CIRJE-F-86

Nonstationary Nonlinear Heteroskedasticity: An Alternative to ARCH

JoonY.Park
Seoul National University/CIRJE, University of Tokyo

July 2000

Nonstationary Nonlinear Heteroskedasticity: An Alternative to ARCH [1]

Joon Y. Park
School of Economics
Seoul National University

Abstract

In this paper, we consider time series with the conditional heteroskedasticities that are given by nonlinear functions of integrated processes. Such time series are said to have nonlinear nonstationary heteroskedasticity (NNH), and the functions generating conditional heterogeneity are called heterogeneity generating functions (HGF's). Various statistical properties of time series with NNH are investigated for a wide class of HGF's. For NNH models with a variety of HGF's, volatility clustering and leptokurticity, which are common features of ARCH type models, are manifest. In particular, it is shown that the sample autocorrelations of their squared processes vanish only very slowly, or do not even vanish at all, in the limit. Volatility clustering is therefore well expected. The NNH models with certain types of HGF's indeed have sample characteristics that are very similar to those of ARCH type models. Moreover, the sample kurtosis of the NNH model either diverges or has a stable limiting distribution with support truncated on the left by the kurtosis of the innovations. This would well explain the presence of leptokurticity in many observed time series data. To illustrate the empirical relevancy of our model, we analyze the spreads between the forward and spot rates of USD/DM exchange rates. It is found that the conditional variances of the spreads can be well modelled as a nonlinear function of the levels of the spot rates.

This version: July 2000

Key words and phrases: conditional heteroskedasticity, ARCH, nonstationarity, nonlinearity, volatility clustering, leptokurticity.

[1] This paper was prepared for the presentation at Cardiff Conference on Long Memory and Nonlinear Time Series, and completed while I was visiting Center for International Research in the Japanese Economy, University of Tokyo. I like to thank Fumio Hayashi, Naoto Kunitomo and Yoshihiro Yajima for their support, hospitality and helpful discussions. I am also grateful to Jinyoung Lee and Bongju Song for excellent research assistance, and to Yoosoon Chang, James Davidson, Clive Granger, Bruce Hansen, Peter Phillips, Timo Teräsvirta and other participants at the conference for many useful comments.

1. Introduction

Since the seminal work by Engle (1982) and the later extension by Bollerslev (1986), the ARCH type models have widely been used to model volatilities for economic and financial time series data. In the ARCH models, conditional heteroskedasticities are regarded as being generated in autoregressive-moving average fashions by the squared past values of the underlying series. As is well known, the processes generated in such ways necessarily show volatility clustering and leptokurticity, which are observed commonly for many actual economic and financial time series data. The reader is referred to Bollerslev, Engle and Nelson (1993) for a good overview, and to Gouriéroux (1997) for a detailed description of some of basic statistical properties, of the ARCH type models.

This paper introduces a new class of models for volatility, which can be used as an alternative to the ARCH type models. We consider time series having conditional heteroskedasticities given by nonlinear functions of some integrated processes with unit roots.[2] There appear to be many potential examples for such time series. The volatility of a stock return may well be positively related with the level of interest rate or transactions quantity, which are believed to be integrated. We may likewise consider the volatility in the nominal interest rate differentials as a function of inflation, which is also commonly considered to have a unit root. It may also be reasonable to model the volatility of the spread between forward and spot rates as a function of the level of the spot rate.

In the paper, the conditional heteroskedasticity given by a nonlinear function of a nonstationary integrated process is said to have nonstationary nonlinear heteroskedasticity (NNH), and the function generating conditional heterogeneity is called heterogeneity generating function (HGF). The statistical properties of the NNH models depend crucially on the types of HGF's. Here we consider two different classes of functions: integrable and asymptotically homogeneous functions. Integrable functions are the functions that are integrable, and asymptotically homogeneous functions are the functions that behave asymptotically as homogeneous functions. They are the classes of functions that are considered by Park and Phillips (1999, 2000) in their studies on nonlinear models with integrated time series. These two classes include a wide variety of HGF's. With added technical regularity conditions, the integrable (I) and asymptotically homogeneous (H) functions are respectively called I- and H-regular.

We investigate various statistical properties of the time series driven by the NNH models with a wide class of HGF's. For the NNH models with a variety of HGF's, manifest are volatility clustering and leptokurticity, which are the well known features of the ARCH type models. In particular, we show that the sample autocorrelations of the squared processes generated by the NNH models have strong persistence, i.e., they vanish very slowly or do not even vanish for all lags. We may thus expect to observe substantial volatility clusterings for the time series generated by the NNH

[2]Nonstationary volatility was studied earlier by Hansen (1995). His model, however, specifies the volatility as a function of normalized (near-integrated) process, and does not have the main features of our model which are due to the presence of stochastic trend in the process generating volatility.

models. The squared processes from the NNH models with H-regular HGF's have in general sample autocorrelations that have the same random limit for all lags. On the other hand, the NNH models with I-regular HGF's yield the squared processes whose sample autocorrelations decrease very slowly as the lag order increases. The NNH models with certain types of HGF's indeed have sample characteristics that are very close to those of the stationary and nonstationary ARCH type processes.

The NNH models also well explain the observed leptokurticity in the economic and financial time series data. The processes generated by the NNH models with various HGF's all unambiguously predict the presence of leptokurticity. For the NNH models with I-regular HGF's, the sample kurtosis diverges at the rate of $\sqrt{n}$. We would thus normally expect to have large sample kurtosis for any reasonably large samples. For the NNH models with H-regular HGF's, the sample kurtosis has well defined limiting distributions, which have supports truncated on the left by the kurtosis of the innovations. Therefore, the NNH models with H-regular HGF's also produce the data which have leptokurticity. The processes driven by the NNH models may not necessarily be unconditionally homoskedastic. Depending upon the HGF's, they may be unconditionally homoskedastic as for the stationary ARCH models, or unconditionally explosive similarly as for the nonstationary ARCH (such as IGARCH) models. Of course, they can be unconditionally heteroskedastic, but not explosive at the same time. The NNH models thus provide a much more flexible class of volatility models, especially in terms of the moment characteristics, than the existing ARCH type models.

The aforementioned properties of the NNH models largely follow from the two essential characteristics of the model: the nonlinearity of the function and the nonstationarity of the explanatory variable, which generate conditional heterogeneity. Of the two, the latter appears to be much more important. In particular, we demonstrate in the paper that the conditional heterogeneity generated by stationary time series, i.e., stationary nonlinear heteroskedasticity (SNH), does not produce volatility clustering. The squared processes from the SNH models, for all classes of HGF's, have sample autocorrelations which decay exponentially as the order of lags increases. Their typical sample paths indeed show little volatility clusterings, regardless of the classes of HGF's used to generate the data.

For the purpose of illustration, we present and investigate an empirical NNH model. The model specifies the volatility of the spread between the forward and spot USD/DM exchange rates as a function of the level of the lagged spot rates. The spread is clearly shown to have volatility that is an increasing function of the lagged spot exchange rates. It becomes more volatile when the lagged spot rates are higher. For the volatility in the forward-spot spread of the USD/DM exchange rates, the NNH model with an H-regular HGF appears to be quite appropriate with the lagged spot USD/DM exchange rate as an explanatory variable.

The rest of the paper is organized as follows. Section 2 introduces the model with some preliminary theories and concepts. Various statistical properties for the samples from the NNH models are investigated in Section 3. The asymptotic behaviors of the sample statistics such as the sample autocorrelations of the squared processes, as well as the sample variance and kurtosis, are derived. The problem of estimating HGF's is addressed in Section 4. We show in particular that appropriately parametrized HGF's can be consistently and efficiently estimated from the nonlinear regressions of the squared processes. Section 5 presents an empirical application of the NNH model to the volatility in the USD/DM forward-spot exchange rate spreads. 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) (ε_t) *is iid* $(0,1)$ *and adapted to* $(\mathcal{F}_t)$, *and*
(b) (σ_t) *is adapted to* $(\mathcal{F}_{t-1})$.

The conditions introduced here are minimal. Some of our subsequent results in particular require stronger moment conditions for (ε_t) such as $\mathbf{E}|\varepsilon_t|^p < \infty$ for $p \geq 4$ or $p \geq 8$. This will be specified more precisely later.

Under Assumption 1, we have

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

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. It is, however, conditionally heteroskedastic with conditional variance σ_t^2. We consider in the paper conditional heteroskedasticity generated by a nonlinear function of some explanatory variable.

Assumption 2 *Let*

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

for some nonnegative function $f : \mathbf{R} \to \mathbf{R}_+$, *and*

$$x_t = \rho x_{t-1} + w_t \tag{3}$$

with $\rho = 1$. *Assume that* (x_t) *is adapted to* $(\mathcal{F}_{t-1})$.

With $\rho = 1$, the explanatory variable (x_t) defined in (3) becomes a nonstationary integrated process. However, we also consider (x_t) generated as in (3) with $|\rho| < 1$ for the purpose of comparison. In our subsequent discussions, our volatility model given by Assumptions 1 and 2 will be referred to as NNH(Nonstationary Nonlinear Heteroskedasticity). In contrast, the volatility model with $|\rho| < 1$ in (3) is called SNH(Stationary Nonlinear Heteroskedasticity). The function f introduced in (2) will be called the heterogeneity generating function (HGF) throughout the paper.

Our NNH model here is suggested as an alternative to ARCH or its variants such as GARCH and EGARCH. The conditional heteroskedasticity given by (1) and Assumption 1 is routinely modelled using one of such models. Among many ARCH type models, the most commonly used seems to be the GARCH(p,q) model, which we may write as

$$\sigma_t^2 = \mu + \sum_{k=1}^{p} \alpha_k \sigma_{t-k}^2 + \sum_{k=1}^{q} \beta_k y_{t-k}^2 \tag{4}$$

for some μ, α_k's and β_k's ≥ 0. Let $\alpha(z) = \sum_{k=1}^{p} \alpha_k z^k$ and $\beta(z) = \sum_{k=1}^{q} \beta_k z^k$. Normally, we assume $0 < \alpha(1) + \beta(1) < 1$ in (4), which as is well known implies that the squared process (y_t^2) is stationary. If $\alpha(1) + \beta(1) = 1$ in (4), then (y_t^2) becomes an integrated process,[3] and for this reason, the model is called an integrated GARCH(p,q) or IGARCH(p,q). For the model with $p = q = 1$, we will signify the parameters simply by α and β.

The behavior of our NNH model depends crucially on the HGF f in (2). Among the three classes of functions – integrable functions, asymptotically homogeneous functions and exponential functions – introduced by Park and Phillips (1999), we consider only the first two classes. The integrated processes have stochastic trends, and their exponential transformations appear to be too explosive to be useful for practical applications. With added regularity conditions, the functions that are integrable and asymptotically homogeneous will be called respectively I and H-regular. The classes of functions will be introduced below. Here we just give a brief description of each class of functions with required regularities. For more details on the descriptions and regularity conditions for each class of functions, the reader is referred to Park and Phillips (1999, 2000).

