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

James Davidson  Xiaoyu Li
University of Exeter  University of Exeter
March 24, 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 \( z_t \sim \text{i.i.d.}(0, 1) \) and

\[ h_t = \omega + \sum_{j=1}^{\infty} \theta_j \xi_{t-j}^2 \quad (1.1) \]

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) who derived the necessary and sufficient condition

$$E \log(\alpha z^2 + \beta) < 0$$

(1.2)

for strict stationarity in the GARCH(1,1) model

$$h_t = \omega + \alpha z^2_{t-1} + \beta h_{t-1}.$$  

(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($\infty$) class is

$$\log S < -E \log(z^2_1)$$

(1.4)

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

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

(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($\infty$) 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. Finally, 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.

2 Stationarity and Persistence in the ARCH($\infty$) Class

Write (1.1) in the alternative form

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

(2.1)
where $\psi_{jt} = \theta_j z_{t-j}^2$. 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, which was introduced as a tool of econometric analysis by Phillips and Solo (1992).

The BN decomposition is the easily verified identity for polynomials $\lambda(z) = \sum_{j=0}^{\infty} \lambda_j z^j$ having the form

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

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 $\psi_{jt} = \theta_j z_{t-j}^2$ with $\psi_{0t} = \theta_0 = 0$. The BN form of this expression is

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

where

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

and

$$E(\Psi_t) = S.$$ (2.3)

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

$$\psi_{kt}^* = -\sum_{l=k+1}^{\infty} \theta_l z_{t-l}^2 \leq 0.$$ (2.4)

Accordingly write (2.1) as

$$h_t = \omega + \Psi_t h_{t-1} + R_t$$ (2.5)

where

$$R_t = \sum_{k=1}^{\infty} \psi_{kt}^* \Delta h_{t-k}.$$ (2.6)

Note that if $\{h_t\}$ is a stationary process, the terms $\Delta h_t$ 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.

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

$$h_t^* = \omega + \Psi_t h_{t-1}^*$$ (2.7)

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

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

$$h_t = \omega + \sum_{j=1}^{\infty} \psi_j h_{t-j}$$ (2.8)
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

\[
h^*_t = \omega + \left( \sum_{j=1}^{\infty} \psi_j \right) h^*_{t-1}.
\]

(2.9)

Stable solutions of (2.8) and (2.9), if they exist, must both be of the form

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

implying in both cases the necessary and sufficient condition

\[
\sum_{j=1}^{\infty} \psi_j < 1.
\]

(2.10)

Next, consider the stochastic sequence \( \{\psi_{jt}\} \). Let this be randomly drawn at date \( t \), and then a step taken according to either equation (2.1) or equation (2.7). Call this in either case a convergent step if \( \sum_{j=1}^{\infty} \psi_{jt} < 1 \), such that, if the process were allowed to continue with the same fixed drawing, the sequence of steps must approach the particular solution \( \omega/(1 - \sum_{j=1}^{\infty} \psi_{jt}) \), a drawing from the common distribution of stable solutions. Suppose 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 a sufficient condition for \( \{h_t\} \) to be finite almost surely.

To show that it is also sufficient for \( \{h_t\} \) to be finite almost surely, consider the BN form (2.5), and notice that \( \psi^*_{jt} < 0 \) by (2.4). Therefore, if \( h_{t-1} > h_{t-2} > h_{t-3} > \cdots \), for example, then \( h_t < \omega + \Psi_t h_{t-1} \). More generally, it holds for every \( j \geq 1 \) that \( R_t \) increases if and only if \( \Delta h_{t-j} \) decreases. The difference components therefore must always contribute a damping effect, such that the probability of divergence of \( \{h_t\} \) cannot exceed that of \( \{h^*_t\} \).

With this consideration in mind we give the following result, as establishing a sufficient condition for stationarity of \( \{h_t\} \).

**Proposition 2.2** If \( E(\log \Psi_t) = \zeta < 0 \) then \( \{h^*_t\}_{t=-\infty}^{\infty} \) defined by (2.7) is strictly stationary and ergodic.

