Strict stationarity, persistence and volatility forecasting in ARCH(∞) processes

James Davidson
University of Exeter

Xiaoyu Li
University of Exeter

August 27, 2015

Abstract

This paper derives a simple sufficient condition for strict stationarity in the ARCH(∞) class of processes with conditional heteroscedasticity. The concept of persistence in these processes is explored, and is the subject of a set of simulations showing how persistence depends on both the pattern of lag coefficients of the ARCH model and the distribution of the driving shocks. The results are used to argue that an alternative to the usual method of ARCH/GARCH volatility forecasting should be considered.

1 Introduction

The ARCH and GARCH and related classes of volatility models are employed to exploit the fact of local persistence in the volatility of returns processes, so as to predict volatility a number of steps into the future. Notwithstanding the large volume of research that has been devoted to understanding these models since their inception, there remains a degree of mystery surrounding their dynamic properties, and hence the degree to which they assist the effective forecasting of future volatility. Analogies drawn from the theory of linear processes in levels have sometimes been invoked inappropriately in attempts to explain their behaviour, as has been detailed in Davidson (2004) among other commentaries.

This paper considers the ARCH(∞) model of an uncorrelated returns sequence $\{\xi_t\}$ in which, for $-\infty < t < \infty$,

$$\xi_t = \sqrt{h_t} z_t$$

where $z_t \sim$ i.i.d.$(0,1)$ and

$$h_t = \omega + \sum_{j=1}^{\infty} \theta_j \xi_{t-j}^2 \tag{1.1}$$

with $\omega > 0$, $\theta_j \geq 0$ for all j and $S = \sum_{j=1}^{\infty} \theta_j < \infty$. Interest focuses on the three salient features of models of this type: the value of S; the decay rate of the lag coefficients; and the distribution of z_t. Having regard to the first of these features, it is well known that $S < 1$ is a necessary condition for covariance stationarity. Unless this condition applies it is inappropriate to speak of h_t as the 'conditional variance' although it is always well-defined as a volatility indicator. In respect of the second feature, it is also well known that the Bollerslev (1986) GARCH class of models imposes exponential decay rates on the coefficients, and the HYGARCH class due to Davidson (2004) which includes the FIGARCH model of Baillie et al. (1996), embodies hyperbolic decay rates. In respect of the third, the disturbances are often specified to be Gaussian, even though it is a well-known stylized fact that the residuals from estimated GARCH models in financial data can exhibit excess kurtosis.

The question of strict stationarity in covariance nonstationary processes was first examined by Nelson (1990). In the GARCH(1,1) model

$$h_t = \gamma + \alpha\xi_{t-1}^2 + \beta h_{t-1} \tag{1.2}$$

which has the form of (1.1) with $\omega = \gamma/(1-\beta)$, and $\theta_j = \alpha\beta^{j-1}$so that $S = \alpha/(1-\beta)$, he showed the necessary and sufficient condition for strict stationarity to be

$$E\log(\alpha z_1^2 + \beta) < 0. \tag{1.3}$$

Subsequent work on this problem notably includes Bougerol and Picard (1992) who consider the GARCH(p,q) extension of Nelson's result, and emphasize the role of the negativity of the top Lyapunov exponent of a certain sequence of random matrices. Kazakevičius and Leipus (2002) show that a necessary condition for a stationary solution in the ARCH(∞) class is

$$\log S < -E\log(z_1^2) \tag{1.4}$$

while Douc et al. (2008) prove a sufficient condition of the form

$$E|z_1|^{2p}\sum_{j=1}^{\infty}\theta_j^p < 1, \text{ some } p \in (0,1]. \tag{1.5}$$

In this paper we consider conditions for strict stationarity, but also the wider question of the persistence of stationary volatility processes; specifically, how long episodes of high volatility tend to persist, once initiated, and hence how far into the future variations in volatility may feasibly be forecast. This notion of persistence, which is independent of the existence of moments, is made precise in Section 3, where we define it in terms of the (in)frequency of crossings of the median in successive steps. Thus, a process which crosses the median at most a finite number of times in a realization of length T, as $T \to \infty$, is necessarily nonstationary, either converging or diverging. At the other extreme, a serially independent process crosses the median with probability 1/2 at each step, by construction. Conditions for strict stationarity of a process in effect define the boundary beyond which persistence becomes divergence, and there is no reversion tendency defining a stationary distribution. In Section 2, a decomposition of the ARCH(∞) equation is introduced which simplifies the problem of seeing how persistence and stationarity depends on the various model features. We use this representation to derive a new sufficient condition for strict stationarity. In the GARCH(1,1) case where the stationarity boundary in the parameter space is known, we show numerically that our condition is not too far from necessity, in contrast to a strong condition such as (1.5). The properties of these models are shown to be the result of rather complex interactions between the shock distribution and the linear structure. Section 4 reports a comprehensive set of simulations, covering covariance stationary, strictly stationary and nonstationary cases. Section 5 considers the implications of our analysis for the optimal forecasting of volatility, and investigates alternatives to the minimum mean squared error criterion, which is conventional but not necessarily optimal in the context of highly skewed volatility processes. Section 6 contains concluding remarks, and proofs of the propositions stated in Section 2 are gathered in an appendix.

2 Stationarity and Persistence in the ARCH(∞) Class

Write (1.1) in the alternative form

$$h_t = \omega + \sum_{j=1}^{\infty}\psi_{jt}h_{t-j} \tag{2.1}$$

where

$$\psi_{jt} = \theta_j z_{t-j}^2. \tag{2.2}$$

In words, we can describe this as an infinite-order linear difference equation with independently distributed random coefficients.

To focus attention on the persistence properties of (2.1), it is helpful to apply a variant of the so-called Beveridge-Nelson (1981) decomposition (henceforth, BN), which was introduced as a tool of econometric analysis by Phillips and Solo (1992). The BN decomposition is the easily verified identity for polynomials $\lambda(x) = \sum_{j=0}^{\infty} \lambda_j x^j$ having the form

$$\lambda(x) = \lambda(1) + \lambda^*(x)(1-x)$$

where $\lambda_j^* = -\sum_{k=j+1}^{\infty} \lambda_k$. In the present application we consider, for each t, the stochastic polynomial in the lag operator

$$\psi_t(L) = \sum_{j=0}^{\infty} \psi_{jt} L^j$$

where the coefficients are given by (2.2) with $\psi_{0t} = \theta_0 = 0$. The BN form of this expression is

$$\psi_t(L) = \Psi_t + \psi_t^*(L)(1-L)$$

where

$$\Psi_t = \psi_t(1) = \sum_{j=1}^{\infty} \psi_{jt} \tag{2.3}$$

and note that

$$E(\Psi_t) = S. \tag{2.4}$$

The coefficients of $\psi_t^*(L)$ are $\psi_{0t}^* = 0$ and, for $k \geq 1$,

$$\psi_{kt}^* = -\sum_{l=k+1}^{\infty} \theta_l z_{t-l}^2 \leq 0. \tag{2.5}$$

Accordingly write (2.1) as

$$h_t = \omega + \Psi_t h_{t-1} + R_t \tag{2.6}$$

where

$$R_t = \sum_{k=1}^{\infty} \psi_{kt}^* \Delta h_{t-k}. \tag{2.7}$$

Note that if $\{h_t\}$ is a stationary process, the terms Δh_t are negatively autocorrelated and their contribution to the dynamics is therefore high-frequency, in general. That the longer-run persistence and stationarity properties of the process depend critically on the distribution of the sequence $\{\Psi_t\}$ is shown by the following proposition. (Proofs are gathered in the appendix.)

Proposition 2.1 *If the stochastic process $\{h_t^*\}_{t=-\infty}^{\infty}$ where*

$$h_t^* = \omega + \Psi_t h_{t-1}^* \tag{2.8}$$

satisfies a sufficient condition for $P(h_t^* < \infty) = 1$, then $P(h_t < \infty) = 1$ also holds for (2.1).

