Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model

James Davidson*
Cardiff University

April 2003: Forthcoming in *Journal of Business and Economic Statistics*
Keywords: ARCH(∞), FIGARCH, hyperbolic lag, near epoch dependence.
JEL Classification: C22

Abstract

This paper analyses moment and near-epoch dependence properties for the general class of models in which the conditional variance is a linear function of squared lags of the process. It is shown how the properties of these processes depend independently on the sum and rate of convergence of the lag coefficients, the former controlling the existence of moments, and the latter the memory of the volatility process. Conditions are derived for existence of second and fourth moments, and also for the processes to be L_1- and L_2- near epoch dependent (NED), and also to be L_0-approximable, in the absence of moments. The geometric convergence cases (GARCH and IGARCH) are compared with models having hyperbolic convergence rates, the FIGARCH, and a newly proposed generalization, the HYGARCH model. The latter model is applied to 10 daily dollar exchange rates for 1980-1996, with very similar results. When nested in the HYGARCH framework, the FIGARCH model appears as a valid simplification. However, when applied to data for Asian exchange rates over the 1997 crisis period, a distinctively different pattern emerges. The model exhibits remarkable parameter stability across the pre- and post-crisis periods.

1 Introduction

Many variants of Engle's (1982) ARCH model of conditional volatility have been proposed, including GARCH (Bollerslev 1986), IGARCH (Engle and Bollerslev 1986) and FIGARCH (Baillie, Bollerslev and Mikkelsen, 1996, Ding and Granger, 1996). All of these models, and many other cases that might be devised, fall into the class in which the conditional variance at time t is an infinite moving average of the squared realizations of the series up to time $t-1$. Formally, let

$$u_t = \sigma_t e_t \tag{1.1}$$

where $\sigma_t > 0$, $e_t \sim iid(0,1)$ and

$$\sigma_t^2 = \omega + \sum_{i=1}^{\infty} \theta_i u_{t-i}^2 \qquad \theta_i \geq 0, \text{ all } i \tag{1.2}$$

*Email: davidsonje@cf.ac.uk. Research supported by the ESRC under award L138251025. I am most grateful to an anonymous referee for suggestions for simplifying the proofs for Section 3, and also to Soyeon Lee for useful comments, and for providing the Asian crisis data sets used in Section 6.2.

where θ_i are lag coefficients depending typically on a small number of underlying parameters. By adding an error term $v_t = u_t^2 - \sigma_t^2$ to both sides, (1.2) can be viewed as an AR(∞) in the squared series, and hence is commonly called an ARCH(∞) model.

In the well-known case of the GARCH(1,1) model,

$$\sigma_t^2 = \gamma + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{1.3}$$

solves to give (1.2) with $\theta_i = \alpha_1 \beta_1^{i-1}$ for $i \geq 1$, and $\omega = \gamma/(1-\beta_1)$. The stationarity condition is well-known to be $\alpha_1 + \beta_1 < 1$, which is equivalent to

$$\sum_{i=1}^{\infty} \theta_i < 1. \tag{1.4}$$

The IGARCH case is here the variant in which $\alpha_1 + \beta_1 = 1$, and hence the sum of the lag coefficients is also unity, where the θ_i form a convergent geometric series.

Generalizing to the higher order cases, let $\delta(L) = 1 - \delta_1 - \cdots - \delta_p$ and $\beta(L) = 1 - \beta_1 - \cdots - \beta_q$ denote polynomials in the lag operator. The GARCH(p, q) model can be expressed in the "ARMA-in-squares" form

$$\delta(L) u_t^2 = \gamma + \beta(L) v_t \tag{1.5}$$

where $v_t = u_t^2 - \sigma_t^2 = (e_t^2 - 1)\sigma_t^2$, as well as in the more conventional representation

$$\beta(L)\sigma_t^2 = \gamma + (\beta(L) - \delta(L)) u_t^2 \tag{1.6}$$

so that $\alpha_1 = \delta_1 - \beta_1$ in the notation of (1.3). The model is rearranged into the form of (1.2) as

$$\begin{aligned} \sigma_t^2 &= \frac{\gamma}{\beta(1)} + \left(1 - \frac{\delta(L)}{\beta(L)}\right) u_t^2 \\ &= \omega + \theta(L) u_t^2 \end{aligned} \tag{1.7}$$

where $\theta(L) = \sum_{i=1}^{\infty} \theta_i L^i$. Note that $\theta_0 = 0$ by construction, here. The general IGARCH(p, q) can be represented by (1.7) subject to the constraint $\delta(1) = 0$, such that the lag coefficients sum to unity. More explicitly, it might be written in the form

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 - L) \tag{1.8}$$

where $\delta(L)$ is defined appropriately. However, it's important to note the fact that there is no explicit requirement for the roots of $\delta(L)$ to be stable. Nelson (1990) shows in the GARCH(1,1) case that $\delta_1 > 1$ is compatible with strict stationarity, although not covariance stationarity. See Section 3.2 for more on this case.

The FIGARCH(p, d, q) model replaces the simple difference in (1.8) with a fractional difference, such that

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 - L)^d \tag{1.9}$$

for $0 < d < 1$. The FIGARCH is a case where the lag coefficients decline hyperbolically to zero, rather than geometrically, and it is these cases on which we focus particular attention in this paper. This form has been used in a number of recent papers to model financial time series. In addition to Baillie et al. (1996), see for example Beltratti and Morana (1999), Baillie, Cecen and Han (2000), Baillie and Osterberg (2000), and Brunetti and Gilbert (2000). As is well known,

$$(1 - L)^d = 1 - \sum_{i=1}^{\infty} a_j L^j \tag{1.10}$$

where

$$a_j = \frac{d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} = O(j^{-1-d}). \tag{1.11}$$

The present paper focuses attention on these linear-in-the-squares models, in contrast to cases such as EGARCH (Nelson 1991) where the logarithm of the conditional variance is modelled. As will become clear in the sequel, the moment and memory properties of the latter type of model must be analysed in a different way. An important related study by Giraitis, Kokoszka and Leipus (2000) (henceforth, GKL) studies the squared process $\{u_t^2\}$ itself, and some of our results can be seen as complementary to theirs. We focus on the process $\{u_t\}$ itself primarily because, as discussed in Section 3.1, our results have a direct application to the asymptotic analysis of conditionally heteroscedastic series. We consider the existence of moments, and the conditions for limited memory, which is characterized here as near-epoch dependence on the independent process e_t.

Section 2 considers the conditions for second order stationarity, and also sufficient conditions for fourth order stationarity. Section 3.1 addresses the near-epoch dependence question, and in Section 3.2, a modified short-memory property is proved for the class of non-wide sense stationary cases such as the IGARCH, for which the variance does not exist, subject to strict stationarity. Section 4 further discusses some features of the IGARCH and FIGARCH models. Some puzzles and paradoxes that have been discussed in the literature are resolved by noting that independent parameter restrictions control the existence of moments, and the memory of the volatility process. IGARCH and FIGARCH models have been described in the literature as 'long memory', by an implicit analogy with the integrated or fractionally integrated linear model of the conditional mean. However, a conclusion we shall emphasize is that such analogies are generally misleading. It turns out that ARCH(∞) models cannot exhibit long memory by the usual criteria. Both the sequence of lag coefficients, and the autocorrelations of the squared process when these are defined, must be summable, to avoid nonstationary (explosive) solutions.

Section 5 of the paper introduces a new model, the HYGARCH, generalizing the FIGARCH, that can be covariance stationary while exhibiting hyperbolic memory. Section 6 reports some applications of the latter model. Section 6.1 applies it to some familiar series, while Section 6.2 considers Asian exchange rate data covering the 1997-98 crisis period. Section 7 concludes the paper.

2 Moment Properties

Volatility models of the ARCH(∞) class possess two salient features, which we will refer to respectively as the *amplitude* and the *memory.* The amplitude determines how large the variations in the conditional variance can be, and hence the order of existing moments, while the memory determines how long shocks to the volatility take to dissipate. The amplitude is measured by

$$S = \sum_{i=1}^{\infty} \theta_i. \tag{2.1}$$

Regarding the phenomenon of (limited) memory, we recognise two cases. Hyperbolic memory is measured by the parameter δ, such that

$$\theta_i = O(i^{-1-\delta}). \tag{2.2}$$

Geometric memory is measured by the parameter ρ, where

$$\theta_i = O(\rho^{-i}). \tag{2.3}$$

Note that the 'length' of memory varies *inversely* with these parameters. In the geometric-decay GARCH(1,1) model, for example, $S = \alpha_1/(1-\beta_1)$ whereas $\rho = 1/\beta_1$. Whereas in the case where $\rho > 1$ the hyperbolic memory assumes the value $+\infty$, it is more realistic to recognise that these represent two different modes of memory decay, in which the low order lags of one can dominate those of the other, in either case. What is true is that the hyperbolic lags must always dominate the geometric by taking i large enough.

The condition

$$S < 1 \tag{2.4}$$

is generally necessary and sufficient for covariance stationarity. To see this write $M_p = Eu_t^p$, assumed not to depend on t. Then for the case $p = 2$ we have, by the law of iterated expectations,

$$E\sigma_t^2 = Eu_t^2 = \omega + \sum_{i=1}^{\infty} \theta_i E u_{t-i}^2 \tag{2.5}$$

with the stationary solution

$$M_2 = \frac{\omega}{1-S}. \tag{2.6}$$

Next, consider the fourth moment. Letting $\mu_4 = Ee_t^4$, note that $Eu_t^4 = \mu_4 E\sigma_t^4$ where

$$E\sigma_t^4 = \omega^2 + 2\omega \sum_{i=1}^{\infty} \theta_i E u_{t-i}^2 + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \theta_i \theta_j E u_{t-i}^2 u_{t-j}^2. \tag{2.7}$$

Even assuming these expectations do not depend on t, to solve the equality in (2.7) exactly is intractable. However, the Cauchy-Schwarz inequality will imply $Eu_{t-i}^2 u_{t-j}^2 \leq M_4$, and hence,

$$M_4 \leq \mu_4 \left(\omega^2 + \frac{2\omega^2 S}{1-S} + S^2 M_4 \right) \tag{2.8}$$

or equivalently,

$$M_4 \leq \frac{\mu_4 \omega^2 (1+S)}{(1-S)(1-\mu_4 S^2)}. \tag{2.9}$$

The condition

$$\mu_4 S^2 < 1 \tag{2.10}$$

is therefore sufficient, although not necessary, for the existence of M_4 and fourth-order stationarity. Conditions equivalent to (2.4) and (2.10) have also been derived by GKL, who consider the conditions for a weakly stationary solution of the process u_t^2.

