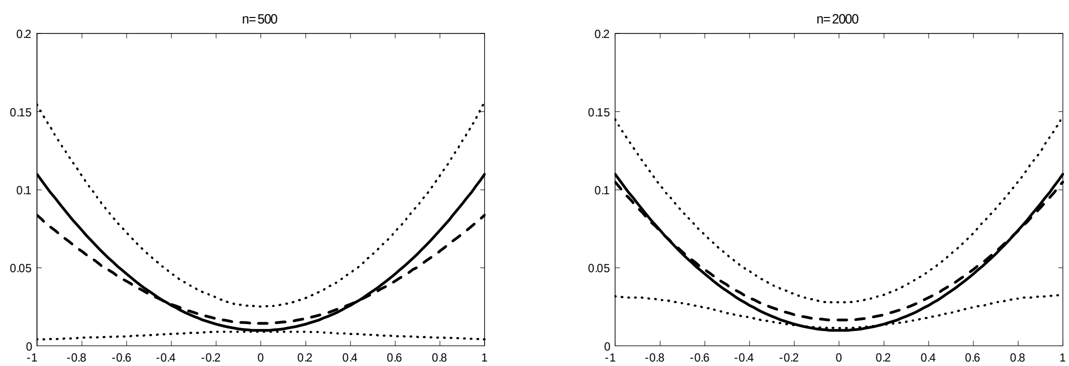

Figure 2. Graphs over the interval [-1,1] of $m(x)$ (solid line), the Monte Carlo estimates of $E(\hat{m}(x))$ (broken line) and 95% estimation bands (dotted line) for the independent case.

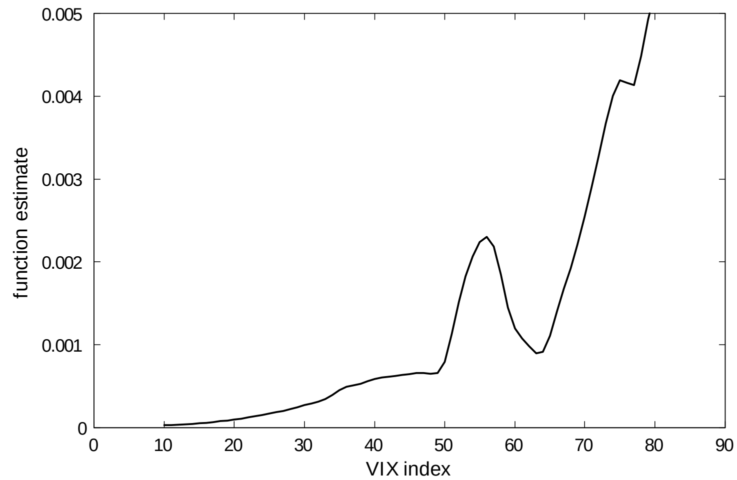

Figure 3. Estimate of $\sigma^2 = m(x_{t-1})$ for the daily S&P 500 index returns from Jan. 1996 to Feb. 2009.