With this consideration in mind we give the following result, establishing a sufficient condition for stationarity of $\{h_t\}$. For convenience of notation, let the symbol ζ denote the constant $E(\log \Psi_t)$, not depending on t since $\{z_t\}$ is i.i.d.

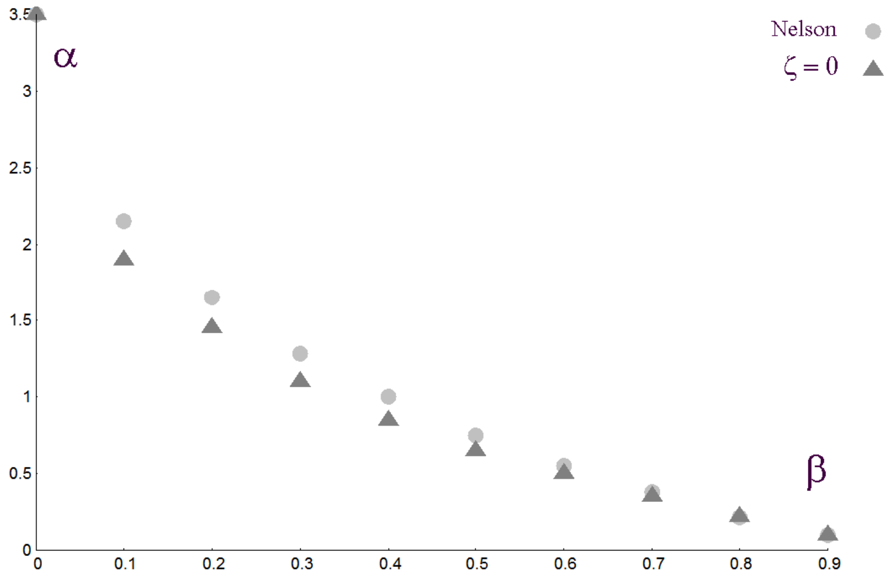

Figure 1: Gaussian GARCH(1,1) model: (α, β) pairs where $\zeta = 0$ and stationarity boundary points (Nelson 1990).

Proposition 2.2 *If*

$$\zeta < 0 \tag{2.9}$$

then $\{h_t^*\}_{t=-\infty}^{\infty}$ *defined by (2.8) is strictly stationary and ergodic.*

Sufficiency of the covariance stationarity condition $S = E(\Psi_t) < 1$ follows from Proposition 2.2 by the Jensen inequality.

Consider this result in the case of the GARCH(1,1) process (1.2). This is a special case because, uniquely among ARCH(∞) processes, it can be expressed exactly in the form (2.8). In other words, we may write the model as

$$h_t = \gamma + \Psi_t^{\dagger} h_{t-1} \tag{2.10}$$

where $\Psi_t^{\dagger} = \alpha z_{t-1}^2 + \beta$ and $\gamma = \omega(1-\beta)$. Proposition 2.2 can be applied directly to (2.10) to obtain condition (1.3), which Nelson (1990) shows to be necessary as well as sufficient. However, writing the model in its ARCH(∞) representation with

$$\Psi_t = \alpha z_{t-1}^2 + \alpha\beta z_{t-2}^2 + \alpha\beta^2 z_{t-3}^2 + \cdots$$

as in (2.3), we obtain

$$\zeta = E(\log \Psi_t) = E[\log(\alpha z_{t-1}^2 + \beta \Psi_{t-1})]. \tag{2.11}$$

In the case $\beta = 0$, so that $S = \alpha$, the conditions (1.3) and (2.9) match. They also match the necessary condition (1.4) which for the GARCH(1,1) case becomes

$$E \log(\alpha z_1^2) < \log(1-\beta).$$

Also, letting $\beta \to 1$ while letting α tend to zero at such a rate as to fix the sum of the coefficients at $S = \alpha/(1-\beta)$, note that condition (2.9) in case (2.11) implies the covariance stationarity condition $S < 1$. This follows because $\Psi_t \to S$ almost surely as $\alpha \to 0$ by the strong law of large numbers, noting that it is a weighted average of i.i.d.random variables with means of unity and weights with finite sum S.

For the intermediate cases with $0 < \beta < 1$, conditions (1.3) and (2.9) do not match but can be compared, giving an opportunity to verify the sharpness of the latter condition. Some numerical experiments with Gaussian shocks are illustrated in Figure 1, showing α-values at which $\zeta \approx 0$ for $\beta = 0, 0.1, 0.2, \ldots, 0.9$. The mean is estimated in each case as the average of 20,000 values of $\log(\Psi_t)$ where Ψ_t is calculated from a generated i.i.d. Gaussian sequence $\{z_t\}$ and the recursion $\Psi_t = \alpha z_{t-1}^2 + \beta\Psi_{t-1}$. The actual stationarity boundary points from (1.3) are shown for comparison, as plotted in Figure 1 of Nelson (1990).[1] By comparison, note that the sufficient condition (1.5) of Douc et al. (2008) is substantially stronger than the bound of Proposition 2.2. For the cases illustrated in Figure 1, the boundary value of $S = \alpha/(1-\beta)$ ranges from 1 at $\beta = 0.9$ up to 2.1 at $\beta = 0.1$. In the Gaussian case, a lower bound on $E|z_1|^{2p}$ is $\sqrt{2/\pi} = 0.798$ at $p = 0.5$, whereas S is a lower bound on the second factor of condition (1.5). For most of these cases, there is no value $p \in (0, 1]$ close to meeting the stated condition.

The way in which these conditions depend on the distribution of z_1^2 can be appreciated by considering Figures 2-4, which show simulated paths ($T = 5000$, with 10,000 presample steps) for three cases of the IGARCH(1,1) model, with $\omega = 1$ and $\beta = 0.9$ in each case. These are among the models studied in Section 4 of the paper. The sole difference between the three cases comes from the shock distributions, which are, respectively, the Student t with 3 degrees of freedom, the Gaussian, and the uniform, in each case normalized to zero mean and unit variance. Estimates of $-E(\log z_1^2)$ (computed as averages of samples of size 20,000) are, respectively, 2.02 for the Student(3), 1.25 for the Gaussian, and 0.87 for the uniform case. These may be compared with $\log(S) = 0$ in the light of the necessary stationarity condition (1.4). The plots show how these characteristics map into differences in persistence, pointing up the somewhat counter-intuitive effect of fat tails on persistence

Turning now to the general ARCH(∞) case, note first that from (2.4) and $\omega > 0$ it follows that the existence of $E(h_t^*)$ requires $S < 1$, mirroring the full model (2.6); in the same case, observe that $E(R_t) = 0$. Except in the case where $S < 1$, stationarity depends on the distribution of Ψ_t and particularly on the degree of positive skewness which, as a moving average of squared shocks, Ψ_t must exhibit in some degree. If the mass of the distribution of Ψ_t falls below one, the mass of the distribution of $\log\Psi_t$ is in the negative part of the line. While $E(\log\Psi_t) < \log S$ by the Jensen inequality, the logarithm of a positive and positively skewed random variable has a more nearly symmetric distribution than the variable itself. Hence, $E(\log\Psi_t)$ lies correspondingly closer to $\text{Median}(\log\Psi_t) = \log(\text{Median}(\Psi_t))$, which in turn lies further below $\log S$, as the skewness is greater. In terms of the dynamics of the process, to the extent that Ψ_t is symmetrically distributed about its mean S, and $S \geq 1$, the probability that a step is convergent, in the sense of Proposition 2.1, is relatively small. The stochastic difference equation defined by (2.8) must, with the complementary probability, behave like either a unit root process with positive drift or an explosive process. However, skewness will increase the proportion of the realizations falling below the mean, yielding stationary behaviour on more frequent occasions, compensated by less frequent but larger excursions above the mean.

