

# MPRA

Munich Personal RePEc Archive

## **Time series properties of ARCH processes with persistent covariates**

Han, Heejoon and Park, Joon Y.

May 2006

Online at <https://mpra.ub.uni-muenchen.de/5199/>

MPRA Paper No. 5199, posted 08 Oct 2007 UTC



## Time Series Properties of ARCH Processes with Persistent Covariates \*

Heejoon Han and Joon Y. Park

RMI Working Paper No. 06/11

Submitted: December 5, 2006

### Abstract

We consider ARCH processes with persistent covariates and provide asymptotic theories that explain how such covariates affect various characteristics of volatility. Specifically, we propose and study a volatility model, named ARCH-NNH model, that is an ARCH(1) process with a nonlinear function of a persistent, integrated or nearly integrated, explanatory variable. Statistical properties of time series given by this model are investigated for various volatility functions. It is shown that our model generates time series that have two prominent characteristics: high degree of volatility persistence and leptokurtosis. Due to persistent covariates, the time series generated by our model has the long memory property in volatility that is commonly observed in high frequency speculative returns. On the other hand, the sample kurtosis of the time series generated by our model either diverges or has a well-defined limiting distribution with support truncated on the left by the kurtosis of the innovation, which successfully explains the empirical finding of leptokurtosis in financial time series. We present two empirical applications of our model. It is shown that the default premium (the yield spread between Baa and Aaa corporate bonds) predicts stock return volatility, and the interest rate differential between two countries accounts for exchange rate return volatility. The forecast evaluation shows that our model generally performs better than GARCH(1,1) and FIGARCH at relatively lower frequencies.

**Keywords:** ARCH, nonstationarity, nonlinearity, NNH, volatility persistence, leptokurtosis.

Heejoon Han  
Department of Economics  
National University of Singapore  
5 Arts Link, AS2 04-36  
Singapore 117570  
Tel: (65) 65166258  
Fax: (65) 6775 2646  
E-mail: [ecshhj@nus.edu.sg](mailto:ecshhj@nus.edu.sg)

Joon Y. Park  
Department of Economics  
Texas A&M University  
Allen 3084  
4228 TAMU, College Station  
TX 77843  
Tel: (979) 845-7309  
E-mail: [jpark@tamu.edu](mailto:jpark@tamu.edu)

\* This paper was circulated previously under the title 'ARCH models with nonstationary covariates'. We would like to thank Yoosoon Chang and Bradley S. Paye for their helpful comments and data. We also thank seminar participants at the 15th annual meeting of Midwest Econometric Group (Carbondale), the 16th annual Asian Finance Association Conference (Kuala Lumpur), the 10th Texas Camp Econometrics (Conroe), Rice University, Indiana University-Bloomington, National University of Singapore, and Bates White, LLC. for helpful comments and suggestions.

# 1 Introduction

ARCH type models have widely been used to model the volatility of economic and financial time series since the seminal work by Engle (1982) and the extension made by Bollerslev (1986). The processes generated by these models successfully show volatility clustering and leptokurtosis, which are commonly observed for many economic and financial time series. However, most of the volatility characteristics of these models have been univariate, relating the volatility of time series only to the information contained in its own past history. Hence, they shed little light on a source of volatility and exclude the possibility that conditional heteroskedasticity may be accounted for by some economic variables.

As an effort to provide some structural or economic explanations for volatility, many previous works considered ARCH type models with exogenous variables. Typically, GARCH(1,1) model with an exogenous variable (mostly in linear form) was used for this purpose. Engle and Patton (2001) tested a GARCH(1,1) model with three month U.S. Treasury bill rates for stock return volatility, and Gray (1996) added the level of interest rates to explain conditional variance in his generalized regime-switching model of short-term interest rates. Likewise, forward-spot spreads and interest rate differentials between countries were used as covariates respectively by Hodrick (1989) and Hagiwara and Herce (1999) to model exchange rate return volatility. Moreover, Lamoureux and Lastrapes (1990) included trading volumes in their stock return volatility model.

The exogenous variables such as interest rates and interest rate differentials between countries used in these works are known to be highly persistent, and may well be modeled as time series having an exact or near unit root. It is therefore natural to expect the persistent covariates would affect the degree of persistence in volatility. However, there has been no theoretical study in the literature to rigorously investigate the effects of the presence of persistent covariates on volatility persistence. They have been largely ignored, and researchers usually focus on the ARCH effects to analyze the degree of volatility persistence. The main motivation of our research is to fill this gap, by providing some important asymptotic theories for ARCH models with persistent covariates. In particular, our theories make it clear how persistent covariates affect volatility persistence and leptokurtosis.

Recently, Park (2002) introduced *nonstationary nonlinear heteroskedasticity* (NNH). NNH models specify conditional variance of given time series as a nonlinear function of an integrated process. The function generating conditional heterogeneity is called *heterogeneity generating function* (HGF). It is shown that volatility clustering and leptokurtosis are manifest for NNH models and, in contrast, *stationary nonlinear heteroskedasticity* (SNH)<sup>2</sup> does not produce volatility clustering.

We introduce a volatility model which combines an ARCH(1) model with the NNH model. We call this model ARCH-NNH. Unlike Park (2002), we allow the NNH part of our models to be generated by a covariate which has a near unit root as well as an exact unit root. Therefore, our models represent a wide class of time series with volatilities

---

<sup>2</sup>In contrast to the NNH model, the conditional heterogeneity is generated by a stationary process in the SNH model.

driven by the autoregression with persistent covariates. The statistical properties of ARCH-NNH models depend crucially on the type of HGF. Two different classes of functions are considered: *integrable* ( $f \in \mathbb{I}$ ) and *asymptotically homogeneous* ( $f \in \mathbb{H}$ ) functions, following Park and Phillips (1999, 2001) who introduced them in their studies on nonlinear models with integrated time series.

We investigate various statistical properties of time series driven by ARCH-NNH models. The asymptotic behavior of sample autocorrelation of squared process generated by ARCH-NNH models shows that volatility clustering is expected for ARCH-NNH models with various HGF's; the autocorrelation is very persistent, i.e., they vanish very slowly or do not even vanish for all lags.

Ding *et al* (1993) found that it is possible to characterize the power transformation of stock return to be long memory. The sample autocorrelation function of squared speculative returns (especially high frequency data) is known to have a typical trend that it decreases fast at first and remains significantly positive for larger lags. The fractionally integrated models such as Long Memory ARCH model by Ding and Granger (1996) and FIGARCH model by Baillie *et al* (1996) are known to capture this long memory property in volatility. Recently, several studies have shown that a number of nonlinear short memory volatility models can also produce spurious long memory characteristics in volatility. One example of such models is the volatility component model by Engle and Lee (1999). And theoretical works in structural change (Mikosch and Starica (2004)), switching regime (Diebold and Inoue (2001)), and occasional breaks (Granger and Hyung (2004)) have shown that any of these events is capable of producing the long memory property.

Our theory shows that ARCH-NNH can also generate the long memory property in volatility. Asymptotically, the autocorrelation function of squared process of ARCH-NNH model with  $f \in \mathbb{I}$  decreases to zero at a hyperbolic rate as the lag order increases. While the autocorrelation function of squared process of GARCH (1,1) vanishes at an exponential rate, that of ARCH-NNH model does at a hyperbolic rate like the fractionally integrated models. On the other hand, the asymptotic autocorrelation function of squared process of ARCH-NNH model with  $f \in \mathbb{H}$  decreases fast (exponentially) at first and converges to some positive random limit. Regardless of the function class of HGF, ARCH-NNH models generate the long memory property in volatility due to persistent covariates.

Similarly as those for the volatility component model by Engle and Lee (1999), the time series generated by ARCH-NNH models can be decomposed into the permanent or long-run component (NNH term) and the transitory or short-run (ARCH term) component. Only a shock to the persistent covariates has a long-run effect on volatility. Additionally, we compare the autocorrelation functions of simulated ARCH-NNH, GARCH(1,1), and FIGARCH processes, which shows that ARCH-NNH processes mimic the movement of real data very well.

Time series generated by ARCH-NNH models with various HGF's all unambiguously predict the presence of leptokurtosis. The sample kurtosis of ARCH-NNH models with  $f \in \mathbb{I}$  diverges at the rate of  $\sqrt{n}$ , which means it is expected to have a large sample kurtosis for any reasonably large sample. The sample kurtosis of ARCH-NNH models with  $f \in \mathbb{H}$  has a random limit bigger than the kurtosis of innovation, which also implies leptokurtosis.

We present two empirical applications of the ARCH-NNH model with HGF given by a

simple power function. One is a stock return volatility model and the persistent covariate used for the model is the default premium (the yield spread between Baa and Aaa corporate bonds). It shows that stock return volatility is well explained by the default premium, which confirms the earlier work by Schwert (1989). The other application is the model for exchange rate return volatility, where we use as a covariate the interest rate differentials between two countries. We could see the modulus of the interest rate differential is well explaining exchange rate return volatility, as observed by Hagiwara and Herce (1999). In both cases, the ARCH-NNH model appears to be quite appropriate and this indicates that the default premium is a proper source of stock return volatility and the interest rate differential is an adequate source of exchange rate return volatility.

Finally, we evaluate the forecasting ability of the ARCH-NNH model. We obtained out-of-sample forecasts at the monthly, weekly and daily frequencies. Using ‘realized volatility’ as a proxy for actual volatility, we employ the regression-based method and the mean absolute error with the forecast accuracy tests by Diebold and Mariano (1995). It is shown that the ARCH-NNH model is not only practically useful, but also outperforms GARCH(1,1) and FIGARCH at relatively lower frequencies.

The rest of the paper is organized as follows. Section 2 introduces the model with some preliminary concepts. Various statistical properties for the samples from ARCH-NNH models are investigated in Section 3. The asymptotic behavior of sample statistics such as sample autocorrelation of squared process, as well as sample variance and kurtosis, is derived. Section 4 presents empirical applications of ARCH-NNH model. The forecasting ability of ARCH-NNH model is evaluated in Section 5. Section 6 concludes the paper, and Appendices A and B contain mathematical proofs for the technical results in the paper.

## 2 The Model and Preliminaries

We write our volatility model as

$$y_t = \sigma_t \varepsilon_t \tag{1}$$

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

**Assumption 1** Assume that

- (a)  $(\varepsilon_t)$  is iid (0,1) and adapted to  $(\mathcal{F}_t)$
- (b)  $(\sigma_t)$  is adapted to  $(\mathcal{F}_{t-1})$

Under assumption 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  $\sigma_t^2$ .

**Assumption 2** Let

$$\sigma_t^2 = \alpha y_{t-1}^2 + f(x_t) \quad (2)$$

for some nonnegative function  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  and

$$x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + v_t \quad (3)$$

assuming that  $(x_t)$  is adapted to  $(\mathcal{F}_{t-1})$  where  $c \geq 0$ .

Assumption 1 and 2 define our volatility model. Under Assumption 1 and 2,  $(y_t)$  is an ARCH process with persistent covariates. The volatility model given by  $\sigma_t^2 = f(x_t)$  is referred to as *Nonstationary Nonlinear Heteroskedasticity* (NNH), which is introduced by Park (2002). Park (2002) considered only the case in which  $(x_t)$  has an exact unit root. We add this NNH term in our volatility model with an extension that allows  $(x_t)$  to have a *near unit root* as well. The specification in (3) allows that  $(x_t)$  has not only an exact unit root but also a near unit root. Since our volatility model is a combination of ARCH(1) and NNH models, it will be referred to as *ARCH-NNH* in our subsequent discussions. Notice that our model does not include  $\sigma_{t-1}^2$  term. Our theory will show that this term is not necessary and instead persistent covariates mainly explain volatility persistence.

The function  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  will be referred to as *heterogeneity generating function* (HGF) in what follows. Clearly,  $f$  must be a nonlinear function, since it has to be nonnegative. More specifically, we consider two classes of functions: *integrable* and *asymptotically homogeneous* functions. These function classes were introduced by Park and Phillips (1999) in their study on the asymptotics of nonlinear transformations of integrated time series. As it was done by Park (2002), the functions that are integrable and asymptotically homogeneous will be called  $\mathbb{I}$  and  $\mathbb{H}$ -regular with added regularity conditions. They will be denoted respectively by  $\mathbb{I}$  and  $\mathbb{H}$ .

To derive the asymptotics for the ARCH-NNH models with  $f \in \mathbb{I}$ , we assume that  $(v_t)$  in (3) is either iid sequence with  $\mathbb{E}|v_t|^q < \infty$  for some  $q > 4$  (as in Assumption 3S), or a stationary linear process driven by an iid sequence  $(\eta_t)$  such that  $\mathbb{E}|\eta_t|^q < \infty$  for some  $q > 4$  (as in Assumption 4S). In the following definition, we let  $q$  be the number that will be given later by such moment conditions.

**Definition 2.1** A transformation  $f$  on  $\mathbb{R}$  is called  $\mathbb{I}$ -regular ( $f \in \mathbb{I}$ ) if  $f$  is bounded, integrable and piecewise Lipschitz, i.e.,

$$|f(x) - f(y)| \leq c|x - y|^\ell$$

on each piece of its support, for some constant  $c$  and  $\ell > 6/(q - 2)$ .

To define  $\mathbb{H}$ -regular functions, we introduce some classes of transformations on  $\mathbb{R}$ . We denote by  $F_B$  the class of all bounded transformations on  $\mathbb{R}$ , and denote by  $F_{LB}$  the class of locally bounded transformations on  $\mathbb{R}$ . Define  $F_B^0$  be the class of all bounded functions vanishing at infinity, and let  $F_{LB}^0$  be the subset of  $F_{LB}$  consisting of  $T$  such that  $T(x) = O(\exp(c|x|))$  as  $|x| \rightarrow \infty$  for some constant  $c$ .

**Definition 2.2** A transformation  $f$  on  $\mathbb{R}$  is called  $\mathbb{H}$ -regular ( $f \in \mathbb{H}$ ) if  $f$  can be written as

$$f(\lambda x) = \kappa(\lambda)\bar{f}(x) + R(x, \lambda)$$

for large  $\lambda$  uniformly in  $x$  over any compact interval, where  $\bar{f}$  is locally Riemann integrable and  $R$  satisfies

$$R(x, \lambda) = a(\lambda)p(x) \quad \text{or} \quad b(\lambda)p(x)q(\lambda x)$$

with  $a$  and  $b$  such that  $a(\lambda)/\kappa(\lambda) \rightarrow 0$  and  $b(\lambda)/\kappa(\lambda) < \infty$  as  $\lambda \rightarrow \infty$ , and  $p$  and  $q$  such that  $p \in F_{LB}^0$  and  $q \in F_B^0$ . For  $f \in \mathbb{H}$ , We call  $\kappa$  and  $\bar{f}$ , respectively, the *asymptotic order* and *limit homogeneous function* of  $f$ .

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  $1\{x \geq 0\}$ .

Standard terminologies and notations in probability and measure theory are used throughout the paper. In particular, notations for various convergences such as  $\rightarrow_{a.s.}$ ,  $\rightarrow_p$  and  $\rightarrow_d$  frequently appear. The notation  $=_d$  signifies equality in distribution. Some theoretic tools are introduced in the following. In the next section, we are going to introduce assumptions for  $(v_t)$ , which make the time series  $(x_t)$  in (3) become a general linear process with a near unit root or an exact unit root. 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. Either Assumption 3 or Assumption 4 in the next section satisfies conditions of  $(v_t)$  for the followings. We define

$$V_{cn}(r) = n^{-1/2}x_{[nr]}$$

for  $r \in [0, 1]$ , where  $[z]$  denotes the largest integer which does not exceed  $z$ . And we let

$$V_c(r) = \int_0^r \exp(-c(r-s)) dV_0(s)$$

where  $r \in [0, 1]$  and  $V_0$  is the standard Brownian Motion. Then, it is well known that

$$V_{cn} \rightarrow_d V_c$$

as  $n \rightarrow \infty$ .  $V_c$  is an Ornstein-Uhlenbeck process, generated by the stochastic differential equation

$$dV_c(r) = -cV_c(r)dr + dV_0(r)$$

with the initial condition  $V_c(0) = 0$ . See Phillips (1987) and Stock (1994).