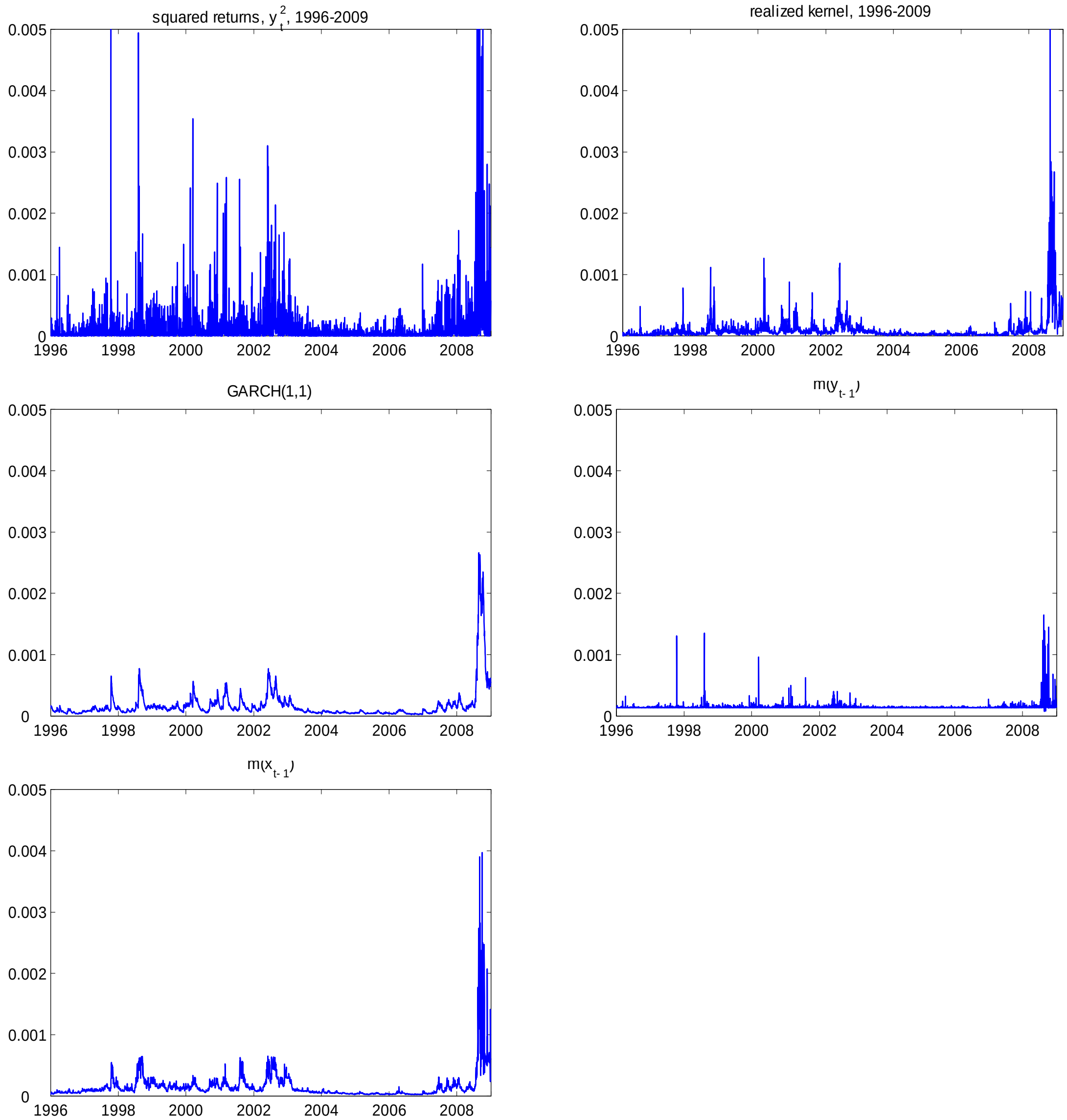

Figure 4. Within-sample fitted values of volatility models for Jan. 1996 - Feb. 2009