It is of interest to evaluate the bound in (2.10) for a case where the exact necessary condition for fourth-order stationarity is known. The GARCH(1,1) in (1.3) has $S = \alpha_1/(1-\beta_1)$. Straightforward manipulations show that[1]

$$M_4 = \frac{\mu_4 \gamma^2 (1-\beta_1)^2 (1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-\mu_4\alpha_1^2 - 2\alpha_1\beta_1 - \beta_1^2)} \tag{2.11}$$

subject to second-order stationarity, and satisfaction of the extra inequality

$$\alpha_1^2 \mu_4 < 1 - 2\alpha_1\beta_1 - \beta_1^2. \tag{2.12}$$

[1] This result may be derived as a special case of the one given in Davidson (2002a) for the GARCH(p,p). For another version of the general formula see He and Terasvirta (1999), and also Karonassos (1999) for the Gaussian case.

Note that (2.10) can be rearranged as

$$\alpha_1^2\mu_4 < 1 - 2\beta_1 + \beta_1^2. \tag{2.13}$$

The majorants of (2.12) and (2.13) differ by $2\beta_1(1-\alpha_1-\beta_1)$, and therefore the latter condition binds as this quantity approaches 0. In this example, the sufficient condition imposes too tough a constraint on the kurtosis of the shocks when the variance is not too big. However, note that the two conditions are identical in the ARCH(1) model, when $\beta_1 = 0$. They are also similar in the region where the second-order stationarity condition is tending to bind, and eventually coincide, although note that to fall in this region requires μ_4 close to 1.

3 Memory Properties

3.1 Near Epoch Dependence

There are a number of ways to measure the memory of a process, some specific to the model structure, such as the rate of decay of the weights of a linear process, and some model-independent, such as the various mixing conditions. The correlogram only measures one facet of the memory of a nonlinear process, although the correlogram of the squared process supplies additional information, relevant to conditional heteroscedasticity in particular. The analysis of GKL is germane to this case. The motivation for studying memory properties is sometimes related to forecastability at long range, but more often we are concerned to check the validity of applying averaging operations to a time series, to estimate parameters and undertake statistical inference. As is well known, the validity of the central limit theorem (CLT) and laws of large numbers depends critically on remote parts of a sequence being independent of each other, in an appropriate sense.

Uncorrelatedness at long range is not a sufficient condition to validate the CLT, and while the mixing property is often invoked, it is difficult to verify. However, the property of near-epoch dependence on a mixing process can suffice. This is the property that the error in the best predictor of the process, based on only the 'near epoch' of an underlying mixing process, is sufficiently small. Thus, letting $\mathcal{F}_s^t = \sigma(e_s, \dots, e_t)$, the sigma-field generated by the collection $\{e_j, s \le j \le t\}$, a process u_t is said to be L_p-near-epoch dependent (L_p-NED) on $\{e_t\}$ of size $-\lambda_0$ if

$$\|u_t - E(u_t|\mathcal{F}_{t-m}^{t+m})\|_p \le d_t m^{-\lambda} \tag{3.1}$$

for $\lambda > \lambda_0$. In the general definition, d_t is a sequence of positive constants, but subject to stationarity, as here, we can write simply $d_t = d < \infty$. See Davidson (1994, Chapter 17) among other references, for additional details. If $m^{-\lambda}$ can be replaced with ρ^{-m} in (3.1), then we will say that the process is geometrically NED. There is no accepted 'size' terminology associated with this case, but obviously we can speak of 'geometric size ρ' in a consistent manner, if convenient to do so.

In the present application, the process $\{e_t\}$ will be taken as the driving process in (1.1). Since this is not merely mixing but i.i.d., by assumption, the condition in (3.1) alone constrains the memory of the process. The application of this approach to a range of nonlinear processes, including GARCH processes, is studied in Davidson (2002). Following the same approach, we now derive conditions for u_t defined by (1.1) plus (1.2) to be, respectively, L_1 or L_2-near-epoch dependent (NED) on $\{e_t\}$. In the following result for the hyperbolic memory case, we formalise the model just to the extent of specifying the lag coefficients to be bounded by a regularly varying function. This strengthens the summability requirement, but only slightly.

Theorem 3.1 **(a)** *If $0 \le \theta_i \le Ci^{-1-\delta}$ for $i \ge 1$ for $C > 0$ and $\delta > \lambda_0 \ge 0$, and $S < 1$, then u_t is L_1-NED on $\{e_t\}$, of size $-\lambda_0$.*

(b) *If in addition* $S < \mu_4^{-1/2}$*, then* u_t *is* L_2*-NED of size* $-\lambda_0$.

The proof, given in the Appendix, follows GKL in working with a Volterra-type series expansion of the process to construct the near-epoch based predictor and bound its residual.

GKL have shown that the process $\{u_t^2\}$ has absolutely summable autocovariances, subject only to the condition $S < \mu_4^{-1/2}$. No separate constraint on the rate of convergence of the lag coefficients is specified in their result, although summability obviously requires $\delta > 0$, so that their conditions match those for L_2-NED of size 0. These authors also prove a central limit theorem for the process $\{u_t^2 - Eu_t^2\}$ subject only to the same condition on the sum. The CLT for L_2-NED processes of De Jong (1997), such as might be applied to $\{u_t\}$ using the present result, calls for $\lambda_0 = \frac{1}{2}$. This provides what to the author's knowledge is the best CLT currently available for ARCH(∞) processes. Extending the result of GKL to the same case is not trivial, since the reverse mapping from u_t^2 to u_t is not single valued, but there is the strong suggestion that still sharper conditions for the CLT for $\{u_t\}$ might be obtainable by exploiting the properties of the process more directly. Essentially, even with decay rates slower than $-\frac{1}{2}$, the restriction on the sum of the coefficients may force them to be individually so small that negligibility arguments can be applied to the tail of the lag distribution. This is an interesting direction for further research.

The following is the corresponding result for the geometric memory case.

Theorem 3.2 **(a)** *If* $0 \leq \theta_i \leq C\rho^{-i}$ *for* $i \geq 1$*, with* $0 < C < \rho$ *and* $\rho > 1$*, and* $S < 1$*, then* u_t *is geometrically* L_1*-NED on* $\{e_t\}$.

(b) *If in addition* $C < \rho\mu_4^{-1/2}$ *and* $S < \mu_4^{-1/2}$*, then* u_t *is geometrically* L_2*-NED.*

The GARCH(p, q) is the leading example of the geometric case, and this result may be compared with Proposition 2.3 of Davidson (2002). Note that since $\theta_1 \leq C\rho^{-1}$ and necessarily $\theta_1 < S$, the sufficient restrictions on C in Theorem 3.2 are minimal, in view of the restrictions on S. In effect, they forbid isolated influential lags of higher order. Inspection of the proof will show that they could be relaxed, at the cost of more complex or specialised conditions. The important point to note about both of these results is that the existence of second (fourth) moments is necessary and nearly sufficient for the L_1-NED (L_2-NED) property, respectively.

3.2 The Nonstationary Geometric Lag Case

Nelson (1990) gives an insightful analysis of persistence (memory) in the GARCH(1,1) model. The key condition he derives for limited persistence (what he would call non-persistence) is

$$E \ln(\beta_1 + \alpha_1 e_t^2) < 0 \tag{3.2}$$

and the Jensen inequality easily shows that the condition $\alpha_1 < 1 - \beta_1$ is sufficient for (3.2). This condition is necessary and sufficient for the process to be strictly stationary and ergodic (Nelson 1990, Theorem 2).

Necessary conditions for strict stationarity of the GARCH(1,1) depend upon the distribution of e_t, and are shown for the standard Gaussian case in Nelson's (1990) Figure 8.1, as a nonlinear trade-off between the values of α_1 and β_1. In that case, note that strict stationarity is compatible with non-existence of second moments, and Nelson's figure shows that, for example, $\alpha_1 > 3$ is permitted when β_1 is close enough to 0.

The NED measure of memory is unavailable without first moments, but an alternative is provided by Pötscher and Prucha's (1991) notion of L_0-approximability (see also Davidson (1994) Chapter 17.4). This is the condition that there exists a locally measurable (finite-lag) approximation to u_t, which is a uniform mixing process, given that e_t is independent (e.g. Davidson

1994 Theorem 14.1). Let h_t^m denote a $\mathcal{F}_{t-m}^{t+m}$-measurable approximation function (depending only on $e_{t-m}, \ldots, e_t$ in the present case). h_t^m is defined to be a geometrically L_0-approximator of σ_t^2 if

$$P(|\sigma_t^2 - h_t^m| > d_t\delta) = O(\rho^{-m}) \tag{3.3}$$

for $\rho > 1$ and all $\delta > 0$ where, subject to stationarity as assumed here, we may set $d_t = 1$. The following result can now be obtained.

Theorem 3.3 *Let u_t be a strictly stationary process, and $0 \leq \theta_i \leq C\rho^{-i}$ for $i \geq 1$ with $\rho > 1$. In either of the following cases, σ_t^2 is geometrically L_0-approximable.*

(a) $C < \rho$.

(b) $C < \rho(\rho - 1)$, and $\log(C^{1+\varepsilon}/(\rho^{1+\varepsilon} - 1)) < \zeta$ for some $\varepsilon > 0$, where $\zeta = E(-\log e_t^2)$.

Note that $S \leq C/(\rho - 1)$, and hence the restriction on C in part (b), and hence also that in part (a), implies that $S < \rho$. However, S can substantially exceed 1 in either case, if ρ is large enough. In addition, inspection of the proof will show that these conditions are only sufficient, and the L_0-approximability property still obtains in numerous cases where (a) and (b) are violated, but are awkward to state compactly.

Taking the GARCH(1,1) as an example, we find $\rho = \beta_1^{-1}$ and $C = \alpha_1/\beta_1$, so the condition to be satisfied in (a) is $\alpha_1 < 1$, and that in (b) is $\alpha_1 < 1/\beta_1 - 1$. It is an interesting question to relate the conditions of the theorem to conditions for strict stationarity such as (3.2). Kazakevičius and Leipus (2002, Theorem 2.3) have shown that $\log(C/(\rho - 1)) < \zeta$ is a necessary condition for strict stationarity, although not sufficient, since in the case of the GARCH(1,1) this is actually a weaker condition than (3.2). When $C < \rho$, then $C\rho^{-j} < 1$ for all $j \geq 1$, and also note that $C^{1+\varepsilon}/(\rho^{1+\varepsilon} - 1) < C/(\rho - 1)$ for $\varepsilon > 0$, which guarantees that the second condition in (b) holds in a stationary process. However, when $C \geq \rho$, condition (b) can evidently fail in a stationary process. Note that while the present proof establishes independence of initial conditions, it makes use of the stationarity assumption, and therefore cannot provide a proof of stationarity as such.

The conceptual importance of this result is chiefly to show the way in which short memory is a feature of the strictly stationary case, whether moments exist or not. From a more practical viewpoint, though, the property might be used in conjunction with mixing limit theorems to show that, for example, a law of large numbers applies to integrable transformations of the process, such as truncations. See Pötscher and Prucha (1991) and Davidson (1994) for more details of this approach.