As in Park (2003), our subsequent theory for the case of integrable HGFs involves the *local time of Ornstein-Uhlenbeck process*. The local time for  $V_c$  is then defined as

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

Roughly speaking,  $2\varepsilon$  times  $L_c(t, s)$  measures the actual time spent by  $V_c$  in the  $\varepsilon$ -neighborhood of  $s$  up to time  $t$ . The local time yields the *occupation time formula*

$$\int_0^t T(V_c(r)) dr = \int_{-\infty}^{\infty} T(s) L_c(t, s) ds$$

for any  $T : \mathbb{R} \rightarrow \mathbb{R}$  locally integrable. For each  $t$ , the occupation time formula allows us to evaluate the time integral of a nonlinear function of an Ornstein-Uhlenbeck process by means of the integral of the function itself weighted by the local time.

### 3 Statistical Properties of ARCH-NNH

We investigate the statistical properties of ARCH-NNH models. In particular, the asymptotic behavior of the sample autocorrelation function of squared process and other sample moments such as sample variance and sample kurtosis of process generated by ARCH-NNH models are derived.

#### 3.1 Sample Autocorrelation of Squared Process

Ding *et al* (1993) investigated the long memory property of stock returns and they found out that it is possible to characterize the power transformation of stock returns to be long memory. This long memory property is known to be commonly observed in high frequency data. Figure 1 shows the sample autocorrelation function of daily squared returns on S&P 500 index and this confirms the long memory property; the autocorrelation function decreases fast at first and remains significantly positive for larger lags. Hitherto the fractionally integrated models such as Long Memory ARCH model by Ding and Granger (1996) and FIGARCH model by Baillie *et al* (1996) are known to capture this property.

Recently, several studies have shown that a number of non-linear short memory volatility models can also produce spurious long memory characteristic in volatility. One example of such models is the volatility component model by Engle and Lee (1999). And, theoretical works in structural change (Mikosch and Starica (2004)), switching regime (Diebold and Inoue (2001)) and occasional breaks (Granger and Hyung (2004)) have shown that any of these events is capable of producing the long memory property. See Hyung *et al* (2005). In this section, we are going to examine if ARCH-NNH models also produce this long memory property in volatility.



Define the sample autocorrelation of  $(y_t^2)$  by

$$R_{nk}^2 = \frac{\sum_{t=k+1}^n (y_t^2 - \bar{y}_n^2) (y_{t-k}^2 - \bar{y}_n^2)}{\sum_{t=1}^n (y_t^2 - \bar{y}_n^2)^2},$$

where  $\bar{y}_n^2$  denotes the sample mean of  $(y_t^2)$ . To precisely characterize the asymptotic behavior of  $R_{nk}^2$  under ARCH-NNH models, we make the following additional assumptions.

**Assumption 3S** Assume

- (a)  $(v_t)$  are iid.
- (b)  $\mathbb{E}f^2(x + v_{kt}) < \infty$  for all  $x \in \mathbb{R}$  and  $k \geq 1$ , where  $v_{kt} = v_{t+1} + \dots + v_{t+k}$ .
- (c)  $\mathbb{E}|\varepsilon_t|^p < \infty$  for some  $p \geq 8$ .
- (d)  $(\varepsilon_t)$  and  $(v_t)$  are independent.
- (e)  $(v_t)$  has distribution absolutely continuous with respect to Lebesgue measure, characteristic function  $\phi(t)$  such that  $t^r \phi(t) \rightarrow 0$  as  $t \rightarrow \infty$  for some  $r > 0$ , and  $\mathbb{E}|v_t|^q < \infty$  for some  $q > 4$ .

**Assumption 3W** Assume (a)-(e) of Assumption 3S with  $q > 2$ .

Assumption 3W is weaker than Assumption 3S, where ‘W’ and ‘S’ stand for *weak* and *strong* respectively. Whenever the distinction is unnecessary, we will just refer to Assumption 3. Under Assumption 3S,  $(v_{kt})$  has density with respect to Lebesgue measure on  $\mathbb{R}$ , and we signify the density by  $p_k$ . Also, we denote the kurtosis of  $(\varepsilon_t)$  by  $\kappa_\varepsilon^4$  throughout the paper.

**Theorem 1** Let Assumptions 1 and 2 hold, and let  $k \geq 1$ . Assume that  $0 < \alpha < 1$  and  $\alpha^2 \kappa_\varepsilon^4 < 1$ .

(a) If  $f \in \mathbb{I}$ , then under Assumption 3S

$$R_{nk}^2 \xrightarrow{p} \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(x+y) \sum_{j=0}^{k-1} \sum_{i=0}^{\infty} \alpha^{i+j} p_{k+i-j}(y) dx dy}{\frac{2\kappa_\varepsilon^4}{1-\alpha^2 \kappa_\varepsilon^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(x+y) \sum_{i=1}^{\infty} \alpha^i p_i(y) dx dy + \frac{\kappa_\varepsilon^4}{1-\alpha^2 \kappa_\varepsilon^4} \int_{-\infty}^{\infty} f^2(s) ds} + \alpha^k$$

as  $n \rightarrow \infty$ .

(b) If  $f \in \mathbb{H}$  with limit homogeneous function  $\bar{f}$ , then under Assumption 3W

$$R_{nk}^2 \xrightarrow{d} \frac{(1 - \alpha^k) \left[ \int_0^1 \bar{f}^2(V_c(r)) dr - \left( \int_0^1 \bar{f}(V_c(r)) dr \right)^2 \right]}{\frac{(1-\alpha^2)}{1-\alpha^2 \kappa_\varepsilon^4} \kappa_\varepsilon^4 \int_0^1 \bar{f}^2(V_c(r)) dr - \left( \int_0^1 \bar{f}(V_c(r)) dr \right)^2} + \alpha^k$$

as  $n \rightarrow \infty$ .

Theorem 1 shows the asymptotic behavior of the sample autocorrelations of squared process generated by ARCH-NNH models. Regardless of function classes, probability limits of  $R_{nk}^2$  contain  $\alpha^k$ . Considering that  $R_{nk}^2$  of ARCH(1) converges to  $\alpha^k$ , we can tell that  $\alpha^k$  term comes from the ARCH component and the other term of each case is originated from the NNH component. (Recall that ARCH-NNH model is a combination of ARCH(1) and NNH.) In the following context, we can see that persistent covariates mainly explain volatility persistence and also generate the long memory property in volatility.

The part (a) of Theorem 1 shows that, for ARCH-NNH models with  $f \in \mathbb{I}$ ,  $R_{nk}^2$  converges in probability to a nonrandom limit, which as a function of  $k \geq 1$  we may regard as the asymptotic autocorrelation function of squared process and we denote  $R_k^2$  hereafter. The actual value of  $R_k^2$  is determined by the distribution of  $(v_t)$  as well as HGF. In order to explain volatility persistence,  $R_k^2$  should at least decrease at a slow rate as  $k \rightarrow \infty$ . As it was done by Park (2002), let us consider the case in which the distribution of  $(v_t)$  is Gaussian. Since we have  $v_{kt} =_d \sqrt{k}v_t$  in the case, it follows that  $p_k(x) = \frac{1}{\sqrt{k}}p\left(\frac{x}{\sqrt{k}}\right)$  where  $p$  is the normal density. Since the normal density is continuous at the origin, we have

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(x+y)p_k(y)dx dy &= \frac{1}{\sqrt{k}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(x+y)p\left(\frac{y}{\sqrt{k}}\right)dx dy \\ &= \frac{1}{\sqrt{k}}p(0) \left( \int_{-\infty}^{\infty} f(x)dx \right)^2 \rightarrow 0 \end{aligned}$$

as  $k \rightarrow \infty$ .

Using this, we can show that  $R_k^2$  in part (a) of Theorem 1 decreases to zero as  $k \rightarrow \infty$ . When  $(v_t)$  is Gaussian, we have

$$\sum_{j=0}^{k-1} \sum_{i=0}^{\infty} \alpha^{i+j} p_{k+i-j}(y) = \sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{\infty} \alpha^i \frac{1}{\sqrt{k+i-j}} p\left(\frac{y}{\sqrt{k+i-j}}\right).$$

Let  $c$  be an arbitrary number such that  $0 < c < k - 1$ . For  $j = c$ , as  $k \rightarrow \infty$

$$\alpha^j \sum_{i=0}^{\infty} \alpha^i \frac{1}{\sqrt{k+i-j}} p\left(\frac{y}{\sqrt{k+i-j}}\right) \rightarrow 0$$

because  $\frac{1}{\sqrt{k+i-j}}$  will dominate and the rate of decay is  $k^{-1/2}$ . For  $j = k - c$ , then  $\alpha^j$  will dominate and we have an exponential convergence rate. Since  $c$  is arbitrary, we can divide  $\sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{\infty} \alpha^i \frac{1}{\sqrt{k+i-j}} p\left(\frac{y}{\sqrt{k+i-j}}\right)$  into two parts; one part decays at a hyperbolic rate  $k^{-1/2}$  and the other part does at an exponential rate  $\alpha^k$ . So, we can deduce that

$$\sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{\infty} \alpha^i \frac{1}{\sqrt{k+i-j}} p\left(\frac{y}{\sqrt{k+i-j}}\right) \rightarrow 0$$

as  $k \rightarrow \infty$  and its decay rate is hyperbolic.

The fact that  $R_k^2$  converges to zero is similar with the behavior of GARCH(1,1) model. However, there is an important difference in their respective decreasing patterns. Under GARCH(1,1),  $R_k^2$  decreases at an exponential rate as  $k$  increases.<sup>3</sup> However,  $R_k^2$  of ARCH-NNH models with  $f \in \mathbb{I}$  decreases at a hyperbolic rate. This means that ARCH-NNH models with  $f \in \mathbb{I}$  produce the long memory property in volatility as the fractionally integrated models do. In other words, volatility persistence is well expected in ARCH-NNH models with  $f \in \mathbb{I}$ .

On the other hand, the part (b) of Theorem 1 gives the asymptotic limit of  $R_{nk}^2$  for ARCH-NNH models with  $f \in \mathbb{H}$ , and it is very different from that of the previous  $f \in \mathbb{I}$  case. Since an Ornstein-Uhlenbeck process is included,  $R_k^2$  is random and it is not affected by the distribution of  $(v_t)$ .  $R_k^2$  of ARCH-NNH models with  $f \in \mathbb{H}$  decreases as  $k \rightarrow \infty$  but it does not converge to zero. Let

$$A = \frac{\frac{1}{(1-\alpha)} \left[ \int_0^1 \bar{f}^2(V(r))dr - \left( \int_0^1 \bar{f}(V(r))dr \right)^2 \right]}{\frac{(1+\alpha)\kappa_\varepsilon^4}{1-\alpha^2\kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V(r))dr - \frac{1}{(1-\alpha)} \left( \int_0^1 \bar{f}(V(r))dr \right)^2}.$$

Then,

$$R_k^2 = A + \alpha^k (1 - A).$$

As  $k \rightarrow \infty$ ,  $R_k^2$  decreases exponentially at first and finally converges to  $A$  which is a random constant clearly smaller than unity and positive unless limit homogeneous function  $\bar{f}$  is constant. This trend of  $R_k^2$  is compatible with the sample autocorrelation of the real data in Figure 1. Like the previous  $f \in \mathbb{I}$  case, ARCH-NNH models with  $f \in \mathbb{H}$  also capture the long memory property in volatility. Therefore, it is also expected that ARCH-NNH models with  $f \in \mathbb{H}$  would properly explain volatility persistence.

When  $f \in \mathbb{H}$ ,  $R_k^2$  of ARCH-NNH models behaves similarly as that of NNH models. However,  $R_k^2$  of ARCH-NNH models with  $f \in \mathbb{H}$  is dependent on  $k$  and decreases as  $k \rightarrow \infty$ , which is different from  $R_k^2$  of NNH models. In case of NNH models with  $f \in \mathbb{H}$ ,  $R_k^2$  is independent of  $k$  and given by a random constant for all values of the lag order  $k \geq 1$ .

Additionally, the result in part (b) of Theorem 1 implies that if  $f$  has constant limit homogeneous functions then the sample autocorrelation of squared process by ARCH-NNH models converges in probability to  $\alpha^k$ . Suppose that  $f(x_t) = c + g(x_t)$  where  $c$  is constant and  $g(x_t) \in \mathbb{I}$ . Then  $f$  is asymptotically homogeneous and its limit homogeneous function is constant. This means  $R_{nk}^2 \rightarrow_p \alpha^k$  just like ARCH(1). Note that if an ARCH-NNH model has  $f(x_t) = c$  then it is exactly the ARCH(1) model. Therefore, if  $f(x_t)$  consists of constant and  $g(x_t) \in \mathbb{I}$ , the integrable function  $g(x_t)$  does not affect volatility persistence asymptotically.

Theorem 1 shows that, due to persistent covariates, ARCH-NNH process explains volatility persistence very well and, especially, produces the long memory property in volatility.

<sup>3</sup>The sample autocorrelation of the squared process of stationary GARCH(1,1) has probability limit given by  $(\alpha + \beta)^{k-1} \frac{\alpha(1-\alpha\beta-\beta^2)}{1-2\alpha\beta-\beta^2}$  for  $k \geq 1$  if  $\alpha^2\kappa_\varepsilon^4 + 2\alpha\beta + \beta^2 < 1$ .

The following text gives us better understanding about how persistent covariates play a role in volatility persistence.

### 3.1.1 Decomposition of Volatility

In the volatility component model by Engle and Lee (1999), volatility is decomposed into a permanent or long-run component and a transitory or short-run component. Similarly, an ARCH-NNH model has two components, and an NNH term represents a permanent/long-run component while an ARCH term explains a transitory/short-run component. We follow the way done by Ding and Granger (1996), who showed that, for the IARCH(1) process, a shock may permanently affect the ‘expectation’ of a future conditional variance process, but it does not permanently affect the ‘true’ conditional process itself.