In this context, we can appreciate the rather complex role played by the rate of decay of the nonnegative sequence $\{\theta_j\}_{j=1}^{\infty}$, given its fixed sum $S = E(\Psi_1)$. First, note that the skewness of Ψ_1 derives from and is bounded by the skewness in the distribution of the increments $\{z_s^2, s \leq 0\}$. Hence, the necessary condition (1.4) can be understood as the minimal condition for non-divergence when $S \geq 1$. This condition would also be sufficient in the case $\theta_j = 0$ for $j > 1$ and $S = \theta_1 = 1$ (the IARCH(1) model), in which case the distributions of Ψ_1 and z_1^2 match. However, when Ψ_1 is a moving average of the $\{z_s^2\}$ process, the distribution of Ψ_1 depends critically on the distribution of the lag coefficients. Since the lag weights have a finite sum S, the effects

[1] Note that the axes in our figure are interchanged relative to Nelson's figure.

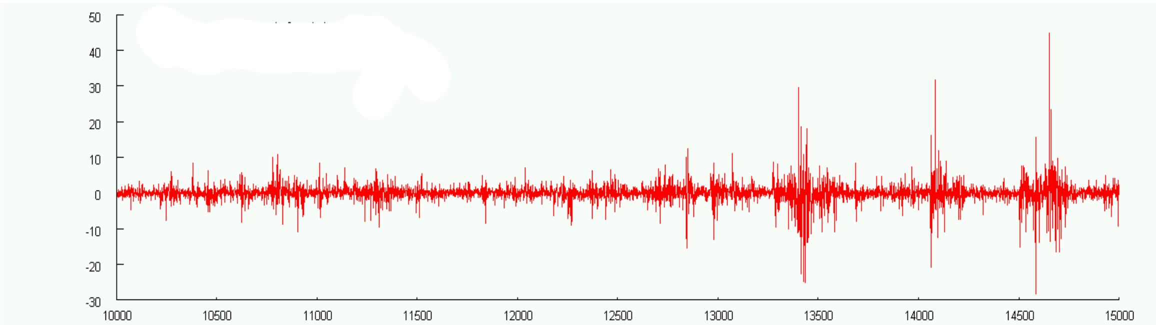

Figure 2: Simulation of IGARCH(1,1) with $\beta = 0.9$ and Student(3) shocks.

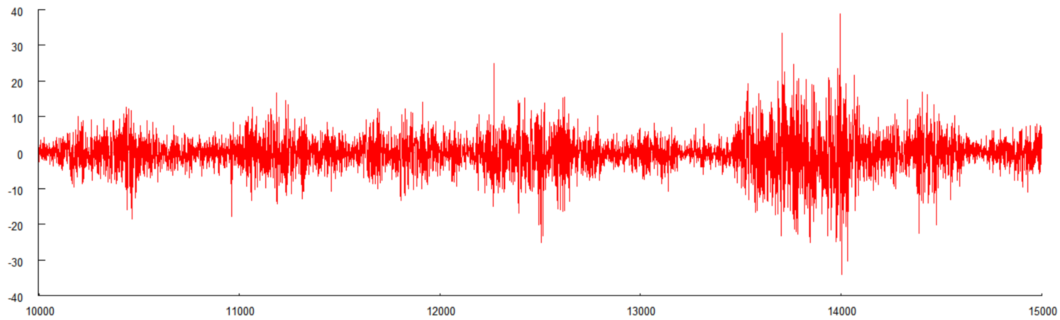

Figure 3: Simulation of IGARCH(1,1) with $\beta = 0.9$ and Gaussian shocks.

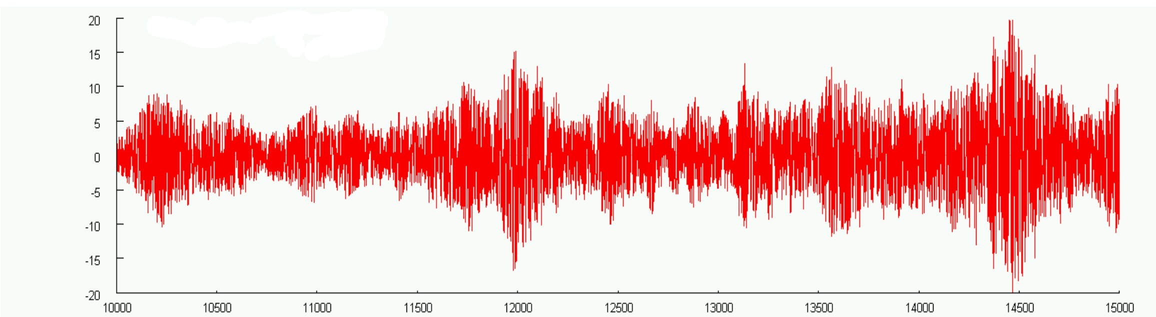

Figure 4: Simulation of IGARCH(1,1) with $\beta = 0.9$ and uniform shocks

of a longer or shorter average lag are to introduce different degrees of averaging of the squared shocks. The somewhat complex nature of this relation depends on the existence of a trade-off between two countervailing effects. Assuming that z_1 possesses a fourth moment, the central limit theorem implies that Ψ_1 is attracted to the normal distribution, with skewness increasingly attenuated, as lag decay gets slower. At the same time, the law of large numbers implies that the variance of Ψ_1 is smaller. The first of these effects is tending to increase the persistence of the $\{h_t^*\}$ process, while the second is tending to lower the influence of h_t^* on the volatility of $\xi_t^* = \sqrt{h_t^*}z_t$, simply because the noise contribution from z_t becomes more dominant as the variations in h_t^* are attenuated. It is therefore difficult to predict the effect of changing the lag decay in any given case.

To summarize: if the contribution of the term R_t in (2.6) to the persistence properties can be largely discounted, as we argue, the persistence and stationarity of the ARCH(∞) process can be related, through the distribution of Ψ_1, to the three key factors: S, the rate of decay of the lag coefficients, and the marginal distribution of z_1. Greater/smaller kurtosis of z_1 implies greater/smaller positive skewness in the distribution of z_1^2, and hence gives rise to less/more persistence in $\{h_t\}$, other things equal. A longer average lag can, counterintuitively, imply a lesser degree of persistence in the observed process, virtually the opposite of the role of lag decay in models of levels, where the sum of the lag coefficients is not constrained in the same way, and shocks are viewed implicitly as having a symmetric distribution. Finally, it is most important to note that the distinction between exponential and hyperbolic decay rates has quite different implications here than in models of levels. There is no counterpart to so-called long memory in levels, otherwise called fractional integration. The dynamics are nonlinear and there is no simple parallel with linear time series models. The closest analogy is with a single autoregressive root which in the covariance nonstationary cases is local to unity.

In the remainder of the paper, we report some simulations to throw light on the volatility persistence properties of alternative simple cases of the ARCH(∞) class. However before that is possible we need a framework for comparing persistence in general time series processes. The next section considers some alternative approaches.

3 Measuring the Persistence of Stationary Time Series

The persistence, or equivalently memory, of a strictly stationary process can be thought of heuristically in terms of the degree to which the history of the process contains information to predict its future path, more accurately than by simple knowledge of the marginal distribution. In the context of univariate forecasting, forecastability must entail that changes in the level of the process are relatively sluggish. It is customary to measure this type of property with reference to the autocovariance sequence, but this is not a valid approach in the absence of second moments.

We resort instead to the idea that the key indicator of persistence is the (in)frequency of reversion towards a point of central tendency. We may formalize this notion by defining the persistence of an arbitrary sequence $\{X_t\}_{t=1}^T$ specifically in terms of the number of occasions on which the series crosses its median point. The direct measure of this property, which is well defined and comparable in any sample sequence whatever, is the relative median-crossing frequency, although it's more convenient to consider the complementary relative frequency of non-crossings. We therefore define

$$J_T = \frac{1}{T}\sum\nolimits_{t=2}^{T} I((X_t - M_T)(X_{t-1} - M_T) > 0) \tag{3.1}$$

where T is sample length, $I(.)$ denotes the indicator of its argument and M_T is the sample median. J_T measures the persistence of a sample as a point in the unit interval. When the sequence is serially independent, $J_T \to 1/2$ as $T \to \infty$, almost surely, by construction. In other words, under independence half of the pairs of successive drawings must fall on different sides of the median on average. The extreme cases are $J_T \to 0$ (anti-persistence) and $J_T \to 1$ (persistence). In the latter case, at most a finite number of median crossings as $T \to \infty$ implies that the sequence either converges, or diverges to infinity. In neither case can it be strictly stationary. The condition limsup $J_T < 1$ is evidently necessary for strict stationarity.