**Proof** The solution of (2.7) is

\[
h^*_t = \omega \left( 1 + \sum_{m=1}^{\infty} \prod_{k=0}^{m-1} \Psi_{t-k} \right).
\]

(2.11)

Since \( \sum_{j=0}^{\infty} \theta_j < \infty \) and the sequence \( \{\sum_{j=1}^{m} \theta_j r_{t-j}, m \geq 1\} \) is monotone, \( \Psi_t \) 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. It follows by the ergodic theorem that

\[
\frac{1}{m} \sum_{k=0}^{m-1} \log \Psi_{t-k} \xrightarrow{a.s.} \zeta.
\]

(2.12)
Hence, with probability one,

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

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

$$h_{1t}^* = \omega \left( 1 + \sum_{m=1}^{N} \prod_{k=0}^{m-1} \Psi_{t-k} \right). \quad (2.13)$$

The remainder term can be made as small as desired by taking $N$ large enough, and (2.13) is a measurable function of $\{z_s, -\infty < s \leq t\}$ by (e.g.) Davidson (1994) Theorem 3.25. Strict stationarity and ergodicity of $\{h_t^*, -\infty < t < \infty\}$ follows, completing the proof. 

Consider this result for the case of the GARCH(1,1) process (1.3). In this case, $\zeta = E \log(\Psi_t)$, $\zeta = E \log(\alpha z_{t-1}^2 + \alpha \beta z_{t-2}^2 + \alpha \beta^2 z_{t-3}^2 + \cdots)$, $\zeta = E \log(\alpha z_{t-1}^2 + \beta \Psi_{t-1})$ (2.14)

which can be compared with condition (1.2). The conditions agree directly for the case $\beta = 0$, and in this case, since $S = \alpha$, also match the necessary condition (1.4). Also, letting $\beta \to 1$ while letting $\alpha$ tend to zero at such a rate as to fix the sum of the coefficients at $S = \alpha/(1 - \beta)$, note that the condition $\zeta < 0$ in (2.14) yields the correct result $S < 1$, since $\Psi_t \to S$ almost surely, by the strong law of large numbers as $\alpha \to 0$. The limiting case of (2.7) when $\alpha$ is small enough solves as $h = \omega/(1 - S)$, setting stability condition $\beta < \alpha + \beta < 1$, and the condition (1.2) also returns $\beta < 1$. For intermediate cases the ranking is ambiguous, but we expect $\zeta$ to dominate $E \log(\alpha z_{t-1}^2 + \beta)$ since (1.2) is known to be necessary as well as sufficient. Some numerical experiments with Gaussian shocks are illustrated in Figure 1, showing $\alpha$-values at which $\zeta \approx 0$. 

Figure 1: Gaussian GARCH(1,1) model: ($\alpha, \beta$) pairs where $\zeta = 0$ and stationarity boundary points (Nelson 1990).
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 $\Psi_t$ is calculated from a generated i.i.d. Gaussian sequence $\{z_t\}$ and the recursion indicated in the last member of (2.14). The actual stationarity boundary points from (1.2) are shown for comparison, as plotted in Figure 1 of Nelson (1990). 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 $z_t^2$ 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 $-E(\log z_t^2)$ (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($\infty$) case, note first that from (2.3) and $\omega > 0$ it follows that the existence of $E(h_t)$ requires $S < 1$, mirroring the full model (2.5); in the same case, observe that $E(R_t) = 0$. Except in the case where $S < 1$, stationarity depends on the distribution of $\Psi_t$ and particularly on the degree of positive skewness which, as a moving average of squared shocks, $\Psi_t$ must exhibit in some degree. If the mass of the distribution of $\Psi_t$ 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 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 $\Psi_t$ 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 (2.7) 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 $S = E(\Psi_t)$. First, note that the skewness of $\Psi_t$ derives from and is bounded by the skewness in the distribution of the increments $z_t^2$. 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 $\Psi_1$ and $z_1^2$ match. However, when $\Psi_t$ is a moving average of the $\{z_t^2\}$ process, the distribution of $\Psi_1$ depends critically on the distribution of the lag coefficients. Since the lag weights have a finite sum $S$, the effects 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 $z_t$ possesses a fourth moment, the central limit theorem implies that $\Psi_1$

\footnote{Note that the axes in our figure are interchanged relative to Nelson’s figure.}
Figure 2: Simulation of IGARCH(1,1) with Student(3) shocks.