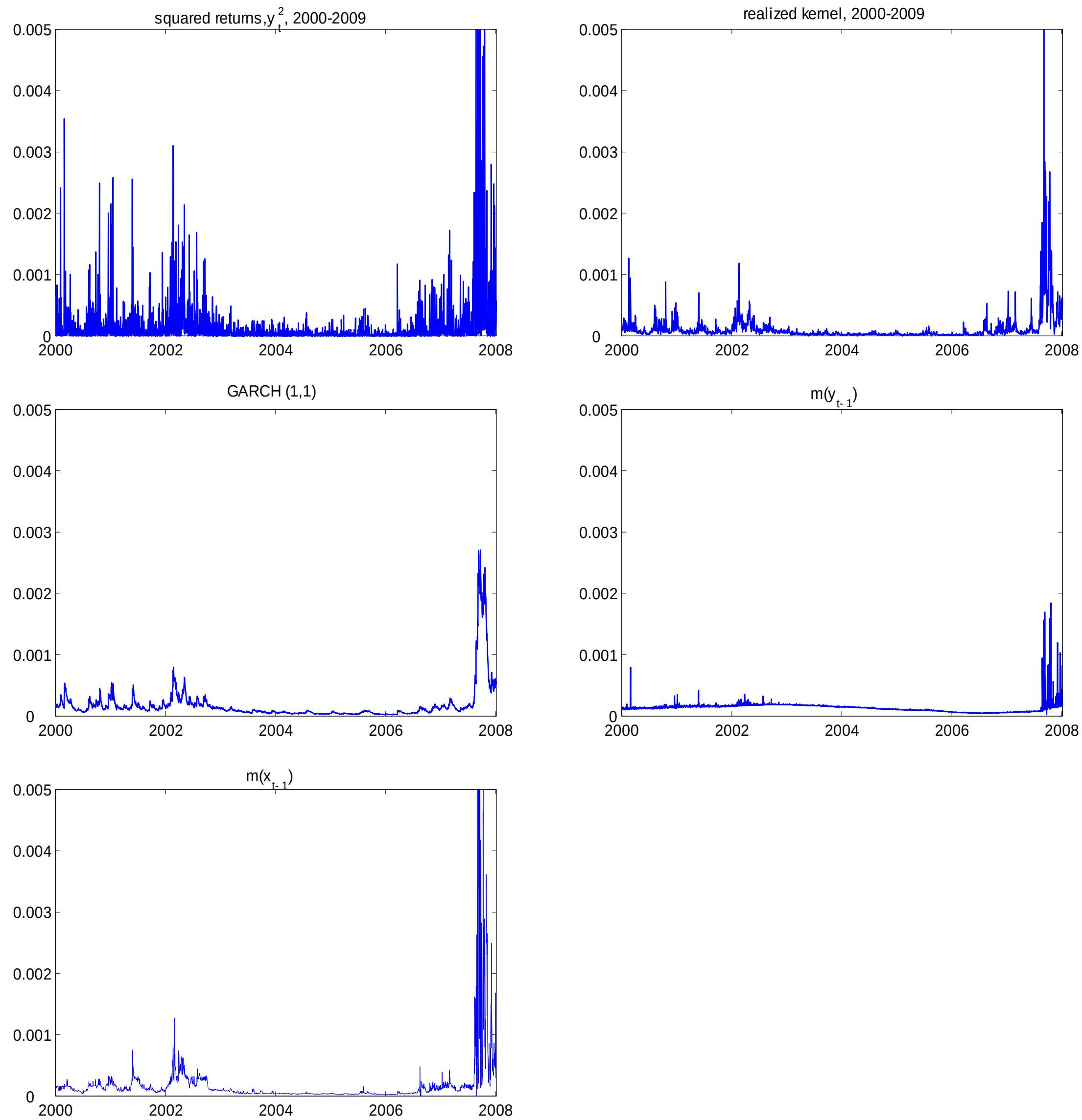

Figure 5. Out-of-sample forecasts of volatility models for Feb. 2000 - Feb. 2009

References

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), Modeling and forecasting realized volatility, *Econometrica,* 71, 529-626.

Audrino, F. and Bühlmann, P. (2001), Tree-structured GARCH models, *Journal of Royal Statistical Society*, 63, 727-744.

Avramidis, P. (2002), Local maximum likelihood estimation of volatility function, Manuscript, LSE.

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996), Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics,* 74, 3-30.

Barndorff-Nielsen, O.E. and Shephard, N. (2002), Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society B*, 64, 253-280.

Barndorff-Nielsen, O.E. P.R. Hansen, A. Lunde and N. Shephard, 2008, Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise, *Econometrica*, 76, 1481-1536.

Blair, B.J., Poon, S.H. and Taylor, S.J. (2001), Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns, *Journal of Econometrics,* 105, 5-26.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics,* 31, 307-327.

Bollerslev, T. and Mikkelsen, H.O. (1996), Modeling and pricing long-memory in stock market volatility, *Journal of Econometrics,* 73, 151-184.

Chiras, D.P. and Manaster, S. (1978), The information content of option prices and a test for market efficiency, *Journal of Financial Economics*, 6, 213-234.

Diebold, F.X. and Inoue, A. (2001), Long memory and regime switching, *Journal of Econometrics,* 105, 131-159.

Diebold, F.X. and Mariano, R.S. (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics,* 13, 253-263.

Ding, Z. and Granger, C.W.J. (1996), Modeling volatility persistence of speculative returns: A new approach, *Journal of Econometrics,* 73, 185-215.

Ding, Z., Granger, C.W.J., and Engle, R.F. (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance,* 1, 83-106.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. Inflation, *Econometrica*, 50, 987-1008.

Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variance, *Econometric Reviews*, 5, 1-50, 81-87.

Engle, R.F. and Lee, G.J. (1999), A long-run and short-run component model of stock return volatility, in Engle R. and H. White ed. *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Chapter 10, 475-497, Oxford University Press.

Engle, R.F. and Ng, V.K. (1993), Measuring and testing the impact of news on volatility, *Journal of Finance*, 48, 1749-1778.

Engle, R.F. and Rangel, J.G. (2008), The spline-GARCH model for low-frequency volatility and its global macroeconomic causes, *Review of Financial Studies*, 21, 1187-1222.