J_T in (3.1) applied to a given sequence measures what we may designate persistence in levels. Persistence in volatility is measured by the statistic analogous to J_T for the squared or (equivalently) absolute values of the series. From the standpoint of returns it is second order persistence, so defined, that is our interest in the present analysis. The J_T statistic can be computed for arbitrary transformations of the variables, and a necessary and sufficient condition for strict stationarity would appear to be that the sequences $\{J_T, T \geq 2\}$ are bounded below 1 for all such variants. However, the two leading cases mentioned appear the important ones in the usual time series context.

J_T is an ordinal measure that is well defined regardless of the existence of moments and is also invariant under monotone transformations. Thus, the cases $X_t = \xi_t^2$ and $X_t = |\xi_t|$ must yield the same value of J_T. More interestingly, it is invariant under the operation of forming the normalized ranks of the series, $\{x_t\}_{t=1}^T$. Letting $\hat{F}_T$ denote the empirical distribution function

$$\hat{F}_T(z) = T^{-1} \sum\nolimits_{s=1}^{T} I(X_s \leq z),$$

$x_t = \hat{F}_T(X_t)$ denotes the relative position of X_t in the sorted sequence $X_{(1)}, \ldots, X_{(T)}$. The sample median of the normalized ranks tends to 1/2 by construction, and when the sample is large enough, J_T must have the same value for $\{x_t\}_{t=1}^T$ as it does for the original series $\{X_t\}_{t=1}^T$. The ranks are also invariant under monotone transformations of the series, so yielding the same values for $X_t = \xi_t^2$ and $X_t = |\xi_t|$ in particular.

Conventional approaches to measuring persistence, for levels or squares/absolute values as the case may be, are based on the autocovariance sequence. There is particular interest in the property of absolute summability of this sequence, often called weak dependence, with strong dependence defining the non-summable case.[2] Popular persistence measures based on the autocovariance sequence are the so-called GPH log-periodogram regression estimators (for different bandwidths) of the fractional persistence parameter d, originally due to Geweke and Porter Hudak (1983). In principle, GPH estimators provide a test of the null hypothesis of weak dependence, although they are well-known to be subject to finite sample bias except under the null of white noise.

Our present interest is due to the fact that the long memory paradigm has proved popular in volatility modelling, and GPH estimation can be validly performed on the normalized ranks of a series regardless of the covariance stationarity property. The particular problem faced in the context of nonstationary volatility is the existence of excessively influential outlying observations, which may invalidate the usual assumptions for valid inference. Rank autocorrelations are free of these influences and may focus more specifically on measuring persistence as characterized here. We should emphasize, though, that our present concerns are not primarily hypothesis testing, but rather to compare and rank different models according to their persistence characteristics.

To calibrate the performance of these alternative measures, we generated some pure fractional series, otherwise known as I(d) processes, for a range of values of d, in samples of size $T = 10,000$, with 5000 pre-sample observations. However, the driving shocks were generated to have an α-stable distribution with $\alpha = 1.8$ and $\beta = 1$, where β is the skewness parameter. The series

---

[2] The well-known difficulty of discriminating between these cases in a finite sample has recently been studied in detail by one of the present authors, see Davidson (2009).

d	J_T	$\hat{d}$	$\hat{d}^R$
0	0.498 (0.004)	-0.033 (0.061)	-0.002 (0.065)
0.3	0.663 (0.009)	0.281 (0.061)	0.330 (0.063)
0.5	0.835 (0.024)	0.496 (0.069)	0.544 (0.068)
0.7	0.948 (0.016)	0.718 (0.078)	0.741 (0.075)
0.9	0.985 (0.006)	0.921 (0.013)	0.986 (0.006)
1	0.992 (0.004)	0.985 (0.056)	0.976 (0.065)

Table 1: Persistence measures in a fractional linear time series, T=10,000. (Means of 100 replications with standard errors in parentheses.)

so constructed do not have second moments and superficially resemble volatility series (after centring) while having a conventional and well-understood linear dependence structure.

Three statistics were computed for these series: J_T in (3.1), the GPH estimator with bandwidth $\sqrt{T}$ for the original series, and also the same GPH estimator for the series of normalized ranks. The simulations were repeated 100 times and the means and standard deviations (in parentheses) of the replications are recorded in Table 1, where $\hat{d}^R$ denotes GPH for the ranked data.

The J_T statistics discriminate rather clearly between the independent case at one end of the dependence spectrum and the strictly nonstationary unit root at the other. The GPH estimates for the raw data in fact behave like consistent estimates of d, while the rank correlation-based estimator appears biased upwards. This is a slightly counter-intuitive result that may or may not be specific to the example considered. However, in our application we are seeking only to rank models, in contexts where a parameter d with the usual linear property is not typically well defined. (In particular, it does not correspond to the 'd' appearing in FIGARCH and HYGARCH models.) We carry this alternative along, chiefly, in a spirit of curiosity about the performance of a seemingly natural measure in the context of an exploration of "long memory in volatility".

4 Some simulation experiments

In this section, we evaluate and compare the properties discussed in Section 2 in the GARCH(1,1) and the "pure" HYGARCH/FIGARCH model. The respective data generation processes are of the form $\xi_t = \sqrt{h_t} z_t$ where $z_t \sim$ i.i.d.$(0,1)$ and either

$$h_t = \omega + \left(1 - \frac{1-\delta L}{1-\beta L}\right)\xi_t^2 \tag{4.1}$$

where $\delta > 0$ and $0 \leq \beta < \min(1, \delta)$ or

$$h_t = \omega + \alpha(1-(1-L)^d)\xi_t^2 \tag{4.2}$$

where $\alpha > 0$ and $0 < d \leq 1$. (See e.g. Davidson (2004) for the context of these examples.) In (4.1), which matches (1.2) on setting $\delta = \alpha + \beta$, $S = (\delta - \beta)/(1-\beta)$; whereas in (4.2), $S = \alpha$. Setting $\delta = 1$ and $\alpha = 1$, respectively, yields the covariance nonstationary IGARCH and FIGARCH models, whereas setting these parameters strictly less than one implies covariance stationarity.

The simulations set a range of values for each of the parameter pairs (δ, β) and (α, d). Covariance stationary cases are specified having $\delta = 0.8$ and $\alpha = 0.8$ respectively. We also simulate nonstationary cases, with $\delta = 1$, $\delta = 1.2$ and $\alpha = 1$, $\alpha = 1.2$. For each of these cases, three values of β and three values of d are chosen, being careful to note that the degree of volatility persistence varies inversely with d (which is of course to be understood as a differencing parameter, not an integration parameter). For each of the nine parameter pairs selected, three different generation processes for z_t are compared: in decreasing order of kurtosis, these are the normalized Student $t(3)$, $z_{St(3)} = t(3)/\sqrt{3}$; the standard Gaussian, z_G; and the normalized uniform distribution, $z_U = \sqrt{12}(U[0,1] - 1/2)$.

Tables 2 and 3 show the results for samples of size $T = 10,000$, with 5000 pre-sample observations to account for any start-up effects. The reported values are the averages of 100 Monte Carlo replications of the generation process, with the replication standard deviations shown in parentheses as a guide to the stability of these persistence indicators. The rows of the tables show the following: first, the sample mean, sample median, and sample logarithmic mean of the random sequences $\{\Psi_t\}_{t=1}^{T}$ as defined in (2.3); second, the values of J_T for various series defined in Section 2: the squared returns, the conditional volatilities h_t, and also the remainder term $R_t = h_t - \omega - \Psi_t h_{t-1}$. The final columns of the tables show, for an alternative view of the persistence, the GPH estimators based on the rank correlations of the squared returns.