4 The IGARCH and FIGARCH Models

The interesting feature of ARCH(∞) models revealed by the foregoing analysis is that the rate of convergence of the lag coefficients to zero is irrelevant to the stationarity property, provided these are summable. The key constraint is the relationship of their sum to unity. When this is equal to or exceeds unity, no second moments exist regardless of the memory of the process. The familiar example is the IGARCH(1,1) model, which according to Theorem 3.3 is geometrically L_0-approximable, in other words, short memory.

This appears paradoxical, since the IGARCH model is often spoken of in the literature as a 'long memory' model, the volatility counterpart of the unit root model of levels. Consider the k-step-ahead 'volatility forecast' from the model represented by

$$\sigma_t^2 = \frac{\gamma}{1 - \beta_1} + (1 - \beta_1) \sum_{j=1}^{\infty} \beta_1^j u_{t-j}^2$$

$$= \gamma + (1 - \beta_1)u_{t-1}^2 + \beta_1\sigma_{t-1}^2. \tag{4.1}$$

Applying the law of iterated expectations would appear to yield the solution

$$E_t u_{t+k}^2 = k\gamma + u_t^2 \tag{4.2}$$

which although diverging, remains dependent on current conditions even at long range. Thus, it appears that u_t^2 fulfils the condition to be long memory proposed by Granger and Terasvirta (1993, page 49), that is, to be forecastable in mean at long range. But this is a paradox, since it is clear that σ_t^2 depends on only the recent past. Theorem 3.3(a) shows that σ_t^2 can be reconstructed from the shock history, $\{e_{t-1}, e_{t-2,} \dots e_{t-m}\}$, with an error that vanishes at an exponential rate as m increases, so clearly it cannot be forecast at long range.

Ding and Granger (1996) discuss this apparent paradox of memory by considering the extreme case in which $\beta_1 = 0$ and $\gamma = 0$, so that (4.2) reduces to $E_t u_{t+k}^2 = u_t^2$. This model can be written in the form

$$u_t = e_t|u_{t-1}| \tag{4.3}$$

While a succession of larger-than-average independent shocks (e_ts) may produce very large deviations of the observed process, such that their variance is infinite, the e_t are still drawn independently from a distribution centred on zero. Note how a single 'small deviation' of e_t (having the highest probability density of occurrence, in general) kills a 'run' of high volatility instantly. That the probability of such an event occurring in (4.3) converges rapidly to 1, is the essential message of Theorem 3.3. Nelson (1990) showed that this particular process converges to zero in a finite number of steps, with probability 1.

The present results allow consideration of a still more extreme case, that of $\sigma_t^2 = \alpha u_{t-1}^2$ for $\alpha > 1$. Some substitutions yield

$$\sigma_t^2 = \alpha e_{t-1}^2 \sigma_{t-1}^2 = \cdots = \alpha^m e_{t-1}^2 e_{t-2}^2 \cdots e_{t-m}^2 \sigma_{t-m}^2. \tag{4.4}$$

Noting that in this case the sum of the lag coefficents is α (this can be treated as the limiting case as $\rho \to \infty$ and $C = \alpha\rho$), applying Theorem 3.3(b) shows that the steady state solution is $\sigma_t^2 = 0$ whenever $\log\alpha < -E\log e_t^2$. In this case, the right-hand side of (4.4) converges to zero in probability (in fact, with probability 1) as $m = t$ increases, starting from any fixed $\sigma_0^2 > 0$.

The straightforward solution to the paradox presented by these cases is that while σ_t^2 in (4.1) or (4.4) is a natural indicator of conditional volatility, depending on the near epoch, it is *not* the conditional variance. Since the unconditional variance does not exist in these cases, the conditional variance is not a well-defined random variable. Note that the application of the law of iterated expectations is not valid here, so that (4.2) has no meaningful interpretation. These examples highlight the important distinction to be maintained, between the moment and memory properties of a sequence.

The FIGARCH model defined by (1.9) is a generalization of the IGARCH, of particular interest not least because this is the one application to date employing hyperbolic lag weights. Note that $\sum_{i=1}^{\infty} a_j = 1$ in (1.10) for any value of d, and this therefore belongs to the same "knife-edge-nonstationary" class represented by the IGARCH, with which it coincides for $d = 1$. However, note the interesting and counterintuitive fact that the length of the memory of this process is *increasing* as d approaches zero.[2] This is of course the opposite of the role of d in the fractionally integrated process in levels. Note that when $d = 1$, then $a_1 = 1$, and $a_i = 0$

[2]Note the error in Baillie et. al. (1996), where the lag coefficients are said to be of $O(k^{d-1})$. (page 11, line 6). This should read $O(k^{-d-1})$.

for $i > 1$. In this particular case, of amplitude $S = 1$, the memory (measured by $-\delta$ in (2.2)) is discontinuous, jumping to $-\infty$ at the point where it attains -1.

At the other extreme, as d approaches 0, the lag weights are approaching non-summability. However, again because of the restriction $S = 1$, the individual a_i are all approaching 0. The limiting case $d = 0$ is actually another short-memory case, in this case the stable GARCH rather than the IGARCH represented by $d = 1$. At $d = 0$ the memory jumps from 0 to $-\infty$, and the amplitude is also discontinuous at this point, jumping from a fixed value of 1 to some value strctly below 1. The characterisation of the FIGARCH model as an intermediate case between the stable GARCH and the IGARCH, just as the I(d) process in levels is intermediate between I(0) and I(1), is therefore misleading. It in fact possesses more memory than either of these models, but behaves oddly owing to the rather arbitrary restriction of holding the amplitude to 1 (the knife-edge value) while the memory increases.

The term 'long memory' has been applied to the FIGARCH model by several authors, for understandable reasons, but our discussion has made clear that the analogy with models of the conditional mean is also misleading in this respect. To illustrate the dangers of taking the "AR-in-squares" characterization of these models too literally, consider the simplest FIGARCH model

$$\sigma_t^2 = \omega + (1 - (1 - L)^d)u_t^2 \tag{4.5}$$

rearranged as

$$(1 - L)^d u_t^2 = \omega + v_t \tag{4.6}$$

with $v_t = (e_t^2 - 1)\sigma_t^2$ as before. This equation might appear to represent u_t^2 as a classic fractionally integrated process. However, just as, in the absence of second moments, the temptation to write $E(v_t) = 0$ must be resisted, so it is important not to confuse this formal representation (in which v_t does not represent a forcing process, and is serially dependent) with the data generation process. Indeed, were we to replace v_t in (4.6) by (say) a zero-mean, independent disturbance for $t > 0$, and by zero for $t \leq 0$, we should actually obtain a nonstationary *trending* process, having expected value ωt^d for $t > 0$ (see Granger (2002)). This clearly contradicts what we know about the actual characteristics of the u_t^2 process.

As remarked above, GKL show that whenever fourth moments exist, the autocovariances of the squared process are always summable. Kazakevičius and Leipus (2002) further show that summability of the ARCH(∞) lag weights is a necessary condition for stationarity. Long memory in mean, characterized by nonsummable autocovariances, does not appear to have a well-defined counterpart in the ARCH(∞) framework, whether or not moments exist, because in any such cases the processes must rapidly diverge. The term 'hyperbolic memory' is therefore to be preferred, to distinguish FIGARCH from the geometric memory cases such as GARCH and IGARCH.

5 The HYGARCH Model

The unexpected behaviour of the FIGARCH model may be due less to any inherent paradoxes than to the fact that the unit-amplitude restriction, appropriate to a model of levels, has been transplanted into a model of volatility. In a more general framework there are good reasons to embed it in a class of models where such restrictions can be tested, and also to adhere to the approach of modelling amplitude and memory as separate phenomena, just as is done in the ordinary GARCH model.

These considerations lead us to propose the 'hyperbolic GARCH', or HYGARCH model. Consider, for comparability with the previous cases, the form

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 + \alpha((1 - L)^d - 1)) \qquad \alpha \geq 0,\ d \geq 0. \tag{5.1}$$

Note that provided $d > 0$,

$$S = 1 - \frac{\delta(1)}{\beta(1)}(1-\alpha). \tag{5.2}$$

The FIGARCH and stable GARCH cases correspond to $\alpha = 1$ and $\alpha = 0$ respectively, and in principle, the hypothesis of either of these two pure cases might be tested. However, in the latter case the parameter d is unidentified, which poses a well-known problem for constructing hypothesis tests. Therefore, also note that when $d = 1$, (5.1) reduces to

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1-\alpha L) \qquad \alpha \geq 0. \tag{5.3}$$

In other words, when $d = 1$ the parameter α reduces to an autoregressive root, and hence the model becomes either a stable GARCH or IGARCH, depending on whether $\alpha < 1$ or $\alpha = 1$. For this reason, testing the restriction $d = 1$ is the natural way to test for geometric versus hyperbolic memory. Also note that $\alpha > 1$ is a legitimate case of nonstationarity. For example, in the case where $\delta(L) = 1$ and $\beta(L) = 1 - \beta_1 L$, the model reduces when $d = 1$ to the covariance-nonstationary GARCH(1,1) discussed in Section 3.2, with α corresponding to $\alpha_1 + \beta_1$, in the notation adopted there.

When d is not too large, this model will correspond closely to the case

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1-\alpha\phi(L)) \tag{5.4}$$

where

$$\phi(L) = \zeta(1+d)^{-1}\sum_{j=1}^{\infty} j^{-1-d}L^j \qquad d > 0. \tag{5.5}$$

and $\zeta(.)$ is the Riemann zeta function. However, note that the models behave quite differently when d is close to 1. In (5.1) $d > 1$ gives rise to negative coefficients and so is not permitted, whereas in (5.5) d can take any positive value, and the model approaches the GARCH case only as $d \to \infty$. It can therefore encompass a range of hyperbolic lag behaviour excluded by (5.1). In practice this is probably not a serious restriction, because it will become increasingly difficult to discriminate between hyperbolic decay, and geometric decay represented by $\delta(L)/\beta(L)$, when d is very large. Nonetheless, in this context it appears an arbitrary restriction to use the hyperbolic decay pattern implied by (1.11) rather than to use weights directly proportional to j^{-1-d}. The chief motivation for using (5.1) must be to nest the FIGARCH and IGARCH cases, but should d be found close to 1, the option of comparing GARCH with (5.4) might be considered.

If the GARCH component observes the usual covariance stationarity restrictions, which imply $\delta(1)/\beta(1) > 0$, then with $\alpha < 1$ these processes are covariance stationary and L_1-NED of size $-d$, according to Theorem 3.1. They are also L_2-NED of size $-d$ if $(1-\alpha)\delta(1)/\beta(1) > 1 - \mu_4^{-1/2}$. For example, with Gaussian disturbances we have $1 - \mu_4^{-1/2} = 0.422$. Therefore, noting the discussion of Section 3.1, the central limit theorem holds at least for $d > \frac{1}{2}$ in that case.

6 Applications

Two applications of the HYGARCH model are discussed in this section. The first is a rather conventional one whose aim is to relate these models to the substantial existing literature on modelling exchange rates. The second is more unusual, and possibly controversial, in which our aim is to argue that these models may have a distinctive and important role to play in more difficult cases.