To derive the asymptotics for the NNH models with I-regular HGF's, we assume that $(w_t), w_t = x_t - x_{t-1}$, is either an iid sequence with $\mathbf{E}|w_t|^q < \infty$ for some $q > 4$ (as in Assumption 3S), or a linear process driven by an iid sequence (η_t) such that $\mathbf{E}|\eta_t|^q < \infty$ for some $q > 4$ (as in Assumptions 4S and 5S). In the following definition, we let q be the number that will be given later by such moment conditions.

Definition 1 *A transformation f on $\mathbf{R}$ is called* I-regular *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 $\ell > 6/(q-2)$.

Roughly, I-regular functions are the bounded and integrable functions which are piecewise smooth. As stated in Definition 1, the required smoothness depends on the order of existing moments for the innovation (w_t) of the explanatory variable (x_t). If the higher moments of (w_t) exists, then we may allow for less smooth functions in the class of I-regular transformations. Examples of I-regular functions include indicators

[3] This, of course, does not imply that the process (y_t) itself is nonstationary. It is indeed shown by Nelson (1990) that the IGARCH(1,1) process is strictly stationary without finite second moment.

on compact intervals and smooth bounded integrable functions such as the Laplacian function, $e^{-|x|}$, and the Gaussian function, e^{-x^2}.

To define H-regular functions, we first introduce some classes of transformations on $\mathbf{R}$. We denote by $\mathcal{T}_{\rm LB}$ the class of locally bounded transformations on $\mathbf{R}$, and define $\mathcal{T}_{\rm LB}^{\circ}$ to be its subset consisting of T such that $T(x) = O(e^{c|x|})$ as $|x| \to \infty$ for some constant c. Also, we let $\mathcal{T}_{\rm B}^{\circ}$ be the class of all bounded functions vanishing at infinity, i.e., $T(x) \to 0$ as $|x| \to \infty$, which is a subset of the class $\mathcal{T}_{\rm B}$ of all bounded transformations on $\mathbf{R}$.

Definition 2 *A transformation f on $\mathbf{R}$ is called* H-regular *if f can be written as*

$$f(\lambda x) = \nu(\lambda)h(x) + r(x, \lambda)$$

where h is a homogeneous function on $\mathbf{R}$, and where r satisfies

$$r(x, \lambda) = a(\lambda)p(x) \quad or \quad b(\lambda)p(x)q(\lambda x)$$

with a and b such that $a(\lambda)/\nu(\lambda) \to 0$ and $b(\lambda)/\nu(\lambda) < \infty$ as $\lambda \to \infty$, and p and q such that $p \in \mathcal{T}_{LB}^{\circ}$ and $q \in \mathcal{T}_{B}^{\circ}$.

The H-regular functions behave asymptotically as homogeneous functions. The conditions in Definition 2 guarantee that the remainder term $r(x, \lambda)$ becomes negligible as $\lambda \to \infty$. Therefore, it follows that $f(\lambda x) \approx \nu(\lambda)h(x)$ for λ large. We will call ν and h respectively the asymptotic order and the limit homogeneous function of f. The class of H-regular functions includes a wide variety of functions. It includes, for instance, constant functions, all distribution function-like functions, logarithmic functions and all functions that behave asymptotically as polynomials.

For a bounded H-regular function f, the corresponding limit homogeneous function h becomes homogeneous of degree zero, and is given by

$$h(x) = c_1 1\{x \geq 0\} + c_2 1\{x < 0\} \tag{5}$$

where $c_1, c_2 \in \mathbf{R}$ are some constants. Such an H-regular function has the asymptotic order $\nu(\lambda) = 1$. The class of such H-regular functions includes in particular constant functions and distribution functions, which have limit homogeneous function h given as in (5) respectively with $c_1 = c_2 = c$ for some constant c and with $c_1 = 0$ and $c_2 = 1$. As shown in Park and Phillips (1999), the logarithmic functions are also H-regular with the asymptotic order $\nu(\lambda) = \log \lambda$ and the limit homogeneous function h in (5) for $c_1 = c_2 = 1$. The H-regular function f that behaves asymptotically as a polynomial of order p has the limit homogeneous function h given by

$$h(x) = c|x|^p \tag{6}$$

for some constant c, which is homogeneous of degree $p > 0$.[4] For the H-regular functions having h in (6) as the limit homogeneous function, we have $\nu(\lambda) = \lambda^p$. We

[4] We may consider H-regular functions with the limit homogeneous functions having negative degrees of homogeneity, like the reciprocal function, as in Park and Phillips (1999, 2000). Such functions, however, appear to be of little empirical relevance here, and thus not considered in the paper.

refer to as H_0- and H_p-regular functions, respectively, the H-regular functions with the asymptotic homogenous functions given by (5) and (6).

Our subsequent asymptotic theory is presented using various functionals of standard Brownian motion defined on $[0,1]$, which will be signified by W throughout the paper. It also involves the local time L of Brownian motion W. The Brownian local time L is a stochastic process with two parameters, t and s, say, which can be defined as

$$L(t,s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t 1\{|W(r) - s| < \epsilon\}\, dr$$

Roughly, it may be interpreted as the time spent by W in an immediate neighborhood of s up to time t. The Brownian local time is known to have a version continuous both in t and s, so we may assume L is given as such. The reader is referred to Chung and Williams (1990) for an elementary introduction to the concept of local time. The asymptotics for the NNH models with I-regular HGF's, in particular, include the Brownian local time at $t = 1$ and $s = 0$, i.e., $L(1,0)$. It is well known that $L(1,0)$ has the same distribution as $|W(1)|$, i.e., the modulus of the standard normal variate. See, e.g., Revuz and Yor (1994). The distribution of $L(1,0)$ is thus given by the truncated standard normal supported on the positive half of $\mathbf{R}$.

3. Statistical Properties of NNH

In this section, we investigate the statistical properties of the NNH model. In particular, the asymptotic behaviors of the sample autocorrelation function of the squared process and other sample moments such as the sample variance and the sample kurtosis of the process generated by the NNH model are derived, and compared with those from the competing models such as the GARCH and the SNH models introduced in the previous section. The sample paths for the NNH and SNH with some selected HGF's, and the GARCH models with several parameter values, are simulated and presented in Figures 1 – 3. They will be referred frequently as we explain our statistical results given below.[5]

3.1 Sample Autocorrelations of the Squared Process

We first consider the sample autocorrelations of the squared processes generated from the NNH models and compare them with those generated by other volatility models.

[5]For all the simulated sample paths, the innovations (ε_t) in (1) are generated as independent normals. For the NNH and SNH models, the innovations (w_t) in (3) are drawn independently of (ε_t) as independent standard normals with $\rho = 1$ and 0.5, respectively. The parameter values for the GARCH models are chosen so that $\alpha + \beta = 1$ and 0.5, to make their results roughly more comparable to those from the NNH and SNH models.

Define the sample autocorrelations for (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 the NNH models, we make the following assumptions, in addition to Assumptions 1 and 2.

Assumption 3S *Assume*
(a) (w_t) *are iid.*
(b) $\mathbf{E}f^2(x + w_{kt}) < \infty$ *for all* $x \in \mathbf{R}$ *and* $k \geq 1$, *where* $w_{kt} = w_{t+1} + \cdots + w_{t+k}$.
(c) $\mathbf{E}|\varepsilon_t|^p < \infty$ *for some* $p \geq 8$.
(d) (ε_t) *and* (w_t) *are independent.*
(e) (w_t) *has distribution absolutely continuous with respect to Lebesque measure, characteristic function* $\phi(t)$ *such that* $t^r \phi(t) \to 0$ *as* $t \to \infty$ *for some* $r > 0$, *and* $\mathbf{E}\,|w_t|^q < \infty$ *for some* $q > 4$.

Assumption 3W *Assume* (a) – (d) *of* Assumption 3S, *and*
(e) $\mathbf{E}w_t^2 < \infty$.

Clearly, Assumption 3W is weaker than Assumption 3S, where 'W' and 'S' respectively stand for *weak* and *strong*. Whenever the distinction is unnecessary, we will just refer to Assumption 3. Under Assumption 3S, (w_{kt}) has density with respect to the Lebesque measure on $\mathbf{R}$, and we signify the density by p_k. Also, we denote by σ_w^2 the variance of (w_t), and by κ_ε^4 the kurtosis of (ε_t) throughout the paper.

The classical Donsker's theorem applies to (w_t) under Assumption 3. Therefore, if we let $[nr]$ be the integral part of nr for $r \in [0, 1]$ and define

$$W_n(r) = \frac{1}{\sigma_w \sqrt{n}} \sum_{t=1}^{[nr]} w_t$$

then it follows that

$$W_n \to_d W \tag{7}$$

where W is the standard Brownian motion on the unit interval $[0, 1]$. Here and elsewhere in the paper, $\to_d$ denotes the convergence in distribution. Likewise, we use $\to_p$ in the paper to signify the convergence in probability.

We now present the asymptotic results for R_{nk}^2 under the NNH models.

Theorem 1 *Let* Assumptions 1 *and* 2 *hold, and let* $k \geq 1$.
(a) *If* f *is* I-*regular, then under* Assumption 3S

$$R_{nk}^2 \to_p \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(x+y)p_k(y)dxdy}{\kappa_\varepsilon^4 \int_{-\infty}^{\infty} f^2(x)dx}$$

as $n \to \infty$.
(b) *If* f *is* H-*regular with limit homogeneous function* h, *then under* Assumption 3W

$$R_{nk}^2 \to_d \frac{\int_0^1 h^2(W(r))\,dr - \left(\int_0^1 h(W(r))\,dr\right)^2}{\kappa_\varepsilon^4 \int_0^1 h^2(W(r))\,dr - \left(\int_0^1 h(W(r))\,dr\right)^2}$$

as $n \to \infty$.

It is expected that the results in Theorem 1 hold under weaker conditions than we impose here. Some of the conditions in Assumption 3 are indeed not absolutely necessary. The independency assumption between (ε_t) and (w_t) in Assumption 3(d) is not necessary. Only the orthogonality between (w_t) and certain higher order functions of (ε_t), such as $(\varepsilon_t^2\varepsilon_{t-k}^2)$, is required. We may thus allow for the presence of correlation between (ε_t) and (w_t) without affecting our results in Theorem 1. See the proof of Theorem 1 in Appendix B. Moreover, we may substantially weaken Assumption 3S(e), which was introduced earlier by Park and Phillips (1999) to develop the asymptotics for the integrable transformations of the integrated time series. This will be shown in a later work.

Theorem 1 shows the asymptotic behaviors of the sample autocorrelation functions of the squared processes generated by the NNH models. For the NNH model with I-regular HGF, 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 the squared process. The actual values of the asymptotic autocorrelation function are determined by the distribution of (w_t), as well as the HGF. A quite different picture emerges if the NNH model is given by an H-regular HGF. For the NNH model with H-regular HGF, R_{nk}^2 has a random limit, in sharp contrast with the NNH model with I-regular HGF. Moreover, the limit is independent of k and given by a random constant for all values of the lag order $k \geq 1$. It is also not affected by the distribution of (w_t).