The salient points of interest in these experimental results seem to us to be the following. First, the relationships between the proximity of the mean of Ψ_t (measuring S) to the corresponding median,[3] and also the proximity of the logarithmic mean to zero, and the measured persistence of the squared returns. Second, we note that the measured persistence of R_t is in general much lower than that of h_t, confirming the fact that Ψ_t is the key determinant of persistence. Third, we draw attention to the relative persistence of the squared returns and of the volatility series. In the former case, for given δ (or α), and given shock distribution, the median-crossing frequencies (measured by $1 - J_T$) actually rise as the lag decay rates decrease, either through β increasing, or d decreasing. In other words, longer average lags imply less persistence. The reason for this phenomenon has been discussed in Section 2, and the interesting observation is that this effect is large enough to counteract the increased persistence in volatility, h_t, which is also observed.

Finally, we draw attention to the cases with $\delta = 1.2$ and $\alpha = 1.2$, where instances of the logarithmic mean exceeding zero are recorded. In the GARCH case, there is clearly a close correspondence between this occurrence and the evidence that stationarity is violated, in the sense that the median is crossed fewer than ten times in 10,000 steps. The necessary condition (1.4) can also be checked out. Compare the estimated values of $-E(\log z_t^2)$ for the three distributions, as reported in Section 2. When $S = 3$ so that $\log(S) = 1.09$, which is the GARCH case corresponding to $\delta = 1.2$ and $\beta = 0.9$, only the uniform distribution case actually violates the necessary condition, but all the distribution alternatives appear nonstationary. All the HYGARCH examples appear stationary, although the uniform case with $d = 0.5$ appears the closest to divergent.

The estimates of the fractional integration parameter in the last column of the tables are of interest in reflecting the persistence measured by J_T quite closely, increasing across the range with β, but are non-monotone with respect to d. Observe that, for the normal and uniform cases in Table 3, the values obtained for $d = 0.5$ are generally greater than those for either $d = 0.9$ or $d = 0.1$. When the volatility is covariance nonstationary these measures can be quite large, and when it is strictly nonstationary, they fall close to unity. In a series of insightful papers, Mikosch and Stărică (2003, 2004) argue that long range dependence of volatility in financial data should

[3] The medians are much better determined than the skewness coefficients, which were also computed, but not reported since they convey a very similar picture to the mean-median gaps.

Model			Ψ_t			J_T			d^R
δ	$\boldsymbol{\beta}$	Dist'n	Mean	Median	MeanLog	ξ_t^2	h_t	R_t	ξ_t^2
0.8	0.1	$St(3)$	0.772 (0.119)	0.204 (0.004)	−1.630 (0.021)	0.571 (0.005)	0.634 (0.005)	0.468 (0.005)	−0.004 (0.070)
		N	0.777 (0.011)	0.411 (0.007)	−1.015 (0.018)	0.613 (0.005)	0.662 (0.005)	0.460 (0.004)	−0.002 (0.076)
		U	0.778 (0.007)	0.605 (0.011)	−0.746 (0.014)	0.639 (0.006)	0.677 (0.006)	0.469 (0.005)	0.006 (0.071)
	0.4	$St(3)$	0.663 (0.065)	0.265 (0.005)	−1.286 (0.023)	0.553 (0.005)	0.748 (0.004)	0.634 (0.004)	−0.009 (0.078)
		N	0.667 (0.009)	0.480 (0.009)	−0.788 (0.015)	0.576 (0.006)	0.751 (0.006)	0.615 (0.004)	0.014 (0.062)
		U	0.666 (0.006)	0.594 (0.007)	−0.618 (0.011)	0.585 (0.006)	0.739 (0.006)	0.574 (0.005)	0.013 (0.061)
	0.7	$St(3)$	0.329 (0.038)	0.182 (0.004)	−1.638 (0.021)	0.517 (0.005)	0.835 (0.004)	0.767 (0.004)	0.004 (0.072)
		N	0.333 (0.004)	0.289 (0.004)	−1.262 (0.015)	0.519 (0.005)	0.809 (0.005)	0.729 (0.005)	−0.002 (0.058)
		U	0.333 (0.003)	0.323 (0.003)	−1.176 (0.009)	0.521 (0.006)	0.774 (0.005)	0.675 (0.005)	0.014 (0.065)
1	0.1	$St(3)$	1.020 (0.262)	0.262 (0.006)	−1.379 (0.020)	0.588 (0.005)	0.650 (0.005)	0.474 (0.005)	0.002 (0.059)
		N	1.001 (0.015)	0.530 (0.011)	−0.761 (0.014)	0.647 (0.007)	0.699 (0.006)	0.477 (0.004)	0.017 (0.056)
		U	1.001 (0.010)	0.778 (0.015)	−0.495 (0.014)	0.693 (0.008)	0.740 (0.008)	0.495 (0.005)	0.049 (0.056)
	0.5	$St(3)$	1.068 (0.857)	0.442 (0.010)	−0.769 (0.019)	0.575 (0.006)	0.809 (0.004)	0.693 (0.005)	0.008 (0.060)
		N	0.997 (0.012)	0.768 (0.012)	−0.298 (0.016)	0.627 (0.008)	0.841 (0.006)	0.680 (0.005)	0.061 (0.073)
		U	0.999 (0.008)	0.927 (0.009)	−0.158 (0.009)	0.671 (0.010)	0.866 (0.007)	0.642 (0.005)	0.189 (0.060)
	0.9	$St(3)$	0.975 (0.098)	0.693 (0.018)	−0.290 (0.031)	0.553 (0.010)	0.953 (0.005)	0.900 (0.004)	0.279 (0.073)
		N	0.998 (0.015)	0.954 (0.016)	−0.049 (0.013)	0.594 (0.022)	0.965 (0.005)	0.873 (0.005)	0.615 (0.074)
		U	0.999 (0.008)	0.992 (0.008)	−0.022 (0.010)	0.629 (0.032)	0.971 (0.006)	0.844 (0.005)	0.729 (0.072)
1.2	0.1	$St(3)$	1.209 (0.180)	0.319 (0.007)	−1.176 (0.019)	0.607 (0.005)	0.669 (0.005)	0.480 (0.004)	0.004 (0.069)
		N	1.224 (0.016)	0.648 (0.011)	−0.565 (0.015)	0.685 (0.007)	0.739 (0.007)	0.495 (0.005)	0.005 (0.068)
		U	1.223 (0.012)	0.951 (0.018)	−0.294 (0.013)	0.760 (0.009)	0.807 (0.008)	0.517 (0.006)	0.050 (0.067)
	0.5	$St(3)$	1.396 (0.188)	0.619 (0.013)	−0.428 (0.024)	0.617 (0.007)	0.843 (0.006)	0.708 (0.005)	0.054 (0.068)
		N	1.400 (0.020)	1.080 (0.018)	0.037 (0.015)	0.840 (0.028)	0.952 (0.010)	0.714 (0.006)	0.574 (0.094)
		U	1.399 (0.010)	1.299 (0.013)	0.178 (0.009)	0.994 (0.003)	0.998 (0.001)	0.779 (0.025)	0.999 (0.025)
	0.9	$St(3)$	3.025 (0.381)	2.090 (0.055)	0.810 (0.034)	0.999 (0.001)	1.000 (0.001)	1.000 (0.001)	0.959 (0.099)
		N	2.999 (0.044)	2.867 (0.048)	1.045 (0.016)	1.000 (0)	1.000 (0)	1.000 (0)	0.991 (0.047)
		U	2.998 (0.026)	2.976 (0.028)	1.078 (0.009)	1.000 (0)	1.000 (0)	1.000 (0)	1.019 (0.010)