6.1 Dollar Exchange Rates 1980-96

Table 1 summarises estimates of the HYGARCH model for a collection of the (logarithms of) major dollar exchange rates.[3] The data are in each case daily for the period 1st January 1980 to 30th September 1996 (4370 observations).[4] The model fitted to all the series is a first-order ARFI-HYGARCH, taking the form

$$(1-L)^{d_{ARF}}(1-\phi_1 L)Y_t = \mu + u_t \tag{6.1}$$

$$h_t = \omega + \Big(1 - \frac{1-\delta_1 L}{1-\beta_1 L}(1+\alpha((1-L)^{d_{FG}}-1))\Big)u_t^2 \tag{6.2}$$

Robust standard errors are shown in parentheses. The estimates of μ and ω have been omitted to save space. In the interests of comparability, the same model is fitted to all the series, even though in some cases parameters are insignificant.

It may appear surprising to model exchange rates with long memory in mean but this turns out, with d_{ARF} suitably small, to be a good parsimonious representation of the autocorrelation. This is not negligible, but is not concentrated at low orders of lag, so that the geometric memory decay of ARMA components cannot capture it. In view of the characteristic incidence of outliers in these data, the Student's t distribution is assumed for the disturbances, rather than the normal. The criterion function for estimation is the Student's t log-likelihood

$$L_T = T\log\frac{\Gamma((\nu+1)/2)}{\sqrt{\pi(\nu-2)}\Gamma(\nu/2)} - \frac{1}{2}\sum_{t=1}^{T}\Big(\log h_t + (\nu+1)\log\Big(1+\frac{u_t^2}{(\nu-2)h_t}\Big)\Big).$$

As a practical matter, observe that small innovations u_t contribute to this criterion in much the same way as to the Gaussian log-likelihood, but large innovations (such that $u_t^2/(\nu-2)h_t \gg 1$) make a much smaller contribution to the aggregate than in the Gaussian case, depending on the size of ν.

The last two columns of the table show the Box-Pierce (1979) $Q(r)$ statistic for $r = 25$ lags, and the Q^2 which is the Q statistic computed from the squared residuals. This test was proposed by McLeod and Li (1983) and studied by Li and Mak (1994) for application to testing neglected heteroscedasticity in ARCH residuals. The latter authors show that using the nominal chi-squared distribution with r degrees of freedom would give an excessively conservative test, similarly to the Box-Pierce result for ARMA residuals. The asymptotic distribution of these statistics, for the cases of hyperbolic lags in mean and variance respectively, has not yet been studied, so both must be treated with caution, as diagnostic tests. What can be remarked is that an examination of the residual correlograms in each case tends to show the largest (absolute) values at rather high lags (10 or 15 is typical). The neglected autocorrelation, in levels or squares, cannot therefore be accounted for by simply adding terms to the ARMA or GARCH components, a conclusion reinforced by conventional significance tests.

Caution must also be observed in interpreting conventional confidence intervals, since although the samples are large, the asymptotic properties of the estimates are not yet well established. Lumsdaine (1996) and Lee and Hansen (1994) have considered the IGARCH(1,1) case, and hence shown that covariance stationarity of the processes is not a necessary condition for consistency and asymptotic normality of the usual QML estimator. However, note that the conjecture of Baillie et al. (1996, page 9), to the effect that the properties of the FIGARCH model are subsumed under those of the IGARCH model, is in doubt in view of the analysis of this paper.

---

[3] The estimation has been carried out using the Ox 3.20 package Time Series Modelling Version 3.0 (called Long Memory Modelling in earlier releases). See Davidson (2003) and Doornik (1999).

[4] All the series used in this study are sourced from Datastream.

Currency	d_{FG}	α	δ_1	β_1	d_{ARF}	ϕ_1	$\nu^{1/2}$	$Q(25)$	$Q^2(25)$
Danish Kroner	0.600 (0.081)	0.962 (0.021)	0.188 (0.045)	0.722 (0.051)	0.056 (0.017)	−0.055 (0.021)	2.31 (0.092)	25	34
Deutschmark	0.681 (0.077)	0.975 (0.017)	0.190 (0.047)	0.789 (0.038)	0.045 (0.022)	−0.038 (0.022)	2.25 (0.082)	27	23
Finnish Mark	0.714 (0.079)	0.946 (0.023)	0.177 (0.054)	0.782 (0.046)	0.013 (0.015)	−0.043 (0.020)	2.22 (0.109)	29	1.24
GB Pound	0.656 (0.103)	0.991 (0.017)	0.229 (0.056)	0.795 (0.056)	0.016 (0.016)	0.005 (0.022)	2.27 (0.0907)	32	31
Irish Punt	0.641 (0.087)	0.991 (0.018)	0.241 (0.057)	0.787 (0.048)	0.030 (0.017)	−0.049 (0.022)	2.23 (0.087)	27	15
Italian Lire	0.556 (0.075)	0.991 (0.025)	0.264 (0.052)	0.699 (0.064)	0.038 (0.015)	−0.0045 (0.021)	2.20 (0.077)	37	32
Japanese Yen	0.564 (0.144)	0.952 (0.041)	0.253 (0.071)	0.696 (0.091)	0.045 (0.015)	−0.081 (0.020)	1.96 (0.059)	34	23
Port. Escudo	0.613 (0.162)	0.986 (0.029)	0.291 (0.081)	0.747 (0.099)	0.026 (0.013)	−0.079 (0.019)	1.99 (0.074)	28	1.44
Spanish Peseta	0.515 (0.060)	1.04 (0.021)	0.258 (0.044)	0.697 (0.047)	0.039 (0.020)	−0.067 (0.020)	2.20 (0.083)	25	13
Swiss Franc	0.819 (0.092)	0.956 (0.017)	0.100 (0.059)	0.835 (0.046)	0.027 (0.017)	−0.015 (0.022)	2.38 (0.097)	23	16

Table 1: The ARFI-HYGARCH model of Exchange Rates.

Even with these caveats in mind, these results show a remarkable degree of uniformity. The point estimates of each parameter seem to differ by hardly more than the sampling error to be expected from identical data generation processes. Since these are rates of exchange determined in closely related markets, this is not perhaps unexpected. Some of these currencies were of course in the Exchange Rate Mechanism for some part of the sample period, and to the extent that they were tied together, may be expected to move similarly against the dollar. However, the exceptional cases (Yen, Swiss franc) do not appear to diverge from the general pattern. We therefore conjecture that the similarity of these structures goes deeper than the fact of some correspondence in their movements in levels

Looking now at the estimates themselves, we note that while the d_{ARF} estimates are small, they are generally significant. On the other hand the hyperbolic memory in variance, measured by d_{FG}, is generally pronounced. The amplitude parameter α is in most cases not significantly different from 1, while generally a little below it. The FIGARCH model will clearly explain these data pretty well. Also, note that the estimate of d for the Deutschmark is similar to the FIGARCH estimations reported in Baillie et al. (1996), and also Beltratti and Morana (1999). A noteworthy feature is that the Student's t degrees of freedom parameter, ν, is generally close to its lower bound, corresponding to $\nu^{1/2} > 1.414$.

6.2 The Asian Crisis

The second application we consider is to the dollar exchange rates for three Asian currencies, for periods covering the Asian crisis of 1997-98. The series in question, in logarithms, are shown in the panels labelled (i) of Figures 1–3. At first sight, it might appear that these data represent two quite distinct regimes. Prior to the crisis, the Won and the Rupiah, at least, appear to be following a creeping peg to the dollar. After the crisis, they are floating and subject to violent fluctuations. The hypothesis that the same time series model might account for both periods is evidently a strong one.

However, it is not wholly unreasonable. We may view these models as representing mechanisms by which exchange markets filter new information, in the process of forming a price. The new information takes the form, by hypothesis (or by definition, even) of an independent random

	13/12/94–15/6/00 (1424 obs.)	13/12/94–16/10/97 (730 obs.)	17/10/97–15/6/00 (694 obs.)
d_{FG}	0.669 (0.046)	0.667 (0.121)	0.686 (0.066)
α	1.252 (0.149)	1.265 (0.275)	1.226 (0.177)
β_1	0.339 (0.092)	0.318 (0.143)	0.363 (0.140)
$\omega^{1/4}$	0.0184 (0.0018)	0.0186 (0.0026)	0.021 (0.0091)
d_{ARF}	0.073 (0.031)	0.072 (0.038)	0.076 (0.059)
ϕ_1	0.116 (0.039)	0.110 (0.051)	0.122 (0.068)
ϕ_2	−0.097 (0.028)	−0.044 (0.038)	−0.152 (0.044)
$\nu^{1/2}$	1.73 (0.080)	1.71 (0.113)	1.74 (0.116)
Kurtosis	8.65	7.16	7.93
$Q(25)$	44	21	34
$Q_{sq}(25)$	39	17	47
LM(δ_1)	0.204	–	–
QLL	7421.85	4314.09	3109.97
	LR Statistic(8 d.f.) = 4.42		

Table 2: Korean Won

sequence. The distribution of this sequence, and the time series model, are distinct contributing factors in the formation of the series. It may be that when unusual events occur the model changes, but a simpler hypothesis that it does not.

Tables 2–4 show estimated models for the three currencies. The models were selected by individual specification searches on the complete samples, and parameters not shown in the tables were restricted to zero. LM statistics for the exclusion of some additional dynamic parameters are shown, to justify these choices. The intercepts in the mean processes were never significantly different from zero, when fitted, and are constrained to zero in these estimates.[5]

As well as the full sample, the same models were also fitted to "pre-crisis" and "post-crisis" subsamples. The break-points were in each case chosen by eye, at the point just preceding the first large fall of the currency.[6] These are marked by the vertical lines in the figures, which show, in panels (ii), (iii) and (iv) respectively, the estimated series $\hat{u}_t$, $\hat{u}_t/\hat{\sigma}_t$ and $\hat{\sigma}_t^2$ from the HYGARCH model, in each case.

Three conclusions emerge from examination if these results. First, the three structures estimated are not wholly dissimilar, but each has distinctive features. In particular, the Won exhibits quite a complex structure of autocorrelation, although this may be due to the fact that the higher leptokurtosis of the other two shock series has the effect of masking autocorrelation that may be present. Second, however, they more closely resemble each other than the currencies analysed in Table 1. The most noteworthy feature is of course the large values of the α parameter in each case, but especially in the cases of the Rupiah and Taiwan dollar.

Third, and perhaps most remarkable, is the stability of these models across the pre- and post-crisis regimes. In all three cases the large α value is common to both periods, and the other parameters are also generally close. The last line of each table shows the likelihood ratio statistic

[5] Note that the fourth root of ω has been estimated. Since ω is small in these models, and hence very close to the boundary of the parameter space, this transformation is found to improve the numerical stability of the estimation algorithm and, likewise, the difference approximations to derivatives used in standard error calculations.

