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

James DAVIDSON
Cardiff Business School, Cardiff CF10 3EU, U.K. (*davidsonje@cardiff.ac.uk*)

This article analyses the statistical properties of that general class of conditional heteroscedasticity models in which the conditional variance is a linear function of squared lags of the process. GARCH, IGARCH, FIGARCH, and a newly proposed generalization, the HYGARCH model, belong to this class. Conditions are derived for the existence of second and fourth moments, and for the limited memory condition of near-epoch dependence. The HYGARCH model is applied to 10 daily dollar exchange rates, and also to data for Asian exchange rates over the 1997 crisis period. In the latter case, the model exhibits notable stability across the pre-crisis and post-crisis periods.

KEY WORDS: ARCH(infinity); FIGARCH; Hyperbolic lag; Near-epoch dependence; Exchange rate.

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

$$u_t = \sigma_t e_t, \tag{1}$$

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

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

where θ_i are lag coefficients typically depending on a small number of underlying parameters. By adding an error term, $v_t = u_t^2 - \sigma_t^2$, to both sides, (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,

$$\sigma_t^2 = \gamma + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3}$$

solves to give (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

$$\sum_{i=1}^{\infty} \theta_i < 1. \tag{4}$$

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

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

$$\delta(L) u_t^2 = \gamma + \beta(L) v_t, \tag{5}$$

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

$$\beta(L)\sigma_t^2 = \gamma + \big(\beta(L) - \delta(L)\big)u_t^2, \tag{6}$$

so that $\alpha_1 = \delta_1 - \beta_1$ in the notation of (3). The model is rearranged into the form of (2) as

$$\sigma_t^2 = \frac{\gamma}{\beta(1)} + \left(1 - \frac{\delta(L)}{\beta(L)}\right)u_t^2 = \omega + \theta(L)u_t^2, \tag{7}$$

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 (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

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 - L), \tag{8}$$

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

The FIGARCH(p, d, q) model replaces the simple difference in (8) with a fractional difference, such that

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 - L)^d \tag{9}$$

for $0 < d < 1$. The FIGARCH is a case where the lag coefficients decline hyperbolically, rather than geometrically, to 0, and it is on these cases that this article focuses. This form has been used in a number of recent works to model financial time series (see, e.g., Baillie et al. 1996; Beltratti and Morana 1999; Baillie, Cecen, and Han 2000; Baillie and Osterberg 2000; Brunetti and Gilbert 2000). As is well known,

$$(1 - L)^d = 1 - \sum_{i=1}^{\infty} a_j L^j, \tag{10}$$

where

$$a_j = \frac{d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} = O(j^{-1-d}). \tag{11}$$

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

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 Section 3.2 proves a modified short-memory property 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 that is emphasized 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 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, and Section 6.2 considers Asian exchange rate data covering the 1997–1998 crisis period. Section 7 concludes the article.

2. MOMENT PROPERTIES

Volatility models of the ARCH(∞) class have two salient features, which in this article will be referred to as, respectively, 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, whereas the memory determines how long shocks to the volatility take to dissipate. The amplitude is measured by

$$S = \sum_{i=1}^{\infty} \theta_i. \tag{12}$$

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

$$\theta_i = O(i^{-1-\delta}). \tag{13}$$

Geometric memory is measured by the parameter ρ, where

$$\theta_i = O(\rho^{-i}). \tag{14}$$

Note that the "length" of memory varies *inversely* with these parameters. In the geometric-decay GARCH(1, 1) model, for example, $S = \alpha_1/(1-\beta_1)$, and $\rho = 1/\beta_1$. Although in the case where $\rho > 1$, the hyperbolic memory assumes the value $+\infty$, it is more realistic to recognize 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

$$S < 1 \tag{15}$$

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

$$E\sigma_t^2 = Eu_t^2 = \omega + \sum_{i=1}^{\infty} \theta_i Eu_{t-i}^2, \tag{16}$$

with the stationary solution

$$M_2 = \frac{\omega}{1-S}. \tag{17}$$

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

$$E\sigma_t^4 = \omega^2 + 2\omega \sum_{i=1}^{\infty} \theta_i Eu_{t-i}^2 + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \theta_i \theta_j Eu_{t-i}^2 u_{t-j}^2. \tag{18}$$

Even assuming that these expectations do not depend on t, to solve the equality in (18) exactly is intractable. However, the Cauchy–Schwarz inequality will imply $Eu_{t-i}^2 u_{t-j}^2 \le M_4$, and hence

$$M_4 \le \mu_4 \left(\omega^2 + \frac{2\omega^2 S}{1-S} + S^2 M_4 \right) \tag{19}$$

or, equivalently,

$$M_4 \le \frac{\mu_4 \omega^2 (1+S)}{(1-S)(1-\mu_4 S^2)}. \tag{20}$$

The condition

$$\mu_4 S^2 < 1 \tag{21}$$

is therefore sufficient, although not necessary, for the existence of M_4 and fourth-order stationarity. Conditions equivalent to (15) and (21) were also derived by GKL, who considered the conditions for a weakly stationary solution of the process u_t^2.

It is of interest to evaluate the bound in (21) for a case where the exact necessary condition for fourth-order stationarity is known. The GARCH(1, 1) in (3) has $S = \alpha_1/(1-\beta_1)$. Straight-

forward manipulations show that

$$M_4 = \frac{\mu_4 \gamma^2 (1 - \beta_1)^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \mu_4 \alpha_1^2 - 2\alpha_1 \beta_1 - \beta_1^2)}, \qquad (22)$$

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

$$\alpha_1^2 \mu_4 < 1 - 2\alpha_1 \beta_1 - \beta_1^2. \qquad (23)$$

This result may be derived as a special case of that given by Davidson (2002) for the GARCH(p, p). (For another version of the general formula, see He and Teräsvirta 1999, and for the Gaussian case see Karonasos 1999.) Note that (21) can be rearranged as

$$\alpha_1^2 \mu_4 < 1 - 2\beta_1 + \beta_1^2. \qquad (24)$$

The majorants of (23) and (24) 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 large. 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 that μ_4 be 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 (e.g., the rate of decay of the weights of a linear process) and some model-independent, (e.g., the various mixing conditions). The correlogram measures only 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 concern focuses on checking 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 law 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 although 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 $\mathcal{F}_s^t = \sigma(e_s, \dots, e_t)$ be 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

$$\|u_t - E(u_t | \mathcal{F}_{t-m}^{t+m})\|_p \le d_t m^{-\lambda} \qquad (25)$$

for $\lambda > \lambda_0$. In the general definition, d_t is a sequence of positive constants, but subject to stationarity, as here, one may simply write $d_t = d < \infty$ (see Davidson 1994, chap. 17, among other references, for additional details). If $m^{-\lambda}$ can be replaced

with ρ^{-m} in (25), then the process will be said to be geometrically NED. There is no accepted "size" terminology associated with this case, but obviously one can speak of "geometric size ρ" in a consistent manner, if it is convenient to do so.