According to assumption 2,  $x_t$  is adapted to  $(\mathcal{F}_{t-1})$ . To make it simple, let us consider the case in which  $x_t$  is adapted to  $(\mathcal{F}_t)$ ,  $\varepsilon_t \sim N(0, 1)$  and  $\frac{\partial f(x)}{\partial x} \neq 0$ . Then we have

$$y_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, 1) \quad (4)$$

$$\sigma_t^2 = \alpha y_{t-1}^2 + f(x_{t-1}) \quad (5)$$

$$x_t = \left(1 - \frac{c}{n}\right) x_{t-1} + v_t \quad (6)$$

In ARCH-NNH models, a shock to the system at time  $t$  comes from  $\varepsilon_t$  or  $v_t$ , and this shock will not affect  $\sigma_t^2$  because  $\sigma_t^2$  depends only on the past information. Since

$$\sigma_{t+k}^2 = \prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2 \alpha^k y_t^2 + \sum_{j=0}^{k-1} \prod_{i=1}^j \varepsilon_{t+k-i}^2 \alpha^j f(x_{t+k-1-j})$$

(we let  $\prod_{i=1}^0 \varepsilon_{t+k-i}^2 = 1$  here), we have

$$\mathbb{E}(y_{t+k}^2) = \mathbb{E}(\sigma_{t+k}^2) = \alpha^k y_t^2 + \alpha^{k-1} f(x_t) + \sum_{j=0}^{k-2} \alpha^j \mathbb{E}(f(x_{t+k-1-j})).$$

A shock at time  $t$  to  $y_t^2$ , from  $\varepsilon_t$ , and a shock to  $x_t$ , from  $v_t$ , will permanently change  $\mathbb{E}(y_{t+k}^2)$  and  $\mathbb{E}(\sigma_{t+k}^2)$ , i.e., both shocks affect the ‘expectation’ of the future squared process and the future conditional variance process. However, we have different situation in case of the ‘true’  $y_{t+k}^2$  and  $\sigma_{t+k}^2$ . The real impact of a shock from  $\varepsilon_t$  (a change in  $y_t^2$ ) to  $\sigma_{t+k}^2$  is

$$\frac{\partial \sigma_{t+k}^2}{\partial y_t^2} = \prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2 \alpha^k \quad (7)$$

and the real impact of a shock from  $v_t$  (a change in  $x_t$ ) to  $\sigma_{t+k}^2$  is

$$\begin{aligned}\frac{\partial \sigma_{t+k}^2}{\partial x_t} &= \sum_{j=0}^{k-1} \prod_{i=1}^j \varepsilon_{t+k-i}^2 \alpha^j \frac{\partial f(x_{t+k-1-j})}{\partial x_t} \\ &= \sum_{j=0}^{k-1} \prod_{i=1}^j \varepsilon_{t+k-i}^2 \alpha^j \left(1 - \frac{c}{n}\right)^{k-1-j} \frac{\partial f(x_{t+k-1-j})}{\partial x_{t+k-1-j}}\end{aligned}$$

because  $\frac{\partial f(x_{t+k-1-j})}{\partial x_t} = \frac{\partial f(x_{t+k-1-j})}{\partial x_{t+k-1-j}} \frac{\partial x_{t+k-1-j}}{\partial x_t} = \frac{\partial f(x_{t+k-1-j})}{\partial x_{t+k-1-j}} \left(1 - \frac{c}{n}\right)^{k-1-j}$ .

**Corollary 2** Given (4)-(6), let  $k \geq 1$ ,  $0 < \alpha \leq 1$ , and  $\frac{\partial f(x)}{\partial x} \neq 0$ . As  $k \rightarrow \infty$ ,

- (a)  $\frac{\partial \sigma_{t+k}^2}{\partial y_t^2} \rightarrow_{a.s.} 0$ .
- (b)  $\left| \frac{\partial \sigma_{t+k}^2}{\partial x_t} \right| \rightarrow_{a.s.} h$  for some  $h > 0$  if  $c = 0$ .
- (c)  $\frac{\partial \sigma_{t+k}^2}{\partial x_t} \rightarrow 0$  at a slower rate than  $\frac{\partial \sigma_{t+k}^2}{\partial y_t^2}$  if  $c > 0$  and  $\alpha^2 \kappa_\varepsilon^4 < 1$ .

Note in the part (a) of Corollary 2 that even if  $\alpha = 1$ ,  $\frac{\partial \sigma_{t+k}^2}{\partial y_t^2} \rightarrow 0$  almost surely as  $k \rightarrow \infty$ . This is because, as shown in Nelson (1990),  $\prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2$  in (7) converges to zero almost surely. Similarly, it should be noticed that the part (b) of Corollary 2 holds even when  $\alpha = 1$ . Since  $y_{t+k}^2 = \sigma_{t+k}^2 \varepsilon_{t+k}^2$ , we have the similar result for  $y_{t+k}^2$ . Therefore, the part (a) and (b) of Corollary 2 indicate that while the real impact of a shock from  $\varepsilon_t$  will converge to zero, a shock to  $v_t$  will permanently affect the ‘true’ process of  $\sigma_{t+k}^2$  and  $y_{t+k}^2$  if  $x_t$  is an  $I(1)$  process. A shock to  $\varepsilon_t$  is not persistent in  $\sigma_{t+k}^2$  and  $y_{t+k}^2$ , but a shock to  $v_t$  is persistent in  $\sigma_{t+k}^2$  and  $y_{t+k}^2$ . If  $c > 0$ , then the real impact of a shock from  $v_t$  will disappear eventually. However, the part (c) of Corollary 2 shows that a shock from  $v_t$  has a longer effect than a shock from  $\varepsilon_t$ . Hence, we can consider a shock from  $\varepsilon_t$  as a short-run shock and a shock from  $v_t$  as a long-run shock.

### 3.1.2 Simulated Autocorrelation Functions

Now we are going to look at the behavior of simulated ARCH-NNH processes and examine how much it mimics the movement of real data. Figure 2-4 show the autocorrelation functions of squared return series on S&P 500 index and some simulated series. The sample sizes are 1014, 1079 and 3938 for the monthly, weekly and daily frequencies respectively. We use these data for our first empirical application in the section 4 and the sample period for each frequency is given there. The solid curves indicate the autocorrelation functions of the real data and the dotted curves are those of simulated series.

We generated data using the estimates obtained from our first empirical application in the section 4 and drew autocorrelation functions for the ARCH-NNH model with a simple power function, GARCH (1,1) model and FIGARCH(1,d,q)<sup>4</sup> model based on 5,000 iteration. The upper and lower dotted curves indicate the 5%- and 95%-quantile of the distribution of the autocorrelations at a fixed lag and the middle dotted curves correspond to the mean of those distributions.

Overall, the simulations show that ARCH-NNH processes mimic the movements of the real data very well. For the monthly case, the simulated ARCH-NNH process performs very well in following the movement of the real data. Even if GARCH(1,1) and FIGARCH are also doing fine, their simulated processes are a little higher than the autocorrelation function of the real data. For the weekly and daily cases, both our model and FIGARCH are doing well, but it is obvious that GARCH(1,1) is not proper in following the movement of the real data.

### 3.2 Sample Variance and Kurtosis

We now investigate the asymptotic behaviors of other sample moments such as sample variance and kurtosis. The sample variance of  $(y_t)$  is defined by

$$S_n^2 = \frac{1}{n} \sum_{t=1}^n y_t^2.$$

We introduce additional assumptions for the asymptotics of the sample variance.

**Assumption 4S** Assume

- (a)  $\mathbb{E}|\varepsilon_t|^p < \infty$  for some  $p \geq 4$ .
- (b)  $(v_t)$  is generated by

$$v_t = \pi(L)\eta_t = \sum_{k=0}^{\infty} \pi_k \eta_{t-k} \tag{8}$$

where  $\pi_0 = 1$ ,  $\pi(1) \neq 0$  with  $\sum_{k=0}^{\infty} k|\pi_k| < \infty$ , and  $(\eta_t)$  are iid and has distribution absolutely continuous with respect to Lebesgue measure, characteristic function  $\varphi(t)$  such that  $t^r \varphi(t) \rightarrow 0$  as  $t \rightarrow \infty$  for some  $r > 0$ , and  $\mathbb{E}|\eta_t|^q < \infty$  for some  $q > 4$ .

**Assumption 4W** Assume (a) of Assumption 4S and (b) of Assumption 4S with  $q > 2$ .

---

<sup>4</sup> $q = 1$  for the monthly frequency and  $q = 0$  for the weekly and daily frequencies. See the estimation part in the section 4.

**Theorem 3** *Let Assumptions 1 and 2 hold. Assume that  $0 < \alpha < 1$ .*

(a) *If  $f \in \mathbb{I}$ , then under Assumption 4S*

$$\sqrt{n}S_n^2 \rightarrow_d \frac{1}{1-\alpha} L_c(1,0) \int_{-\infty}^{\infty} f(s)ds$$

as  $n \rightarrow \infty$ .

(b) *If  $f \in \mathbb{H}$  with limit homogeneous function  $\bar{f}$  and asymptotic order  $\kappa$ , then under Assumption 4W*

$$\kappa(\sqrt{n})^{-1}S_n^2 \rightarrow_d \frac{1}{1-\alpha} \int_0^1 \bar{f}(V_c(r))dr$$

as  $n \rightarrow \infty$ .

The asymptotics for the sample variance are given in Theorem 3. The results are exactly same as those of NNH models except that  $\frac{1}{1-\alpha}$  is multiplied. The sample variance of ARCH-NNH processes with  $f \in \mathbb{I}$  converges in probability to zero as  $n \rightarrow \infty$ . The behavior of the sample variance of ARCH-NNH processes with  $f \in \mathbb{H}$  depends on the asymptotic order of HGF. If the asymptotic order  $\kappa$  is unity (for example, bounded  $f \in \mathbb{H}$ ), the asymptotic variance is finite. If  $\kappa(\sqrt{n}) \rightarrow \infty$  as  $n \rightarrow \infty$  (for example, power functions), the asymptotic variance would be infinite like IGARCH models.

It is well known that many financial series are leptokurtic. In order to see if the process generated by ARCH-NNH models is leptokurtic, we investigate the asymptotic behavior of sample kurtosis. We define the sample kurtosis of  $(y_t)$  by

$$K_n^4 = \frac{1}{n} \sum_{t=1}^n y_t^4 \bigg/ \left( \frac{1}{n} \sum_{t=1}^n y_t^2 \right)^2.$$

We introduce additional assumptions for the asymptotics of the sample kurtosis.

**Assumption 5S** Assume (a)  $\mathbb{E}|\varepsilon_t|^p < \infty$  for some  $p \geq 8$ , and (b) of Assumption 4S.

**Assumption 5W** Assume (a) of Assumption 5S and (b) of Assumption 4W.

**Theorem 4** *Let Assumptions 1 and 2 hold. Assume that  $0 < \alpha < 1$  and  $\alpha^2 \kappa_\varepsilon^4 < 1$ .*

(a) *If  $f \in \mathbb{I}$ , then under Assumption 5S*

$$\frac{1}{\sqrt{n}}K_n^4 \rightarrow_d \frac{\frac{2\kappa_\varepsilon^4}{1-\alpha^2\kappa_\varepsilon^4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(x+y) \sum_{i=1}^{\infty} \alpha^i p_i(y) dx dy + \frac{\kappa_\varepsilon^4}{1-\alpha^2\kappa_\varepsilon^4} \int_{-\infty}^{\infty} f^2(s)ds}{L_c(1,0) \left( \frac{1}{1-\alpha} \int_{-\infty}^{\infty} f(s)ds \right)^2}$$

as  $n \rightarrow \infty$ .

(b) If  $f \in \mathbb{H}$  with limit homogeneous function  $\bar{f}$ , then under Assumption 5W

$$K_n^4 \rightarrow_d \frac{\frac{1-\alpha^2}{1-\alpha^2\kappa_\varepsilon^4}\kappa_\varepsilon^4 \int_0^1 \bar{f}^2(V_c(r))dr}{\left(\int_0^1 \bar{f}(V_c(r))dr\right)^2}$$

as  $n \rightarrow \infty$ .

The asymptotics for the sample kurtosis of  $(y_t)$  are given in Theorem 4. The result for ARCH-NNH models with  $f \in \mathbb{I}$  shows that the sample kurtosis  $(y_t)$  diverges at the rate of  $\sqrt{n}$  as  $n \rightarrow \infty$ . Therefore, it is expected to have a larger sample kurtosis as sample size increases, which explains leptokurtosis observed in many economic and financial data.

On the other hand, the sample kurtosis of ARCH-NNH models with  $f \in \mathbb{H}$  converges to a random constant. However, the limit of the sample kurtosis is bigger than the kurtosis of innovation  $(\varepsilon_t)$ ,

$$\kappa_\varepsilon^4 < \frac{\frac{1-\alpha^2}{1-\alpha^2\kappa_\varepsilon^4}\kappa_\varepsilon^4 \int_0^1 \bar{f}^2(V(r))dr}{\left(\int_0^1 \bar{f}(V(r))dr\right)^2},$$

because  $\left(\int_0^1 \bar{f}(V(r))dr\right)^2 \leq \int_0^1 \bar{f}^2(V(r))dr$  by Cauchy-Schwarz inequality and  $1 < \frac{1-\alpha^2}{1-\alpha^2\kappa_\varepsilon^4}$ . Therefore, leptokurtosis is naturally expected for time series generated by ARCH-NNH models with  $f \in \mathbb{H}$ . Note that the inequality is strict even if  $\bar{f}$  is constant.<sup>5</sup> In case of ARCH(1),  $K_n^4 \rightarrow_d \frac{1-\alpha^2}{1-\alpha^2\kappa_\varepsilon^4}\kappa_\varepsilon^4$ . Unless  $\bar{f}$  is constant, asymptotically, ARCH-NNH models with  $f \in \mathbb{H}$  will have bigger kurtosis than ARCH (1) by Cauchy-Schwarz inequality.

## 4 Empirical Applications

We investigate two empirical applications in this section. One is for stock return volatility and the other is for exchange rate return volatility. We consider the ARCH-NNH model with a simple HGF,  $f(x) = \beta |x|^\rho$ , in both cases.

### 4.1 Stock Return Volatility

Schwert (1989) found out that the difference between yields on bonds of different quality is directly related to subsequently observed stock return volatility. This leads us to expect that a proper function of the yield spreads can predict stock return volatility. We selected to work with the S&P 500 Index return series.<sup>6</sup> The sample period for the monthly frequency

<sup>5</sup>This is another difference between ARCH-NNH and NNH models. For the NNH model,  $K_n^4 \rightarrow_p \kappa_\varepsilon^4$  if  $S$  is constant.

<sup>6</sup>We obtained the monthly indexes, average of daily indexes in the month, from Dr.Robert J. Shiller's website. The weekly indexes, from Yahoo.Finance, are sampled on every Friday. The daily indexes are from the Center for Research in Security Prices (CRSP).



is from January 1919 to June 2003 (1014 observations). It is from 23 October 1982 to 27 June 2003 at the weekly frequency (1079 observations) and from 2 November 1987 to 30 June 2003 at the daily frequency (3938 Observations).

The default premium (the spread between the Moody's Baa and Aaa corporate bond yields) is used for the ARCH-NNH model and Table 1 shows the results of unit root tests for the series. We consider two alternative autoregressive specifications for the series: with and without a linear deterministic trend. Let us consider the monthly and weekly cases first. The estimated autoregressive coefficients are between 0.974 and 0.982. Phillips-Perron  $Z_t$  rejects the null hypothesis of a unit root in most cases. However, the KPSS test rejects the null hypothesis of stationarity at 1% significance level in every case, which suggests that there exists strong evidence in favor of the nonstationary alternative. Considering the strong results of KPSS tests and that the estimated autoregressive coefficients are close to unity, we conclude that there exists at least a near unit root for the monthly and weekly cases. For the daily case, unit root tests strongly support presence of a unit root. The estimated autoregressive coefficients are 0.996 and Phillips-Perron tests are unable to reject the null hypothesis of a unit root while KPSS tests reject the null hypothesis of stationarity at 1% significance level.

We estimate the following model

$$\begin{aligned} y_t &= \mu + \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha(y_{t-1} - \mu)^2 + \beta |x_{t-1}|^\rho \end{aligned} \quad \text{ARCH-NNH}$$

where  $y_t$  denotes the stock return series and  $x_t$  is the default premium (Baa - Aaa). We also estimate GARCH(1,1) and FIGARCH models for comparison.

$$\begin{aligned} \sigma_t^2 &= c + \alpha(y_{t-1} - \mu)^2 + \beta\sigma_{t-1}^2 && \text{GARCH}(1,1) \\ \sigma_t^2 &= c + \beta\sigma_{t-1}^2 + \left[1 - \beta L - (1 - \phi L)(1 - L)^d\right] (y_t - \mu)^2 && \text{FIGARCH}(1, d, 1) \\ \sigma_t^2 &= c + \beta\sigma_{t-1}^2 + \left[1 - \beta L - (1 - L)^d\right] (y_t - \mu)^2 && \text{FIGARCH}(1, d, 0) \end{aligned}$$

For the above three models, we use the quasi-maximum likelihood estimation method that is the standard way to estimate ARCH type models. Refer to Han and Park (2006) for the asymptotic distribution theory of the quasi-maximum likelihood estimator in ARCH-NNH models, which establishes the consistency and asymptotic mixed normality. The estimation results for the models are summarized and presented in Tables 2-4.