[6] Using the methodology of Lavielle and Moulines (2000). Andreou and Ghysels (2002) detect multiple breaks in the volatility dynamics of stock market indices during the Asian crisis. In this instance, however, there is inevitably a moment at which the monetary authorities allow the currencies to float freely, leading, under the conditions of the crisis, to precipitate devaluations. It is these events that we take to mark the regime switch dates.

	01/1/96–31/12/99 (1045 obs.)	01/1/96–11/7/97 (400 obs.)	14/7/97–31/12/99 (645 obs.)
d_{FG}	0.496 (0.042)	0.588 (0.104)	0.54 (0.054)
α	2.94 (0.85)	1.81 (1.49)	2.92 (1.59)
$\omega^{1/4}$	0.0099 (0.0015)	0.011 (0.002)	0.037 (0.007)
ϕ_1	0.047 (0.032)	−0.038 (0.055)	0.097 (0.041)
$\nu^{1/2}$	1.52 (0.038)	1.54 (0.120)	1.51 (0.071)
Kurtosis	14.6	20.4	10.1
$Q(25)$	37	22	29
$Q_{sq}(25)$	27	19	22
LM(d_{ARF})	0.012	–	–
LM(δ_1)	2.31	–	–
QLL	5889.68	2978.46	2917.85
	LR Statistic(5 d.f.) = 12.18		

Table 3: Indonesian Rupiah

	03/01/94–15/6/00 (1683 obs.)	03/01/94–15/10/97 (988 obs.)	16/10/97–15/6/00 (695 obs.)
d_{FG}	0.860 (0.079)	1.001 (0.010)	0.667 (0.073)
α	2.96 (0.466)	2.956 (0.466)	2.946 (0.877)
δ_1	0.242 (0.138)	0.009 (0.187)	0.568 (0.221)
β_1	0.635 (0.043)	0.606 (0.042)	0.726 (0.136)
$\omega^{1/4}$	0.021 (0.006)	0.022 (0.006)	0.000037 (0.004)
ϕ_1	−0.075 (0.024)	−0.131 (0.031)	-0.007 (0.037)
$\nu^{1/2}$	1.47 (0.010)	1.46 (0.010)	1.46 (0.019)
Kurtosis	336	21	184
$Q(25)$	9.40	49	5.27
$Q_{sq}(25)$	0.21	23	0.11
LM(d_{ARF})	0.037	–	–
QLL	8631.47	5263.64	3380.61
	LR Statistic(7 d.f.) = 25.56		

Table 4: Taiwan Dollar

for the test of model stability across the sample. Note that this cannot be interpreted as an asymptotic chi-squared test, because the break points have been chosen with reference to the data — that is, the most extreme contrast has been drawn in each case. Therefore, the correct null distribution of this statistic is the distribution of the maximum log-likelihood ratio over all break points. These critical values must exceed the nominal chi-squared values. The statistic for the Won is actually within the nominal acceptance region for the 5% test, and that for the Rupiah only slightly outside it. Overall, these results provide little evidence for changes of the model following the crisis, and the residual plots in panel (iii) of the figures (from the model fitted to the full sample) provide another view of this evidence. In two out of three cases, at least, it would appear impossible to detect the break point with confidence 'by eye'.

One further piece of evidence on the performance of these models is presented in Figure 4. This is a simulation using the model of the Korean Won, driven by shocks randomly resampled from the residuals of the same model, as shown in panel (iii) of the figure. The data shown were generated after letting the process run for 2000 pre-sample periods, to remove dependence on initial conditions. A "crisis" was introduced by inserting into the (otherwise randomly drawn) sequence a succession of five positive shocks, beginning at period 801. The values arbitrarily chosen were 4.2, 6.0, 3.3, 2, and 5.1, expressed in standard deviations since that of the shock distribution is 1 by construction. Such a realization would be a fairly rare event under random resampling, though major exchange crises are similarly rare, so this is not inappropriate.

There is, of course, no suggestion that the model (essentially, a heteroscedastic random walk) always generates runs of this appearance. Several repetitions of the experiment were required to produce the case illustrated, selected for its resemblance to the observed data. The point to be made here is merely that the observed data, taken as a whole, are *compatible* with this type of data generation process. Specifically, the pre-crisis 'pegged-rate' segments of the series in Figures 1(i)–3(i), while they might appear 'stationary', are actually well explained by a $I(1+d)$ process, provided the innovations are small enough. To switch to the post-crisis behaviour, all that is required are some unusually large shocks, and a conditional variance process with hyperbolic memory and large amplitude. As can be seen, the resulting pattern of high volatility can persist without further external stimulus, for scores and even hundreds of periods.

This analysis points to the possibility that the behaviour of currency markets as filters of new information, could be simpler in structure than many observers seem to believe. The natural rivals for the type of model presented here feature exogenous variables, either measured variables or dummies indicating the new environment, or alternatively are Markov-switching (SWARCH) models in which the *deus ex machina* takes the form of an autonomous stochastic process, to provide the switching mechanism (see for example Hamilton and Susmel 1994). What we aim to show is that, while any of these may be the true explanations, there is no *necessity* to introduce them. The crisis behaviour can be well described by a very simple endogenous mechanism, driven solely by the information contained in the shock process itself.

7 Conclusion

In this paper, conditions have been derived for the existence of moments and near-epoch dependence of the general class of ARCH(∞) processes. This class includes the GARCH, IGARCH and FIGARCH models, among other alternatives. It has been argued that the properties of these processes should be represented as varying in the two dimensions of amplitude and memory, relating respectively to the magnitude of the sum of the lag coefficients and their rate of convergence. The proposed HYGARCH model, generalizing the FIGARCH model, permits both the existence of second moments on the one hand, and on the other hand, even more extreme amplitudes than the simple IGARCH and FIGARCH models permit. The application of the model to exchange