For the NNH model with I-regular HGF, the asymptotic dependency of R_{nk}^2 on the lag order k can further be investigated by introducing some additional assumption. Let the distribution of (w_t) be stable, and assume, in particular, that $w_{kt} =_d c_k w_t$ for some numerical sequence (c_k) such that $c_k \to \infty$ as $k \to \infty$, where $=_d$ denotes the distributional equivalence. Then it follows that

$$p_k(x) = c_k^{-1}p(c_k^{-1}x)$$

where p is the density of (w_t) with respect to the Lebesque measure, and we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(x+y)p_k(y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x)f(x+c_k y)p(y)dxdy \to 0$$

Consequently, the asymptotic autocorrelation function of the squared process decreases down to zero as $k \to \infty$. It is thus expected that R^2_{nk} for the NNH model with I-regular HGF has small values for large k, at least for large enough samples. The asymptotic autocorrelation function of the squared process for the Gaussian NNH model with HGF $f(x) = e^{-x^2}$ is given in the second column of Table 1 for some selected values of k.

It is illuminating to compare the behaviors of the sample autocorrelations of the squared processes for the data generated by the NNH and the GARCH models. Consider the simple, but most popular GARCH(1,1) model with parameters α and β. For the stationary GARCH model with $0 < \alpha+\beta < 1$, the theoretical autocorrelation of the squared process decreases at an exponential rate. One may also easily see that the sample autocorrelation of the squared process has probability limit given for $k \geq 1$ by

$$(\alpha+\beta)^{k-1}\frac{\alpha+\alpha\beta^2+\beta^3}{1-\alpha\beta}$$

which is just the k-th autocorrelation of the stationary process (y_t^2). When $\alpha+\beta = 1$, the process (y_t^2) becomes an integrated process and R^2_{nk} converges in probability to unity at all values of k.

Both for the NNH model with I-regular HGF and the stationary GARCH process, the sample autocorrelation of the squared process is expected to decrease, at least for the large enough samples, as k increases and to ultimately vanish as k tends to infinity. It seems, however, that there is an important difference in their respective decreasing patterns. Under the stationary GARCH model, the probability limits of R^2_{nk} decrease at an exponential rate monotonically as k increases. On the other hand, when the data are generated by the NNH model with I-regular HGF, the limiting values of R^2_{nk} remain significantly different from zero for very large values of k. See the second column of Table 1. The typical sample paths of the processes driven by the NNH model with an I-regular HGF would thus be expected to show more volatility clusterings. In contrast, the volatility clusterings for the samples generated from the stationary GARCH model are supposedly not very conspicuous. This is clearly demonstrated in Figures 1 and 2.

The limiting densities for the sample autocorrelations of the squared processes from the Gaussian NNH models with H-regular HGF's having limit homogeneous functions $h(x) = 1\{x \geq 0\}$ and $h(x) = |x|$ are given in Figure 4. It is interesting to note that the limit distributions may vary substantially across different HGF's. For the models with HGF's having limit homogeneous function $h(x) = 1\{x \geq 0\}$, the squared processes are likely to have serial correlations that could be either very small or very large. In contrast, if the squared processes are generated by the models with HGF's having limit homogeneous function $h(x) = |x|$, the serial correlations of the squared processes are most likely to have moderate values. For the former, the

Table 1: Asymptotic Autocorrelations of Squared Processes

	NNH	SNH			
k	e^{-x^2}	e^{-x^2}	$1+e^{-x^2}$	$e^x/(1+e^x)$	$\lvert x\rvert$
1	.2357	.0218	.0040	.0395	.0344
2	.1925	.0050	.0009	.0197	.0085
3	.1667	.0012	.0002	.0098	.0021
4	.1491	.0003	.0001	.0049	.0005
5	.1361	.0000	.0000	.0025	.0000
10	.1005	.0000	.0000	.0001	.0000
100	.0332	.0000	.0000	.0000	.0000

sample paths may show either little or heavy volatility clusterings. The latter in most cases generates samples with moderate volatility clusterings.

The expected behaviors of the NNH models with H-regular HGF's are more comparable to those of the IGARCH models. Indeed, some of their typical sample paths look quite similar, as one may see from Figures 1 and 2. The sample autocorrelations of the squared processes generated by both models are not expected to decrease as k increases, though they have different probability limits, i.e., unity for the IGARCH models and some random numbers for the NNH models with H-regular HGF's. For both models, the probability limits of R_{nk}^2's do not depend upon the lag length k. There is, however, one important difference between the two models. For the NNH model with H-regular HGF, we expect R_{nk}^2 to be in large samples close to unity at $k = 0$ and stay at the same value below unity for all $k \geq 1$. On the other hand, for the IGARCH process, R_{nk}^2 is expected to take values close to unity at all lags for large enough samples. At least as far as the sample autocorrelations of the squared processes are concerned, the NNH models with H-regular HGF's seem to be practically more relevant. In many cases, the sample autocorrelations of the squared processes are quite persistent, but their values are unambiguously distant from unity. The sample autocorrelations of the squared returns computed from various stock price indices are given in Table 2.[6]

We now look at the SNH models, and compare them with the NNH models. Under SNH, the sample autocorrelations of the squared process are expected to decrease as k increases. Let (x_t) be generated as in (3) with $|\rho| < 1$. Assume that (w_t) is iid and independent of (ε_t) and that (ε_t) satisfies a certain moment condition, as in Assumptions 3(a), (c) and (d). Also, denote by q and q_k respectively the densities of (x_t) and (x_t^k), where $x_t^k = w_t + \rho w_{t-1} + \cdots + \rho^{k-1} w_{t-k+1}$, and assume

[6] The data were obtained for the period of 1970.1.1 – 2000.6.30, and the returns were calculated as the first differences of logged stock price indices.

Table 2: Sample Autocorrelations of Squared Stock Returns

	Daily		Weekly		Monthly	
k	DJ	SP500	DJ	SP500	DJ	SP500
1	.0933	.1152	.1547	.1352	.0237	.0878
2	.1457	.1505	.0926	.1096	.0267	.0622
3	.0665	.0772	.0651	.0896	.0306	.0651
4	.0172	.0234	.0581	.0881	.0101	.0254
5	.1058	.1367	.0596	.0492	−.0108	−.0088
10	.0084	.0141	.0832	.1043	.0158	.0305
100	−.0010	−.0004	−.0250	−.0303	−.0595	−.0504

that $\mathbf{E}f^2(x_t) < \infty$ and $\mathbf{E}f^2(x_t^k) < \infty$ for all values of k. Then we have

$$R_{nk}^2 \to_p \frac{\mathbf{E}y_t^2 y_{t-k}^2 - \left(\mathbf{E}y_t^2\right)^2}{\mathbf{E}y_t^4 - \left(\mathbf{E}y_t^2\right)^2}$$

where

$$\begin{aligned}\mathbf{E}\, y_t^2 &= \mathbf{E}f(x_t) = \int_{-\infty}^{\infty} f(x)q(x)\,dx \\ \mathbf{E}\, y_t^4 &= \kappa_\varepsilon^4\, \mathbf{E}f^2(x_t) = \kappa_\varepsilon^4 \int_{-\infty}^{\infty} f^2(x)q(x)\,dx\end{aligned}$$

and

$$\begin{aligned}\mathbf{E}\, y_t^2 y_{t-k}^2 &= \mathbf{E}\, f(x_t)f(x_{t-k}) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\rho^k x + y)f(x)q(x)q_k(y)dxdy\end{aligned}$$

Since we have $|\rho|^k \to 0$ and $q_k \to q$ as $k \to \infty$, we may easily see that the autocorrelation of (y_t^2) vanishes as the order k of lags increases up to infinity.

For the sample autocorrelations of the squared processes, the SNH models thus yield results similar to the NNH models with I-regular HGF's. It seems, however, that the sample autocorrelations of the squared processes from the SNH models diminish much faster than those from the NNH models with I-regular HGF's. For the Gaussian case, the former quickly goes to zero after $k = 1$, while the latter shows rather strong persistence though they ultimately vanish as k tends to infinity. This is observed for a wide class of functions. For the SNH models, the diminishing patterns of R_{nk}^2 are indeed not significantly different across different classes of functions, which is in sharp contrast to the NNH models. The patterns of R_{nk}^2 for the processes generated from the SNH models are roughly close to those for the stationary GARCH processes that we considered earlier. The probability limits of R_{nk}^2 for the processes generated by the Gaussian SNH models are tabulated in Table 1 for a selected set of functions. See Figure 3 for some typical sample paths generated by the SNH models.

The result in part (b) of Theorem 1 implies, in particular, that the sample autocorrelations of the squared processes for the NNH models with H-regular HGF's having constant limit homogeneous functions converge in probability to zero at all lags. To further investigate their asymptotic behaviors, we write

$$f(x) = c + g(x) \tag{8}$$

for some constant c. In what follows, p_k is defined as in Theorem 1 and $N_k(0,1)$'s denote a sequence of independent standard normal random variates.

Corollary 2 *Let* Assumptions 1 *and* 2 *hold, and let* $k \geq 1$. *If* f *is given as in* (8) *and* g *is* I-*regular, then we have under* Assumption 3S

$$\sqrt{n}\, R_{nk}^2 \to_d \frac{L(1,0)}{\sigma_w(\kappa_\varepsilon^4 - 1)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)g(x+y)p_k(y)dxdy + N_k(0,1)$$

as $n \to \infty$, *where* $N_k(0,1)$*'s are independent of* $L(1,0)$, *and given also independently across all* $k \geq 1$.

The H-regular functions with constant limit homogeneous functions are normally expected to be representable as the sum of the constants and integrable functions as in (8). The result in Corollary 2 therefore should be applicable for a wide range of the NNH models with H-regular HGF's having constant limit homogeneous functions.

The sample autocorrelations of the squared processes from the NNH models with H-regular HGF's having constant limit homogeneous functions should have small values at all lags, especially when the sample size is large. Heavy volatility clustering is therefore not expected. Corollary 2 gives some interesting further characterizations on their distributions with respect to the lags for the NNH models with H-regular HGF's having constant limit homogeneous functions. Of the two independent random terms representing the limit distribution of R_{nk}^2, the first term decreases a.s. as the lag order k increases, while the second term is of mean zero and given independently for each k. Notice that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)g(x+y)p_k(y)dxdy \to 0$$

as $k \to \infty$, under the same condition as that we have assumed earlier for the NNH model with I-regular HGF. For the NNH models with H-regular HGF's having constant limit homogeneous functions, the squared processes yield the sample autocorrelations which may be regarded in large samples as independent random draws from a normal distribution realized around a declining trend.

3.2 Other Sample Moments: Sample Variance and Kurtosis

We now investigate the asymptotic behaviors of other sample moments, such as sample variance and kurtosis.[7] The sample variance of (y_t) is defined by

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

We first introduce the assumptions to obtain the asymptotics of the sample variance S_n^2 under the NNH models. Similarly as before for the analysis of the sample autocorrelation function of (y_t^2), two sets of conditions, strong and weak, are introduced.