Table 2 shows the estimation results of the ARCH-NNH model with  $f(x) = \beta |x|^\rho$ .<sup>7</sup> For each frequency, we report estimation results of the model without any constraint on  $\rho$  as well as with constraint of  $\rho = 1$ . At first, we estimate the model without any constraint

---

<sup>7</sup>Notice that the coefficient of the lag of the squared return,  $\alpha$ , is compatible with the asymptotic theory because  $\alpha < 0.58$ . If we assume that  $(\varepsilon_t)$  is iid  $N(0,1)$ , then  $\kappa_\varepsilon^4 = 3$  and we need  $0 < \alpha < \frac{1}{\sqrt{3}}$  ( $= 0.577\dots$ ) in order to have  $0 < \alpha^2 \kappa_\varepsilon^4 < 1$ . See Theorem 2.1.

on  $\rho$  and all coefficients are tested to be significant except for  $\rho$  of the weekly frequency. The estimates of  $\rho$  for the monthly and daily frequencies are close to unity (1.05 and 1.10 respectively) and we are not able to reject the null hypothesis of  $\rho = 1$ . The estimation results of the model with constraint of  $\rho = 1$  are quite similar to those of the model without constraint, and every estimate of the model without constraint is statistically significant. Hence,  $f(x) = \beta|x|$  appears to be better in describing the volatility of the stock return series, and we are going to use this function for our simulation and forecasting evaluation.

Not only is the ARCH-NNH model statistically appropriate, but the model confirms economic thoughts. As expected, the default premium is positively related to stock return volatility, which is compatible with Schwert (1989). It is not surprising that stock market is more volatile when default risk is high.

The estimation results of GARCH(1,1) are reported in Table 3. Consistent with other empirical findings, the estimated  $\alpha + \beta$  (ARCH effect) is very close to unity, suggestive of the IGARCH behavior. FIGARCH<sup>8</sup> estimation results are given in Table 4. For each frequency, we report both FIGARCH(1,d,1) and FIGARCH(1,d,0) estimation results. For the monthly frequency, FIGARCH(1,d,1) is preferred because estimates of  $d$  and  $\beta$  are tested to be insignificant in FIGARCH(1,d,0). On the other hand, FIGARCH(1,d,0) is preferred for the weekly and daily frequencies because estimates of  $\phi$  are insignificant in FIGARCH(1,d,1). Hence, we are going to use FIGARCH(1,d,1) for the monthly frequency and FIGARCH(1,d,0) for the weekly and daily frequencies in our simulation and forecasting evaluation. The long-run dynamics are modeled by the fractional differencing parameter  $d$ , and it is estimated 0.67, 0.37 and 0.30 for the selected FIGARCH model at the monthly, weekly and daily frequencies respectively. The null hypothesis of  $d = 0.5$  is rejected in the daily case, but it is not rejected in the monthly and weekly cases. While the null hypothesis of  $d = 1$  is clearly rejected in the weekly and daily cases, it is rejected only at 10% significance level in the monthly case.

## 4.2 Exchange Rate Return Volatility

In their portfolio selection model of exchange rate determination, Hagiwara and Herce (1999) showed that the interest rate differential between countries (absolute value or squared) is related to exchange rate return volatility. We apply this finding to the ARCH-NNH model. Based on the model of Hagiwara and Herce (1999), we estimate the following models;

$$\begin{aligned} y_t &= b_0 + b_1 y_{t-1} + b_2 y_{t-2} + b_3 x_{t-1} + \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha (\sigma_{t-1} \varepsilon_{t-1})^2 + \beta |x_{t-1}|^\rho && \text{ARCH-NNH} \\ \sigma_t^2 &= c + \alpha (\sigma_{t-1} \varepsilon_{t-1})^2 + \beta \sigma_{t-1}^2 && \text{GARCH}(1, 1) \end{aligned}$$

where  $y_t$  denotes the exchange rate return series in percentage form,  $y_t = 100 \times (\ln P_t - \ln P_{t-1})$  with  $P_t = \text{UK pound/US dollar}$ , and  $x_t$  represents eurocurrency interest rate spreads

---

<sup>8</sup>The G@RCH by Laurent and Peters is used for FIGARCH estimation and forecast. We fixed the truncation lag at 1,000 in estimation.

between US and UK ( $x_t = r_{US,t} - r_{UK,t}$ ). The mean equation for  $y_t$  is exactly same as that of Hagiwara & Herce (1999). We obtained weekly observations of the exchange rate for UK pound and one-month eurocurrency interest rate for US and UK.<sup>9</sup> The sample period is from 21 October 1983 to 31 December 2004 (1107 observations).

Table 5 shows that unit root tests for the interest rate differential support the presence of a unit root. The estimated autoregressive coefficient is very close to unity in both cases. Phillips-Perron tests are unable to reject the null hypothesis of a unit root. KPSS tests reject the null hypothesis of stationarity at 1% significance level.

The estimation results are summarized and presented in Table 6. The estimation result of the ARCH-NNH model shows that it works well for the exchange rate return volatility. Most estimates are tested to be significant and the null hypothesis of  $\rho = 1$  is also rejected, which means a nonlinear function is clearly needed in this case. Roughly, the positivity of  $\hat{b}_3$  supports the uncovered interest parity. The absolute values of the interest rate differentials are positively related to exchange rate return volatility, which is compatible with Hagiwara & Herce (1999). On the other hand, GARCH(1,1) performs very poorly and most estimates are insignificant.

## 5 Forecast Evaluation

We evaluate out-of-sample volatility forecasts of three models in the previous stock return volatility application; ARCH-NNH, GARCH(1,1), and FIGARCH(1,d,q)<sup>10</sup>. A rolling forecast procedure is adapted; i.e., each forecast is based on the estimated parameters from the previous (970 monthly, 979 weekly and 3688 daily) observations. We obtained 44 monthly forecasts for the November 1999 to June 2003 out-of-sample period. And, we obtained 100 weekly forecasts for the 3 August 2001 to 27 June 2003 out-of-sample period and 250 daily forecasts for the 1 July 2002 to 30 June 2003 out-of-sample period.

To evaluate the accuracy of volatility forecasts, they have to be compared with actual volatility, which cannot be observed. It is common in practice to define actual volatility as squared observed returns, which for one-day ahead volatility is equal to

$$y_{T+1}^2 = \sigma_{T+1}^2 \varepsilon_{T+1}^2.$$

However, the squared error term  $\varepsilon_{T+1}^2$  will vary widely and this implies that only a relatively small part is attributable to  $\sigma_{T+1}^2$ . An alternative approach which addresses this problem has been suggested. Refer to Andersen *et al* (2003) for a theoretical underpinning for the use of ‘realized volatility’. They employ the theory of quadratic variation to show that realized volatility computed from high-frequency intraperiod returns is an unbiased and effectively error-free measure of return volatility.

---

<sup>9</sup>From the Datastream. The exchange rates are sampled on every Friday and the interest rates are sampled on every Wednesday.

<sup>10</sup>Again,  $q = 1$  for the monthly frequency and  $q = 0$  for the weekly and daily frequencies following the estimation results. We fixed the truncation lag at 900 in forecast.

As a proxy for actual volatility, we use ‘realized volatility’ instead of squared returns. The measure for monthly volatility is the sum of squared daily returns:

$$\bar{\sigma}_t^2 = \sum_{i=1}^{N_t} (y_{it} - \mu_t)^2,$$

where there are  $N_t$  daily returns  $y_{it}$  in month  $t$  and  $\mu_t$  is the average of daily returns  $y_{it}$  in the month. It turned out that subtracting the average has only negligible effect in the monthly case. Since the mean returns are smaller for higher frequencies, we assume that  $\mu = 0$  for the weekly and daily frequencies. Hence, the measure for weekly volatility is the sum of squared daily returns. The daily realized volatility is defined as the sum of the squared overnight, close-to-open, and the cumulative squared 30-minute intraday returns.

In order to assess predictive abilities of models, we use the regression-based method and the mean absolute error (MAE). As the first method, we report on  $R^2$  as calculated from the OLS regression

$$\bar{\sigma}_{T+1}^2 = a + b\hat{\sigma}_{T+1}^2 + \varepsilon, \quad (9)$$

where  $\hat{\sigma}_{T+1}^2$  denotes one-period ahead volatility forecasts. This appears to be the most commonly used method in the literature when measures for volatility are computed with ‘realized volatility’ as the sum of squared intraperiod returns. If volatility forecasts are unbiased, then  $a = 0$  and  $b = 1$ . We test these hypotheses using the standard regression method with adjustment followed by Andrews and Monahan (1992) to account for the error covariance. The quadratic spectral kernel with automatic bandwidth selection is used to obtain heteroskedasticity and autocorrelation consistent covariance estimates.

Table 7 presents forecasting performance results evaluated on the regression-based method. It shows that the ARCH-NNH model outperforms GARCH (1,1) and FIGARCH at the monthly and weekly frequencies. The null hypothesis of  $b = 1$  is not rejected in every case. The estimates for  $a$  are very small and close to zero, but the null hypothesis of  $a = 0$  is rejected in some cases. The comparison of forecasting ability between models is done by the  $R^2$  statistic. For the monthly frequency,  $R^2$  of the ARCH-NNH model is 0.12, which is higher than that of GARCH(1,1), 0.05, and FIGARCH, 0.07. The weekly case shows the similar result that the ARCH-NNH model performs better than the other models. For the daily frequency, the FIGARCH model has the highest  $R^2$ , 0.27. The result that  $R^2$ s in the daily case are bigger than those of other frequencies seems to be originated from the fact that the daily realized volatility is constructed by intraday returns. Comparison between GARCH(1,1) and FIGARCH shows that FIGARCH performs slightly better than GARCH(1,1) at every frequency.

We also evaluate the volatility forecasts on the basis of the mean absolute error (MAE) and test the null hypotheses of equal MAE by the Diebold and Mariano (1995) test procedure. The null maintains that the predictive performance of the best performing model relative to another model is not different. First, define loss differential between MAEs from two forecasting models

$$d_t = MAE_{1,t} - MAE_{2,t}.$$

So, the null hypothesis is  $\mathbb{E}(d_t) = 0$ . This hypothesis is evaluated using the statistic

$$D_1 = \frac{\bar{d}}{\sqrt{2\pi\hat{f}_d(0)/N}} \quad (10)$$

where  $\bar{d}$  is the sample mean loss differential  $\left(\bar{d} = \frac{1}{N} \sum_{t=1}^N d_t\right)$ ,  $\hat{f}_d(0)$  is a consistent estimate of the spectral density function of the loss differential at frequency 0, and  $N$  is the number of forecasts. To compute  $\hat{f}_d(0)$ , Diebold and Mariano (1995) suggest the uniform, or rectangular, lag window with bandwidth parameter  $k - 1$  for forecast horizon  $k$ . Since we deal with only one-step ahead forecasts, we have  $2\pi\hat{f}_d(0) = \frac{1}{N} \sum_{t=1}^N (d_t - \bar{d})^2$ . Under the null hypothesis,  $D_1 \sim N(0, 1)$  asymptotically.

However, since  $N$  is relatively small for the monthly case ( $N = 44$ ), we add another statistic  $D_2$  for the exact finite-sample test of the monthly case, which is also developed by Diebold and Mariano (1995). The null hypothesis is a zero median loss differential,  $med(d_t) = 0$ , and this hypothesis is evaluated using the statistic

$$D_2 = \sum_{t=1}^N I_+(d_t) \quad (11)$$

where

$$\begin{aligned} I_+(d_t) &= 1 && \text{if } d_t > 0 \\ &= 0 && \text{otherwise.} \end{aligned}$$

Significance is assessed using a table of the cumulative binomial distribution with parameters  $N$  and  $\frac{1}{2}$  under the null hypothesis.

Forecasting evaluation based on MAE is reported in Table 8. The lowest MAEs come from GARCH(1,1) for the monthly frequency, from the ARCH-NNH model for the weekly frequency, and from FIGARCH for the daily frequency. It is hard to select the best performing model from the evaluation result based on MAE even after we consider the forecast accuracy test. The forecast accuracy tests using  $D_1$  and  $D_2$  show the same result in the monthly case. For the monthly frequency, the forecast accuracy tests show that GARCH(1,1) performs significantly better than the ARCH-NNH model, but its performance is not significantly different from that of FIGARCH. For the weekly frequency, the ARCH-NNH model has the lowest MAE, but there is no significant difference in forecasting ability between models according to the forecast accuracy test. At the daily frequency, the forecast of the FIGARCH model is significantly better than that of GARCH(1,1), but it is not significantly different from that of the ARCH-NNH model.

While the forecast evaluation based on MAE does not give a certain rank on forecasting ability, the regression-based method shows that the ARCH-NNH model performs better than GARCH(1,1) and FIGARCH in the weekly and monthly frequencies. The fact that our model outperforms GARCH(1,1) and FIGARCH at relatively lower frequencies may come from the advantage that our model is structural.

## 6 Concluding Remarks

As explained in the introduction, econometricians have been using ARCH type models with persistent covariates to model the volatility of economic and financial time series. While the effect of persistent covariates on various characteristics of volatility has been ignored, this paper fills this gap and gives theoretic understandings about ARCH models with persistent covariates. We provide the asymptotic theories showing how persistent covariates in ARCH models would affect various characteristics of volatility.

There are two more stylized facts about volatility in financial time series which we did not consider in this paper. One is the leverage effect in stock return series, and the ARCH-NNH model can easily deal with the issue in the same way as Glosten *et al* (1993) used a dummy variable. The other is the co-movement in volatility. If we look at exchange rate returns for different currencies, we observe big movements in one currency being matched by big movements in another. This suggests the importance of multivariate models in modelling cross-correlations in different markets. Multivariate or panel models using ARCH-NNH processes could provide a proper way to explain volatility spillover in foreign exchange markets. This task awaits further research.

The forecast evaluation in this paper shows that the ARCH-NNH model outperforms other standard models at relatively lower frequencies. Hence, if we apply an ARCH-NNH model in a price determination model, it could produce better forecasts of a price level such as the exchange rate or stock return. One of the important issues in finance is whether stock returns are predictable by macroeconomic variables. Much of the literature pertaining to this issue uses a linear function in mean equation of stock return. Nonlinear functions of persistent macroeconomic variables can be used to investigate both the mean and volatility equations of asset price returns. These possibilities will be addressed in future research.

## References

- Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), "Modeling and forecasting realized volatility," *Econometrica*, 71, 529-626.
- Andrews, D.W.K. and Monahan, J. (1992), "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953-966.
- Baillie, R.T., Bollerslev, T., and Mikkelsen, H.O. (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74, 3-30.
- Bartlett, M.S. (1946), "On the theoretical specification of sampling properties of autocorrelated time series," *Journal of the Royal Statistical Society*, B 8, 24-41.
- Berkes, L., Horvath, L., and Kokoszka, P. (2003): "GARCH processes: structure and estimation," *Bernoulli*, 9, 201-207.

- Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T., Engle, R.F., and Nelson, D.B. (1994): "ARCH models," In R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, 2959-3038, Elsevier: Amsterdam.
- Bollerslev, T. and Wooldridge, J.M. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric reviews*, 11, 143-172.
- Chung, H. and Park, J.Y. (2004), "Nonstationary nonlinear heteroskedasticity in regression," mimeograph, Department of Economics, Rice University.
- 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.
- Diebold, F.X. and Mariano, R.S. (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.
- Diebold, F.X. and Inoue, A. (2001), "Long memory and regime switching," *Journal of Econometrics*, 105, 131-159.
- Engle, R.F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. Inflation," *Econometrica*, 50, 987-1008.
- 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 Patton, A.J. (2001), "What good is a volatility model?," *Quantitative Finance*, 1(2), 237-245.
- Glosten, L.R., Jagannathan, R. and Runkle, D. (1993), "On the relation between the expected value and the volatility of nominal excess returns on stocks," *Journal of Finance*, 48, 1779-1801.
- 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.
- Gray, S.F. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics*, 42, 27-62.

- He, C. and Teräsvirta, T. (1999), "Properties of moments of a family of GARCH processes," *Journal of Econometrics*, 92, 173-192.
- Hagiwara, M. and Herce, M.A. (1999), "Endogenous exchange rate volatility, trading volume and interest rate differentials in a model of portfolio selection," *Review of International Economics*, 7(2), 202-218.
- Han, H. and Park, J.Y. (2006), "ARCH models with persistent covariates: asymptotic distribution theory of QMLE and IGARCH behavior," mimeograph, Department of Economics, National University of Singapore.
- Hillebrand, E. (2005), "Neglecting parameter changes in GARCH models," *Journal of Econometrics*, 129, 121-138.
- Hodrick R.J. (1989), "Risk, uncertainty, and exchange rates," *Journal of Monetary Economics*, 23, 433-459.
- Hol, E.M.J.H. (2003), *Empirical studies on volatility in international stock markets*. Dordrecht: Kluwer Academic.
- Hyung, N., Poon, S.H. and Granger, C.W.J. (2006), "A source of long memory in volatility," mimeograph, Manchester Business School, University of Manchester.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, 54, 159-178.
- Lamoureux, C.G. and Lastrapes, W.D. (1990), "Heteroskedasticity in stock return data: volume versus GARCH effects," *Journal of Finance*, 45, 221-229.
- Laurent, S. and Peters, J.P. (2005), "G@RCH 4.0, Estimating and forecasting ARCH models," Timberlake Consultants.
- Lee, S.W. and Hansen, B.E. (1994), "Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator," *Econometric Theory*, 10, 29-52.
- Lumsdaine, R.L. (1996), "Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models," *Econometrica*, 64, 575-596.
- 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.
- Miller, J.I. and Park, J.Y. (2005), "Nonlinearity, nonstationarity, and thick tails: how they interact to generate persistence in memory," mimeograph, Department of Economics, Rice University.



- Nelson, D.B. (1990), "Stationarity and persistence in the GARCH(1,1,) model," *Econometric Theory*, 6, 318-334.
- Park, J.Y. (2002), "Nonstationary nonlinear heteroskedasticity," *Journal of Econometrics*, 110, 383-415.
- Park, J.Y. (2003), "Weak unit root," mimeograph, Department of Economics, Rice University.
- 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.
- Phillips, P.C.B. (1987), "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535-547.
- Phillips, P.C.B. and Perron, P. (1988), "Testing for a unit root in time series regression," *Biometrika*, 75, 335-346.
- Poon, S.H. and Granger, C.W.J. (2003), "Forecasting volatility in financial markets," *Journal of Economic Literature*, 41, 478-539.
- Robinson, P.M. and Zaffaroni, P. (2005), "Pseudo-maximum likelihood estimation of ARCH( $\infty$ ) models," *Annals of Statistics*, forthcoming.
- Schwert, G.W. (1989), "Why does stock market volatility change over time?," *Journal of Finance*, 44, 1115-1154.
- Stock, J.H. (1994): "Unit roots and structural breaks," In R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, 2739-2841, Elsevier: Amsterdam.
- Weiss, A. (1986), "Asymptotic theory for ARCH models," *Econometric Theory*, 2, 107-131.

## Appendix A. Useful lemmas and their proofs

The proofs of the theorems in the paper rely on the results from the following lemmas. For the lemmas and their proofs, we let Assumptions 1 and 2 hold.

**Lemma 1** *Let  $T$  be a transformation on  $\mathbb{R}$ . Define*

$$M_{1n} = \sum_{t=1}^n T(x_t) \quad \text{and} \quad M_{2n} = \sum_{t=1}^n T^2(x_t).$$

(a) *If  $T \in \mathbb{I}$ , we have under Assumption 4S(b)*

$$\begin{aligned} n^{-1/2} M_{1n} &\rightarrow_d L_c(1, 0) \int_{-\infty}^{\infty} T(x) dx \\ n^{-1/2} M_{2n} &\rightarrow_d L_c(1, 0) \int_{-\infty}^{\infty} T^2(x) dx \end{aligned}$$

as  $n \rightarrow \infty$ .

(b) *If  $T \in \mathbb{H}$  with asymptotic order  $\kappa$  and limit homogeneous function  $\bar{f}$  and if we let  $\kappa_n = \kappa(\sqrt{n})$ , we have under Assumption 4W(b)*

$$\begin{aligned} (n\kappa_n)^{-1} M_{1n} &\rightarrow_d \int_0^1 \bar{f}(V_c(r)) dr \\ (n\kappa_n^2)^{-1} M_{2n} &\rightarrow_d \int_0^1 \bar{f}^2(V_c(r)) dr \end{aligned}$$

as  $n \rightarrow \infty$ .

The weak convergences in (a) and (b) hold jointly.

**Proof of Lemma 1.** See Park (2002, 2003). The classes of  $\mathbb{I}$  and  $\mathbb{H}$  transformations are closed under the product operation. If  $T \in \mathbb{I}$ , then  $T^2 \in \mathbb{I}$ . And if  $T \in \mathbb{H}$  with limit homogeneous function  $\bar{f}$ , then  $T^2 \in \mathbb{H}$  with limit homogeneous function  $\bar{f}^2$ . If  $T^2 \in \mathbb{I}$ , then

$$n^{-1/2} M_{2n} = \sqrt{n} \int_0^1 T^2(\sqrt{n}V_{cn}(r)) dr.$$

If  $T^2 \in \mathbb{H}$  with limit homogeneous function  $\bar{f}^2$  and asymptotic order  $\kappa_n^2$ , then

$$(n\kappa_n^2)^{-1} M_{2n} = \frac{1}{n} \sum_{t=1}^n \frac{T^2(x_t)}{\kappa_n^2} \approx \frac{1}{n} \sum_{t=1}^n \bar{f}^2\left(\frac{1}{\sqrt{n}}x_t\right).$$

□

**Lemma 2** *Let  $T$  be a transformation on  $\mathbb{R}$ , and let  $u_t$  be a martingale difference sequence (MDS) with respect to  $F_t$  such that  $\mathbb{E}(u_t^2 | F_{t-1}) = \sigma_u^2$  a.s. for each  $t$  and  $\sup_t \mathbb{E}(|u_t|^{2+\epsilon} | F_{t-1}) < \infty$  a.s. for some  $\epsilon > 0$ . Define*

$$U_{1n} = \sum_{t=1}^n T(x_t) u_t \quad \text{and} \quad U_{2n} = \sum_{t=1}^n T^2(x_t) u_t.$$

- (a) If  $T \in \mathbb{I}$ , then  $U_{1n}, U_{2n} = O_p(n^{1/4})$  respectively under Assumption 4S and 5S.  
(b) If  $T \in \mathbb{H}$  with asymptotic order  $\kappa$ , and if we let  $\kappa_n = \kappa(\sqrt{n})$ , then  $U_{1n} = O_p(n^{1/2}\kappa_n)$  and  $U_{2n} = O_p(n^{1/2}\kappa_n^2)$  respectively under Assumption 4W and 5W.

**Proof of Lemma 2.** See Park (2003). □

For example, let us consider asymptotic limits of  $\sum_{t=1}^n T(x_t)\varepsilon_t^2$  and  $\sum_{t=1}^n T(x_t)\varepsilon_t^4$ . Notice that  $\varepsilon_t^2 - 1$  and  $\varepsilon_t^4 - \kappa_\varepsilon^4$  are MDSs. If  $T \in \mathbb{I}$ , then

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t)\varepsilon_t^2 &= \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t) (\varepsilon_t^2 - 1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t) + \frac{1}{\sqrt{n}} O_p(n^{1/4}) \rightarrow_d L_c(1, 0) \int_{-\infty}^{\infty} T(x) dx \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t)\varepsilon_t^4 &= \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t)\kappa_\varepsilon^4 + \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t) (\varepsilon_t^4 - \kappa_\varepsilon^4) \\ &= \kappa_\varepsilon^4 \frac{1}{\sqrt{n}} \sum_{t=1}^n T(x_t) + \frac{1}{\sqrt{n}} O_p(n^{1/4}) \rightarrow_d \kappa_\varepsilon^4 L_c(1, 0) \int_{-\infty}^{\infty} T(x) dx. \end{aligned}$$

**Lemma 3** Let  $T$  be a transformation on  $\mathbb{R}$  and denote  $p_k$  the density of  $(v_{kt})$  with respect to measure  $m$  on  $\mathbb{R}$ . Define

$$B_n = \sum_{t=k+1}^n T(x_t)T(x_{t-k}).$$

(a) If  $T \in \mathbb{I}$ , we have under Assumption 3S

$$n^{-1/2} B_n \rightarrow_d L_c(1, 0) \int_{-\infty}^{\infty} \mu_k T(x) dx$$

as  $n \rightarrow \infty$ , where  $\mu_k = \int_{-\infty}^{\infty} T(x+y)p_k(y)m(dy)$ .

(b) If  $T \in \mathbb{H}$  with asymptotic order  $\kappa$  and limit homogeneous function  $\bar{f}$  and if we let  $\kappa_n = \kappa(\sqrt{n})$ , we have under Assumption 3W

$$(n\kappa_n^2)^{-1} B_n \rightarrow_d \int_0^1 \bar{f}^2(V_c(r)) dr$$

as  $n \rightarrow \infty$ .

**Proof of Lemma 3.**

$$\begin{aligned} x_t &= \left(1 - \frac{c}{n}\right)^k x_{t-k} + \sum_{i=0}^{k-1} \left(1 - \frac{c}{n}\right)^i v_{t-i} \\ &= x_{t-k} + \sum_{i=0}^{k-1} v_{t-i} + q_1\left(\frac{c}{n}, x_{t-k}\right) + q_2\left(\frac{c}{n}, v_{t-1}, v_{t-2}, \dots, v_{t-k+1}\right) \end{aligned}$$

where

$$q_1\left(\frac{c}{n}, x_{t-k}\right) = \left[\left(1 - \frac{c}{n}\right)^k - 1\right] x_{t-k}$$

and

$$q_2\left(\frac{c}{n}, v_{t-1}, v_{t-2}, \dots, v_{t-k+1}\right) = \sum_{i=1}^{k-1} \left[\left(1 - \frac{c}{n}\right)^i - 1\right] v_{t-i}.$$

Let  $v_{kt} = \sum_{i=0}^{k-1} v_{t-i}$ . Notice that  $q_1, q_2 \rightarrow 0$  as  $n \rightarrow \infty$ . We have

$$\begin{aligned} \sum_{t=k+1}^n T(x_t)T(x_{t-k}) &= \sum_{t=k+1}^n T(x_{t-k} + v_{kt} + q_1 + q_2)T(x_{t-k}) \\ &= \sum_{t=k+1}^n T(x_{t-k} + v_{kt})T(x_{t-k}) + \sum_{t=k+1}^n D_n T(x_{t-k}) \end{aligned}$$

where  $D_n = T(x_{t-k} + v_{kt} + q_1 + q_2) - T(x_{t-k} + v_{kt})$ .

If  $T \in \mathbb{L}$ , then  $D_n \in \mathbb{L}$  and  $D_n T \in \mathbb{L}$ . Since  $|D_n T(x)| \leq |D_1 T(x)|$  for all  $x \in \mathbb{R}$ , we apply Lebesgue's Dominated Convergence Theorem and, slightly abusing notation, obtain

$$\begin{aligned} n^{-1/2} \sum_{t=k+1}^n D_n T(x_{t-k}) &\rightarrow_d \lim_{n \rightarrow \infty} L_c(1, 0) \int_{-\infty}^{\infty} D_n T(x) dx \\ &= L_c(1, 0) \int_{-\infty}^{\infty} \lim_{n \rightarrow \infty} D_n T(x) dx = 0. \end{aligned}$$

Thus, we have

$$\begin{aligned} &n^{-1/2} \sum_{t=k+1}^n T(x_t)T(x_{t-k}) \\ &= n^{-1/2} \sum_{t=k+1}^n T(x_{t-k} + v_{kt})T(x_{t-k}) + o_p(1) \\ &\rightarrow_d L_c(1, 0) \int_{-\infty}^{\infty} \mu_k T(x) dx. \end{aligned}$$

See the proof of Theorem 1 of Park (2002) for the third line.

If  $T \in \mathbb{H}$ , then we have  $D_n \in \mathbb{H}$  where its asymptotic order is  $\kappa$  and its limit homogeneous function is zero. Hence, we have  $D_n T(x) \in \mathbb{H}$  where its asymptotic order is  $\kappa$  and its limit homogeneous function is zero. Thus, we have

$$(n\kappa_n)^{-1} \sum_{t=k+1}^n D_n T(x_{t-k}) \rightarrow_p 0.$$

Therefore,

$$\begin{aligned} & (n\kappa_n^2)^{-1} \sum_{t=k+1}^n T(x_t)T(x_{t-k}) \\ &= (n\kappa_n^2)^{-1} \sum_{t=k+1}^n T(x_{t-k} + v_{kt})T(x_{t-k}) + o_p(1) \\ &\rightarrow_d \int_0^1 \bar{f}^2(V_c(r))dr. \end{aligned}$$

See the proof of Theorem 1 of Park (2002) for the third line.  $\square$

**Lemma 4** Suppose that  $0 < \alpha < 1$  and  $v_n \rightarrow \infty$  monotonically as  $n \rightarrow \infty$ . If  $\frac{1}{v_n} \sum_{k=1}^n f(x_t) \rightarrow_d Q$ , then, as  $n \rightarrow \infty$ ,

$$\frac{1}{v_n} \sum_{k=0}^{n-1} \alpha^k \sum_{t=1}^{n-k} f(x_t) \rightarrow_d \frac{1}{1-\alpha} Q.$$

**Proof of Lemma 4.** Let  $C_n = \frac{1}{v_n} \sum_{k=1}^n f(x_t) - Q$ . Then  $C_n \rightarrow_d 0$  and  $\sum_{k=1}^n f(x_t) = v_n(Q + C_n)$ .

$$\begin{aligned} \frac{1}{v_n} \sum_{k=0}^{n-1} \alpha^k \sum_{t=1}^{n-k} f(x_t) &= \frac{1}{v_n} \sum_{k=0}^{n-1} \alpha^k [v_{n-k}Q + v_{n-k}C_{n-k}] \\ &= Q \frac{1}{v_n} S + \frac{1}{v_n} T \end{aligned}$$

where  $S = \sum_{k=0}^{n-1} \alpha^k v_{n-k}$  and  $T = \sum_{k=0}^{n-1} \alpha^k v_{n-k} C_{n-k}$ .