Figure 3: Simulation of IGARCH(1,1) with Gaussian shocks.

Figure 4: Simulation of IGARCH(1,1) with uniformly distributed shocks.
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 $\Psi_1$ 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 $R_t$ to the persistence properties can be largely discounted, as we argue, three factors independently contribute to the persistence and stationarity of the ARCH($\infty$) process: the size of $S$; the rate of decay of the lag coefficients; and the marginal distribution of $z_t$. Greater/smaller kurtosis of $z_t$ 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($\infty$) 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. We are accustomed to measuring this type of property in terms of 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

$$J_T = \frac{1}{T} \sum_{t=2}^{T} I((X_t - M_T)(X_{t-1} - M_T) > 0)$$  \hspace{1cm} (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, $J_T \rightarrow 1/2$ as $T \rightarrow \infty$, 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 $J_T \rightarrow 0$ (anti-persistence) and $J_T \rightarrow 1$ (persistence). In the latter case, at most a finite number of median crossings as $T \rightarrow \infty$ 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. It is possible for this statistic to converge to 1 even if the levels statistic does not, the obvious case in point being an uncorrelated returns sequence with nonstationary volatility. 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, \( \{ x_t \}_{t=1}^T \). This maps the sequence elements into the unit interval by means of the empirical distribution function

\[
\hat{F}_T(z) = T^{-1} \sum_{t=1}^T I(X_t \leq z).
\]

In other words, \( x_t \) denotes the position of \( X_t \) in the sorted sequence \( X(1), \ldots, X(T) \), divided by sample size \( T \). 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.\(^2\) A popular persistence measure based on the autocovariance sequence is the log-periodogram regression estimator of the fractional persistence parameter \( d \) known as GPH, originally due to Geweke and Porter Hudak (1983). In principle GPH provides a test of the null hypothesis of weak dependence, although it is 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 concerns here 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 \( \alpha \)-stable distribution with \( \alpha = 1.8 \) and \( \beta = 1 \), where \( \beta \) is the skewness parameter. The series so constructed do not have second moments and superficially resemble volatility series (after centring) while having a conventional and well-understood linear dependence structure.

\(^2\)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).
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 of the replications (in parentheses) are recorded in Table 1, where $d^R$ 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".

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

<table>
<thead>
<tr>
<th>$d$</th>
<th>$J_T$</th>
<th>$d$</th>
<th>$d^R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.498</td>
<td>-0.033</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.061)</td>
<td>(0.065)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.663</td>
<td>0.281</td>
<td>0.330</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.061)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.835</td>
<td>0.496</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.069)</td>
<td>(0.068)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.948</td>
<td>0.718</td>
<td>0.741</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.078)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.985</td>
<td>0.921</td>
<td>0.986</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.013)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>1</td>
<td>0.992</td>
<td>0.985</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.056)</td>
<td>(0.065)</td>
</tr>
</tbody>
</table>

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 $\xi_t = h_t^{1/2} z_t$ where $z_t \sim \text{i.i.d.}(0, 1)$ and either

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

(4.1)

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

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

(4.2)

where $\alpha > 0$ and $0 < d \leq 1$. (See e.g. Davidson (2004) for the context of these examples.) In (4.1), $S = (\delta - \beta)/(1 - \beta)$ 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 $(\delta, \beta)$ and $(\alpha, d)$. 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 $\beta$ 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
the relationship between the proximity of the mean of persistence, the GPH estimators based on the rank correlations of the squared returns. Second, note that the measured persistence of median, \( R_t \) in Section 2, the squared returns, the conditional volatilities \( z_t \) reported since they convey a very similar picture to the mean-median gaps. 