In the present application, the process $\{e_t\}$ will be taken as the driving process in (1). Because this is not merely mixing but iid, by assumption, the condition in (25) alone constrains the memory of the process. The application of this approach to a range of nonlinear processes, including GARCH processes, has been studied earlier (Davidson 2002). Following the same approach, the present work now derives conditions for u_t defined by (1) plus (2) to be L_1 or L_2-NED on $\{e_t\}$. In the following result for the hyperbolic memory case, the model is formalized 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 1. a. If $0 \le \theta_i \le C i^{-1-\delta}$ 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 proved a CLT for the process $\{u_t^2 - E u_t^2\}$ 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 $\lambda_0 = 1/2$. 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, because 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 $-1/2$, 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 2. a. If $0 \le \theta_i \le C\rho^{-i}$ for $i \ge 1$, 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 because $\theta_1 \le C\rho^{-1}$ and necessarily $\theta_1 < S$, the sufficient restrictions on C in Theorem 2 are minimal, in view of the restrictions on S. In effect, they forbid isolated influential lags of higher order. Inspection of the proof show that they could be relaxed, at the cost of more complex or specialized 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.

3.2 The Nonstationary Geometric Lag Case

Nelson (1990) gave an insightful analysis of persistence (memory) in the GARCH(1, 1) model. The key condition that he derived for limited persistence (what he would call "nonpersistence") is

$$E \ln(\beta_1 + \alpha_1 e_t^2) < 0, \tag{26}$$

and the Jensen inequality easily shows that the condition $\alpha_1 < 1 - \beta_1$ is sufficient for (26). This condition is necessary and sufficient for the process to be strictly stationary and ergodic (Nelson 1990, thm. 2).

Necessary conditions for strict stationarity of the GARCH(1, 1) depend on 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 nonexistence 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 the notion of L_0-approximability due to Pötscher and Prucha (1991) (see also Davidson 1994, chap. 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, thm 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. Then define h_t^m to be a geometrically L_0-approximator of σ_t^2 if

$$P(|\sigma_t^2 - h_t^m| > d_t \delta) = O(\rho^{-m}) \tag{27}$$

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. Let u_t be a strictly stationary process, and $0 \le \theta_i \le C\rho^{-i}$ for $i \ge 1$ 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 $S \le C/(\rho - 1)$, and hence the restriction on C in part (b), as well as 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 shows 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 (26). Kazakevičius and Leipus (2002, thm. 2.3) have shown that $\log(C/(\rho - 1)) < \zeta$ is a necessary condition for strict stationarity, although not sufficient, because in the case of the GARCH(1, 1) this is actually a weaker condition than (26). When $C < \rho$, then $C\rho^{-j} < 1$ for all $j \ge 1$, 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. But, when $C \ge \rho$, condition (b) can evidently fail in a stationary process. Note that although the present proof establishes independence of initial conditions, it makes use of the stationarity assumption and thus 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 0 is irrelevant to the covariance stationarity property, provided that 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, is geometrically L_0-approximable, or in other words, short memory.

This appears to be paradoxical, because 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

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

$$= \gamma + (1 - \beta_1)u_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{28}$$

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

$$E_t u_{t+k}^2 = k\gamma + u_t^2, \tag{29}$$

which, although diverging, remains dependent on current conditions even at long range. Thus it appears that u_t^2 fulfils the condition for long memory proposed by Granger and Terasvirta (1993, p. 49)—that is, to be forecastable in mean at long range. But this is a paradox, because σ_t^2 clearly depends only on the recent past. Theorem 3a shows that σ_t^2 can be reconstructed from the shock history, $\{e_{t-1}, e_{t-2}, \ldots, e_{t-m}\}$, 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) discussed this apparent paradox of memory by considering the extreme case in which $\beta_1 = 0$ and $\gamma = 0$, so that (29) reduces to $E_t u_{t+k}^2 = u_t^2$. This model can be written in the form

$$u_t = e_t|u_{t-1}|. \tag{30}$$

Although a succession of larger-than-average independent shocks (e_t's) may produce very large deviations of the observed process, such that their variance is infinite, the e_t's are still drawn independently from a distribution centered on 0. 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 (30) converges rapidly to 1, is the essential message of Theorem 3. Nelson (1990) showed that this particular process converges to 0 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

$$\sigma_t^2 = \alpha e_{t-1}^2 \sigma_{t-1}^2 = \cdots = \alpha^m e_{t-1}^2 e_{t-2}^2 \cdots e_{t-m}^2 \sigma_{t-m}^2. \tag{31}$$

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 3b shows that the steady state solution is $\sigma_t^2 = 0$ whenever $\log \alpha < -E \log e_t^2$. In this case, the right side of (31) converges to 0 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 although σ_t^2 in (28) or (31) is a natural indicator of conditional volatility, depending on the near epoch, it is *not* the conditional variance. Because 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 (29) 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 (9) is a generalization of the IGARCH, of particular interest because this is the one application to date using hyperbolic lag weights. Note that $\sum_{i=1}^{\infty} a_j = 1$ in (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 0. [Note the error in Baillie et al. 1996, p. 11, line 6, where the lag coefficients are said to be of $O(k^{d-1})$. This should read $O(k^{-d-1})$.] 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 $a_i = 0$ for $i > 1$. In this particular case, of amplitude $S = 1$, the memory [measured by $-\delta$ in (13)] 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 nonsummability. However, again because of the restriction $S = 1$, the individual a_i's 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 strictly below 1. The characterization 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. In fact, it has 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,

$$\sigma_t^2 = \omega + \left(1 - (1 - L)^d\right)u_t^2, \tag{32}$$

rearranged as

$$(1 - L)^d u_t^2 = \omega + v_t, \tag{33}$$

with $v_t = (e_t^2 - 1)\sigma_t^2$ as earlier. This equation might appear to represent u_t^2 as a classic fractionally integrated process. However, just as the temptation to write $E(v_t) = 0$ must be resisted, in the absence of second moments, so it is important to not 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 one to replace v_t in (33) by (say) an independent disturbance with mean of 0 for $t > 0$, and by 0 for $t \le 0$, then one would actually obtain a nonstationary *trending* process with expected value ωt^d for $t > 0$ (see Granger 2002). This clearly contradicts what is known about the actual characteristics of the u_t^2 process.