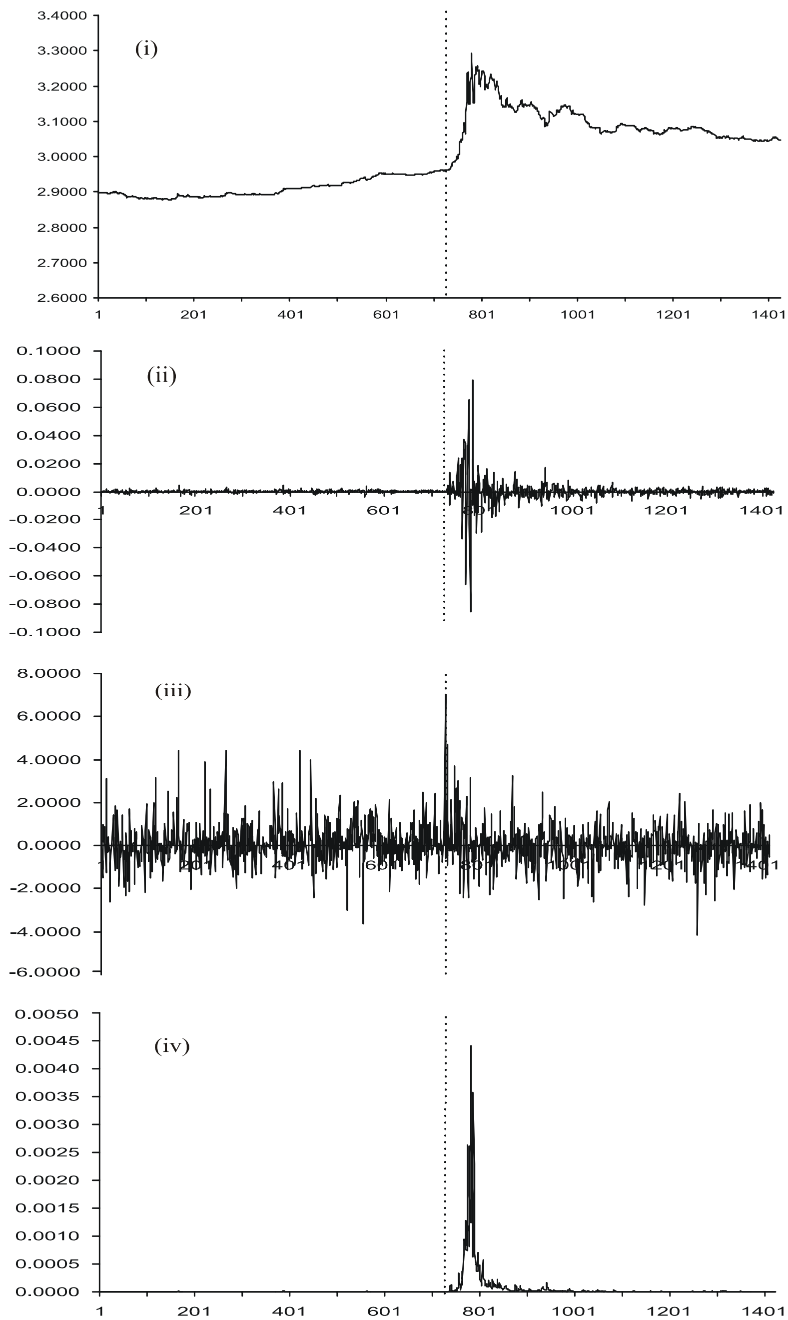

Figure 1: Korean Won

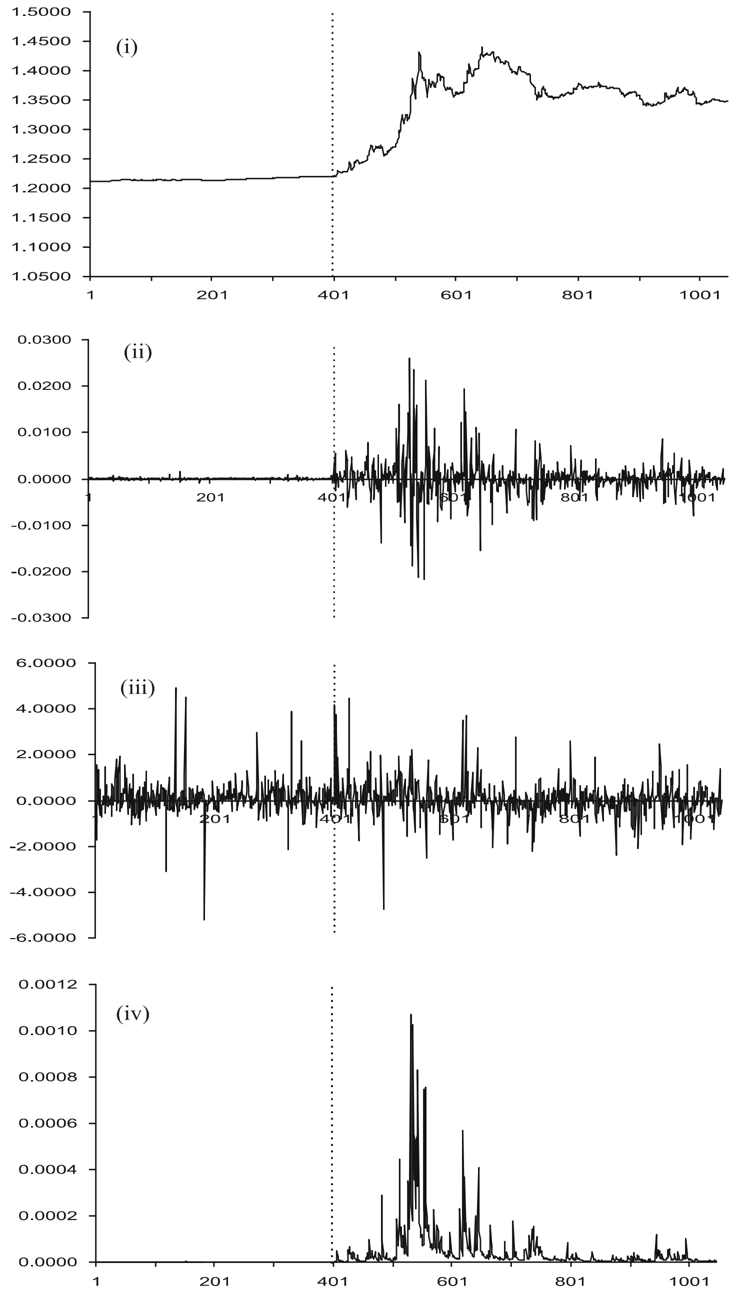

Figure 2: Indonesian Rupiah

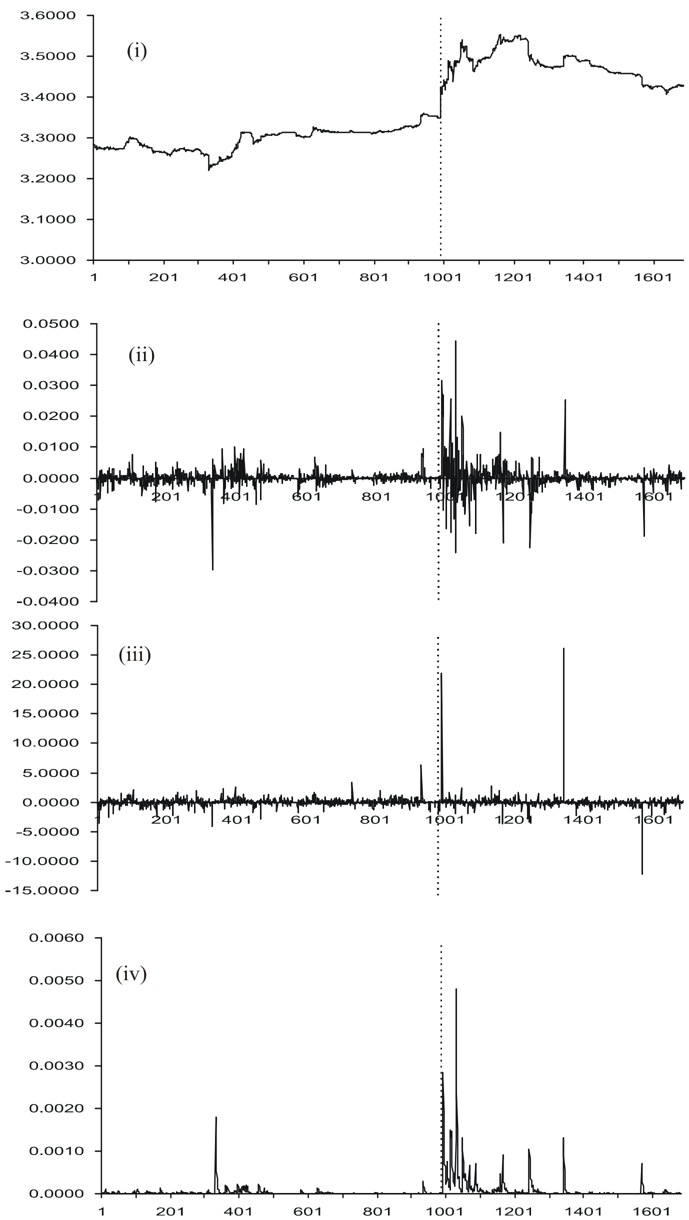

Figure 3: Taiwan Dollar

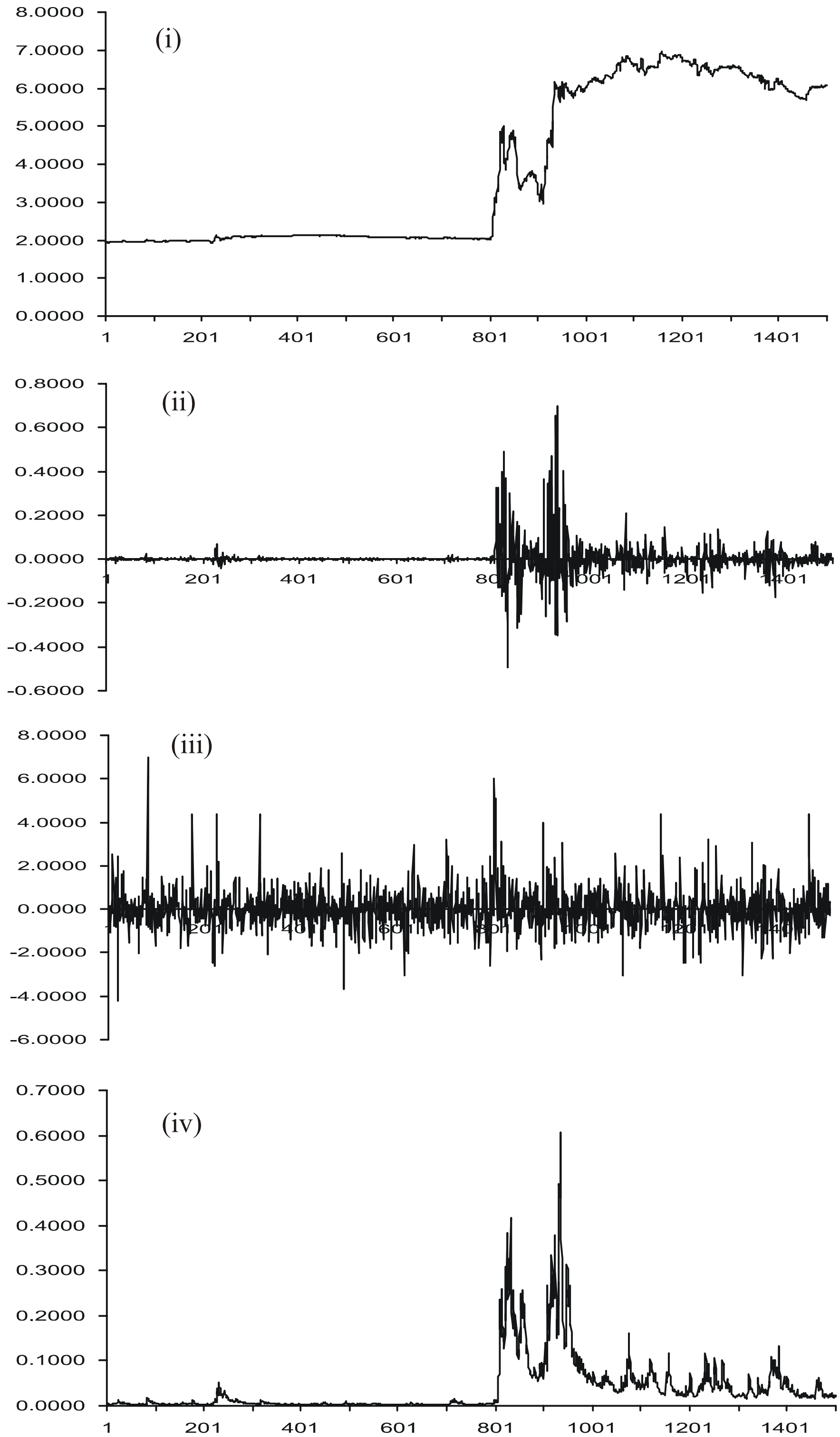

Figure 4: Korean Won model, simulation

rates is illustrated by two contrasting sets of examples.

An important implication of the results is the danger of pressing too far the analogy between integrated and fractional models in levels, and models whose conditional variances have an apparently similar dynamic structure. The relationship between the degree of persistence and the wide-sense stationarity of the process obeys very different rules in the two cases. The IGARCH is a short memory process having no variance, the FIGARCH(d) model has shortest memory with d closest to 1, and in general ARCH(∞) processes can be persistent and yet wide-sense stationary.

A further implication is the inability of the ARCH(∞) class to represent the degree of persistence commonly called 'long memory'. An interesting reference on this question is Andersen and Bollerslev (1997). These authors consider a semiparametric model in which the spectrum of absolute returns is treated as unbounded at the origin, diverging like $|\omega|^{-2d}$ for a parameter $d > 0$, as $\omega \to 0$. Their d is estimated by the Geweke-Porter-Hudak (1983) procedure from high frequency data under varying degrees of time aggregation. It is important to stress that this d parameter is *different* from the d parameter defined by our equations (1.9) or (5.1). A spectrum diverging at the origin implies non-summable autocorrelations, which are not permitted in the ARCH(∞) framework. If true long memory in variance is to be represented parametrically, this will have to be in the context of a different class of models.

The exponential ARCH(∞) class, in which $\log h_t$ is modelled by a distributed lag of some appropriate indicator of realized volatility, could be a plausible candidate for this role. Andersen and Bollerslev (1997) derive an exponential model, but also argue that persistence characteristics should be preserved under monotone increasing transformations, so that the spectrum of absolute returns should capture the same long memory characteristics. The EGARCH model of Nelson (1991) may be taken as a case in point. Here, $\log h_t$ is represented as an infinite moving average of a function $g(e_t)$, where e_t is the i.i.d. driving process. If we were to allow equation (5.1) to represent (in a purely formal way) the lag structure in this model, the interpretation of the coefficients α and d would of course be entirely different. In particular, α would be irrelevant to the existence of moments, and $d < 0$ (such that the lag coefficients are not absolutely summable) is not necessarily incompatible with stationarity. One might even speculate that the parameter estimated by Andersen and Bollerslev (1997) corresponds to $-d$, in this setup. To investigate these issues goes well beyond the scope of the present paper, but represents an interesting avenue for further research.

A Appendix

A.1 Proof of Theorem 3.1

Since $\sigma_t^2 \geq \omega$ and $E_{t-m}^{t+m}\sigma_t^2 \geq \omega$, the inequality

$$\|u_t - E_{t-m}^{t+m}u_t\|_p \leq \omega^{-1/2}\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\|_p$$

follows by a minor extension of Lemma 4.1 of Davidson (2002), replacing 2 by $p \geq 1$. Therefore, in view of stationarity, it suffices to prove the inequalities

$$\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\|_p \leq C_p m^{-\delta} \tag{A-1}$$

for $C_p > 0$, for $p = 1$ and $p = 2$.