Assumption 4S *Assume*
(a) $\mathbf{E}|\varepsilon_t|^p < \infty$ *for some* $p \geq 4$.
(b) (w_t) *is generated by*

$$w_t = \varphi(L)\,\eta_t = \sum_{k=0}^{\infty} \varphi_k \eta_{t-k} \tag{9}$$

where $\varphi_0 = 1$, $\varphi(1) \neq 0$ *with* $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$, *and* (η_t) *are iid and has distribution absolutely continuous with respect to Lebesque measure, characteristic function* $\varphi(t)$ *such that* $t^r\varphi(t) \to 0$ *as* $t \to \infty$ *for some* $r > 0$, *and* $\mathbf{E}|\eta_t|^q < \infty$ *for some* $q > 4$.

Assumption 4W *Assume* (a) *of* Assumption 4S *and*
(b) (w_t) *is generated by* (9) *where* $\varphi_0 = 1$, $\varphi(1) \neq 0$ *with* $\sum_{k=0}^{\infty} k^{1/2}|\varphi_k| < \infty$, *and* (η_t) *are iid* $(0, \sigma_\eta^2)$.

The asymptotics for the sample variance are given in Theorem 3 below. The limiting distribution of the sample variance for the NNH model with I-regular HGF involves the Brownian local time, which we denote as earlier by $L(t,s)$.

Theorem 3 *Let* Assumptions 1 *and* 2 *hold.*
(a) *If* f *is* I-*regular, then under* Assumption 4S

$$\sqrt{n}\,S_n^2 \to_d (1/\sigma_w)\,L(1,0)\int_{-\infty}^{\infty} f(x)\,dx$$

as $n \to \infty$.
(b) *If* f *is* H-*regular with limit homogeneous function* h *and asymptotic order* ν, *then under* Assumption 4W

$$\frac{1}{\nu(\sqrt{n})}\,S_n^2 \to_d \int_0^1 h(\sigma_w W(r))\,dr$$

as $n \to \infty$.

[7] To define the sample moments here, we assume that $\mathbf{E}y_t = 0$ is known. Our subsequent results, however, hold also for the sample moments defined in terms of the demeaned (y_t), i.e., $(y_t - \bar{y})$, without such an assumption.

The sample variances of the time series generated by the NNH models with I-regular HGF's converge in probability to zero, as the sample size increases. The behaviors of the sample variances for the time series from the NNH models with H-regular HGF's are more diverse, depending upon the asymptotic orders of the HGF's. The asymptotic order ν is unity for the bounded H-regular HGF's and the time series generated by the NNH models with such HGF's have finite, though random in general, asymptotic variances, as for stationary ARCH processes. For the NNH models with logarithmic or polynomial HGF's, $\nu(\lambda) \to \infty$ as $\lambda \to \infty$, and the sample variances of the generated time series would thus diverge as the sample size increase. They have infinite variances in the limit, and therefore, they are more comparable to nonstationary ARCH, such as IGARCH, models.

Many financial series like stock returns are known to be leptokurtic. It is thus interesting to investigate the asymptotic behavior of the sample kurtosis of the process generated by the NNH model. We define the sample kurtosis of (y_t) by

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

and introduce the assumptions needed for its asymptotic analysis.

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

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

The asymptotic theory for the sample kurtosis K_n^4 of (y_t) is presented below.

Theorem 4 *Let* Assumptions 1 *and* 2 *hold.*
(a) *If* f *is* I-*regular, then under* Assumption 5S

$$\frac{1}{\sqrt{n}} K_n^4 \to_d \frac{\sigma_w \kappa_\varepsilon^4 \int_{-\infty}^{\infty} f^2(x)dx}{L(1,0)\left(\int_{-\infty}^{\infty} f(x)dx\right)^2}$$

as $n \to \infty$.
(b) *If* f *is* H-*regular with limit homogeneous function* h, *then under* Assumption 5W

$$K_n^4 \to_d \frac{\kappa_\varepsilon^4 \int_0^1 h^2(W(r))\,dr}{\left(\int_0^1 h(W(r))\,dr\right)^2}$$

as $n \to \infty$.

Theorem 4 summarizes the asymptotic behaviors of the sample kurtosis for the NNH models. For the NNH model with I-regular HGF, the sample kurtosis diverges as $n \to \infty$. It is therefore expected to have larger values as the sample size increases. The leptokurticity observed in many economic and financial data may thus well be explained.

The sample kurtosis does not, on the other hand, diverge if we consider the NNH model with H-regular HGF. Instead, it has a random limit. However, for all H-regular HGF's, the limiting distribution of the sample kurtosis has support truncated on the left by the kurtosis κ_ε^4 of the innovations (ε_t). To see this, note that

$$\left(\int_0^1 h(W(r))\,dr\right)^2 \leq \int_0^1 h^2(W(r))\,dr$$

which holds for all h, due to Cauchy-Schwarz inequality. The inequality is strict unless h is a constant function. The leptokurticity is therefore also naturally expected for the time series generated by the NNH models with H-regular HGF's.

For the NNH models with H-regular HGF's, the actual limiting distribution of the sample kurtosis is determined by the limit homogeneous function of the HGF. If, for instance, the HGF is given by a distribution function such as the logistic function $f(x) = e^x/(1+e^x)$, then its limit homogeneous function becomes $h(x) = 1\{x \geq 0\}$. As is well known, $\int_0^1 1\{W(r) \geq 0\}\,dr$ has arcsine law with density $1/(\pi\sqrt{x(1-x)})$ on the unit interval $(0,1)$. The limiting distribution of the sample kurtosis in this case is therefore given by a constant multiple of the reciprocal of arcsine law, which has the density $1/(\pi x\sqrt{x-1})$ over the support $(1,\infty)$.

The densities of the limit distributions of K_n^4 for the Gaussian NNH models with H-regular HGF's having limit homogeneous functions $h(x) = 1\{x \geq 0\}$ and $h(x) = |x|$ are given in Figure 5. If we compare the two classes of models with $h(x) = 1\{x \geq 0\}$ and $h(x) = |x|$, the latter are more likely to yield leptokurtic observations. Note that we have

$$K_n^4 \to_p \kappa_\varepsilon^4$$

if and only if h is a constant function. The asymptotic kurtosis of the observed time series becomes identical to the kurtosis of the innovations, only for the NNH models with H-regular HGF's having constant limit homogeneous functions.

4. Estimation of HGF

Let the HGF f in (2) be specified in a parametric form

$$f(x) = g(x, \theta_0)$$

with some known function g. We may then write

$$y_t^2 = g(x_t, \theta_0) + u_t \tag{10}$$

where

$$u_t = f(x_t)(\varepsilon_t^2 - 1) \tag{11}$$

The unknown parameter value θ_0 in nonlinear regression (10) can then be consistently estimated using the standard nonlinear least squares method.

The nonlinear regression with integrated regressor such as the one in (10) has been studied by Park and Phillips (2000). Our regression here, however, is different from theirs, since ours has errors that are conditionally heteroskedastic and given as in (11). We define

$$\dot{g}(x,\theta) = \frac{\partial}{\partial\theta} g(x,\theta) \tag{12}$$

Similarly as in Park and Phillips (2000), we may show that the nonlinear least squares estimator $\hat{\theta}_n$ of θ_0 in (10) behaves asymptotically as the least squares estimator in the linear regression

$$y_t^2 = \theta' \dot{g}(x_t, \theta_0) + u_t \tag{13}$$

under suitable regularity conditions. The asymptotics for $\hat{\theta}_n$ may therefore be deduced easily from the linear regression (13).

The regularity conditions required for the asymptotic equivalence between the nonlinear least squares regression (10) and the linear least squares regression (13) are given in Park and Phillips (2000). The regression function g is a function of the parameter, as well as the explanatory variable. The required conditions involve some additional assumptions other than we introduced earlier in Definitions 1 and 2, to control the behavior of g as a function of the parameter. The conditions are fairly weak and hold for virtually all functions used in practical nonlinear analysis. In what follows, we simply assume that g satisfies the conditions in Park and Phillips (2000) modified in an obvious way to accommodate the conditional heterogeneity in our model, and only consider the functional properties of

$$g = g(\cdot, \theta_0) \quad \text{and} \quad \dot{g} = \dot{g}(\cdot, \theta_0)$$

The following theorem presents the asymptotic distributions of $\hat{\theta}_n$. Here we signify by $N(0,1)$ a standard normal random variate independent of $L(0,1)$.

Theorem 5 *Let* Assumptions 1 *and* 2 *hold.*
(a) *If g and $\dot{g}$ are* I-*regular, then under* Assumption 4S

$$\sqrt[4]{n}(\hat{\theta}_n - \theta_0) \to_d \left(\frac{\kappa_\varepsilon^4 - 1}{L(1,0)}\right)^{1/2} P^{-1/2} Q^{1/2} P^{-1/2} N(0,1)$$

where

$$P = \int_{-\infty}^{\infty} (\dot{g}\dot{g}')(x)dx \quad \text{and} \quad Q = \int_{-\infty}^{\infty} (g^2 \dot{g}\dot{g}')(x)dx$$

as $n \to \infty$.
(b) *If g and $\dot{g}$ are* H-*regular respectively with limit homogeneous function and asymptotic order (ν, h) and $(\dot{\nu}, \dot{h})$, then under* Assumption 4W

$$\frac{\sqrt{n}\dot{\nu}(\sqrt{n})'}{\nu(\sqrt{n})}(\hat{\theta}_n - \theta_0) \to_d (\kappa_\varepsilon^4 - 1)^{1/2} R^{-1/2} S^{1/2} R^{-1/2} N(0,1)$$

where

$$R = \int_0^1 (\dot{h}\dot{h}')(\sigma_w W(r))dr \quad \text{and} \quad S = \int_0^1 (h^2\dot{h}\dot{h}')(\sigma_w W(r))dr$$

as $n \to \infty$.

The cases we consider in Theorem 5 are not exhaustive. In particular, it may well be the case that g is asymptotically homogeneous and $\dot{g}$ is integrable, as for the regression function $g(x,\theta) = e^{\theta x}/(1+e^{\theta x})$. For such a regression function, however, we normally expect $g\dot{g}$ to be I-regular. As one may easily see from the proof, the result in part (a) of Theorem 5 applies in this case.

The limiting distribution of $\hat{\theta}_n$ for our regression (10) with I-regular regression function (having I-regular derivative) is mixed normal and has the convergence rate $\sqrt[4]{n}$, exactly as in the model with homogeneous errors considered in Park and Phillips (2000). However, our limiting distribution is different from theirs, reflecting the errors being conditionally heteroskedastic. The presence of conditional heterogeneity has a more noticeable effect on the limiting distribution of $\hat{\theta}_n$ when the regression function is H-regular (and has H-regular derivative). Here we have in general a differing convergence rate, which is slower by the factor of $\nu(\sqrt{n})$ compared to the same type of regression with homoskedastic errors. The conditional heterogeneity in the errors is therefore more detrimental in this case. The limiting distribution for the regression with H-regular regression function (having H-regular derivative) is also mixed normal in our case, since we are effectively looking at the strictly exogenous case here.