As remarked earlier, GKL showed that whenever fourth moments exist, the autocovariances of the squared process are always summable. Kazakevičius and Leipus (2002) further showed 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 diverge rapidly. The term "hyperbolic memory" is therefore preferable to distinguish FIGARCH from the geometric memory cases such as GARCH and IGARCH.

5. THE HYGARCH MODEL

The unexpected behavior 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 in which such restrictions can be tested, and also to adhere to the approach of modeling amplitude and memory as separate phenomena, just as is done in the ordinary GARCH model.

In view of these considerations a new model is here proposed, the "hyperbolic GARCH," or HYGARCH model. Consider, for comparability with the previous cases, the form

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}\left(1 + \alpha((1 - L)^d - 1)\right), \tag{34}$$

where $\alpha \ge 0$, $d \ge 0$. Note that, provided that $d > 0$,

$$S = 1 - \frac{\delta(1)}{\beta(1)}(1 - \alpha). \tag{35}$$

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 $d = 1$, (34) reduces to

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 - \alpha L), \qquad \alpha \geq 0. \qquad (36)$$

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 memory 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

$$\theta(L) = 1 - \frac{\delta(L)}{\beta(L)}(1 - \alpha\phi(L)), \qquad (37)$$

where

$$\phi(L) = \zeta(1 + d)^{-1} \sum_{j=1}^{\infty} j^{-1-d} L^j, \qquad d > 0 \qquad (38)$$

and $\zeta(\cdot)$ is the Riemann zeta function. Note, however, that the models behave quite differently when d is close to 1. In (34), $d > 1$ gives rise to negative coefficients and so is not permitted, whereas in (38), 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 behavior excluded by (34). 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. In this context, it is in fact an arbitrary choice to assume the hyperbolic decay pattern implied by (11) rather than use weights directly proportional to j^{-1-d}. The chief motivation for using (34) must be to nest the FIGARCH and IGARCH cases, but should d be found close to 1, then the option of comparing GARCH with (37) might be considered.

If the GARCH component observes the usual covariance stationarity restrictions, which imply that $\delta(1)/\beta(1) > 0$, then with $\alpha < 1$, these processes are covariance stationary and L_1-NED of size $-d$, according to Theorem 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, $1 - \mu_4^{-1/2} = .422$. Therefore, noting the discussion of Section 3.1, the CLT holds at least for $d > 1/2$ in that case.

6. APPLICATIONS

This section discusses two applications of the HYGARCH model. The first application is a rather conventional one with the aim of relating these models to the substantial existing literature on modeling exchange rates. The second is more unusual, and possibly controversial, in which the aim is to argue that these models may play a distinctive and important role in more difficult cases.

6.1 Dollar Exchange Rates, 1980–1996

Table 1 summarizes estimates of the HYGARCH model for a collection of the (logarithms of) major dollar exchange rates, obtained using the Ox package Time Series Modeling 3.1 (Davidson 2003; Doornik 1999). The data, sourced from Datastream, are in each case daily for the period January 1, 1980–September 30, 1996 (4,370 observations). The model fitted to all of the series is a first-order ARFI-HYGARCH, taking the form

$$(1 - L)^{d_{ARF}}(1 - \phi_1 L)Y_t = \mu + u_t \qquad (39)$$

and

$$h_t = \omega + \left(1 - \frac{1 - \delta_1 L}{1 - \beta_1 L}\left(1 + \alpha\left((1 - L)^{d_{FG}} - 1\right)\right)\right)u_t^2. \qquad (40)$$

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 the 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 t, rather than the normal, distribution is assumed for the disturbances. The criterion function for estimation is the Student t log-likelihood,

$$L_T = T \log \frac{\Gamma(\nu + 1)/2)}{\sqrt{\pi(\nu - 2)}\Gamma(\nu/2)}$$

$$- \frac{1}{2}\sum_{t=1}^{T}\left(\log h_t + (\nu + 1)\log\left(1 + \frac{u_t^2}{(\nu - 2)h_t}\right)\right).$$

As a practical matter, observe that small innovations, u_t, contribute to this criterion in much the same way as to

Table 1. The ARFI-HYGARCH Model of Exchange Rates

Currency	d_{FG}	α	δ_1	β_1	d_{ARF}	ϕ_1	$\nu^{1/2}$	$Q(25)$	$Q_{sq}(25)$
Danish Kroner	.600 (.081)	.962 (.021)	.188 (.045)	.722 (.051)	.056 (.017)	−.055 (.021)	2.31 (.092)	25	34
Deutschmark	.681 (.077)	.975 (.017)	.190 (.047)	.789 (.038)	.045 (.022)	−.038 (.022)	2.25 (.082)	27	23
Finnish Mark	.714 (.079)	.946 (.023)	.177 (.054)	.782 (.046)	.013 (.015)	−.043 (.020)	2.22 (.109)	29	1.24
GB Pound	.656 (.103)	.991 (.017)	.229 (.056)	.795 (.056)	.016 (.016)	.005 (.022)	2.27 (.0907)	32	31
Irish Punt	.641 (.087)	.991 (.018)	.241 (.057)	.787 (.048)	.030 (.017)	−.049 (.022)	2.23 (.087)	27	15
Italian Lire	.556 (.075)	.991 (.025)	.264 (.052)	.699 (.064)	.038 (.015)	−.0045 (.021)	2.20 (.077)	37	32
Japanese Yen	.564 (.144)	.952 (.041)	.253 (.071)	.696 (.091)	.045 (.015)	−.081 (.020)	1.96 (.059)	34	23
Port. Escudo	.613 (.162)	.986 (.029)	.291 (.081)	.747 (.099)	.026 (.013)	−.079 (.019)	1.99 (.074)	28	1.44
Spanish Peseta	.515 (.060)	1.04 (.021)	.258 (.044)	.697 (.047)	.039 (.020)	−.067 (.020)	2.20 (.083)	25	13
Swiss Franc	.819 (.092)	.956 (.017)	.100 (.059)	.835 (.046)	.027 (.017)	−.015 (.022)	2.38 (.097)	23	16

NOTE: Robust standard errors are given in parentheses.