Table 2: Series properties and persistence measures for the GARCH(1,1) model

Model			Ψ_t			J_T			d^R
α	d	Dist'n	Mean	Median	MeanLog	ξ_t^2	h_t	R_t	ξ_t^2
0.8	0.9	$St(3)$	0.805 (0.140)	0.218 (0.005)	−1.418 (0.027)	0.571 (0.005)	0.634 (0.007)	0.656 (0.016)	0.018 (0.073)
		N	0.800 (0.011)	0.412 (0.008)	−0.898 (0.016)	0.614 (0.005)	0.655 (0.006)	0.608 (0.007)	0.030 (0.073)
		U	0.801 (0.008)	0.622 (0.012)	−0.678 (0.012)	0.642 (0.006)	0.669 (0.007)	0.570 (0.006)	0.018 (0.068)
	0.5	$St(3)$	0.797 (0.091)	0.422 (0.012)	−0.738 (0.030)	0.556 (0.006)	0.785 (0.012)	0.738 (0.020)	0.172 (0.093)
		N	0.800 (0.011)	0.614 (0.010)	−0.432 (0.016)	0.577 (0.006)	0.754 (0.008)	0.656 (0.007)	0.157 (0.078)
		U	0.798 (0.007)	0.710 (0.009)	−0.345 (0.010)	0.585 (0.006)	0.722 (0.007)	0.599 (0.005)	0.134 (0.070)
	0.1	$St(3)$	0.765 (0.079)	0.611 (0.030)	−0.400 (0.056)	0.523 (0.007)	0.873 (0.017)	0.809 (0.012)	0.224 (0.101)
		N	0.778 (0.010)	0.730 (0.010)	−0.279 (0.014)	0.524 (0.005)	0.797 (0.010)	0.732 (0.006)	0.143 (0.066)
		U	0.779 (0.006)	0.763 (0.007)	−0.266 (0.008)	0.525 (0.005)	0.753 (0.013)	0.666 (0.005)	0.149 (0.062)
1	0.9	$St(3)$	0.980 (0.120)	0.271 (0.005)	−1.199 (0.020)	0.588 (0.005)	0.656 (0.007)	0.673 (0.016)	0.043 (0.073)
		N	1.002 (0.013)	0.516 (0.011)	−0.674 (0.016)	0.648 (0.007)	0.702 (0.010)	0.648 (0.014)	0.105 (0.083)
		U	1.001 (0.008)	0.777 (0.013)	−0.453 (0.014)	0.695 (0.009)	0.741 (0.011)	0.619 (0.011)	0.166 (0.062)
	0.5	$St(3)$	0.993 (0.117)	0.528 (0.016)	−0.507 (0.035)	0.575 (0.009)	0.818 (0.017)	0.760 (0.023)	0.291 (0.086)
		N	0.999 (0.014)	0.767 (0.011)	−0.207 (0.014)	0.622 (0.019)	0.832 (0.024)	0.696 (0.015)	0.410 (0.087)
		U	0.999 (0.009)	0.887 (0.012)	−0.122 (0.009)	0.651 (0.022)	0.838 (0.026)	0.635 (0.010)	0.468 (0.085)
	0.1	$St(3)$	0.995 (0.306)	0.769 (0.047)	−0.172 (0.067)	0.536 (0.013)	0.893 (0.022)	0.820 (0.016)	0.352 (0.107)
		N	0.976 (0.013)	0.914 (0.012)	−0.057 (0.013)	0.537 (0.006)	0.855 (0.021)	0.744 (0.006)	0.304 (0.058)
		U	0.973 (0.007)	0.953 (0.007)	−0.042 (0.009)	0.538 (0.006)	0.839 (0.024)	0.682 (0.006)	0.344 (0.060)
1.2	0.9	$St(3)$	1.183 (0.129)	0.326 (0.007)	−1.013 (0.023)	0.606 (0.005)	0.682 (0.009)	0.693 (0.020)	0.083 (0.089)
		N	1.198 (0.015)	0.617 (0.012)	−0.496 (0.015)	0.702 (0.027)	0.780 (0.033)	0.716 (0.034)	0.305 (0.116)
		U	1.201 (0.010)	0.932 (0.015)	−0.271 (0.014)	0.830 (0.041)	0.895 (0.033)	0.751 (0.022)	0.562 (0.101)
	0.5	$St(3)$	1.211 (0.242)	0.631 (0.021)	−0.322 (0.033)	0.620 (0.032)	0.871 (0.030)	0.792 (0.024)	0.431 (0.121)
		N	1.193 (0.018)	0.916 (0.015)	−0.023 (0.014)	0.929 (0.034)	0.979 (0.013)	0.721 (0.010)	0.966 (0.047)
		U	1.193 (0.010)	1.062 (0.013)	0.063 (0.008)	0.977 (0.010)	0.992 (0.005)	0.683 (0.009)	1.004 (0.013)
	0.1	$St(3)$	1.148 (0.104)	0.917 (0.041)	0.010 (0.057)	0.574 (0.047)	0.925 (0.026)	0.827 (0.011)	0.454 (0.130)
		N	1.167 (0.015)	1.094 (0.014)	0.125 (0.013)	0.643 (0.029)	0.956 (0.013)	0.788 (0.014)	0.643 (0.058)
		U	1.168 (0.009)	1.144 (0.010)	0.141 (0.009)	0.686 (0.024)	0.964 (0.010)	0.792 (0.019)	0.695 (0.056)

Table 3: Series properties and persistence measures for the HY/FIGARCH model

be attributed to structural breaks in the unconditional variance, rather than to GARCH-type dynamics. However, it is clear that apparent long range dependence can be observed in the stationary cases simulated here. We would agree with these authors that the evidence of long-range dependence is spurious, in the sense that it is not generated by a fractionally integrated structure, as it is in Table 1 for example. However, our diagnosis of the cause does not invoke structural breaks. Rather, we see it as a phenomenon analogous to having an autoregressive root local to unity in a levels process, leading to Ornstein-Uhlenbeck-type dynamics which are easily confused with long memory in finite samples. However, the analogy is necessarily a loose one in view of the special features of the volatility process which we have detailed in Section 2.

5 Implications for volatility forecasting

When using models of the ARCH/GARCH class for volatility forecasting two or more steps ahead, the usual methodology is to apply the standard recursion for a minimum mean squared error (MSE) forecast, with ξ_{T+j}^2 for $j > 0$ replaced by its (assumed) conditional expectation. Among many references describing this technique see for example Poon (2005) page 39 and also the Eviews 8 User Guide (2013), page 218, for a practical implementation.

In other words, if h_t is defined by (1.1) (and implicitly assuming the parameters are replaced by appropriate estimates) we would replace ξ_t^2 by $E_{t-1}\xi_t^2 = h_t$, and so set[4]

$$\hat{h}_{t+1|t-1} = \omega + \theta_1 h_t + \sum\nolimits_{j=2}^{\infty} \theta_j \xi_{t-j+1}^2. \tag{5.1}$$

The volatility forecast error accordingly has the form

$$\begin{aligned} f_{t+1|t-1} &= h_{t+1} - \hat{h}_{t+1|t-1} \\ &= \theta_1(\xi_t^2 - h_t) \\ &= \theta_1 h_t (z_t^2 - 1). \end{aligned} \tag{5.2}$$

In the general k-step ahead case,

$$\hat{h}_{t+k|t-1} = \omega + \sum\nolimits_{j=1}^{k-1} \theta_j \hat{h}_{t-j+k|t-1} + \theta_k h_t + \sum\nolimits_{j=k+1}^{\infty} \theta_j \xi_{t-j+k}^2. \tag{5.3}$$

and so

$$f_{t+k|t-1} = \sum\nolimits_{j=1}^{k-1} \theta_j f_{t-j+k|t-1} + \sum\nolimits_{j=1}^{k} \theta_j h_{t-j+k}(z_{t-j+k}^2 - 1) \tag{5.4}$$