By repeated substitution we obtain, for given m, the decomposition

$$\sigma_t^2 = \omega + \sum_{j=1}^{\infty} \theta_j e_{t-j}^2 \sigma_{t-j}^2$$

$$= \omega\Bigg(1 + \sum_{j_1=1}^{\infty} \theta_{j_1} e^2_{t-j_1} + \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty} \theta_{j_1}\theta_{j_2} e^2_{t-j_1} e^2_{t-j_1-j_2} + \cdots$$

$$\cdots + \sum_{j_1=1}^{\infty} \cdots \sum_{j_m=1}^{\infty} \theta_{j_1}\cdots\theta_{j_m} e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_m}\Bigg)$$

$$+ \sum_{j_1=1}^{\infty} \cdots \sum_{j_{m+1}=1}^{\infty} \theta_{j_1}\cdots\theta_{j_{m+1}} e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_{m+1}} \sigma^2_{t-j_1-\cdots-j_{m+1}}. \qquad \text{(A-2)}$$

To prove part (a), first note that

$$E|e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_p} - E(e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_p}|\mathcal{F}^{t+m}_{t-m})| \begin{cases} = 0 & j_1+\cdots+j_p \le m \\ \le 2 & \text{otherwise} \end{cases}$$

using the Jensen inequality, and law of iterated expectations in the second case. Similarly,

$$E|e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_{m+1}}\sigma^2_{t-j_1-\cdots-j_{m+1}} - E(e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_{m+1}}\sigma^2_{t-j_1-\cdots-j_{m+1}}|\mathcal{F}^{t+m}_{t-m})| \le 2M_2$$

where M_2 is defined following (2.4). Next, define

$$\begin{aligned} T_p &= \sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty} 1_{\{j_1+\cdots+j_p>m\}}(j_1,\ldots,j_p)\theta_{j_1}\cdots\theta_{j_p} \\ &\le \Bigg(\sum_{j=m/p+1}^{\infty}\theta_j\Bigg)S^{p-1} \\ &\le C\Bigg(\sum_{j=m/p+1}^{\infty} j^{-1-\delta}\Bigg)S^{p-1} \\ &= O(p^\delta m^{-\delta} S^{p-1}) \end{aligned}$$

where the first inequality uses the fact that $\max\{j_1,...,j_p\} > m/p$ when $j_1+\cdots+j_p > m$. Since $S<1$, applying the triangle inequality yields

$$\begin{aligned} E|\sigma^2_t - E^{t+m}_{t-m}\sigma^2_t| &\le 2\omega\sum_{p=1}^{m} T_p + 2M_2 S^{m+1} \\ &= O\Bigg(m^{-\delta}\sum_{p=1}^{m} p^\delta S^{p-1}\Bigg) \\ &= O(m^{-\delta}). \end{aligned}$$

To prove part (b), note similarly that

$$\|e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_p} - E(e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_p}|\mathcal{F}^{t+m}_{t-m})\|_2 \begin{cases} = 0 & j_1+\cdots+j_p \le m \\ \le 2\mu_4^{p/2} & \text{otherwise} \end{cases}$$

whereas

$$\begin{aligned} \|e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_{m+1}}\sigma^2_{t-j_1-\cdots-j_{m+1}} - E(e^2_{t-j_1}\cdots e^2_{t-j_1-\cdots-j_{m+1}}\sigma^2_{t-j_1-\cdots-j_{m+1}}|\mathcal{F}^{t+m}_{t-m})\|_2 \\ \le 2\mu_4^{(m+1)/2} M_4^{1/2}. \end{aligned}$$

Therefore, Minkowski's inequality gives

$$\begin{aligned}\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\|_2 &\leq 2\omega\sum_{p=1}^{m} T_p\mu_4^{p/2} + 2\mu_4^{(m+1)/2}M_4^{1/2}S^{m+1}\\ &= O\Big(m^{-\delta}\sum_{p=1}^{m} p^{\delta}(\mu_4^{1/2}S)^{p-1}\Big)\\ &= O(m^{-\delta}).\end{aligned}$$

∎

A.2 Proof of Theorem 3.2

The proof of Theorem 3.1 is modified as follows. For part (a), note that since $\rho > 1$ and $S < 1$, there exists $\varepsilon > 0$ such that

$$\tilde{S} = \sum_{j=1}^{\infty}\theta_j^{1-\varepsilon} < 1. \tag{A-3}$$

There is no loss of generality in setting $1 < C < \rho$. In this case, defining $\tilde{\rho} = \rho^{\varepsilon} > 1$ and $\tilde{C} = C^{\varepsilon}$, note that $1 < \tilde{C} < \tilde{\rho}$. Then, note that

$$\begin{aligned}T_p &= \sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty} 1_{\{j_1+\cdots+j_p>m\}}(j_1,\ldots,j_p)\theta_{j_1}\cdots\theta_{j_p}\\ &= \sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty} 1_{\{j_1+\cdots+j_p>m\}}(j_1,\ldots,j_p)|\theta_{j_1}\cdots\theta_{j_p}|^{\varepsilon}|\theta_{j_1}\cdots\theta_{j_p}|^{1-\varepsilon}\\ &\leq \tilde{C}^p\tilde{\rho}^{-m}\tilde{S}^p\\ &= \tilde{C}^{p-m}\tilde{S}^p(\tilde{C}\tilde{\rho}^{-1})^m.\end{aligned} \tag{A-4}$$

Therefore,

$$\begin{aligned}E|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2| &\leq 2\omega\sum_{p=1}^{m} T_p + 2M_2|\theta_{j_1}\cdots\theta_{j_{m+1}}|^{\varepsilon}\tilde{S}^{m+1}\\ &= O(m\max\{\tilde{C}^{-m},\tilde{S}^m\}(\tilde{C}\tilde{\rho}^{-1})^m\}).\end{aligned} \tag{A-5}$$

To prove part (b), choose $\varepsilon > 0$ such that $\tilde{S} = \sum_{j=1}^{\infty}\theta_j^{1-\varepsilon} < \mu_4^{(\varepsilon-1)/2}$. On the assumptions, C can be chosen w.l.o.g. such that $1 < \mu_4^{\varepsilon/2}\tilde{C} < \tilde{\rho}$. Therefore,

$$\begin{aligned}\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\|_2 &\leq 2\omega\sum_{p=1}^{m} T_p\mu_4^{p/2} + 2\mu_4^{(m+1)/2}M_4^{1/2}|\theta_{j_1}\cdots\theta_{j_{m+1}}|^{\varepsilon}\tilde{S}^{m+1}\\ &= O(m\max\{(\mu_4^{\varepsilon/2}\tilde{C})^{-m},(\tilde{S}\mu_4^{(1-\varepsilon)/2})^m\}(\mu_4^{\varepsilon/2}\tilde{C}\tilde{\rho}^{-1})^m).\end{aligned}$$

∎

A.3 Proof of Theorem 3.3

Let

$$h_t^m = \omega\Big(1 + \sum_{j_1=1}^{m}\theta_{j_1}e_{t-j_1}^2 + \sum_{j_1=1}^{m}\sum_{j_2=1}^{m-j_1}\theta_{j_1}\theta_{j_2}e_{t-j_1}^2e_{t-j_1-j_2}^2 + \cdots$$

$$+\sum_{j_1=1}^{m}\cdots\sum_{j_m=1}^{m-j_1-\cdots j_{m-1}}\theta_{j_1}\cdots\theta_{j_m}e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_m}^2\Big).$$

Then, using (A-2),

$$\begin{aligned}\sigma_t^2-h_t^m=\omega\Big(&\sum_{j_1=m+1}^{\infty}\theta_{j_1}e_{t-j_1}^2+\sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}1_{\{j_1+j_2>m\}}(j_1,j_2)\theta_{j_1}\theta_{j_2}e_{t-j_1}^2e_{t-j_1-j_2}^2+\cdots\\ &+\sum_{j_1=1}^{\infty}\cdots\sum_{j_m=1}^{\infty}1_{\{j_1+\cdots+j_m>m\}}(j_1,\ldots,j_m)\theta_{j_1}\cdots\theta_{j_m}e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_m}^2\Big)\\ &+\sum_{j_1=1}^{\infty}\cdots\sum_{j_{m+1}=1}^{\infty}\theta_{j_1}\cdots\theta_{j_{m+1}}e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2\\ =\ &\omega(U_1+\cdots+U_m)+V_m\end{aligned}$$

where the last equality defines $U_1,\ldots,U_m$ and V_m. The dependence of these terms on t is implicit, but not indicated for ease of notation. By subadditivity,

$$P(|\sigma_t^2-h_t^m|>\delta)\le P(\omega|U_1+\cdots+U_m|>\delta/2)+P(V_m>\delta/2)$$

and the approach is to bound each term separately. First, note that $E|U_p|=T_p$, as defined in (A-4). Under the assumptions,

$$\begin{aligned}T_p&\le C^p\rho^{-m}\sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty}1_{\{j_1+\cdots+j_m>m\}}(j_1,\ldots,j_m)\rho^{m-j_1-\cdots-j_p}\\ &=O(S^p\rho^{-m})\end{aligned}\tag{A-6}$$

and note that $S\le C/(\rho-1)$. Therefore, by subadditivity and the Markov inequality,

$$\begin{aligned}P\Big(\omega\sum_{p=1}^{m}|U_p|>\delta/2\Big)&\le P\Big(\bigcup_{p=1}^{m}\Big\{U_p>\frac{\delta\omega}{2m}\Big\}\Big)\\ &\le\sum_{p=1}^{m}P\Big(U_p>\frac{\delta\omega}{2m}\Big)\\ &\le\frac{2m}{\delta\omega}\sum_{p=1}^{m}T_p=O\Big(m^2\Big(\frac{C}{\rho(\rho-1)}\Big)^m\Big).\end{aligned}$$

Next, consider V_m. Let

$$\tilde S=\sum_{j=1}^{\infty}\theta_j^{1+\varepsilon}\tag{A-7}$$

where $\varepsilon>0$ is to be chosen.[7] By subadditivity note that

$$\begin{aligned}&P(V_m>\delta/2)\\ &\quad\le P\Big(\bigcup_{j_1=1}^{\infty}\cdots\bigcup_{j_{m+1}=1}^{\infty}\{\tilde S^{m+1}|\theta_{j_1}\cdots\theta_{j_{m+1}}|^{-\varepsilon}e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2>\delta/2\}\Big)\end{aligned}$$

[7] Note that while (A-7) is similar to the expression in (A-3), here the sign on ε is reversed, and in this case the sum may exceed 1.

$$\leq \sum_{j_1=1}^{\infty} \cdots \sum_{j_{m+1}=1}^{\infty} P(\tilde{S}^{m+1}|\theta_{j_1} \cdots \theta_{j_{m+1}}|^{-\varepsilon} e^2_{t-j_1} \cdots e^2_{t-j_1-\cdots-j_{m+1}} \sigma^2_{t-j_1-\cdots-j_{m+1}} > \delta/2). \quad \text{(A-8)}$$