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 Table 1 show the Box–Pierce (1979) $Q(r)$ statistic for $r = 25$ lags, as well as the Q_{sq}, 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. Li and Mak showed that using the nominal chi-squared distribution with r degrees of freedom would give an excessively conservative test, similar to the Box–Pierce result for ARMA residuals. The asymptotic distributions of these statistics for the cases of hyperbolic lags in mean and variance have not yet been studied, so both must be treated with caution, as diagnostic tests. What can be stated is that 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, thus cannot 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, because although the samples are large, the asymptotic properties of the estimates are not yet well established. Lumsdaine (1996) and Lee and Hansen (1994) considered the IGARCH(1, 1) case and showed that covariance stationarity of the processes is not a necessary condition for consistency and asymptotic normality of the usual quasi-maximum likelihood. However, note that the conjecture of Baillie et al. (1996, p. 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 article.

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. Because these are rates of exchange determined in closely related markets, this perhaps is not 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, they 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. It therefore can be conjectured that the similarity of these structures goes deeper than the fact of some correspondence in their movement in levels.

Looking now at the estimates themselves, note that although 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. In most cases the amplitude parameter α is not significantly different from 1, while generally a little below it. The FIGARCH model explains these data pretty well. Also, note that the estimate of d for the Deutschmark is similar to the FIGARCH estimations reported by Baillie et al. (1996) and also Beltratti and Morana (1999). A noteworthy feature is that the Student 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 considered is to the dollar exchange rates for three Asian currencies, for periods covering the Asian crisis of 1997–1998. The series in question, in logarithms, are shown in panels (a) of Figures 1–3. At first sight, it might appear that these data represent two quite distinct regimes. Before 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. These models may be seen 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 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 is 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 0. Lagrange multiplier (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 0 when fitted,
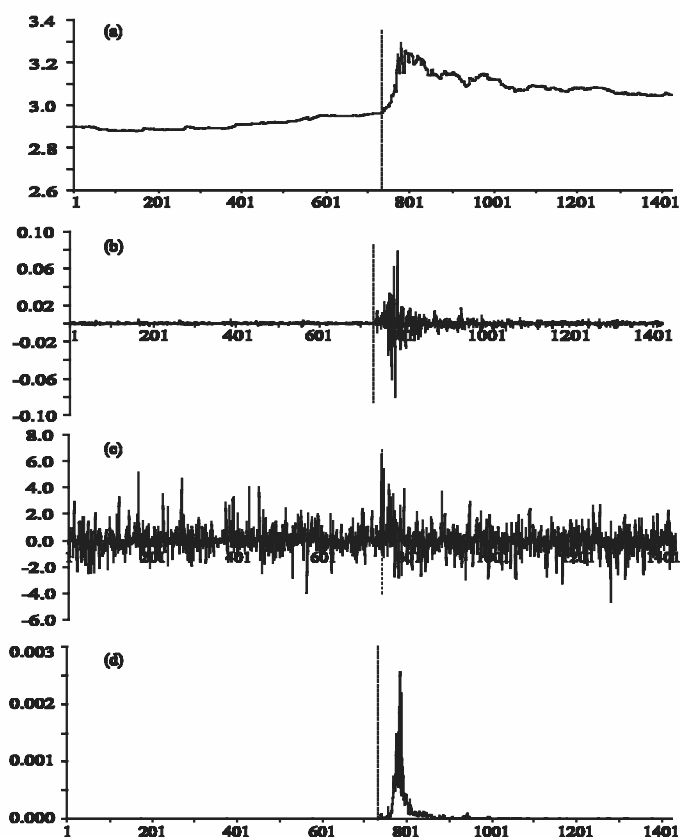

Figure 1. Korean Won. (a) Korean Won/US Dollar exchange rate (logarithms), observed time series; (b) unadjusted residuals ($\hat{u}_t$), model in Table 2, Column 1; (c) adjusted residuals ($\hat{u}_t/\hat{\sigma}_t$), model in Table 2, Column 1; and (d) conditional variances ($\hat{\sigma}_t^2$), model in Table 2, Column 1.
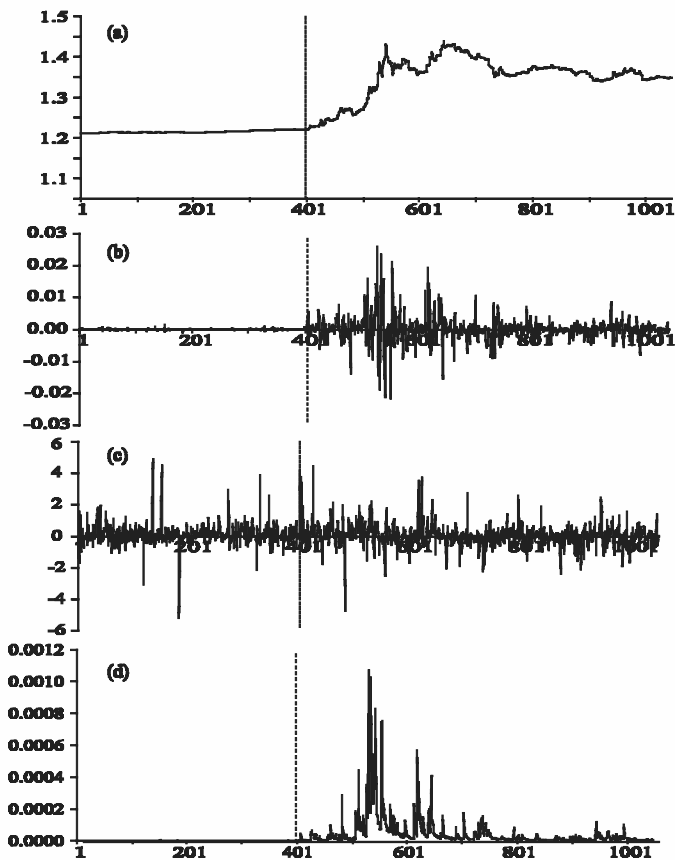
Figure 2. Indonesian Rupiah. (a) Indonesian Rupiah/US Dollar exchange rate (logarithms), observed time series; (b) unadjusted residuals ($\hat{u}_t$), model in Table 3, Column 1; (c) adjusted residuals ($\hat{u}_t/\hat{\sigma}_t$), model in Table 3, Column 1; and (d) conditional variances ($\hat{\sigma}_t^2$), model in Table 3, Column 1.
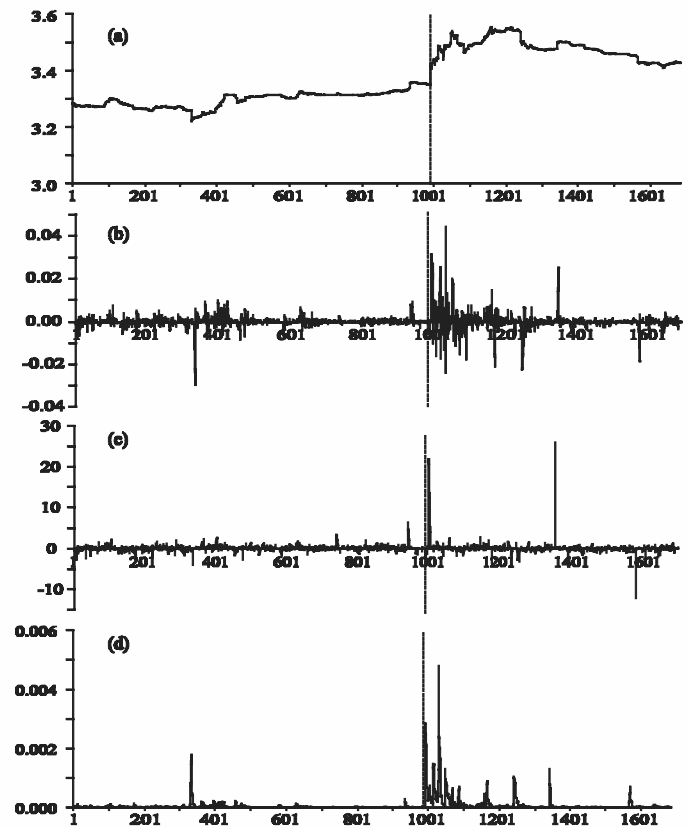