In the subsequent empirical application, we consider the nonlinear regression with the regression function given by $g(x,\theta) = \alpha x^\beta$ with $\theta = (\alpha, \beta)$. For such regression function g, we have $\nu(\lambda) = \alpha_0\lambda^{\beta_0}$ and $h(x) = x^{\beta_0}$, and

$$\dot{\nu}(\lambda) = \begin{pmatrix} \lambda^{\beta_0} & 0 \\ \alpha_0\lambda^{\beta_0}\log\lambda & \alpha_0\lambda^{\beta_0} \end{pmatrix} \quad \text{and} \quad \dot{h}(x) = \begin{pmatrix} x^{\beta_0} \\ x^{\beta_0}\log x \end{pmatrix}$$

We have, in particular, $\lambda\dot{\nu}(\lambda)/\nu(\lambda) \to \infty$ as $\lambda \to \infty$, and the nonlinear least squares estimator $\hat{\theta}_n$ is consistent.

The nonlinear regression given in (10) has conditionally heteroskedastic errors. It is thus well expected that the usual nonlinear least squares estimator is not efficient. To obtain an efficient estimator, the generalized least squares correction for heteroskedasticity is required. It can be done by minimizing the weighted sum of squares

$$\sum_{t=1}^n \frac{(y_t^2 - g(x_t,\theta))^2}{g^2(x_t,\theta_0)}$$

We denote by $\tilde{\theta}_n$ the resulting estimator. For the actual implementation, we of course replace $g(x_t,\theta_0)$ with $g(x_t,\hat{\theta}_n)$. The estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$ will subsequently be referred to respectively as the ordinary nonlinear least squares (ONLS) estimator and the weighted nonlinear least squares (WNLS) estimator.

The asymptotics for the WNLS estimator $\tilde{\theta}_n$ are given below in Corollary 6. Here we use the same notation as Theorem 5, which present the asymptotics for the ONLS estimator $\hat{\theta}_n$.

Corollary 6 *Let* Assumptions 1 *and* 2 *hold.*
(a) *If g and $\dot{g}$ are* I-*regular, then under* Assumption 4S

$$\sqrt[4]{n}(\tilde{\theta}_n - \theta_0) \to_d \left(\frac{\kappa_\varepsilon^4 - 1}{L(1,0)}\right)^{1/2} \left(\int_{-\infty}^{\infty} (\dot{g}\dot{g}'/g^2)(x)dx\right)^{-1/2} N(0,1)$$

as $n \to \infty$.
(b) *If g and $\dot{g}$ are* H-*regular respectively with limit homogeneous function and asymptotic order (h,ν) and $(\dot{h},\dot{\nu})$, and if $\lambda\dot{\nu}(\lambda)/\nu(\lambda) \to \infty$ as $\lambda \to \infty$, then under* Assumption 4W

$$\frac{\sqrt{n}\dot{\nu}(\sqrt{n})}{\nu(\sqrt{n})}(\tilde{\theta}_n - \theta_0) \to_d (\kappa_\varepsilon^4 - 1)^{1/2} \left(\int_0^1 (\dot{h}\dot{h}/h^2)(\sigma_w W(r))dr\right)^{-1/2} N(0,1)$$

as $n \to \infty$.

It is easy to see that the WNLS estimator $\tilde{\theta}_n$ is more efficient than the ONLS estimator $\hat{\theta}_n$. If we let

$$A = \int_{-\infty}^{\infty} (\dot{g}\dot{g}/g^2)(x)dx \quad \text{and} \quad B = \int_0^1 (\dot{h}\dot{h}/h^2)(\sigma_w W(r))dr$$

and let P, Q, R and S be defined as in Theorem 5, then it indeed readily follows that

$$A \geq PQ^{-1}P \quad \text{and} \quad B \geq RS^{-1}R \ a.s.$$

which establishes the relative efficiency of $\tilde{\theta}_n$ over $\hat{\theta}_n$.

5. Empirical Evidence

In this section, we consider the NNH model to explain the volatilities in the spreads between forward and spot exchange rates. The weekly USD/DM exchange rates for the period of 1986.1.3 – 2000.6.30 (757 observations) are used, and the one-month forward and spot spread volatilities of them are specified as a function of the lagged spot rates. The unit root tests for the spot USD/DM exchange rates all unambiguously support the presence of a unit root. We do not report the details to save space.

The relationships between the spreads and spot rates for the USD/DM exchange rates are given in Figure 6. It is clearly seen from Figure 6 that the forward-spot spreads have volatilities that are positively related to the lagged spot levels for the USD/DM exchange rates. The NNH model with the lagged spot rates as the explanatory variable seems appropriate for the volatilities in the USD/DM forward-spot spreads.[8] Figure 6 also presents the relationship between the squared forward-spot

[8] we also have the similar results when we use the USD as the numeraire and consider DM/USD exchange rates. The relationship between the volitilities of the forward-spot spreads and the lagged spot rates becomes, however, somewhat weaker. There seems to be a compelling reason. Lothian (1997) indeed points out that the behavior of the USD has been dominated by one episode, i.e., the large appreciation and depreciation in the 1980s, and much more irregular than the DM whose variation has been caused by a number of episodes.

Table 3: Estimation Results for NNH Model

	WNLS		ONLS	
	α	β	α	β
estimates	2.276^5	3.858	2.336^5	3.917
standard errors	1.257^5	1.002	1.843^5	1.521
t-values	1.811	3.852	1.267	2.576

Table 4: Estimation Result for ARCH and GARCH Models

	ARCH(1)		GARCH(1,1)		
	μ	α	μ	α	β
estimates	7.865^8	1.695	5.642^9	1.318^1	1.555
standard errors	1.750^8	9.858^2	2.464^9	4.148^2	1.189^1
t-values	4.495	17.19	2.290	3.178	13.07

spreads and the lagged spot levels for the USD/DM exchange rates. The HGF is subsequently fitted to a parametric model $f(x) = g(x, \theta) = \alpha x^\beta$ using the weighted nonlinear least squares (WNLS) method introduced in the previous section.[9] For comparison, we also fit the ARCH(1) and the GARCH(1,1) models to the squared forward-spot spreads. The estimation results for the NNH model, and ARCH and GARCH models are summarized and presented in Tables 3 and 4.[10]

It seems that the NNH model with HGF given by $f(x) = g(x, \theta) = \mu + \alpha x^\beta$ fits the USD/DM exchange rate data reasonably well. The fitted errors in the squared forward-spot spreads become large as the spot levels increase. For the large values of the spot levels, in particular, the fit does not appear to be particularly good. This, however, should not be interpreted as the evidence for the lack of fit. It is expected from the presence of conditional heterogeneity in the errors of our HGF estimation regression, which is given by (10) and (11). Both the coefficients α and β are tested to be significant, if the inference is based on the efficient WNLS estimation. The estimated HGF implies infinite variance, leptokurticity and persistence in volatility for the forward-spot spreads.

For the USD/DM forward-spot spreads, all the estimated coefficients in the ARCH(1) and GARCH(1,1) models are positive and significant. However, both the fitted ARCH(1) and GARCH(1,1) models are not stationary or integrated. The sums

[9] We tried other specifications, especially the given function with constant, i.e., $f(x) = g(x, \theta) = \mu + \alpha x^\beta$. The constant term, however, turns out to be very small and insignificant.

[10] For simplicity, we denote by a^b to denote the number $a \times 10^{-b}$ in the tables.

of their coefficients, except for the constant term, are 1.695 and 1.674 respectively for the ARCH(1) and GARCH(1,1) models. It is thus strongly suggested that the squared forward-spot spreads do not have finite second moments, consistently with our result from the fitted NNH model. For the Gaussian ARCH(1) and GARCH(1,1) models, however, our estimates for the coefficients imply that the forward-spot spreads are strictly stationary. See Nelson (1990).

6. Concluding Remarks

In this paper, we propose a new class of volatility models. The models simply set the conditional heteroskedasticities as functions of integrated explanatory variables. Unlike the volatility driven by a stationary process, our models generate time series showing high degrees of volatility clusterings, infinite variance and leptokurticity, which are common features of many observed economic and financial time series. Yet the models are quite flexible, and have very diverse characteristics depending upon the classes of functions generating heterogeneity. We analyze the volatility, for the USD/DM exchange rates, in the forward-spot spread as a function of the lagged spot, and show the empirical relevancy of our models.

Our models appear to have a wide applicability in modelling volatilities for many important economic and financial time series, which are used to be fitted using the ARCH type models. Our models offer more flexibilities. It seems not rare that none of the ARCH type models fit the data satisfactorily, in which case our models can be an attractive alternative. Some of our models have characteristics that are very similar to both the stationary and nonstationary ARCH type models. It is therefore expected that both the ARCH type model and our models may often provide fits equally good. Even in this case, our models have some clear advantage in that they are more structural and, if appropriately modelled, they may yield much more precise volatility forecasts in some contexts.

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 A1 *Let T be a transformation on* $\mathbf{R}$. *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 is* I*-regular, we have under* Assumption 4S(b)

$$n^{-1/2} M_{1n} \to_d (1/\sigma_w)\, L(1,0) \int_{-\infty}^{\infty} T(x)dx$$
$$n^{-1/2} M_{2n} \to_d (1/\sigma_w)\, L(1,0) \int_{-\infty}^{\infty} T^2(x)dx$$

(b) *If T is* H-*regular with asymptotic order κ and limit homogeneous function H and if we let $\kappa_n = \kappa(\sqrt{n})$, then we have under* Assumption 4W(b)

$$(n\kappa_n)^{-1}M_{1n} \to_d \int_0^1 H(\sigma_w W(r))\,dr$$
$$(n\kappa_n^2)^{-1}M_{2n} \to_d \int_0^1 H^2(\sigma_w W(r))\,dr$$

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

Proof of Lemma A1 See Park and Phillips (1999, 2000). The proofs here are essentially identical to theirs, though the settings and regularity conditions are different. In particular, it follows from Lemma A6 of Park and Phillips (2000), as a special case and under weaker conditions, that the classes of I- and H-regular transformations are closed under the product operation. The square T^2 of I-regular transformation T is therefore I-regular. Moreover, if T is H-regular with limit homogeneous function H, then T^2 is also H-regular with limit homogeneous function H^2. ∎

Lemma A2 *Let T be a transformation on* $\mathbf{R}$, *and let $\varepsilon_{1t} = \varepsilon_t^2 - 1$ and $\varepsilon_{2t} = \varepsilon_t^4 - \kappa_\varepsilon^4$. Define*

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

(a) *If T is* I-*regular, then $U_{1n}, U_{2n} = O_p(n^{1/4})$ respectively under* Assumptions 4S *and* 5S.
(b) *If T is* H-*regular with asymptotic order κ, 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* Assumptions 4W *and* 5W.

Proof of Lemma A2 The stated results follow immediately from Park and Phillips (1999). ∎

In what follows, we let $\mathcal{T}_{\mathrm{B}}^{\circ}$ and $\mathcal{T}_{\mathrm{LB}}^{\circ}$ be the classes of functions defined as earlier in the text. Therefore, $T \in \mathcal{T}_{\mathrm{B}}^{\circ}$ implies that T is bounded and $T(x) \to 0$ as $|x| \to \infty$, and $S \in \mathcal{T}_{\mathrm{LB}}^{\circ}$ implies that S is locally bounded and $S(x) = O(e^{c|x|})$ with some constant c for all large $|x|$.