For example, consider the GARCH(1,1) model in (4.1) which rearranges as

$$h_{t+1} = \omega(1-\beta) + [(\delta - \beta) z_t^2 + \beta] h_t.$$

If z_t^2 is replaced by $E_{t-1} z_t^2 = 1$ to construct the forecast, (5.2) reduces to

$$f_{t+1|t-1} = (\delta - \beta) h_t (z_t^2 - 1).$$

The problem with this formulation, as the preceding analysis demonstrates, is that due to the skewness of the distribution of z_t^2, the mean may not be the best measure of central tendency. The persistence of the process, and hence its forecastability, will be exaggerated by this choice. In effect, the problem is closely allied to that of forecasting in model (2.8) by using S as the forward projection for unobserved Ψ_t. S is not the value that Ψ_t is close to with highest probability, and

[4] We call this expression the two-step volatility forecast since h_t itself is of course the one-step forecast.

hence the one that will deliver an accurate projection with high probability. The majority of volatility forecasts will be "overshoots", balanced by a smaller number of more extreme "undershoots". The forecast is unbiased in the sense $E(f_{t+k|t-1}) = 0$ when this expectation is defined, but this condition excludes the IGARCH and FIGARCH and other nonstationary cases. Even if the mean squared forecast error is defined, in this context, it is not clear that the MSE is an appropriate loss function.

We investigated this issue experimentally with the results reported in Tables 4 and 5 for the GARCH(1,1) and pure HY/FIGARCH models respectively. We studied the distribution of errors in the two-step forecasts constructed under different assumptions about the appropriate measure of central tendency of the shocks, denoted by M in the definition

$$f_{t+1|t-1} = \theta_1 h_t (z_t^2 - M). \tag{5.5}$$

The median absolute values (MAVs) of the variables defined in (5.5) were computed for six choices of M. In the tables, the minimum value of the MAV in each row is indicated in boldface. Note that in only two of these cases does M exceed 0.5 and in both, the difference from the adjacent lower value is minimal. The rule that $M = 0.1$ gives the best result for the Student(3) case, $M = 0.3$ for the Gaussian case and $M = 0.5$ for the uniform case appears to hold quite generally. The implication may be that future volatility is significantly overstated by conventional procedures.

We can reasonably assume that the optimal M values are those closest to the modes of the respective distributions. While estimating the mode of an empirical distribution is not a straightforward procedure, constructing medians is easy and the medians of our squared normalized distributions, estimated from samples of size 10,000, are 0.763 for the uniform, 0.423 for the Gaussian and 0.176 for the Student(3). In default of a more precise analysis, a rough and ready rule of thumb would be to estimate the MAV-minimizing M by 2/3 times the sample median of the normalized residuals. This corresponds to computing the k-step volatility forecasts by the recursion

$$\hat{h}_{t+k|t-1} = \omega + \tfrac{2}{3}\mathrm{Median}(z_t^2) \sum\nolimits_{j=1}^{k} \theta_j \hat{h}_{t-j+k|t-1} + \sum\nolimits_{j=k+1}^{\infty} \theta_j \xi_{t-j+k}^2. \tag{5.6}$$

where $\hat{h}_{t|t-1} = h_t$.

A more extensive simulation study than the present one would be needed to confirm this recommendation. We do note, however, that the rule would apply successfully in both the covariance stationary and the covariance nonstationary cases that have been simulated here. Although h_t has the interpretation of a conditional variance only in the stationary case, note that the problem we highlight is not connected with the non-existence of moments. It is entirely a matter of adopting a minimum MSE estimator of a highly skewed distribution, such that the outcome is overestimated in a substantially higher proportion of cases than it is underestimated.

6 Concluding Remarks

In this paper we have investigated the dynamics of certain conditional volatility models with a view to understanding their propensity to predict persistent patterns of high or low volatility. Understanding how persistence depends on the various model characteristics, while intriguing and often counterintuitive, is perhaps a matter of mainly theoretical interest. However, there is also an important message here for practitioners. Conventional forecasting methodologies that are optimal under the assumption of symmetrically distributed shocks may be viewed as overstating the degree of future volatility. This is, of course, an issue essentially of the preferred choice of loss function. Practitioners may validly elect to favour the unbiasedness and minimum MSE

Model			M					
δ	$\boldsymbol{\beta}$	Dist'n	1	0.9	0.7	0.5	0.3	0.1
0.8	0.1	$St(3)$	0.070	0.063	0.049	0.035	0.020	***0.010***
		N	0.078	0.070	0.054	0.041	***0.032***	0.047
		U	0.091	0.083	0.071	***0.066***	0.085	0.121
	0.4	$St(3)$	0.171	0.153	0.118	0.084	0.052	***0.028***
		N	0.204	0.185	0.148	0.115	***0.090***	0.111
		U	0.232	0.216	0.184	***0.162***	0.176	0.246
	0.7	$St(3)$	0.073	0.065	0.050	0.036	0.022	***0.010***
		N	0.076	0.069	0.055	0.041	***0.028***	0.034
		U	0.076	0.070	0.058	0.046	***0.045***	0.065
1.0	0.1	$St(3)$	0.094	0.084	0.064	0.045	0.028	***0.015***
		N	0.122	0.110	0.088	0.069	***0.061***	0.085
		U	0.184	0.174	***0.160***	0.161	0.193	0.257
	0.5	$St(3)$	0.338	0.303	0.234	0.168	0.108	***0.069***
		N	0.693	0.637	0.536	0.446	***0.386***	0.410
		U	1.198	1.150	1.076	***1.034***	1.067	1.248
	0.9	$St(3)$	0.246	0.221	0.173	0.127	0.085	***0.054***
		N	1.118	1.033	0.876	0.734	0.621	***0.592***
		U	2.352	2.267	2.109	1.984	***1.932***	2.103

Table 4: MAV 2-step forecast error in GARCH(1,1), against M (see(5.5))

Model			M					
α	d	Dist'n	1	0.9	0.7	0.5	0.3	0.1
0.8	0.9	$St(3)$	0.038	0.034	0.026	0.018	0.011	***0.005***
		N	0.043	0.039	0.030	0.023	***0.017***	0.026
		U	0.050	0.046	0.039	***0.036***	0.048	0.068
	0.5	$St(3)$	0.156	0.140	0.108	0.078	0.049	***0.027***
		N	0.216	0.196	0.157	0.122	***0.094***	0.114
		U	0.244	0.226	0.192	***0.166***	0.182	0.257
	0.1	$St(3)$	0.053	0.047	0.036	0.026	0.016	***0.007***
		N	0.055	0.049	0.039	0.029	***0.019***	0.024
		U	0.053	0.049	0.041	0.032	***0.031***	0.045
1.0	0.9	$St(3)$	0.050	0.045	0.034	0.024	0.014	***0.008***
		N	0.073	0.066	0.053	0.041	***0.036***	0.050
		U	0.116	0.110	***0.101***	***0.101***	0.120	0.161
	0.5	$St(3)$	0.300	0.269	0.208	0.150	0.097	***0.060***
		N	1.382	1.271	1.060	0.874	***0.734***	0.751
		U	3.862	3.688	3.372	***3.164***	3.198	3.838
	0.1	$St(3)$	0.429	0.384	0.298	0.216	0.138	***0.076***
		N	0.663	0.603	0.488	0.379	***0.279***	0.314
		U	0.658	0.614	0.526	0.440	***0.419***	0.599

Table 5: MAV 2-step forecast error in HY/FIGARCH, against M (see(5.5))

properties over minimizing the MAV. They should nonetheless not overlook the fact that the usual rationale for the former criterion implicitly assumes a Gaussian framework, and is arguably inappropriate in the context of predicting volatility.