Figure 3. Taiwan Dollar. (a) Taiwan Dollar/US Dollar exchange rate (logarithms), observed time series; (b) unadjusted residuals ($\hat{u}_t$), model in Table 4, Column 1; (c) adjusted residuals ($\hat{u}_t/\hat{\sigma}_t$), model in Table 4, Column 1; and (d) conditional variances ($\hat{\sigma}_t^2$), model in Table 4, Column 1.

and are constrained to 0 in these estimates. Note that the fourth root of ω has been estimated. Because ω is close to 0 in these models, this transformation is found to improve numerical stability, probably by giving better numerical approximations to derivatives.

Along with the full sample, the same models were also fitted to "pre-crisis" and "post-crisis" subsamples. The breakpoints were in each case chosen by eye, at the point just preceding

the first large fall of the currency. Using the methodology of Lavielle and Moulines (2000), Andreou and Ghysels (2002) detected multiple breaks in the volatility dynamics of stock market indices during the Asian crisis. However, in this instance there is inevitably a moment in the crisis at which the monetary authorities allow the currencies to float freely, leading to precipitate devaluations. It is these events that are taken to mark the regime switch dates. These are marked by the vertical lines in the figures, which show in panels (b), (c), and (d) respectively,

Table 2. Korean Won

	12/13/94–6/15/00 (1,424 observations)	12/13/94–10/16/97 (730 observations)	10/17/97–6/15/00 (694 observations)
d_{FG}	.669 (.046)	.667 (.121)	.686 (.066)
α	1.252 (.149)	1.265 (.275)	1.226 (.177)
β_1	.339 (.092)	.318 (.143)	.363 (.140)
$\omega^{1/4}$	.0184 (.0018)	.0186 (.0026)	.021 (.0091)
d_{ARF}	.073 (.031)	.072 (.038)	.076 (.059)
ϕ_1	.116 (.039)	.110 (.051)	.122 (.068)
ϕ_2	−.097 (.028)	−.044 (.038)	−.152 (.044)
$\nu^{1/2}$	1.73 (.080)	1.71 (.113)	1.74 (.116)
Kurtosis	8.65	7.16	7.93
$Q(25)$	44	21	34
$Q_{sq}(25)$	39	17	47
$LM(\delta_1)$	.204	—	—
Log likelihood	7,421.85	4,314.09	3,109.97
		LR statistic (8 df) = 4.42	

NOTE: Robust standard errors are given in parentheses.

Table 3. Indonesian Rupiah

	1/01/96–12/31/99 (1,045 observations)	1/01/96–7/11/97 (400 observations)	7/14/97–12/31/99 (645 observations)
d_{FG}	.496 (.042)	.588 (.104)	.54 (.054)
α	2.94 (.85)	1.81 (1.49)	2.92 (1.59)
$\omega^{1/4}$	.0099 (.0015)	.011 (.002)	.037 (.007)
ϕ_1	.047 (.032)	−.038 (.055)	.097 (.041)
$\nu^{1/2}$	1.52 (.038)	1.54 (.120)	1.51 (.071)
Kurtosis	14.6	2.4	1.1
$Q(25)$	37	22	29
$Q_{sq}(25)$	27	19	22
$LM(d_{ARF})$	.012	—	—
$LM(\delta_1)$	2.31	—	—
Log likelihood	5,889.68	2,978.46	2,917.85
	LR statistic (5 df) = 12.18		

NOTE: Robust standard errors are given in parentheses.

the estimated series $\hat{u}_t$, $\hat{u}_t/\hat{\sigma}_t$, and $\hat{\sigma}_t^2$ from the HYGARCH model in each case.

An examination of these results leads to three conclusions. 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 any autocorrelation that may be present. Second, however, they more closely resemble each other than the currencies analyzed in Table 1. The most noteworthy feature is, of course, the large values of the α parameter in each case, especially for the Rupiah and Taiwan dollar.

Third, and perhaps most remarkable, is the stability of these models across the pre-crisis 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 for the test of model stability across the sample. Note that this cannot be interpreted as an asymptotic chi-squared test, because the breakpoints 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 breakpoints. 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 is only slightly outside it. Overall, these results provide little evidence for changes in the

model following the crisis, and the residual plots in panels (c) of Figures 1–3 (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 breakpoint 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 Figure 4(c). The data shown were generated after letting the process run for 2,000 presample 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, because that of the shock distribution is 1 by construction. Such a realization would be a fairly rare event under random resampling, although 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

Table 4. Taiwan Dollar

	01/03/94–6/15/00 (1,683 observations)	01/03/94–10/15/97 (988 observations)	10/16/97–6/15/00 (695 observations)
d_{FG}	.860 (.079)	1.001 (.010)	.667 (.073)
α	2.96 (.466)	2.956 (.466)	2.946 (.877)
δ_1	.242 (.138)	.009 (.187)	.568 (.221)
β_1	.635 (.043)	.606 (.042)	.726 (.136)
$\omega^{1/4}$	.021 (.006)	.022 (.006)	.000037 (.004)
ϕ_1	−.075 (.024)	−.131 (.031)	−.007 (.037)
$\nu^{1/2}$	1.47 (.010)	1.46 (.010)	1.46 (.019)
Kurtosis	336	21	184
$Q(25)$	9.40	49	5.27
$Q_{sq}(25)$	.21	23	.11
$LM(d_{ARF})$	.037	—	—
Log likelihood	8,631.47	5,263.64	3,38.61
	LR statistic (7 df) = 25.56		