Lemma A3 *Suppose that* Assumption 4W(b) *holds, and let $(u_t, \mathcal{F}_t)$ be a martingale difference sequence such that $\mathbf{E}(u_t^2|\mathcal{F}_{t-1}) = \sigma_u^2 < \infty$. Define*

$$A_n = \sum_{t=1}^{n} T(x_t)u_t \quad \text{and} \quad B_n = \sum_{t=1}^{n} S\left(\frac{x_t}{\sqrt{n}}\right)u_t$$

where T and S are transformations on $\mathbf{R}$.
(a) *If $T \in \mathcal{T}_B^{\circ}$, then $A_n = o_p(n^{1/2})$.*
(b) *If $S \in \mathcal{T}_{LB}^{\circ}$, then $B_n = o_p(n^{1/2+\delta})$ for any $\delta > 0$.*

Proofs of Lemmas A3 The proof of part (a) is in Park and Phillips (1999). To prove the result stated in part (b), we redefine W_n, up to the distributional equivalence, and assume $W_n \to_{a.s.} W$, as in Park and Phillips (1999,2000). This is always possible due to the Skorohod representation theorem. Define

$$s_{\mathrm{M}} = \sup_{1 \le r \le 1} |W(r)|$$

Since $W_n \to_{a.s.} W$, we have

$$\sup_{0 \le r \le 1} |W_n(r)| \le s_{\mathrm{M}} + 1$$

a.s. for all large n. We therefore have

$$\left| \frac{1}{n} \sum_{t=1}^{n} S^2\left(\frac{x_t}{\sqrt{n}}\right) \right| = \int_0^1 |S^2(\sigma_w W_n(r))\, dr| \le \sup_{|x| \le s_{\mathrm{M}}+1} |S^2(\sigma_w x)| \tag{14}$$

for all large n. However,

$$\mathbf{E} \sup_{|x| \le s_{\mathrm{M}}+1} |S^2(\sigma_w x)| < \infty \tag{15}$$

since s_{M} has Gaussian distribution (truncated and restricted to the nonnegative half of the real line) and $S^2(x)$ is of order $O_p(e^{c|x|})$, at most with some constant c, for all large $|x|$. Due to (14) and (15), we may now apply the dominate convergence to deduce

$$\mathbf{E}\left(n^{-1/2-\delta} B_n\right)^2 = n^{-2\delta} \mathbf{E}\left(\frac{1}{n} \sum_{t=1}^{n} S^2\left(\frac{x_t}{\sqrt{n}}\right)\right) \to 0$$

from which we have $n^{-1/2-\delta} B_n \to_p 0$, as was to be shown. ∎

Let parts (a) and (b) of Assumption 3 hold, and denote by p_k the density of (w_{kt}) with respect to measure m on $\mathbf{R}$. Then we write

$$f(x_t) = \mu_k(x_{t-k}) + \sigma_k(x_{t-k}) v_{kt}$$

where

$$\begin{aligned} \mu_k(x) &= \int_{-\infty}^{\infty} f(x+y) p_k(y)\, m(dy) \\ \sigma_k^2(x) &= \int_{-\infty}^{\infty} f(x+y)^2 p_k(y)\, m(dy) - \left(\int_{-\infty}^{\infty} f(x+y) p_k(y)\, m(dy)\right)^2 \end{aligned}$$

Note that

$$\mu_k(x_{t-1}) = \mathbf{E}\left(f(x_t) | \mathbf{F}_{t-k}\right) \quad \text{and} \quad \sigma_k^2(x_{t-1}) = \mathrm{var}\left(f(x_t) | \mathcal{F}_{t-k}\right)$$

Since

$$v_{kt} = \frac{f(x_t) - \mu_k(x_{t-k})}{\sigma_k(x_{t-k})}$$

it can be easily deduced that

$$\mathbf{E}\left(v_{kt}^2 \middle| \mathcal{F}_{t-k}\right) = 1$$

for all $k \ge 1$.

Lemma A4 *Suppose that* Assumptions 3(a) *and* (b) *hold, and let* μ_k *and* σ_k^2 *be defined as above.*
(a) *If* f *is* I-*regular, then so are* μ_k *and* σ_k^2.
(b) *If* f *is* H-*regular with limit homogeneous function* h, *then so is* μ_k. *For* H_0-*regular* f, $\sigma_k^2 \in \mathcal{T}_{\mathrm{B}}^{\circ}$. *On the other hand, we have* $\sigma_k^2 \in \mathcal{T}_{LB}^{2p-1}$, *if* f *is* H_p-*regular.*

Proof of Lemma A4 For part (a), note that if f is bounded and smooth, then so are μ_k and σ_k^2. It is indeed straightforward to show that μ_k and σ_k^2 satisfy the I-regularity conditions in Park and Phillips (2000). To prove the result for H_0-regular in part (b), we first consider the simplest H_0-regular function given by

$$f(x) = 1\{x \geq 0\}$$

For such f, it can be easily seen that μ_k is given by the distribution function P_k, say, of $(-w_{kt})$, which is H_0-regular with limit homogeneous function $1\{x \geq 0\}$. Moreover, we may deduce that σ_k^2 is given by $P_k(1 - P_k)$, which is in class $\mathcal{T}_{\mathrm{B}}^{\circ}$. The extension to the general H_0-regular function given by

$$f(x) = c_1 1\{x \geq 0\} + c_2 1\{x < 0\} + r(x)$$

with $r \in \mathcal{T}_{\mathrm{B}}^{\circ}$ is straightforward.

To complete the proof for part (b), we consider

$$f(x) = |x|^p$$

It is immediate from Assumption 3 that

$$\begin{aligned}
\sup_{|x| \leq c} |\mu_k(x)| &\leq \int_{-\infty}^{\infty} (c + |y|)^p p_k(y)\, m(dy) < \infty \\
\sup_{|x| \leq c} \left|\sigma_k^2(x)\right| &\leq \int_{-\infty}^{\infty} (c + |y|)^{2p} p_k(y)\, m(dy) < \infty
\end{aligned}$$

for any constant $c \in \mathbf{R}$. Therefore, μ_k and σ_k^2 are locally bounded. Moreover, we have for each fixed y

$$\left|1 + \frac{y}{|x|}\right|^p, \quad \left|1 + \frac{y}{|x|}\right|^{2p} = 1 + O\left(\frac{1}{|x|}\right)$$

as $|x| \to \infty$, and it follows from the dominated convergence that

$$\int_{-\infty}^{\infty} \left|1 + \frac{y}{|x|}\right|^p p_k(y)\, m(dy), \quad \left|1 + \frac{y}{|x|}\right|^{2p} p_k(y)\, m(dy) = 1 + O\left(\frac{1}{|x|}\right)$$

for $|x|$ large. Consequently, we have

$$\mu_k(x) = |x|^p + O(|x|^{p-1}) \quad \text{and} \quad \sigma_k^2(x) = O(|x|^{2p-1})$$

as one may easily check. It is obvious that we have the same result for the general H_p-regular functions. ∎

Appendix B: Proofs of the Main Results

Proof of Theorem 1 To deduce the stated results, we need to analyze the following three sample moments:

$$\sum_{t=1}^{n} y_t^2, \ \sum_{t=1}^{n} y_t^4, \ \sum_{t=k+1}^{n} y_t^2 y_{t-k}^2$$

The first two sample moments are easy to handle. We only need to write them as

$$\sum_{t=1}^{n} y_t^2 = \sum_{t=1}^{n} f(x_t) + \sum_{t=1}^{n} f(x_t)(\varepsilon_t^2 - 1) \tag{16}$$

$$\sum_{t=1}^{n} y_t^4 = \kappa_\varepsilon^4 \sum_{t=1}^{n} f^2(x_t) + \sum_{t=1}^{n} f^2(x_t)(\varepsilon_t^4 - \kappa_\varepsilon^4) \tag{17}$$

and apply Lemmas A1 and A2 for each class of functions.

The asymptotics for the third sample moment can be obtained from Lemmas A1, A3 and A4. To see this, we first write

$$\begin{aligned}\sum_{t=k+1}^{n} y_t^2 y_{t-k}^2 &= \sum_{t=k+1}^{n} f(x_t) f(x_{t-k}) \varepsilon_t^2 \varepsilon_{t-k}^2 \\ &= \sum_{t=k+1}^{n} \mu_k f(x_{t-k}) + R_n \end{aligned} \tag{18}$$

where the remainder term R_n is given by

$$R_n = \sum_{t=k+1}^{n} \mu_k f(x_{t-k})\, a_{kt} + \sum_{t=k+1}^{n} \sigma_k f(x_{t-k})\, b_{kt} \tag{19}$$

where in turn

$$a_{kt} = \varepsilon_t^2 \varepsilon_{t-k}^2 - 1 \quad \text{and} \quad b_{kt} = \varepsilon_t^2 \varepsilon_{t-k}^2 v_{kt} \tag{20}$$

It is easy to deduce the asymptotics for the leading term in (18). Recall that Lemma A4 specifies the class of functions that $\mu_k f$ belongs to, and Lemma A1 establishes the asymptotic result for each class of functions. Moreover, the stochastic orders of R_n in (19) can readily be obtained by Lemma A3, given the characteristics of $\mu_k f$ and $\sigma_k f$ provided by Lemma A4. Note that (a_{kt}) and (b_{kt}) introduced in (20) satisfy the conditions for (u_t) in Lemma A3.

For I-regular f, we have

$$\mu_k f, \sigma_k f \in \mathcal{T}_{\mathrm{B}}^{\circ}$$

due to Lemma A4, and therefore, $R_n = o_p(n^{1/2})$ by Lemma A3. If f is H$_0$-regular, then

$$\mu_k f \in \mathcal{T}_{\mathrm{LB}}^{\circ} \quad \text{and} \quad \sigma_k f \in \mathcal{T}_{\mathrm{B}}^{\circ}$$

as shown in Lemma A4. We consequently have $R_n = o_p(n^{1/2+\delta})$ for any $\delta > 0$, due to Lemma A3. For H$_p$-regular f, it follows directly from Lemma A4 that

$$\mu_k f \in \mathcal{T}_{\mathrm{LB}}^{2p} \quad \text{and} \quad \sigma_k f \in \mathcal{T}_{\mathrm{LB}}^{2p-1/2}$$

and we have $R_n = o_p(n^{p+1/2+\delta})$ for any $\delta > 0$ because of Lemma A3.