A Appendix: Proofs

A.1 Proof of Proposition 2.1

First, consider the case of where $\{\psi_{jt}\}$ is replaced by $\{\psi_j\}$, a nonstochastic sequence of coefficients. Then

$$h_t = \omega + \sum_{j=1}^{\infty} \psi_j h_{t-j} \tag{A-1}$$

with $\omega > 0$ and $\psi_j \geq 0$ for all $j \geq 1$ has a stable, positive solution if and only if this is true of the equation

$$h_t^* = \omega + \left(\sum_{j=1}^{\infty} \psi_j \right) h_{t-1}^*. \tag{A-2}$$

Stable solutions of (A-1) and (A-2), if they exist, are both of the form

$$\frac{\omega}{1 - \sum_{j=1}^{\infty} \psi_j} > 0$$

implying in both cases the necessary and sufficient condition

$$\sum_{j=1}^{\infty} \psi_j < 1. \tag{A-3}$$

Next, consider the stochastic sequence $\{\psi_{jt}\}$. Let this be randomly drawn at date t_0, as the functional of the random sequence $\{z_{t_0 - j}, j > 0\}$, and then let a step be taken according to either equation (2.1) or equation (2.8). Call this in either case a *convergent step* if $\sum_{j=1}^{\infty} \psi_{jt_0} = \Psi_{t_0} < 1$. That is, if the process is allowed to continue with this *same* fixed drawing, the sequence of steps so generated must approach the particular solution

$$\overline{h}_0 = \frac{\omega}{1 - \Psi_{t_0}}. \tag{A-4}$$

This is a drawing from the common distribution of stable solutions, which are almost surely finite.

Suppose that every step taken is convergent, in this sense. Then, the sequence is always moving so as to reduce its distance from some point in the distribution of stable solutions. It therefore cannot diverge. More generally, let each step have a certain fixed probability of being convergent. The probability that the sequence diverges can be reduced to zero by setting this probability high enough. This is, from elementary considerations, a sufficient condition for $\{h_t^*\}$ to be finite almost surely.

To show that the same condition is sufficient for $\{h_t\}$ generated by (2.1) to be finite almost surely, first note that the step defined by (2.8) can be written for given Ψ_{t_0} in the form

$$\Delta h_t^* = (\Psi_{t_0} - 1)(h_{t-1}^* - \overline{h}_0). \tag{A-5}$$

In this representation, the condition for a convergent step is that Δh_t^* and $h_{t-1}^* - \overline{h}_0$ have different signs. Now write the BN form (2.6) in the equivalent representation, as

$$\Delta h_t = (\Psi_{t_0} - 1)(h_{t-1} - \overline{h}_0) + R_t^0 \tag{A-6}$$

where the remainder, like Ψ_{t_0}, is specified for the particular shock sequence $\{z_{t_0-j}, j > 0\}$ as

$$R_t^0 = \sum_{k=1}^{\infty} \psi_{kt_0}^* \Delta h_{t-k}. \tag{A-7}$$

In this case, $\Psi_{t_0} < 1$ does not imply $\Delta h_t(h_{t-1} - \overline{h}_0) < 0$ since the sign of Δh_t also depends on R_t^0.

For the case $h_{t-1} > \overline{h}_0$, consider the circumstances in which $R_t^0 > 0$. Rearrangement of the sum (A-7) leads to

$$R_t^0 = -\sum_{k=2}^{\infty} \theta_k z_{t_0-k}^2 (h_{t-1} - h_{t-k})$$

so that a necessary condition for $R_t^0 > 0$ is that $h_{t-1} < h_{t-k}$ for at least one value of $k > 1$. This shows that with $\Psi_{t_0} < 1$ a sequence $\{h_t\}$ generated by (A-6) can never diverge, and is almost surely finite. Conversely, if $h_{t-1} < \overline{h}_0$ the necessary condition for $R_t^0 < 0$ is $h_{t-1} > h_{t-k}$ for at least one $k > 1$, although this case is not critical to the property $P(h_t < \infty) = 1$. ∎

A.2 Proof of Proposition 2.2

The solution of (2.8) is

$$h_t^* = \omega \left(1 + \sum_{m=1}^{\infty} \prod_{k=0}^{m-1} \Psi_{t-k} \right). \tag{A-8}$$

Since $\sum_{j=0}^{\infty} \theta_j < \infty$ and the sequence $\left\{ \sum_{j=1}^{m} \theta_j z_{t-j}^2, m \geq 1 \right\}$ is monotone, Ψ_t is a measurable function of $\{z_s, -\infty < s < t\}$ by (e.g.) Davidson (1994), Theorems 3.25 and 3.26. The sequence $\{\Psi_t, -\infty < t < \infty\}$ is therefore strictly stationary and ergodic.[5] It follows by the ergodic theorem that

$$\frac{1}{m} \sum_{k=0}^{m-1} \log \Psi_{t-k} \xrightarrow{\text{a.s.}} \zeta. \tag{A-9}$$

Hence, with probability one,

$$\limsup_{m \to \infty} \mathrm{e}^{-m\zeta} \prod_{k=0}^{m-1} \Psi_{t-k} < \infty$$

for $-\infty < t < \infty$. There therefore exists $N < \infty$ such that $h_t^* = h_{1t}^* + O(\mathrm{e}^{N\zeta})$ with probability 1, where

$$h_{1t}^* = \omega \left(1 + \sum_{m=1}^{N} \prod_{k=0}^{m-1} \Psi_{t-k} \right). \tag{A-10}$$

The remainder term can be made as small as desired by taking N large enough, and (A-10) is a measurable function of $\{z_s, -\infty < s < t\}$ by (e.g.) Davidson (1994) Theorem 3.25. Strict stationarity and ergodicity of $\{h_t^*, -\infty < t < \infty\}$ follows, completing the proof. ∎

[5] Nelson (1990) cites Theorem 3.5.8 of Stout (1974) in support of a comparable assertion to this one. While the conditions do not precisely correspond, Phillips (1988) Section 1.15 provides a concise proof for the general case of doubly-infinite sequences.

References

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

Beveridge, S., and C. R. Nelson, (1981), A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle, *Journal of Monetary Economics* 7, 151–174.

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

Bougerol, P., and N. Picard (1992): "Stationarity of GARCH processes and some nonnegative time series," *Journal of Econometrics*, 52, 115–127.

Davidson, J. (1994) *Stochastic Limit Theory* Oxford: Oxford University Press.

Davidson, J. (2004) Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model *Journal of Business & Economic Statistics* 22,16-29.

Davidson, J. (2009) When is a time series I(0)? Chapter 13 of *The Methodology and Practice of Econometrics*, eds. J. Castle and N. Shepherd, Oxford: Oxford University Press.

Douc, R., F. Roueff and P. Soulier (2008) On the existence of some ARCH(∞) processes, *Stochastic Processes and their Applications* 118, 755-761.

Eviews 8 User Guide (2013), Quantitative Micro Software LLC, Irvine, CA.

Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221-38.

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

Kazakevičius, V. and R. Leipus (2002) On Stationarity in the ARCH(∞) Model, *Econometric Theory* 18, 1-16.

Mikosch, T and C Stărică (2003) Long range dependence and ARCH modelling, in *Theory and Applications of Long Range Dependence*, edited by P. Doukhan, G. Oppenheim and M. S. Taqqu Boston: Birkhauser.

Mikosch, T and C Stărică (2004) Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects, *Review of Economics and Statistics* 86 (1) 378-390.

Nelson, D. (1990) Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory*, 6, 318–334.

Phillips, P. C. B. (1988) *Lectures on Stationary and Nonstationary Time Series*, Institute of Advanced Studies, Vienna. Text at http://korora.econ.yale.edu/phillips/teach/notes/1988-lectures.pdf

Phillips, P. C. B. and V. Solo (1992) Asymptotics for linear processes, *Annals of Statistics* 20, 971-1001

Poon, S-H. (2005) *A Practical Guide to Forecasting Financial Market Volatility*, Chichester, John Wiley and Sons Ltd.

Stout, W. F. (1974) *Almost Sure Convergence*, New York: Academic Press