At first, we are going to show that  $\frac{1}{v_n} S \rightarrow 0$ . Since

$$\alpha S - S = -v_n + \sum_{i=0}^{n-2} \alpha^{1+i} (v_{n-i} - v_{n-1-i}) + \alpha^n v_1,$$

we have

$$\begin{aligned} \frac{1}{v_n} S &= \frac{1}{1-\alpha} + \frac{1}{v_n} \frac{1}{\alpha-1} \left[ \sum_{i=0}^{n-2} \alpha^{1+i} (v_{n-i} - v_{n-1-i}) + \alpha^n v_1 \right] \\ &\rightarrow \frac{1}{1-\alpha}. \end{aligned}$$

The last line follows because

$$\begin{aligned} & \frac{1}{v_n} \left[ \sum_{i=0}^{n-2} \alpha^{1+i} (v_{n-i} - v_{n-1-i}) + \alpha^n v_1 \right] \\ & \leq \frac{1}{v_n} V^* [\alpha + \alpha^2 + \dots + \alpha^n] = \frac{1}{v_n} V^* \frac{\alpha(1 - \alpha^n)}{1 - \alpha} \rightarrow 0 \end{aligned}$$

where  $V^* = \max_t (v_t - v_{t-1}, v_1)$  for  $2 \leq t \leq n$ .

Now, We need to show that  $\frac{1}{v_n} T \rightarrow_p 0$ . Recall that  $\rightarrow_d$  and  $\rightarrow_p$  are identical when the convergence is to a nonrandom limit. Therefore,  $C_n \rightarrow_d 0$  implies  $C_n \rightarrow_p 0$ . The stated result follows because

$$\begin{aligned} \left| \frac{1}{v_n} T \right| & \leq \sum_{k=0}^{n-1} \alpha^k \frac{v_{n-k}}{v_n} |C_{n-k}| \\ & < \sum_{k=0}^{n-1} \alpha^k |C_{n-k}| \rightarrow_p 0. \end{aligned}$$

Here, the convergence to zero follows because  $C_n \rightarrow_p 0$  and  $0 < \alpha < 1$ . This completes the proof.  $\square$

## Appendix B. Proofs of the main results

**Proof of Theorem 1.** Suppose that  $v_n = \sqrt{n}$  for  $f \in \mathbb{I}$  and  $v_n = n\kappa^2(\sqrt{n})$  for  $f \in \mathbb{H}$ . At first, we need to obtain asymptotic limits for the following three sample moments:

$$\sum y_t^2, \quad \sum y_t^4, \quad \sum y_t^2 y_{t-k}^2.$$

The first sample moment is

$$\begin{aligned} \sum_{t=1}^n y_t^2 & = \sum_{t=1}^n \sigma_t^2 \varepsilon_t^2 = \sum_{t=1}^n (f(x_t) + \alpha y_{t-1}^2) \varepsilon_t^2 \\ & = \sum_{t=1}^n \sum_{j=0}^{n-1} \alpha^j f(x_{t-j}) \prod_{h=0}^j \varepsilon_{t-h}^2 \\ & = \sum_{j=0}^{n-1} \alpha^j \sum_{t=1}^n f(x_{t-j}) + \sum_{j=0}^{n-1} \alpha^j \sum_{t=1}^n f(x_{t-j}) \left( \prod_{h=0}^j \varepsilon_{t-h}^2 - 1 \right) \end{aligned}$$

Since  $\left( \prod_{h=0}^j \varepsilon_{t-h}^2 - 1 \right)$  for  $j = 0, 1, \dots, n-1$  are MDSs,

$$\frac{1}{v_n} \sum_{j=0}^{n-1} \alpha^j \sum_{t=1}^n f(x_{t-j}) \left( \prod_{h=0}^j \varepsilon_{t-h}^2 - 1 \right) = o_p(1) \quad (\text{A1})$$

by lemma 2 and Lemma 4. Therefore,

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{t=1}^n y_t^2 &= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \alpha^j \sum_{t=1}^n f(x_{t-j}) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \alpha^j \sum_{t=1}^{n-j} f(x_t) + o_p(1).\end{aligned}$$

Applying Lemmas 1 and 4 gives us

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n y_t^2 \rightarrow_d \frac{1}{1-\alpha} L_c(1, 0) \int_{-\infty}^{\infty} f(s) ds.$$

Similarly, for  $f \in \mathbb{H}$ , we have

$$\begin{aligned}\frac{\kappa(\sqrt{n})^{-1}}{n} \sum_{t=1}^n y_t^2 &= \frac{\kappa(\sqrt{n})^{-1}}{n} \sum_{j=0}^{n-1} \alpha^j \sum_{t=1}^{n-j} f(x_t) + o_p(1) \\ &\rightarrow_d \frac{1}{1-\alpha} \int_0^1 \bar{f}(V_c(r)) dr.\end{aligned}$$

The second sample moment is

$$\begin{aligned}\sum_{t=1}^n y_t^4 &= \sum_{t=1}^n f(x_t)^2 \varepsilon_t^4 + 2\alpha \sum_{t=1}^n f(x_t) y_{t-1}^2 \varepsilon_t^4 + \alpha^2 \sum_{t=1}^n y_{t-1}^4 \varepsilon_t^4 \\ &= \sum_{t=1}^n f(x_t)^2 \varepsilon_t^4 + 2 \sum_{i=1}^{n-1} \alpha^i \sum_{t=1+i}^n f(x_t) f(x_{t-i}) \varepsilon_t^4 \prod_{h=1}^i \varepsilon_{t-h}^2 \\ &\quad + \alpha^2 \sum_{t=2}^n y_{t-1}^4 \varepsilon_t^4 \\ &= \left[ \kappa_\varepsilon^4 \sum_{t=1}^n f(x_t)^2 + 2\kappa_\varepsilon^4 \sum_{i=1}^{n-1} \alpha^i \sum_{t=1+i}^n f(x_t) f(x_{t-i}) \right] + \alpha^2 \sum_{t=2}^n y_{t-1}^4 \varepsilon_t^4 \\ &\quad + \left[ \begin{aligned} &\sum_{t=1}^n f(x_t)^2 (\varepsilon_t^4 - \kappa_\varepsilon^4) \\ &+ 2 \sum_{i=1}^{n-1} \alpha^i \sum_{t=1+i}^n f(x_t) f(x_{t-i}) \left( \varepsilon_t^4 \prod_{h=1}^i \varepsilon_{t-h}^2 - \kappa_\varepsilon^4 \right) \end{aligned} \right].\end{aligned}$$

Since  $\varepsilon_t^4 \prod_{h=1}^i \varepsilon_{t-h}^2 - \kappa_\varepsilon^4$  are MDSs, similarly as the equation (A1), we have

$$\frac{1}{v_n} \left[ \sum_{t=1}^n f(x_t)^2 (\varepsilon_t^4 - \kappa_\varepsilon^4) + 2 \sum_{i=1}^{n-1} \alpha^i \sum_{t=1+i}^n f(x_t) f(x_{t-i}) \left( \varepsilon_t^4 \prod_{h=1}^i \varepsilon_{t-h}^2 - \kappa_\varepsilon^4 \right) \right] = o_p(1).$$

Therefore, we have

$$\frac{1}{v_n} \sum_{t=1}^n y_t^4 = \frac{\kappa_\varepsilon^4}{v_n} \left[ \sum_{t=1}^n f(x_t)^2 + 2 \sum_{i=1}^{n-1} \alpha^i \sum_{t=1+i}^n f(x_t) f(x_{t-i}) \right] + \frac{\alpha^2}{v_n} \sum_{t=2}^n y_{t-1}^4 \varepsilon_t^4 + o_p(1).$$

Expanding  $\frac{\alpha^2}{v_n} \sum_{t=2}^n y_{t-1}^4 \varepsilon_t^4$  in the same way, we have

$$\begin{aligned} \frac{1}{v_n} \sum_{t=1}^n y_t^4 &= \frac{\kappa_\varepsilon^4}{v_n} \sum_{j=0}^{n-1} (\alpha^2 \kappa_\varepsilon^4)^j \sum_{t=1}^{n-j} f(x_t)^2 \\ &\quad + 2 \sum_{i=1}^{n-1} \alpha^i \frac{\kappa_\varepsilon^4}{v_n} \sum_{j=0}^{n-i-1} (\alpha^2 \kappa_\varepsilon^4)^j \sum_{t=1+i}^{n-j} f(x_t) f(x_{t-i}) + o_p(1). \end{aligned}$$

If  $f \in \mathbb{I}$ , applying Lemmas 1, 3, and 4 gives us

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-i-1} (\alpha^2 \kappa_\varepsilon^4)^j \sum_{t=1+i}^{n-j} f(x_t) f(x_{t-i}) \rightarrow_d \frac{1}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1, 0) \int_{-\infty}^{\infty} \mu_i f(s) ds.$$

Thus,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n y_t^4 \rightarrow_d \frac{2\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1, 0) \sum_{i=1}^{\infty} \alpha^i \int_{-\infty}^{\infty} \mu_i f(s) ds + \frac{\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1, 0) \int_{-\infty}^{\infty} f^2(s) ds.$$

If  $f \in \mathbb{H}$ , then

$$\frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{j=0}^{n-i-1} (\alpha^2 \kappa_\varepsilon^4)^j \sum_{t=i+1}^{n-j} f(x_t) f(x_{t-i}) \rightarrow_d \frac{1}{1 - \alpha^2 \kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V_c(r)) dr$$

similarly by Lemmas 1, 3, and 4. Then, applying Lemma 4 again gives us

$$\begin{aligned} &\frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=1}^n y_t^4 \\ &\rightarrow_d \frac{2a}{1 - \alpha} \frac{\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V_c(r)) dr + \frac{\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V_c(r)) dr \\ &= \frac{1 + \alpha}{1 - \alpha} \frac{\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V_c(r)) dr. \end{aligned}$$



The third sample moment is

$$\begin{aligned}
& \sum_{t=k+1}^n y_t^2 y_{t-k}^2 \\
&= \sum_{t=k+1}^n \left[ \left( \sum_{i=0}^{n-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \sum_{j=0}^{n-k-1} \alpha^j f(x_{t-k-j}) \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) \right] \\
&= \sum_{t=k+1}^n \left[ \left( \sum_{i=0}^{k-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \sum_{j=0}^{n-k-1} \alpha^j f(x_{t-k-j}) \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) \right] \\
&+ \sum_{t=k+1}^n \left[ \left( \sum_{i=k}^{n-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \sum_{j=0}^{n-k-1} \alpha^j f(x_{t-k-j}) \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) \right].
\end{aligned}$$

For the third line, we divide  $y_t^2 = \sum_{i=0}^{n-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2$  into two parts,

$$\sum_{i=0}^{k-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \quad \text{and} \quad \sum_{i=k}^{n-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2,$$

because each term produces different types of MDSs if it is multiplied by  $y_{t-k}^2$ .

At first, since  $0 \leq j \leq n-k-1$ , if  $0 \leq i \leq k-1$  then

$$\mathbb{E} \left( \prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) = 1.$$

Hence, we have

$$\begin{aligned}
& \left( \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \alpha^j f(x_{t-k-j}) \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) \\
&= \alpha^{i+j} f(x_{t-i}) f(x_{t-k-j}) + \alpha^{i+j} f(x_{t-i}) f(x_{t-k-j}) \left[ \prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 - 1 \right]
\end{aligned}$$

where  $\prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 - 1$  is a MDS. Like the equation (A1),

$$\frac{1}{v_n} \sum_{t=k+1}^n \alpha^{i+j} f(x_{t-i}) f(x_{t-k-j}) \left[ \prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 - 1 \right] = o_p(1).$$

Secondly, for  $k \leq i \leq n-1$ ,

$$\mathbb{E} \left( \prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) = (\kappa_\varepsilon^4)^{\min(i-k+1, j+1)}.$$

Therefore,

$$\begin{aligned}
& \left( \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \alpha^j f(x_{t-k-j}) \prod_{h=0}^j \varepsilon_{t-k-h}^2 \right) \\
= & (\kappa_\varepsilon^4)^{\min(i-k+1, j+1)} \alpha^{i+j} f(x_{t-i}) f(x_{t-k-j}) \\
& + \alpha^{i+j} f(x_{t-i}) f(x_{t-k-j}) \left[ \prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 - (\kappa_\varepsilon^4)^{\min(i-k+1, j+1)} \right]
\end{aligned}$$

where  $\prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 - (\kappa_\varepsilon^4)^{\min(i-k+1, j+1)}$  is a MDS. Like the equation (A1),

$$\frac{1}{v_n} \sum_{t=k+1}^n \alpha^{i+j} f(x_{t-i}) f(x_{t-k-j}) \left[ \prod_{h=0}^i \varepsilon_{t-h}^2 \prod_{h=0}^j \varepsilon_{t-k-h}^2 - 1 \right] = o_p(1).$$

Hence, the first part of  $\sum_{t=k+1}^n y_t^2 y_{t-k}^2$  can be written as

$$\begin{aligned}
& \frac{1}{v_n} \sum_{t=k+1}^n \left[ \left( \sum_{i=0}^{k-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \sum_{i=0}^{n-k-1} \alpha^i f(x_{t-k-i}) \prod_{h=0}^i \varepsilon_{t-k-h}^2 \right) \right] \\
= & \frac{1}{v_n} \sum_{t=k+1}^n \left[ \left( \sum_{i=0}^{k-1} \alpha^i f(x_{t-i}) \right) \times \left( \sum_{i=0}^{n-k-1} \alpha^i f(x_{t-k-i}) \right) \right] + o_p(1) \\
= & \sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{n-k-1} \alpha^i \frac{1}{v_n} \sum_{t=k+1}^n f(x_{t-j}) f(x_{t-k-i}) + o_p(1)
\end{aligned}$$

and the second part is

$$\begin{aligned}
& \frac{1}{v_n} \sum_{t=k+1}^n \left[ \left( \sum_{i=k}^{n-1} \alpha^i f(x_{t-i}) \prod_{h=0}^i \varepsilon_{t-h}^2 \right) \times \left( \sum_{i=0}^{n-k-1} \alpha^i f(x_{t-k-i}) \prod_{h=0}^i \varepsilon_{t-k-h}^2 \right) \right] \\
= & 2\alpha^k \kappa_\varepsilon^4 \sum_{i=1}^{n-k-1} \alpha^i \sum_{j=0}^{n-k-i-1} (\alpha^2 \kappa_\varepsilon^4)^j \frac{1}{v_n} \sum_{t=k+1}^n f(x_{t-k-j}) f(x_{t-k-j-i}) \\
& + \alpha^k \kappa_\varepsilon^4 \sum_{j=0}^{n-k-1} (\alpha^2 \kappa_\varepsilon^4)^j \frac{1}{v_n} \sum_{t=k+1}^n f(x_{t-k-j})^2 + o_p(1).
\end{aligned}$$

Combining these two terms, we obtain

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=k+1}^n y_t^2 y_{t-k}^2 \\
\rightarrow_d & \sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{\infty} \alpha^i L_c(1,0) \int_{-\infty}^{\infty} \mu_{k+i-j} f(s) ds \\
& + \frac{2\alpha^k \kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1,0) \sum_{i=1}^{\infty} \alpha^i \int_{-\infty}^{\infty} \mu_i f(s) ds + \frac{\alpha^k \kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1,0) \int_{-\infty}^{\infty} f^2(s) ds
\end{aligned}$$

for  $f \in \mathbb{I}$  in the same way we did for  $\sum_{t=1}^n y_t^4$ , where  $\int_{-\infty}^{\infty} \mu_0 f(s) ds = \int_{-\infty}^{\infty} f^2(s) ds$ .