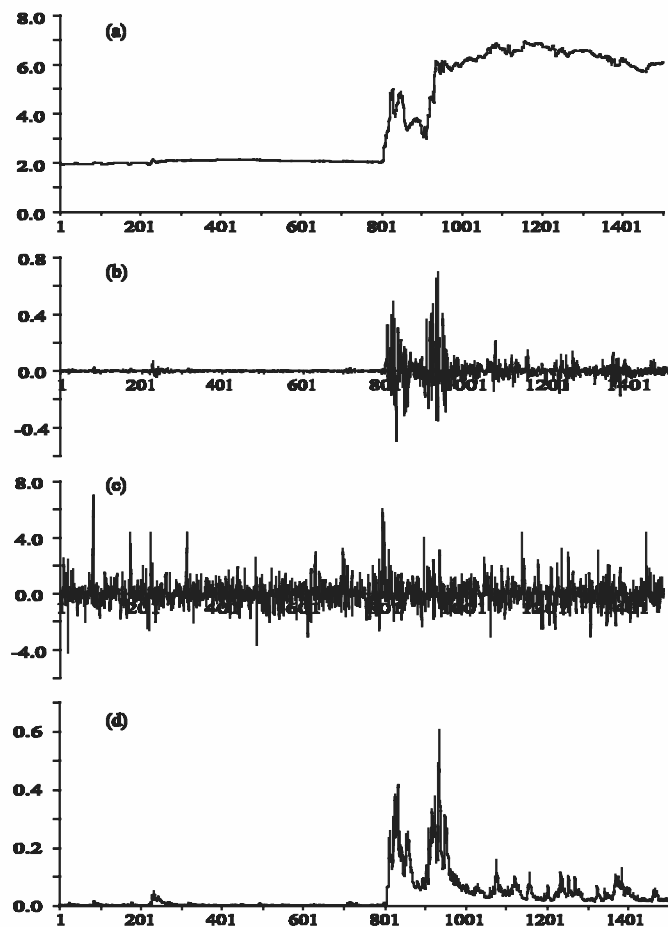NOTE: Robust standard errors are given in parentheses.

Figure 4. Korean Won Simulation. (a) Stochastically simulated series, using model in Table 2, Column 1; (b) HYGARCH shocks of simulated series; (c) shocks randomly resampled from series in Figure 1(c), with inserted "crisis" at period 801; and (d) conditional variances of simulated series.

"pegged-rate" segments of the series in Figures 1(a)–3(a), although possibly appearing "stationary," are actually well explained by a $I(1 + d)$ process, provided that the innovations are small enough. To switch to the post-crisis behavior, 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 behavior 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, e.g., Hamilton and Susmel 1994). What we aim to show is that although any of these may be the true explanations, there is no *need* to introduce them. The crisis behavior 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 article, 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 to the magnitude of the sum of the lag coefficients and their rate of convergence. The proposed HYGARCH model generalizes the FIGARCH model to permit both the existence of second moments on the one hand, and greater extremes of amplitude on the other. Application of the model to exchange 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 (i.e., 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 commentary on this question is provided by Andersen and Bollerslev (1997), who considered 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 was estimated by the method of Geweke and Porter-Hudak (1983) 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 (9) or (34). A spectrum diverging at the origin implies nonsummable autocorrelations, which are not permitted in the ARCH(∞) framework. If true long memory in variance is to be represented parametrically, it will have to be in the context of a different class of models.

The exponential ARCH(∞) class, in which $\log h_t$ is modeled by a distributed lag of some appropriate indicator of realized volatility, could be a plausible candidate for this role. Andersen and Bollerslev (1997) derived an exponential model, but also argued 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 iid driving process. If (34) were allowed to represent (in a purely formal way) the lag structure in this model, then 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) would not necessarily be incompatible with stationarity. One might even speculate that the parameter estimated by Andersen and Bollerslev (1997) corresponds to $-d$ in this setup. Investigating these issues goes well beyond the scope of this article, but represents an interesting avenue for further research.

ACKNOWLEDGMENTS

APPENDIX: PROOFS OF THEOREMS

A.1 Proof of Theorem 1

Because $\sigma_t^2 \geq \omega$ 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 with $p \geq 1$. Therefore, in view of stationarity, it suffices to prove the inequalities

$$\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\|_p \leq C_p m^{-\delta} \qquad (A.1)$$

for $C_p > 0$, for $p = 1$ and $p = 2$. Repeated substitution leads to, for given m, the decomposition

$$
\begin{aligned}
\sigma_t^2 &= \omega + \sum_{j=1}^{\infty}\theta_j e_{t-j}^2\sigma_{t-j}^2 \\
&= \omega\Bigg(1 + \sum_{j_1=1}^{\infty}\theta_{j_1}e_{t-j_1}^2 + \sum_{j_1=1}^{\infty}\sum_{j_2=1}^{\infty}\theta_{j_1}\theta_{j_2}e_{t-j_1}^2 e_{t-j_1-j_2}^2 + \cdots \\
&\qquad + \sum_{j_1=1}^{\infty}\cdots\sum_{j_m=1}^{\infty}\theta_{j_1}\cdots\theta_{j_m}e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_m}^2\Bigg) \\
&\quad + \sum_{j_1=1}^{\infty}\cdots\sum_{j_{m+1}=1}^{\infty}\theta_{j_1}\cdots\theta_{j_{m+1}} \\
&\qquad\qquad \times e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2.
\end{aligned}
$$
$$(A.2)$$

To prove part (a), first note that

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

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

$$E\big|e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2$$
$$- E\big(e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2\big|\mathcal{F}_{t-m}^{t+m}\big)\big| \leq 2M_2,$$

where M_2 is defined after (15). Next, define

$$T_p = \sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty}\mathbb{1}_{\{j_1+\cdots+j_p>m\}}(j_1,\ldots,j_p)\theta_{j_1}\cdots\theta_{j_p}$$

$$\leq \Bigg(\sum_{j=m/p+1}^{\infty}\theta_j\Bigg)S^{p-1}$$

$$\leq C\Bigg(\sum_{j=m/p+1}^{\infty}j^{-1-\delta}\Bigg)S^{p-1}$$
$$= O(p^{\delta}m^{-\delta}S^{p-1}),$$