Rewrite the probabilities in (A-8) in the form $P(e^2_{t-j_1} \cdots e^2_{t-j_1-\cdots-j_{m+1}} > B(j_1, \dots, j_{m+1}))$ where

$$B(j_1, \dots, j_{m+1}) = \frac{\delta|\theta_{j_1} \cdots \theta_{j_{m+1}}|^{\varepsilon}}{2\tilde{S}^{m+1}\sigma^2_{t-j_1-\cdots-j_{m+1}}}.$$

For brevity, write

$$\bar{P}(\cdot) = P(\cdot|\mathcal{F}_{t-j_1-\cdots-j_{m+1}-1})$$

where $\mathcal{F}_t$ represents the sigma-field generated by $\{e_s, s \leq t\}$, and hence $\sigma^2_{t-j_1-\cdots-j_{m+1}}$ may be held conditionally fixed under $\bar{P}$. Consider the sequence

$$\log(e^2_{t-j_1} \cdots e^2_{t-j_1-\cdots-j_{m+1}}) = \sum_{p=1}^{m} \log e^2_{t-j_1-\cdots-j_p}$$

for $m = 1, 2, \dots$ Since the e_t are *i.i.d.* random variables, the central limit theorem implies that, for large m, the distribution of the sum is approximately Gaussian with mean $-m\zeta$ and variance $m\tau^2$, where $\tau^2 = \mathrm{Var}(\log e_t^2)$. Note that by the Jensen inequality,

$$\zeta = -E(\log e_j^2) > -\log E(e_j^2) = 0.$$

Hence we have, for large enough m,

$$\begin{aligned}
&\bar{P}(e^2_{t-j_1} \cdots e^2_{t-j_1-\cdots-j_{m+1}} > B(j_1, \dots, j_{m+1})) \\
&\quad = \bar{P}\Big(\frac{\log(e^2_{t-j_1} \cdots e^2_{t-j_1-\cdots-j_{m+1}}) + m\zeta}{\tau\sqrt{m}} > \frac{\log B(j_1, \dots, j_{m+1}) + m\zeta}{\tau\sqrt{m}}\Big) \\
&\quad = \frac{\exp\Big\{-\dfrac{(\log B(j_1, \dots, j_{m+1}) + m\zeta)^2}{2\tau^2 m}\Big\}}{\sqrt{2\pi}(\log B(j_1, \dots, j_{m+1}) + m\zeta)}(1 + O((\log B(j_1, \dots, j_{m+1}) + m\zeta)^{-2}) \qquad \text{(A-9)}
\end{aligned}$$

Here, the second equality is obtained, assuming $\log B(j_1, \dots, j_{m+1}) + m\zeta > 0$ (to be established below) from the asymptotic expansion of the Gaussian probability function, see 26.2.12 of Abramovitz and Stegun (1965). Note that the error in the expansion is conditionally of $O(m^{-2})$.

Since $\{\log \sigma_t^2\}$ is a stationary sequence by assumption, and hence $O_p(1)$,

$$\frac{\log B(j_1, \dots, j_{m+1})}{m} = \frac{\varepsilon}{m} \log|\theta_{j_1} \cdots \theta_{j_{m+1}}| - \log \tilde{S} + O_p(1/m). \qquad \text{(A-10)}$$

As m increases, the conditional probability expression in (A-9) (suitably renormalised so that it does not vanish) is converging in probability to a nonstochastic limit, which necessarily matches that of the large-m unconditional probability. Henceforth, we work with this formula by neglecting the terms of $O_p(1/m)$. Note first that

$$0 \leq \frac{1}{m} \log|\theta_{j_1} \cdots \theta_{j_{m+1}}| \leq \log \theta_{\max}.$$

where $\theta_{\max} = \max_{j \geq 1} \theta_j$. The denominator in (A-9) is therefore always positive if $\zeta > \log \tilde{S}$, which we henceforth assume. Next note that

$$\exp\Big\{-\frac{(\varepsilon \log|\theta_{j_1} \cdots \theta_{j_{m+1}}| - m \log \tilde{S} + m\zeta)^2}{2\tau^2 m}\Big\} \leq$$

$$|\theta_{j_1}\cdots\theta_{j_{m+1}}|^{\varepsilon\zeta/\tau^2}\exp\Big\{-m\Big(\frac{(\varepsilon\log\theta_{\max}-\log\tilde{S})^2+2\zeta\log\tilde{S}+\zeta^2}{2\tau^2}\Big)\Big\}. \tag{A-11}$$

Let $\check{S}$ be defined by

$$\check{S}^{m+1}=\sum_{j_1=1}^{\infty}\cdots\sum_{j_{m+1}=1}^{\infty}|\theta_{j_1}\cdots\theta_{j_{m+1}}|^{\varepsilon\zeta/\tau^2}.$$

Combining (A-8), (A-9), (A-10) and (A-11) yields, for large enough m,

$$\begin{aligned}P(V_m>\delta/2)&\le\check{S}^{m+1}\frac{\exp\Big\{-m\Big(\frac{(\varepsilon\log\theta_{\max}-\log\tilde{S})^2+2\zeta\log\tilde{S}+\zeta^2}{2\tau^2}\Big)\Big\}}{\sqrt{2\pi}m(\zeta-\log\tilde{S})}\\&=O\left(\exp\Big\{-\frac{m}{2\tau^2}(2\zeta\log\tilde{S}-2\tau^2\log\check{S}+(\varepsilon\log\theta_1-\log\tilde{S})^2+\zeta^2)\Big\}\right).\end{aligned}$$

For the right-hand expression to vanish as $m\to\infty$ requires that the sum of terms in parentheses, in the exponent, be positive. Using $\zeta>\log\tilde{S}$ we obtain the sufficient condition

$$3\log^2\tilde{S}>2\tau^2\log\check{S}. \tag{A-12}$$

Consider the 'worst case' in which $\theta_j=C\rho^{-j}$. Substituting $\tilde{S}=C^{1+\varepsilon}/(\rho^{1+\varepsilon}-1)$ and $\check{S}=C^{\varepsilon\zeta/\tau^2}/(\rho^{\varepsilon\zeta/\tau^2}-1)$ into (A-12) gives

$$3((1+\varepsilon)\log C-\log(\rho^{1+\varepsilon}-1))^2>2\tau^2(\varepsilon\zeta/\tau^2\log C-\log(\rho^{\varepsilon\zeta/\tau^2}-1)$$

By taking ε large enough, this can be made arbitrarily close to

$$\begin{aligned}3((1+\varepsilon)^2(\log C-\log\rho)^2&>2(\varepsilon\zeta\log C-\varepsilon\zeta\log\rho)\\&>2\varepsilon(\log C-\log\rho)^2\end{aligned}$$

which holds for any choice of C and ρ. It follows that $\zeta>\log\tilde{S}$ is sufficient, which proves part (b) of the theorem. In turn, $C<\rho$ ensures that $\log\tilde{S}\le0$ for large enough $\varepsilon>0$, which proves part (a). ∎

References

Abramowitz, M. and I. A. Stegun (1965) *Handbook of Mathematical Functions*, New York: Dover Publications.

Andersen, T. G. and T. Bollerslev (1997) Heterogeneous information arrivals and retuyrn volatility dynamics: uncovering the long run in high volatility returns. Journal of Finance LII, 3, 975–1005

Andreou, E. and E. Ghysels (2002) Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics* 17, 579–600..

Baillie, R. T.,A. A. Cecen and Y.-W. Han (2000) High Frequency Deutsche Mark-US Dollar Returns: FIGARCH Representations and Non Linearities. *Multinational Finance Journal* 4, (3&4) 247–267

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

Baillie, R. T., W. P. Osterberg (2000) Deviations from daily uncovered interest rate parity and the role of intervention, *Journal of International Financial Markets Institutions and Money* 10, 363–379.

Beltratti, A. and C. Morana (1999) Computing value at risk with high frequency data. *Journal of Empirical Finance* 6, 431–455.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–27.

Box, G. E. P., and D. A. Pierce (1970). 'The distribution of residual autocorrelations in autoregressive-integrated moving average time series models' *Journal of the American Statistical Association* 65, 1509–26.

Brunetti, C. and C. L. Gilbert (2000) Bivariate FIGARCH and fractional cointegration. *Journal of Empirical Finance* 7, 509–530.

Davidson, J. (1994) *Stochastic Limit Theory -An Introduction for Econometricians.* Oxford University Press

Davidson, J. (2002) 'Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes' *Journal of Econometrics* 106, 243–269.

Davidson, J. (2003) Time Series Modelling Version 3.0, www.cf.ac.uk/carbs/econ/davidsonje/TSMod30.htm

Davidson, J. and R. M. de Jong (2000) The functional central limit theorem and weak convergence to stochastic integrals II: fractionally integrated processes *Econometric Theory* 16, 5, 643–666.

Doornik, J.A. (1999), *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik

De Jong, R. M. (1997) Central limit theorems for dependent heterogeneous processes, *Econometric Theory* 13, 353–367

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

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Engle, R. F. and T. Bollerslev (1986), Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1–50.

Geweke, J. and S. Porter-Hudak (1983) 'The estimation and application of long memory time series models', *Journal of Time Series Analysis* 4, 221-238

Giraitis, L., P. Kokoszka and R. Leipus (2000) Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* 16(1), 3–22.

Granger, C. W. J. (2002) Long memory, volatility, risk and distribution. Working paper, Department of Economics, University of California San Diego.

Granger, C. W. J. and T. Teräsvirta (1993) *Modelling Nonlinear Economic Relationships,* Oxford University Press

Hamilton, J. D. and R. Susmel (1994) Autoregressive conditional heteroscedasticity and changes in regime, *Journal of Econometrics* 64, 307–333.

He, C. and T. Teräsvirta (1999) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15(6), 824–846.

Karonasos, M. (1999) The second moment and the autocovariance function of the squared errors of the GARCH model. Journal of Econometrics 90, 63–76.

Kazakevičius, V. and R. Leipus (2002) On stationarity in the ARCH(∞) model, *Econometric Theory* 18 c(1), 1–16

Lavielle, M. and E. Moulines (2000) Least squares estimation of an unknown number of shifts in time series. *Journal; of Time Series Analysis* 20, 33–60.

Lee, S.-W. and B. E. Hansen (1994) 'Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator', *Econometric Theory* 10, 29–52.

Li, W. K. and T. K. Mak (1994) On the squared residual autocorrelation in nonlinear time series with conditional heteroscedasticity. *Journal of Time Series Analysis* 15, 627–636.

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(3) 575–596.

McLeod, A. I. and Li, W. K. (1983) 'Diagnostic checking ARMA time series models using squared-residual autocorrelations', *Journal of Time Series Analysis* 4, pp. 269–273.

Nelson, D. B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–34.

Nelson, D. B. (1991) Conditional heteroscedasticity in asset returns: a new approach. *Econometrica* 59, 347–70.

Pötscher, B. M. and Prucha, I. R. (1991), Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part I: Consistency and approximation concepts, *Econometric Reviews* 10, 125–216.