If  $f \in \mathbb{H}$ , then we have

$$\begin{aligned}
& \frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=k+1}^n y_t^2 y_{t-k}^2 \\
\rightarrow_d & \frac{1 - \alpha^k}{(1 - \alpha)^2} \int_0^1 \bar{f}^2(V_c(r)) dr + \frac{1 + \alpha}{1 - \alpha} \frac{\alpha^k \kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V_c(r)) dr \\
= & \left[ \frac{1 - \alpha^k}{(1 - \alpha)} + \alpha^k \frac{(1 + \alpha) \kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} \right] \frac{1}{1 - \alpha} \int_0^1 \bar{f}^2(V_c(r)) dr.
\end{aligned}$$

For the result for  $f \in \mathbb{I}$  in part (a), we first note that  $\bar{y}_n^2 = \frac{1}{n} \sum_{t=1}^n y_t^2 = O_p(n^{-1/2})$  which follows easily from the proof for  $\sum_{t=1}^n y_t^2$ . Therefore,

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=k+1}^n (y_t^2 - \bar{y}_n^2) (y_{t-k}^2 - \bar{y}_n^2) = \frac{1}{\sqrt{n}} \sum_{t=k+1}^n y_t^2 y_{t-k}^2 + O_p(n^{-1/2}) \\
\rightarrow_d & \sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{\infty} \alpha^i L_c(1,0) \int_{-\infty}^{\infty} \mu_{k+i-j} f(s) ds \\
& + \frac{2\alpha^k \kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1,0) \sum_{i=1}^{\infty} \alpha^i \int_{-\infty}^{\infty} \mu_i f(s) ds + \frac{\alpha^k \kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1,0) \int_{-\infty}^{\infty} f^2(s) ds.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{t=1}^n (y_t^2 - \bar{y}_n^2)^2 = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t^4 + O_p(n^{-1/2}) \\
\rightarrow_d & \frac{2\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1,0) \sum_{i=1}^{\infty} \alpha^i \int_{-\infty}^{\infty} \mu_i f(s) ds + \frac{\kappa_\varepsilon^4}{1 - \alpha^2 \kappa_\varepsilon^4} L_c(1,0) \int_{-\infty}^{\infty} f^2(s) ds,
\end{aligned}$$

from which the stated result in part (a) follows easily. Note that we apply Lebesgue's Increasing Convergence Theorem for

$$\sum_{j=0}^{k-1} \alpha^j \sum_{i=0}^{\infty} \alpha^i \int_{-\infty}^{\infty} \mu_{k+i-j} f(s) ds = \int_{-\infty}^{\infty} \sum_{j=0}^{k-1} \sum_{i=0}^{\infty} \alpha^{i+j} \mu_{k+i-j} f(s) ds$$

and

$$\sum_{i=1}^{\infty} \alpha^i \int_{-\infty}^{\infty} \mu_i f(s) ds = \int_{-\infty}^{\infty} \sum_{i=1}^{\infty} \alpha^i \mu_i f(s) ds.$$

In order to prove the result for  $f \in \mathbb{H}$  in part (b), notice that

$$\kappa(\sqrt{n})^{-1} \bar{y}_n^2 = \frac{\kappa(\sqrt{n})^{-1}}{n} \sum_{t=1}^n y_t^2 \rightarrow_d \frac{1}{1-\alpha} \int_0^1 \bar{f}(V_c(r)) dr.$$

which we already proved in the beginning. One may easily deduce that

$$\begin{aligned} & \frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=k+1}^n (y_t^2 - \bar{y}_n^2) (y_{t-k}^2 - \bar{y}_n^2) \\ = & \frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=k+1}^n y_t^2 y_{t-k}^2 - (\kappa(\sqrt{n})^{-1} \bar{y}_n^2)^2 + O_p(n^{-1/2}) \\ \rightarrow_d & \left[ \frac{1-\alpha^k}{(1-\alpha)} + \alpha^k \frac{(1+\alpha)\kappa_\varepsilon^4}{1-\alpha^2\kappa_\varepsilon^4} \right] \frac{1}{1-\alpha} \int_0^1 \bar{f}^2(V_c(r)) dr \\ & - \left( \frac{1}{1-\alpha} \int_0^1 \bar{f}(V_c(r)) dr \right)^2 \end{aligned}$$

and that

$$\begin{aligned} & \frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=1}^n (y_t^2 - \bar{y}_n^2)^2 \\ = & \frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=1}^n y_t^4 - (\kappa(\sqrt{n})^{-1} \bar{y}_n^2)^2 + O_p(n^{-1/2}) \\ \rightarrow_d & \frac{1+\alpha}{1-\alpha} \frac{\kappa_\varepsilon^4}{1-\alpha^2\kappa_\varepsilon^4} \int_0^1 \bar{f}^2(V_c(r)) dr - \left( \frac{1}{1-\alpha} \int_0^1 \bar{f}(V_c(r)) dr \right)^2 \end{aligned}$$

The stated result in part (b) follows immediately.  $\square$

**Proof of Corollary 2.** See Theorem 2 and 5 in Nelson (1990) for (a). Since  $\mathbb{E} [\ln \varepsilon_t^2] < 0$  when  $\varepsilon_t \sim N(0, 1)$ ,  $\prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2 = O(\exp(-\lambda k))$  for  $\lambda = |\mathbb{E} [\ln \varepsilon_t^2]| / 2$  almost surely. Then, we have  $\prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2 \rightarrow_{a.s.} 0$  as  $k \rightarrow \infty$ . Since we let  $\prod_{i=1}^0 \varepsilon_{t+k-i}^2 = 1$ , this means

$$1 < \sum_{j=0}^{k-1} \prod_{i=1}^j \varepsilon_{t+k-i}^2 < \infty \text{ a.s..}$$

This directly leads to

$$\left| \frac{\partial f(x_{t+k-1})}{\partial x_{t+k-1}} \right| < \left| \sum_{j=0}^{k-1} \prod_{i=1}^j \varepsilon_{t+k-i}^2 \frac{\partial f(x_{t+k-1-j})}{\partial x_{t+k-1-j}} \right| < \infty,$$

which proves (b). Note that whether  $\alpha < 1$  or  $\alpha = 1$  does not matter here. Next, if  $c > 0$ , then we have

$$\frac{\partial \sigma_{t+k}^2}{\partial x_t} = \sum_{j=0}^{k-1} \prod_{i=1}^j \varepsilon_{t+k-i}^2 \alpha^j \left(1 - \frac{c}{n}\right)^{k-1-j} \frac{\partial f(x_{t+k-1-j})}{\partial x_{t+k-1-j}} \rightarrow 0.$$

Note that  $\frac{\partial \sigma_{t+k}^2}{\partial x_t}$  consists of two parts; one part decays at a rate of  $\prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2 \alpha^{k-1}$  and the other part does at a rate of  $\left(1 - \frac{c}{n}\right)^{k-1}$ . The stated result in (c) follows if  $\left(1 - \frac{c}{n}\right)^{k-1}$  decays slower than  $\prod_{i=1}^{k-1} \varepsilon_{t+k-i}^2 \alpha^{k-1}$ . Since  $\alpha^2 \kappa_\varepsilon^4 < 1$  and  $\varepsilon_t \sim N(0, 1)$ , we have  $0 < \alpha < 0.58$ . Thus,  $\alpha^k$  vanishes faster than  $\left(1 - \frac{c}{n}\right)^{k-1}$  because  $\left(1 - \frac{c}{n}\right)$  is close to unity. This completes the proof of (c).  $\square$

**Proof of Theorem 3.** The results are proven in the proof of Theorem 1.  $\square$

**Proof of Theorem 4.** If we write

$$\frac{1}{\sqrt{n}} K_n^4 = \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t^4 \left/ \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n y_t^2 \right)^2 \right.$$

the stated result for  $f \in \mathbb{I}$  in part (a) follows directly. The stated result for  $f \in \mathbb{H}$  in part (b) is followed by

$$K_n^4 = \frac{(\kappa(\sqrt{n})^{-1})^2}{n} \sum_{t=1}^n y_t^4 \left/ \left( \frac{\kappa(\sqrt{n})^{-1}}{n} \sum_{t=1}^n y_t^2 \right)^2 \right.$$

$\square$

Table 1. Unit root tests for the default premium<sup>a</sup>

<Monthly>

	with intercept	with intercept and trend
<b>autoregressive coefficient</b>	0.978	0.974
<b>Phillips-Perron test</b>		
$Z_t$ statistic	-3.21*	-3.43*
<b>KPSS test</b>		
statistic	1.12**	0.37**

<Weekly>

	with intercept	with intercept and trend
<b>autoregressive coefficient</b>	0.982	0.980
<b>Phillips-Perron test</b>		
$Z_t$ statistic	-4.50**	-4.14**
<b>KPSS test</b>		
statistic	2.53**	0.81**

<Daily>

	with intercept	with intercept and trend
<b>autoregressive coefficient</b>	0.996	0.996
<b>Phillips-Perron test</b>		
$Z_t$ statistic	-2.56	-2.49
<b>KPSS test</b>		
statistic	2.24**	2.11**

<sup>a</sup>Notes: The critical values for Phillips-Perron test are following; -2.57 (10%), -2.86 (5%), -3.44 (1%) with intercept, and -3.13 (10%), -3.42 (5%), -3.97 (1%) with intercept and trend. 1% critical values of KPSS statistic are 0.74 and 0.22 respectively. \* means that  $H_0$  is rejected by 5%, and \*\* means that  $H_0$  is rejected by 1%.

Table 2. ARCH-NNH estimation results for returns on the S&amp;P 500 index

	Monthly		Weekly		Daily	
	$\rho = 1$		$\rho = 1$		$\rho = 1$	
$\mu$	0.007 (5.01)	0.007 (5.05)	0.002 (3.16)	0.002 (3.06)	0.0006 (3.62)	0.0006 (3.53)
$\alpha$	0.18 (2.30)	0.18 (2.30)	0.18 (3.55)	0.20 (3.73)	0.22 (4.30)	0.22 (4.25)
$\beta$	$1.3 \times 10^{-3}$ (11.77)	$1.3 \times 10^{-3}$ (11.72)	$3.9 \times 10^{-4}$ (10.84)	$4.1 \times 10^{-4}$ (9.52)	$1.1 \times 10^{-4}$ (18.65)	$1.0 \times 10^{-4}$ (21.55)
$\rho$	1.05 (6.75) [0.31]	1.0 –	0.21 (0.97) [–16.42]	1.0 –	1.10 (6.45) [0.57]	1.0 –

$t$ –statistics are reported in parentheses.  $t$ –statistic for  $H_0 : \rho = 1$  is in [ ].

Table 3. GARCH(1,1) estimation results for returns on the S&amp;P 500 index

	Monthly	Weekly	Daily
$\mu$	0.007 (5.95)	0.002 (4.45)	0.0006 (3.90)
$c$	$0.7 \times 10^{-4}$ (2.54)	$0.6 \times 10^{-5}$ (1.37)	$0.1 \times 10^{-5}$ (2.02)
$\alpha$	0.14 (3.27)	0.094 (2.60)	0.047 (3.78)
$\beta$	0.84 (21.85)	0.898 (23.60)	0.949 (69.66)

$t$ –statistics are reported in parentheses.

Table 4. FIGARCH estimation results for returns on the S&amp;P 500 index

	Monthly		Weekly		Daily	
	( <b>1,d,1</b> )	(1,d,0)	(1,d,1)	( <b>1,d,0</b> )	(1,d,1)	( <b>1,d,0</b> )
$\mu$	0.007 (5.94)	0.007 (6.12)	0.002 (4.36)	0.003 (4.63)	0.001 (4.05)	0.001 (4.43)
$c$	$0.5 \times 10^{-4}$ (1.97)	$1.2 \times 10^{-4}$ (1.96)	$0.1 \times 10^{-4}$ (1.05)	$0.3 \times 10^{-4}$ (2.36)	$0.5 \times 10^{-5}$ (1.97)	$0.1 \times 10^{-4}$ (3.71)
$d$	0.67 (3.73)	0.43 (1.42)	0.46 (3.74)	0.37 (3.86)	0.35 (6.98)	0.30 (8.83)
$\beta$	0.73 (8.25)	0.27 (0.74)	0.59 (2.40)	0.25 (2.15)	0.50 (3.87)	0.26 (6.84)
$\phi$	0.25 (1.99)	–	0.26 (1.29)	–	0.20 (1.87)	–

$t$ –statistics are reported in parentheses.

Table 5. Unit root tests for the interest rate differential

	with intercept	with intercept and trend
<b>autoregressive coefficient</b>	0.991	0.989
<b>Phillips-Perron test</b>		
$Z_t$ statistic	-2.31	-2.58
<b>KPSS test</b>		
statistic	1.05**	0.40**

<sup>a</sup>Notes: The critical values for Phillips-Perron test are following; -2.57 (10%), -2.86 (5%), -3.44 (1%) with intercept, and -3.13 (10%), -3.42 (5%), -3.97 (1%) with intercept and trend. 1% critical values of KPSS statistic are 0.74 and 0.22 respectively. \* means that  $H_0$  is rejected by 5%, and \*\* means that  $H_0$  is rejected by 1%.

Table 6. Estimation results for the exchange rate return volatility

	ARCH-NNH		GARCH(1,1)
$b_0$	0.07 (1.30)	$b_0$	0.05 (0.97)
$b_1$	0.14 (3.22)	$b_1$	0.03 (0.94)
$b_2$	0.04 (1.11)	$b_2$	-0.03 (-0.88)
$b_3$	0.04 (1.99)	$b_3$	0.04 (1.71)
$\alpha$	0.46 (4.59)	$c$	0.07 (1.08)
$\beta$	1.19 (12.62)	$\alpha$	0.09 (1.50)
$\rho$	0.23 (3.76) [-12.67]	$\beta$	0.88 (10.68)

$t$ -statistics are reported in parentheses.  $t$ -statistic for  $H_0 : \rho = 1$  is in [ ].



Table 7. Forecasting results evaluated on the OLS regression<sup>a</sup>

	ARCH-NNH	GARCH(1,1)	FIGARCH(1,d,q)
<Monthly>			
<i>a</i>	0.002 (2.01)	0.002 (1.86)	0.002 (2.15)
<i>b</i>	1.23 (0.34)	0.78 (-0.31)	0.82 (-0.29)
$R^2$	<b>0.12</b>	0.05	0.07
<Weekly>			
<i>a</i>	0.0006 (1.81)	0.0006 (1.39)	0.0006 (1.56)
<i>b</i>	0.82 (-0.41)	0.56 (-0.79)	0.59 (-0.76)
$R^2$	<b>0.11</b>	0.06	0.09
<Daily>			
<i>a</i>	$0.6 \times 10^{-4}$ (1.81)	$0.2 \times 10^{-4}$ (0.77)	$0.3 \times 10^{-5}$ (0.10)
<i>b</i>	0.93 (-0.35)	0.77 (-1.54)	0.98 (-0.14)
$R^2$	0.18	0.21	<b>0.27</b>

<sup>a</sup>Notes: Parameter estimates and goodness-of-fit  $R^2$  statistics for the OLS regressions as defined in equation (9). The t-statistics testing for the null hypotheses  $a = 0$  and  $b = 1$  are in parentheses, and they are based on standard errors using the heteroskedasticity and autocorrelation consistent covariance estimates followed by Andrews and Monahan with quadratic spectral kernel and automatic bandwidth selection. FIGARCH(1,d,q) : q=1 for monthly and q=0 for weekly and daily.