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

Fleming, J. (1998), The quality of market volatility forecasts implied by S&P 100 index option prices, *Journal of Empirical Finance*, 5, 317-345.

Giot, P. (2003), The information content of implied volatility in agricultural commodity markets, *Journal of Futures Markets*, 23, 441-454.

Granger, C.W.J. and Hyung, N. (2004), Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns, *Journal of Empirical Finance,* 11, 399-421.

Han, H. and Park, J.Y. (2008), Time series properties of ARCH processes with persistent covariates, *Journal of Econometrics*, 146, 275-292.

Hansen, P.R. and Lunde, A. (2006), Consistent ranking of volatility models, *Journal of Econometrics*, 131, 97-121.

Härdle, W. and Tsybakov, A.B. (1997), Local polynomial estimators of the volatility function, *Journal of Econometrics*, 81, 223-242.

Heber, G., Lunde, A., Shephard, N. and Sheppard, K. (2009), *OMI's realised library*, Version 0.1, Oxford-Man Institute, University of Oxford.

Hurvich, C.M. and Soulier, P. (2009), Stochastic volatility models with long memory, in Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, Th. (eds.), *Handbook of Financial Time Series*, 157-168, Springer.

Jensen, S.T. and A. Rahbek (2004). "Asymptotic inference for nonstationary GARCH," *Econometric Theory*, 20, 1203-1226.

Karlsen, H.A., Myklebust, T. and Tjϕstheim, D. (2007), Nonparametric estimation in a nonlinear cointegration type model, *Annals of Statistics*, 35, 252-299.

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

Latane, H.A. and Rendleman, R.J. (1976), Standard deviations of stock price ratios implied in option prices, *Journal of Finance*, 31, 369-381.

Loretan, M. and Phillips P.C.B. (1994), Testing the covariance stationarity of heavy-tailed time series, *Journal of Empirical Finance,* 1, 211-248.

Linton, O.B. (2009), Semiparametric and nonparametric ARCH modeling, in Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, Th. (eds.), *Handbook of Financial Time Series*, 157-168, Springer.

Linton, O.B. and Mammen, E. (2005), Estimating semiparametric ARCH(∞) models by kernel smoothing methods, *Econometrica*, 73, 771-836.

Masry, E. and Tjϕstheim, D. (1995), Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality, *Econometric Theory*, 11, 258-289.

Mikosch, T. and Starica, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects, *The Review of Economics and Statistics,* 86, 378-390.

Pagan, C.R. and Hong, Y.S. (1991), Nonparametric estimation and the risk premium, in Barnett, W., Powell, J, and Tauchen, G.E. (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press.

Pagan, A.R. and Schwert, G.W. (1990a), Alternative models for conditional stock volatility, *Journal of Econometrics*, 45, 267-290.

Pagan, A.R. and Schwert, G.W. (1990b), Testing for covariance stationarity in stock market data, *Economics Letters*, 33, 165-170.

Park, J.Y. (2002), Nonstationary nonlinear heteroskedasticity, *Journal of Econometrics*, 110, 383-415.

Park, J.Y. and Phillips, P.C.B. (1999), Asymptotics for nonlinear transformations of integrated time series, *Econometric Theory,* 15, 269-298.

Park, J.Y. and Phillips, P.C.B. (2001), Nonlinear regressions with integrated time series, *Econometrica*, 69, 117-161.

Patton, A.J. (2010), Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics*, forthcoming.

Patton, A.J. and Sheppard, K. (2009), Evaluating volatility and correlation forecasts, in Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, Th. (eds.), *Handbook of Financial Time Series*, 801-838, Springer.

Starica, C. and Granger, C.W.J. (2005), Nonstationarities in stock returns, *The Review of Economics and Statistics,* 87, 503-522.

Wang, Q. and Phillips, P.C.B. (2009a), Asymptotic theory for local time density estimation and nonparametric cointegrating regression, *Econometric Theory,* 25, 1-29.

Wang, Q. and Phillips, P.C.B. (2009b), Structural nonparametric cointegrating regression, *Econometrica*, 77, 1901-1948.

West, K.D. (1996), Asymptotic inference about predictive ability, *Econometrica*, 64, 1067-1084.

Yang, L. (2006), A semiparametric GARCH model for foreign exchange volatility, *Journal of Econometrics*, 130,2, 365-384.

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.