We now derive each of the results stated in Theorem 1. For the result for I-regular f in part (a), we first note that

$$\bar{y}_n^2 = O_p(n^{-1/2})$$

which follows directly from (16) and Lemmas A1 and A2. Moreover, due to (18) and Lemmas A1 and A2,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n}(y_t^2-\bar{y}_n^2)(y_{t-k}^2-\bar{y}_n^2) &= \frac{1}{\sqrt{n}}\sum_{t=k+1}^{n} y_t^2 y_{t-k}^2 + O_p(n^{-1/2}) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mu_k f(x_t) + o_p(1) \\
&\to_d (1/\sigma_w)\, L(1,0)\int_{-\infty}^{\infty}\mu_k f(x)dx
\end{aligned}$$

Similarly, we have from (17) and Lemmas A1 and A2

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(y_t^2-\bar{y}_n^2)^2 &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} y_t^4 + O_p(n^{-1/2}) \\
&= \kappa_\varepsilon^4\frac{1}{\sqrt{n}}\sum_{t=1}^{n} f^2(x_t) + O_p(n^{-1/4}) \\
&\to_d (1/\sigma_w)\kappa_\varepsilon^4\, L(1,0)\int_{-\infty}^{\infty} f^2(x)dx
\end{aligned}$$

from which the stated result in part (a) follows easily.

Set $\nu_n = \nu(\sqrt{n})$ in what follows. To prove the result in part (b) for H-regular f, we observe that

$$\begin{aligned}
\nu_n^{-1}\bar{y}_n^2 &= \frac{1}{n\nu_n}\sum_{t=1}^{n} y_t^2 \\
&= \frac{1}{n\nu_n}\sum_{t=1}^{n} f(x_t) + O_p(n^{-1/2}) \\
&\to_d \int_0^1 h(\sigma_w W(r))\, dr
\end{aligned}$$

which follows from (16) and Lemmas A1 and A2. Moreover, one may easily deduce from (16), (17) and (18), and from Lemmas A1 and A2 that

$$\begin{aligned}
\frac{1}{n\nu_n^2}\sum_{t=k+1}^{n}(y_t^2-\bar{y}_n^2)(y_{t-k}^2-\bar{y}_n^2) &= \frac{1}{n\nu_n^2}\sum_{t=k+1}^{n} y_t^2 y_{t-k}^2 - \left(\nu_n^{-1}\bar{y}_n^2\right)^2 + O_p(n^{-1/2}) \\
&= \frac{1}{n\nu_n^2}\sum_{t=1}^{n}\mu_k f(x_t) - \left(\nu_n^{-1}\bar{y}_n^2\right)^2 + o_p(n^{-1/2+\delta}) \\
&\to_d \int_0^1 h^2(\sigma_w W(r))\, dr - \left(\int_0^1 h(\sigma_w W(r))\, dr\right)^2
\end{aligned}$$

and that

$$\begin{aligned}
\frac{1}{n\nu_n^2}\sum_{t=1}^n (y_t^2-\bar{y}_n^2)^2 &= \frac{1}{n\nu_n^2}\sum_{t=1}^n y_t^4 - \left(\nu_n^{-1}\bar{y}_n^2\right)^2 + O_p(n^{-1/2}) \\
&= \kappa_\varepsilon^4 \frac{1}{n\nu_n^2}\sum_{t=1}^n f^2(x_t) - \left(\nu_n^{-1}\bar{y}_n^2\right)^2 + O_p(n^{-1/2}) \\
&\to_d \kappa_\varepsilon^4 \int_0^1 h^2(\sigma_w W(r))\,dr - \left(\int_0^1 h(\sigma_w W(r))\,dr\right)^2
\end{aligned}$$

The stated result in part (b) follows immediately upon noticing that h is a homogeneous function. ∎

Proof of Corollary 2 It follows from the proof of Theorem 1 that

$$\frac{1}{n}\sum_{t=1}^n (y_t^2-\bar{y}_n^2)^2 \to_p c^2(\kappa_\varepsilon^4-1) \tag{21}$$

Under the given specification of f, we may write

$$y_t^2 = c + g(x_t) + c(\varepsilon_t^2-1) + g(x_t)(\varepsilon_t^2-1)$$

It therefore follows from Lemmas A1 and A2 that

$$\begin{aligned}
\frac{1}{n}\sum_{t=k+1}^n y_t^2 y_{t-k}^2 &= c^2 + \frac{1}{n}\sum_{t=1}^n g(x_t)g(x_{t-k}) + \frac{c^2}{n}\sum_{t=1}^n (\varepsilon_t^2-1)(\varepsilon_{t-k}^2-1) \\
&\quad + \frac{2c}{n}\sum_{t=1}^n g(x_t) + \frac{2c^2}{n}\sum_{t=1}^n (\varepsilon_t^2-1) + O_p(n^{-3/4})
\end{aligned} \tag{22}$$

Moreover, we have

$$\frac{1}{n}\sum_{t=1}^n y_t^2 = c + \frac{1}{n}\sum_{t=1}^n g(x_t) + \frac{c}{n}\sum_{t=1}^n (\varepsilon_t^2-1) + O_p(n^{-3/4})$$

and therefore,

$$\left(\frac{1}{n}\sum_{t=1}^n y_t^2\right)^2 = c^2 + \frac{2c}{n}\sum_{t=1}^n g(x_t) + \frac{2c^2}{n}\sum_{t=1}^n (\varepsilon_t^2-1) + O_p(n^{-3/4}) \tag{23}$$

again using the results in Lemmas A1 and A2.

We have from (22) and (23)

$$\begin{aligned}
&\sqrt{n}\left(\frac{1}{n}\sum_{t=k+1}^n y_t^2 y_{t-k}^2 - \left(\frac{1}{n}\sum_{t=1}^n y_t^2\right)^2\right) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^n g(x_t)g(x_{t-k}) + \frac{c^2}{\sqrt{n}}\sum_{t=1}^n (\varepsilon_t^2-1)(\varepsilon_{t-k}^2-1) + O_p(n^{-1/4})
\end{aligned} \tag{24}$$

However, it follows as in the proof of Theorem 1 that

$$\frac{1}{\sqrt{n}}\sum_{t=k+1}^{n} g(x_t)g(x_{t-k}) \to_d (1/\sigma_w)\, L(1,0)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)g(x+y)p_k(y)dxdy \qquad (25)$$

Moreover,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varepsilon_t^2-1)(\varepsilon_{t-k}^2-1) \to_d (\kappa_\varepsilon^4-1)N_k(0,1) \qquad (26)$$

which clearly becomes independent of $L(1,0)$ and is given independently for each $k \geq 1$. The stated result now follows from (21), (24), (25) and (26). ∎

Proof of Theorem 3 The stated result for I-regular f in part (a) follows immediately from (16) and Lemmas A1 and A2. The result for H-regular f in part (b) is proved in the proof of Theorem 1. ∎

Proof of Theorem 4 The stated result for the I-regular f in part (a) follows directly from Lemmas A1 and A2, if we write

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

due to (16) and (17). To prove the result for the H-regular f in part (b), we let $\nu_n = \nu(\sqrt{n})$ and note that

$$K_n^4 = \frac{\dfrac{1}{n\nu_n^2}\displaystyle\sum_{t=1}^{n} y_t^4}{\left(\dfrac{1}{n\nu_n}\displaystyle\sum_{t=1}^{n} y_t^2\right)^2}$$

Once again, the stated result follows immediately from (16) and (17) by Lemmas A1 and A2. ∎

Proof of Theorem 5 Here we assume the regularity conditions in Park and Phillips (2000), properly modified to account for the presence of heterogeneity in our regression, are met so that we have asymptotic equivalence of nonlinear regression (10) and linear regression (13). We may therefore let $\hat{\theta}_n$ be the linear least squares estimator of θ in (13), which is given by

$$\hat{\theta}_n = \theta_0 + \left(\sum_{t=1}^{n}(\dot{g}\dot{g}')(x_t)\right)^{-1}\sum_{t=1}^{n}(\dot{g}g)(x_t)v_t$$

where $v_t = \varepsilon_t^2 - 1$. The stated results now follow immediately as in Park and Phillips (1999). Notice that (v_t) is iid, independent of (x_t), and has variance $\kappa_\varepsilon^4 - 1$. ∎

Proof of Corollary 6 The stated results are obtained similarly as in the proof of Theorem 5. The detailed proofs are therefore omitted. ∎

References

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

Bollerslev, T., R.F. Engle and D. Nelson (1993). "ARCH models," in *Handbook of Econometrics*, Vol. IV, Amsterdam: North-Holland.

Chung, K.L. and Williams, R.J. (1990). *Introduction to Stochastic Integration*, 2nd ed. Birkhäuser, Boston.

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

Gouriéroux, C. (1997). *ARCH Models and Financial Applications.* New York: Springer–Verlag.

Hansen, B.E. (1995). "Regression with nonstationary volatility," *Econometrica*, 63, 1113-1132.

Lothian, J. (1997). "Multi-country evidence on the behavior of purchasing power parity under the current float," mimeographed, Graduate School of Business, Fordham University.

Nelson, D. (1990). "ARCH models as diffusion approximations," *Journal of Econometrics*, 45, 7-38.

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

Park, J.Y. and P.C.B. Phillips (2000). "Nonlinear regressions with integrated time series," forthcoming in *Econometrica.*

Revuz, D. And M. Yor (1994). *Continuous Martingale and Brownian Motion.* New York: Springer–Verlag.