Table 8. Forecasting results evaluated on MAE<sup>a</sup>

		MAE	$D_1 (D_2)$
<Monthly>	ARCH-NNH	$2.53 \times 10^{-3}$	2.39* (30*)
	GARCH(1,1)	<b><math>2.32 \times 10^{-3}</math></b>	
	FIGARCH(1,d,q)	$2.36 \times 10^{-3}$	-1.20 (10)
<Weekly>	ARCH-NNH	<b><math>6.46 \times 10^{-4}</math></b>	
	GARCH(1,1)	$6.93 \times 10^{-4}$	-1.04
	FIGARCH(1,d,q)	$6.74 \times 10^{-4}$	-0.60
<Daily>	ARCH-NNH	$1.34 \times 10^{-4}$	1.61
	GARCH(1,1)	$1.49 \times 10^{-4}$	5.99**
	FIGARCH(1,d,q)	<b><math>1.25 \times 10^{-4}</math></b>	

<sup>a</sup>Notes:  $D_1$  is defined in equation (10) and  $D_1 \sim N(0, 1)$  asymptotically under the null hypothesis.  $D_2$  is defined in equation (11) and has the binomial distribution with parameters  $N$  and  $\frac{1}{2}$  under the null hypothesis. \*\* and \* means rejecting the null hypothesis by 1% and 5% respectively. FIGARCH(1,d,q) : q=1 for monthly and q=0 for weekly and daily.

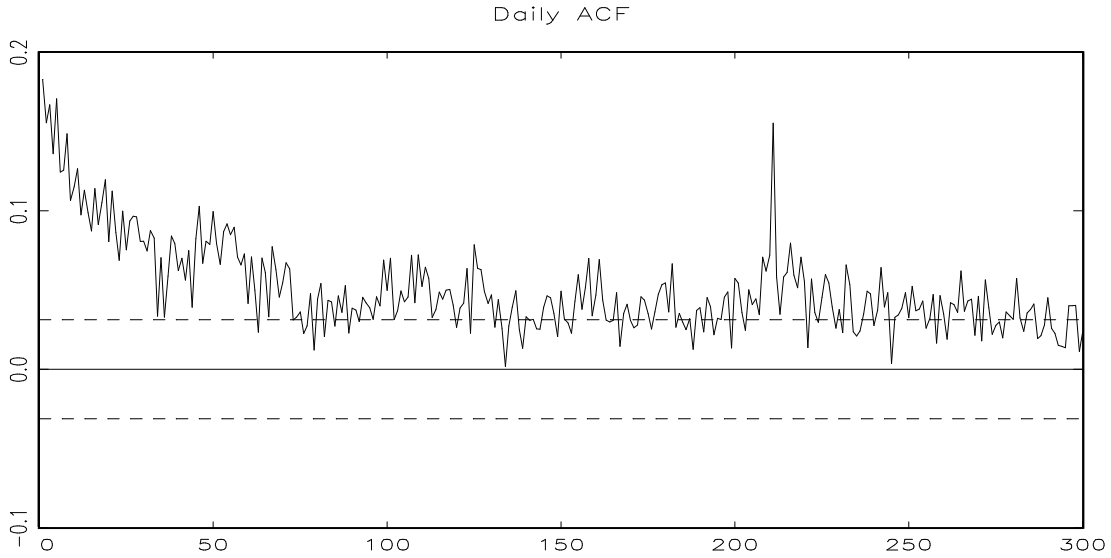


Figure 1. Autocorrelation function of daily squared returns on S&P 500 index from November 2, 1987 to June 30, 2003 (3938 observations). The solid curve indicates the autocorrelation function and the dotted lines show the 95% Bartlett (1946) confidence bands ( $\pm 1.96/\sqrt{T}$ ) for no serial dependence.

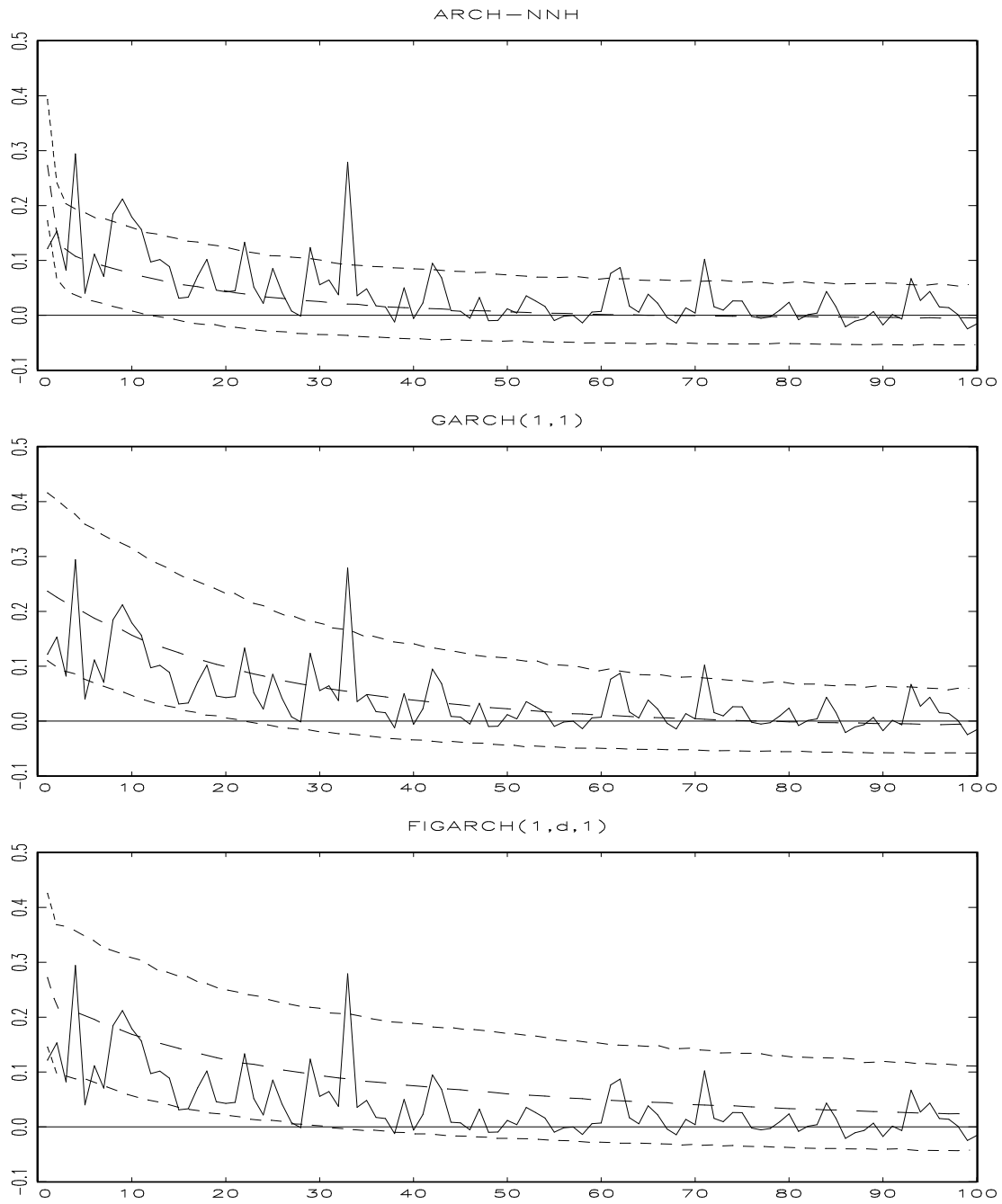


Figure 2. Autocorrelation function of MONTHLY squared returns (S&P 500). The solid curve indicates the sample ACF of the real data and the dotted curves are ACFs of simulated series. The upper and lower dotted curves indicate the 5%- and 95%-quantile of the distribution of the autocorrelations at a fixed lag and the middle dotted curves correspond to the mean of those distributions.

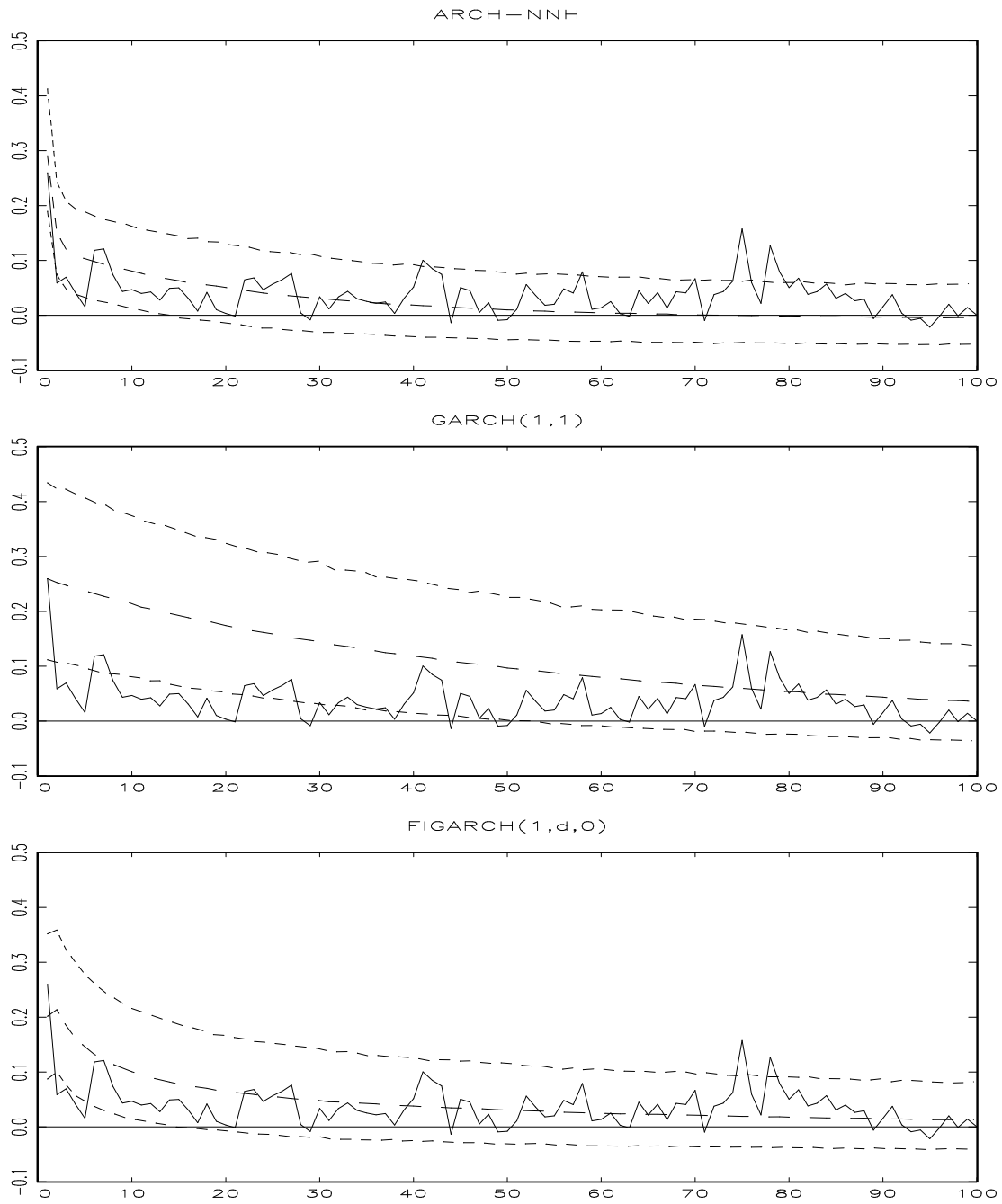


Figure 3. Autocorrelation function of WEEKLY squared returns (S&P 500). The solid curve indicates the sample ACF of the real data and the dotted curves are ACFs of simulated series. The upper and lower dotted curves indicate the 5%- and 95%-quantile of the distribution of the autocorrelations at a fixed lag and the middle dotted curves correspond to the mean of those distributions.

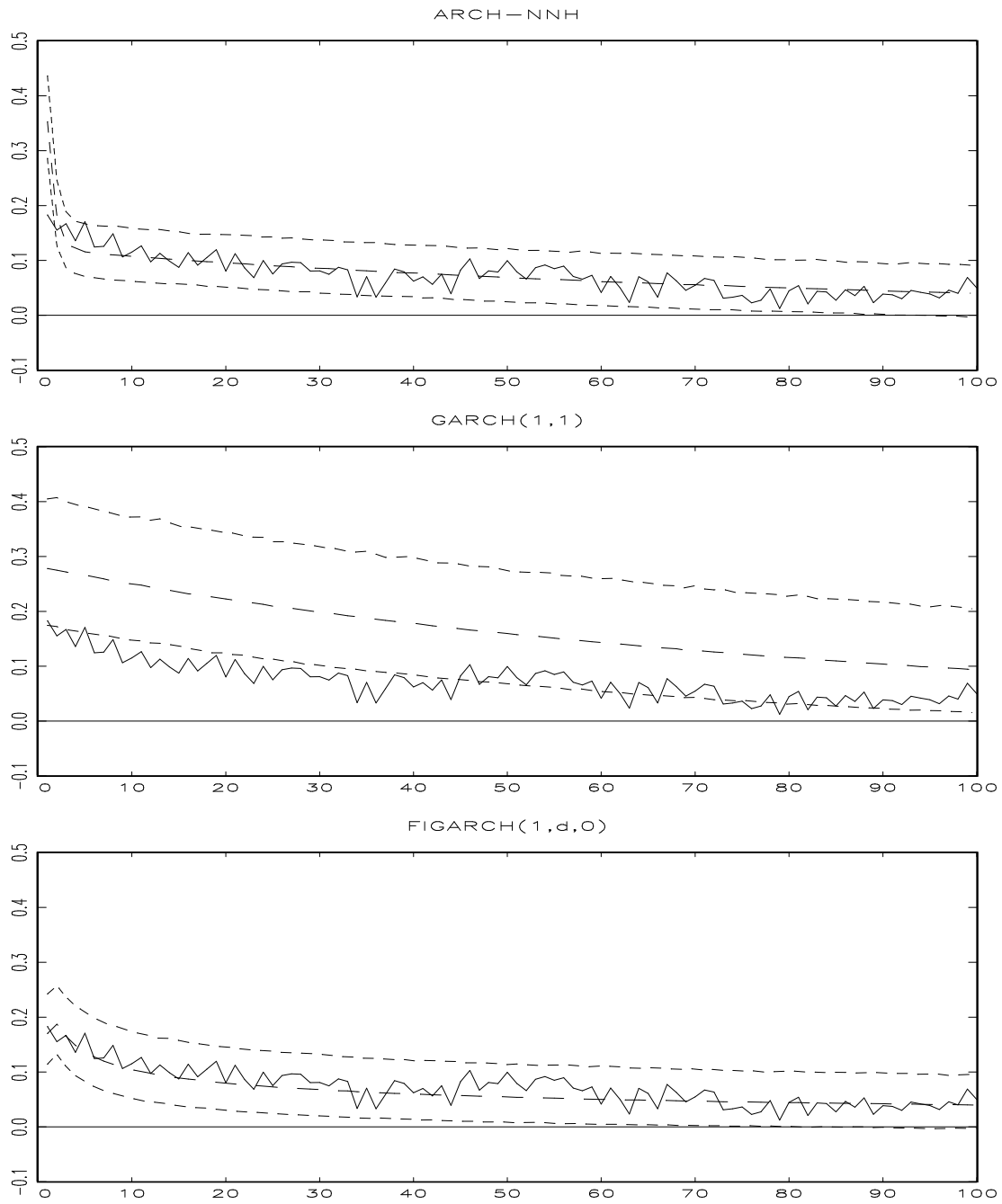


Figure 4. Autocorrelation function of DAILY squared returns (S&P 500). The solid curve indicates the sample ACF of the real data and the dotted curves are ACFs of simulated series. The upper and lower dotted curves indicate the 5%- and 95%-quantile of the distribution of the autocorrelations at a fixed lag and the middle dotted curves correspond to the mean of those distributions.