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

$$E\big|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\big| \leq 2\omega\sum_{p=1}^{m}T_p + 2M_2 S^{m+1}$$

$$= O\Bigg(m^{-\delta}\sum_{p=1}^{m}p^{\delta}S^{p-1}\Bigg)$$

$$= O(m^{-\delta}).$$

To prove part (b), note similarly that

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

whereas

$$\big\|e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2$$
$$- E\big(e_{t-j_1}^2\cdots e_{t-j_1-\cdots-j_{m+1}}^2\sigma_{t-j_1-\cdots-j_{m+1}}^2\big|\mathcal{F}_{t-m}^{t+m}\big)\big\|_2$$
$$\leq 2\mu_4^{(m+1)/2}M_4^{1/2}.$$

Therefore, Minkowski's inequality gives

$$\big\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\big\|_2 \leq 2\omega\sum_{p=1}^{m}T_p\mu_4^{p/2} + 2\mu_4^{(m+1)/2}M_4^{1/2}S^{m+1}$$

$$= O\Bigg(m^{-\delta}\sum_{p=1}^{m}p^{\delta}(\mu_4^{1/2}S)^{p-1}\Bigg)$$

$$= O(m^{-\delta}).$$

A.2 Proof of Theorem 2

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

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

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

$$T_p = \sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty}\mathbb{1}_{\{j_1+\cdots+j_p>m\}}(j_1,\ldots,j_p)\theta_{j_1}\cdots\theta_{j_p}$$

$$= \sum_{j_1=1}^{\infty}\cdots\sum_{j_p=1}^{\infty}\mathbb{1}_{\{j_1+\cdots+j_p>m\}}(j_1,\ldots,j_p)$$

$$\times |\theta_{j_1}\cdots\theta_{j_p}|^{\varepsilon}|\theta_{j_1}\cdots\theta_{j_p}|^{1-\varepsilon}$$

$$\leq \tilde{C}^p \tilde{\rho}^{-m} \tilde{S}^p$$
$$= \tilde{C}^{p-m} \tilde{S}^p (\tilde{C}\tilde{\rho}^{-1})^m. \tag{A.4}$$

Therefore,

$$E\left|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\right| \leq 2\omega \sum_{p=1}^m T_p + 2M_2 \left|\theta_{j_1} \cdots \theta_{j_{m+1}}\right|^\varepsilon \tilde{S}^{m+1}$$

$$= O\big(m \max\{\tilde{C}^{-m}, \tilde{S}^m\}(\tilde{C}\tilde{\rho}^{-1})^m\}\big). \tag{A.5}$$

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 without loss of generality such that $1 < \mu_4^{\varepsilon/2}\tilde{C} < \tilde{\rho}$. Therefore,

$$\left\|\sigma_t^2 - E_{t-m}^{t+m}\sigma_t^2\right\|_2$$

$$\leq 2\omega \sum_{p=1}^m T_p \mu_4^{p/2} + 2\mu_4^{(m+1)/2} M_4^{1/2} |\theta_{j_1} \cdots \theta_{j_{m+1}}|^\varepsilon \tilde{S}^{m+1}$$

$$= O\big(m \max\{(\mu_4^{\varepsilon/2}\tilde{C})^{-m}, (\tilde{S}\mu_4^{(1-\varepsilon)/2})^m\}(\mu_4^{\varepsilon/2}\tilde{C}\tilde{\rho}^{-1})^m\big).$$

A.3 Proof of Theorem 3

Let

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

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

Then, using (A.2),

$$\sigma_t^2 - h_t^m$$

$$= \omega\left(\sum_{j_1=m+1}^\infty \theta_{j_1} e_{t-j_1}^2 + \right.$$

$$+ \sum_{j_1=1}^\infty \sum_{j_2=1}^\infty \mathbb{1}_{\{j_1+j_2>m\}}(j_1, j_2)\theta_{j_1}\theta_{j_2} e_{t-j_1}^2 e_{t-j_1-j_2}^2 + \cdots$$

$$+ \sum_{j_1=1}^\infty \cdots \sum_{j_m=1}^\infty \mathbb{1}_{\{j_1+\cdots+j_m>m\}}(j_1, \ldots, j_m)\theta_{j_1} \cdots \theta_{j_m}$$

$$\left. e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_m}^2 \right)$$

$$+ \sum_{j_1=1}^\infty \cdots \sum_{j_{m+1}=1}^\infty \theta_{j_1} \cdots \theta_{j_{m+1}}$$

$$\times e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_{m+1}}^2 \sigma_{t-j_1-\cdots-j_{m+1}}^2$$

$$= \omega(U_1 + \cdots + U_m) + V_m,$$

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,

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

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,

$$T_p \leq C^p \rho^{-m} \sum_{j_1=1}^\infty \cdots \sum_{j_p=1}^\infty \mathbb{1}_{\{j_1+\cdots+j_m>m\}}(j_1, \ldots, j_m)\rho^{m-j_1-\cdots-j_p}$$

$$= O(S^p \rho^{-m}), \tag{A.6}$$

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

$$P\left(\omega \sum_{p=1}^m |U_p| > \delta/2\right) \leq P\left(\bigcup_{p=1}^m \left\{U_p > \frac{\delta\omega}{2m}\right\}\right)$$

$$\leq \sum_{p=1}^m P\left(U_p > \frac{\delta\omega}{2m}\right)$$

$$\leq \frac{2m}{\delta\omega} \sum_{p=1}^m T_p = O\left(m^2 \left(\frac{C}{\rho(\rho-1)}\right)^m\right).$$

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. [Note that although (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.] By subadditivity, note that

$$P(V_m > \delta/2)$$

$$\leq P\left(\bigcup_{j_1=1}^\infty \cdots \bigcup_{j_{m+1}=1}^\infty \left\{\tilde{S}^{m+1}|\theta_{j_1} \cdots \theta_{j_{m+1}}|^{-\varepsilon}\right.\right.$$

$$\left.\left. \times e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_{m+1}}^2 \sigma_{t-j_1-\cdots-j_{m+1}}^2 > \delta/2\right\}\right)$$

$$\leq \sum_{j_1=1}^\infty \cdots \sum_{j_{m+1}=1}^\infty P\big(\tilde{S}^{m+1}|\theta_{j_1} \cdots \theta_{j_{m+1}}|^{-\varepsilon}$$

$$\times e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_{m+1}}^2 \sigma_{t-j_1-\cdots-j_{m+1}}^2 > \delta/2\big). \tag{A.8}$$