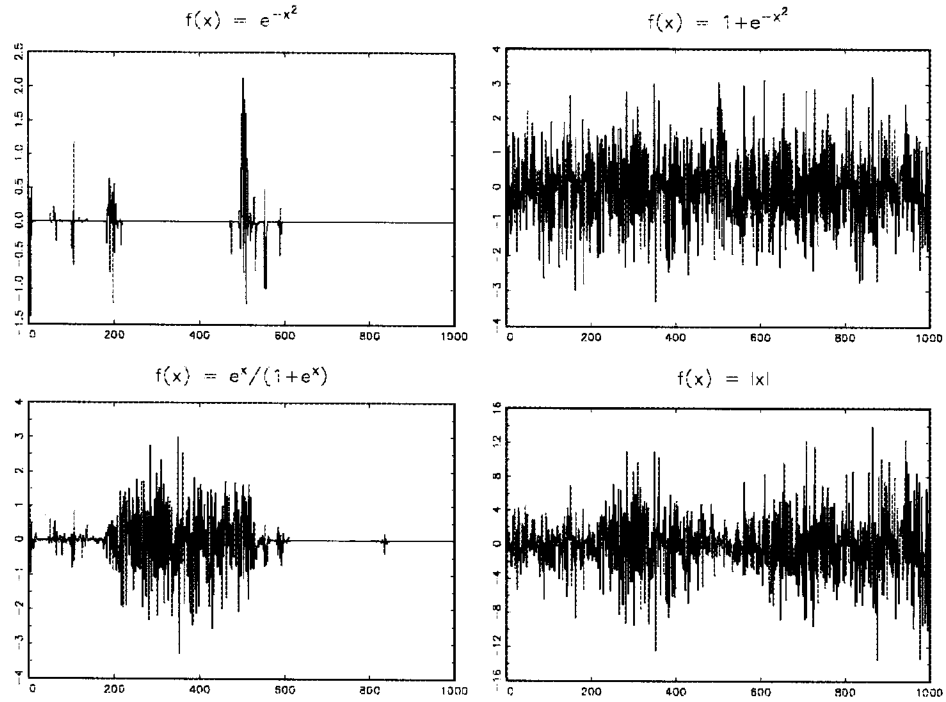

Figure 1: Simulated Sample Paths for NNH

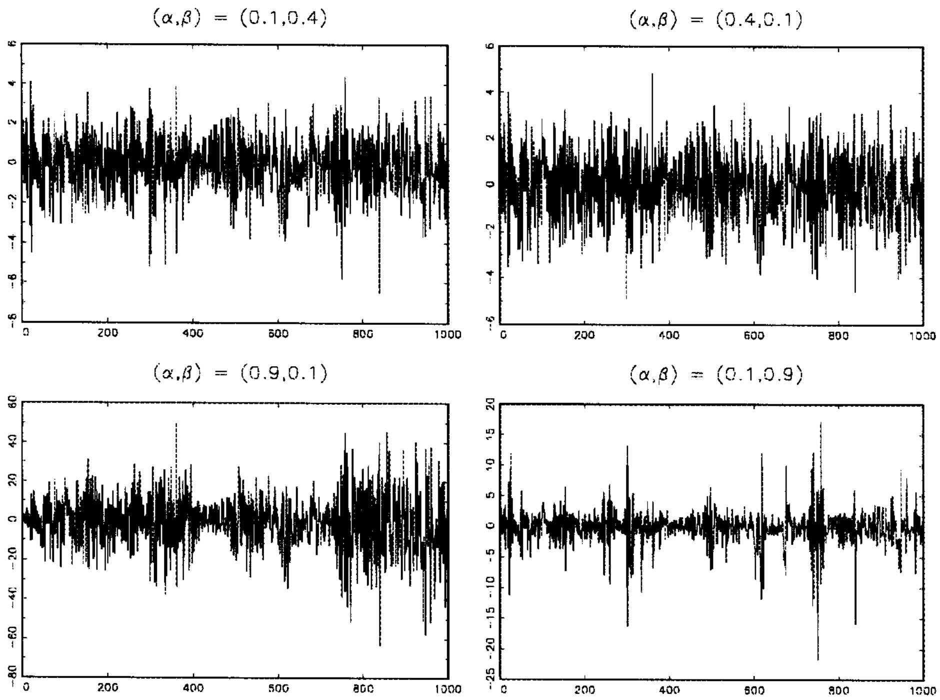

Figure 2: Simulated Sample Paths for GARCH

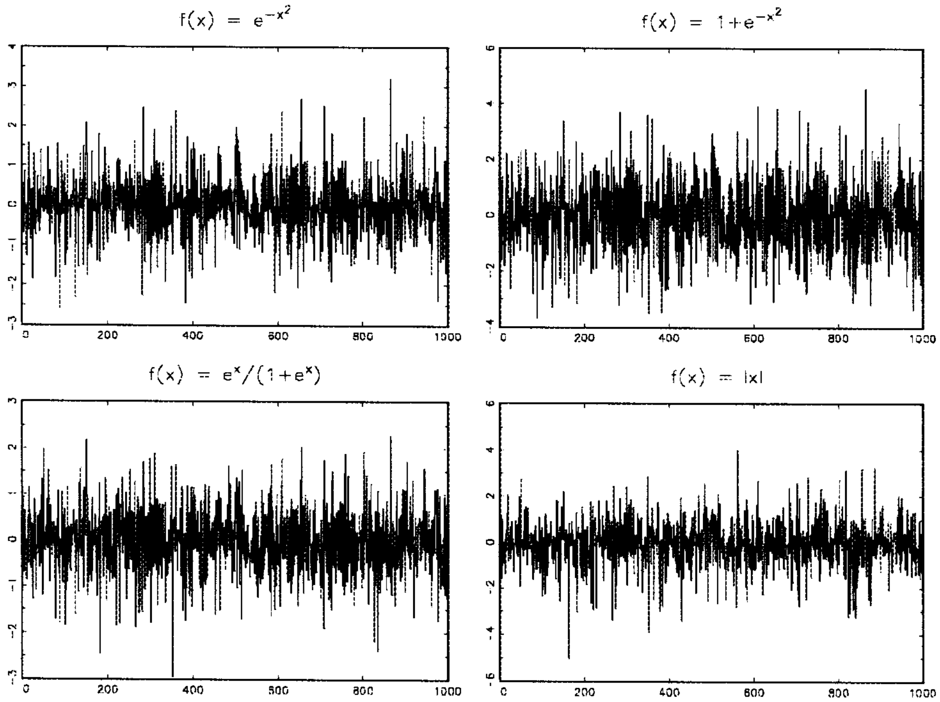

Figure 3: Simulated Sample Paths for SNH

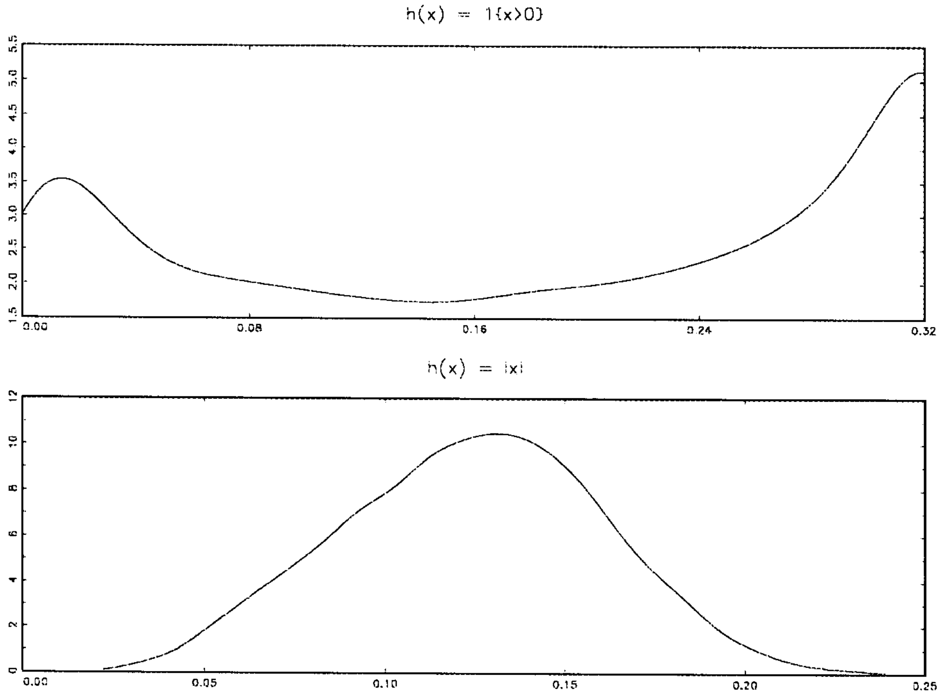

Figure 4: Asymptotic Densities for Sample Autocorrelations of Squared Processes

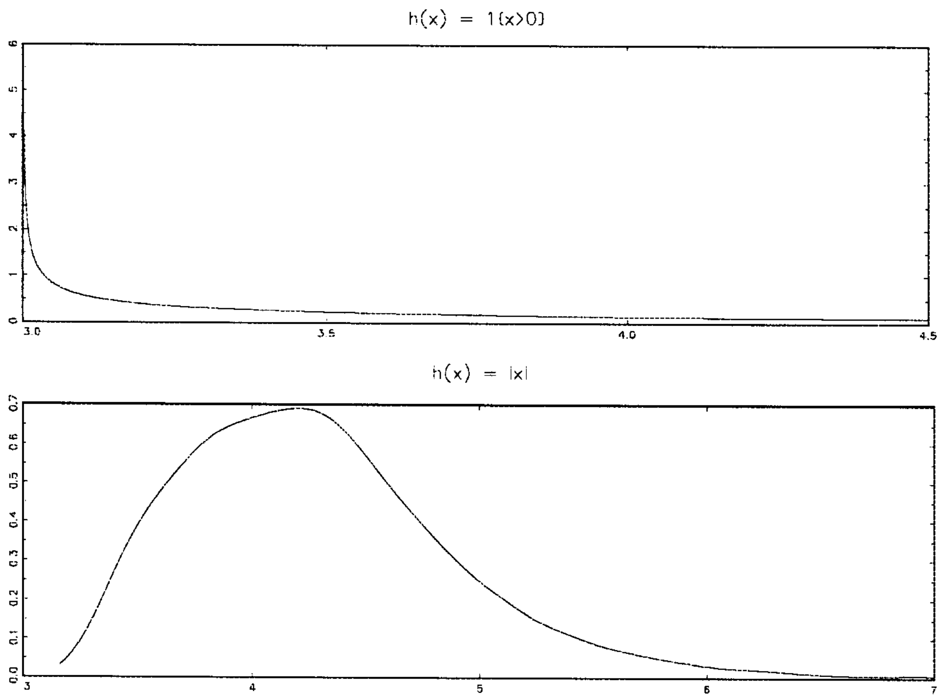

Figure 5: Asymptotic Densities for Sample Kurtoses

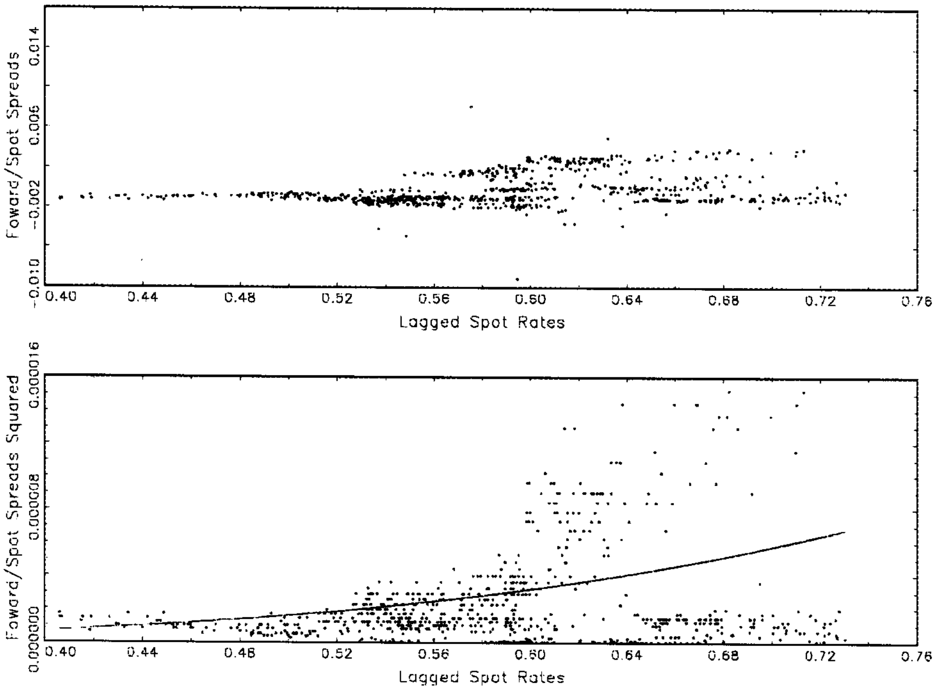

Figure 6: USD/DM Forward/Spot Spread Volatility