Rewrite the probabilities in (A.8) in the form

$$P\big(e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_{m+1}}^2 > B(j_1, \ldots, j_{m+1})\big),$$

where

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

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 $\sigma_{t-j_1-\cdots-j_{m+1}}^2$ may be held conditionally fixed under $\bar{P}$. Consider the sequence

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

for $m = 1, 2, \ldots$. Because the e_t are iid random variables, the CLT implies that for large m, the distribution of the sum is approximately Gaussian with mean $-m\zeta$ and variance $m\tau^2$, where $\tau^2 = \text{var}(\log e_t^2)$. Note that by the Jensen inequality,

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

Hence, for large enough m,

$$\bar{P}\big(e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_{m+1}}^2 > B(j_1, \ldots, j_{m+1})\big)$$

$$= \bar{P}\bigg(\frac{\log(e_{t-j_1}^2 \cdots e_{t-j_1-\cdots-j_{m+1}}^2) + m\zeta}{\tau\sqrt{m}}$$

$$> \frac{\log B(j_1, \ldots, j_{m+1}) + m\zeta}{\tau\sqrt{m}}\bigg)$$

$$= \frac{\exp\big\{-\frac{(\log B(j_1,\ldots,j_{m+1})+m\zeta)^2}{2\tau^2 m}\big\}}{\sqrt{2\pi}(\log B(j_1,\ldots,j_{m+1}) + m\zeta)}$$

$$\times \big(1 + O(\log B(j_1, \ldots, j_{m+1}) + m\zeta)^{-2}\big). \quad (A.9)$$

Here the second equality is obtained, assuming that

$$\log B(j_1, \ldots, j_{m+1}) + m\zeta > 0$$

(to be established below) from the asymptotic expansion of the Gaussian probability function, (see 26.2.12 of Abramowitz and Stegun 1965). Note that the error in the expansion is conditionally of $O(m^{-2})$.

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

$$\frac{\log B(j_1, \ldots, j_{m+1})}{m}$$

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

As m increases, the conditional probability expression in (A.9) (suitably renormalized 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, this formula is modified by neglecting the terms of $O_p(1/m)$. First, note that

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

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

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

$$\le |\theta_{j_1} \cdots \theta_{j_{m+1}}|^{\varepsilon\zeta/\tau^2}$$

$$\times \exp\bigg\{-m\bigg(\frac{(\varepsilon\log\theta_{\max} - \log\tilde{S})^2 + 2\zeta\log\tilde{S} + \zeta^2}{2\tau^2}\bigg)\bigg\}. \quad (A.11)$$

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.11) yields, for large enough m,

$$P(V_m > \delta/2)$$

$$\le \check{S}^{m+1} \frac{\exp\big\{-m\big(\frac{(\varepsilon\log\theta_{\max}-\log\tilde{S})^2 + 2\zeta\log\tilde{S}+\zeta^2}{2\tau^2}\big)\big\}}{\sqrt{2\pi}m(\zeta - \log\tilde{S})}$$

$$= O\bigg(\exp\bigg\{-\frac{m}{2\tau^2}\big(2\zeta\log\tilde{S} - 2\tau^2\log\check{S}$$

$$+ (\varepsilon\log\theta_1 - \log\tilde{S})^2 + \zeta^2\big)\bigg\}\bigg).$$

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

$$3\log^2\tilde{S} > 2\tau^2\log\check{S}, \quad (A.12)$$

is obtained. 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

$$3\big((1+\varepsilon)\log C - \log(\rho^{1+\varepsilon} - 1)\big)^2$$

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

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

$$3\big((1+\varepsilon)^2(\log C - \log\rho)\big)^2 > 2(\varepsilon\zeta\log C - \varepsilon\zeta\log\rho)$$

$$> 2\varepsilon(\log C - \log\rho)^2,$$

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 $\log\tilde{S} \le 0$ for large enough $\varepsilon > 0$, which proves part (a).

[Received June 2001. Revised April 2003.]

REFERENCES

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

Andersen, T. G., and Bollerslev, T. (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long Run in High Volatility Returns," *Journal of Finance*, 3, 975–1005.

Andreou, E., and Ghysels, E. (2002), "Detecting Multiple Breaks in Financial Market Volatility Dynamics," *Journal of Applied Econometrics*, 17, 579–600.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996), "Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 74, 3–30.

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

Baillie, R. T., and Osterberg, W. P. (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 Morana, C. (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, 307–327.

Box, G. E. P., and Pierce, D. A. (1970), "The Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.

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

Davidson, J. (1994), *Stochastic Limit Theory—An Introduction for Econometricians*, Oxford, U.K.: Oxford University Press.

—— (2002), "Establishing Conditions for the Functional Central Limit Theorem in Nonlinear and Semiparametric Time Series Processes," *Journal of Econometrics*, 106, 243–269.

—— (2003), "Time Series Modelling, Version 3.1," available at *www.cf.ac.uk/carbs/econ/davidsonje/software.html*.

De Jong, R. M. (1997), "Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 13, 353–367.

Ding, Z., and Granger, C. W. J. (1996), "Modelling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185–215.

Doornik, J. A. (1999), *Object-Oriented Matrix Programming Using Ox* (3rd ed.), London: Timberlake Consultants Press.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R. F., and Bollerslev, T. (1986), "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long-Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221–238.

Giraitis, L., Kokoszka, P., and Leipus, R. (2000), "Stationary ARCH Models: Dependence Structure and Central Limit Theorem," *Econometric Theory*, 16, 3–22.

Granger, C. W. J. (2002), "Long Memory, Volatility, Risk and Distribution," working paper, University of California San Diego, Dept. of Economics.

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

Hamilton, J. D., and Susmel, R. (1994), "Autoregressive Conditional Heteroscedasticity and Changes in Regime," *Journal of Econometrics*, 64, 307–333.

He, C., and Teräsvirta, T. (1999), "Fourth-Moment Structure of the GARCH(p, q) Process," *Econometric Theory*, 15, 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 Leipus, R. (2002), "On Stationarity in the ARCH(∞) Model," *Econometric Theory*, 18, 1–16.

Lavielle, M., and Moulines, E. (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 Hansen, B. E. (1994), "Asymptotic Theory for the GARCH($1, 1$) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Li, W. K., and Mak, T. K. (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, 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, 269–273.

Nelson, D. B. (1990), "Stationarity and Persistence in the "GARCH($1, 1$) Model," *Econometric Theory*, 6, 318–334.

—— (1991), "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

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.