random sequences show the following: first, the sample mean, sample median, and sample logarithmic mean of the 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 (2.2); 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 relationship between the proximity of the mean of \( \Psi_t \) (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, note that the measured persistence of \( R_t \) is in general much lower than that of \( h_t \), confirming the fact that \( \Psi_t \) is the key determinant of persistence. Third, we draw attention to the relative persistences of the volatility series \( h_t \) and of the squared returns. In the latter case, for given \( \delta \) (or \( \alpha \)), and given shock distribution, the median-crossing frequencies (measured by \( 1 - J_T \)) actually rise as the lag decay rates decrease, either through \( \beta \) 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 HY-GARCH 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 \( \beta \), 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 Ståríc¾ (2003, 2004) argue that long range dependence of volatility in financial data should 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

\(^3\) 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.
<table>
<thead>
<tr>
<th>Model</th>
<th>( \delta )</th>
<th>( \beta )</th>
<th>Dist’n</th>
<th>Mean ( \Psi_t )</th>
<th>Median ( \Psi_t )</th>
<th>MeanLog ( \Psi_t )</th>
<th>( \xi_t^2 )</th>
<th>( h_t )</th>
<th>( R_t )</th>
<th>( d_t^R )</th>
<th>( \xi_t^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>0.1</td>
<td>St(3)</td>
<td>0.772</td>
<td>0.204</td>
<td>−1.630</td>
<td>(0.110) (0.064) (0.021)</td>
<td>0.571</td>
<td>0.634</td>
<td>0.468</td>
<td>−0.004</td>
<td>(0.070)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>0.777</td>
<td>0.411</td>
<td>−1.015</td>
<td>(0.011) (0.007) (0.018)</td>
<td>0.613</td>
<td>0.662</td>
<td>0.460</td>
<td>−0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>0.778</td>
<td>0.605</td>
<td>−0.746</td>
<td>(0.007) (0.011) (0.014)</td>
<td>0.639</td>
<td>0.677</td>
<td>0.469</td>
<td>0.006</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1</td>
<td>St(3)</td>
<td>0.663</td>
<td>0.265</td>
<td>−1.286</td>
<td>(0.065) (0.005) (0.023)</td>
<td>0.553</td>
<td>0.748</td>
<td>0.634</td>
<td>−0.009</td>
<td>(0.078)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>0.667</td>
<td>0.480</td>
<td>−0.788</td>
<td>(0.009) (0.009) (0.015)</td>
<td>0.576</td>
<td>0.751</td>
<td>0.615</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>0.666</td>
<td>0.594</td>
<td>−0.618</td>
<td>(0.006) (0.007) (0.011)</td>
<td>0.585</td>
<td>0.739</td>
<td>0.574</td>
<td>0.013</td>
</tr>
<tr>
<td>0.7</td>
<td>0.1</td>
<td>St(3)</td>
<td>0.329</td>
<td>0.182</td>
<td>−1.638</td>
<td>(0.038) (0.004) (0.021)</td>
<td>0.517</td>
<td>0.835</td>
<td>0.767</td>
<td>0.004</td>
<td>(0.072)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>0.333</td>
<td>0.289</td>
<td>−1.262</td>
<td>(0.004) (0.004) (0.015)</td>
<td>0.519</td>
<td>0.809</td>
<td>0.729</td>
<td>−0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>0.333</td>
<td>0.323</td>
<td>−1.176</td>
<td>(0.003) (0.003) (0.009)</td>
<td>0.521</td>
<td>0.774</td>
<td>0.675</td>
<td>0.014</td>
</tr>
<tr>
<td>1</td>
<td>0.1</td>
<td>St(3)</td>
<td>1.020</td>
<td>0.262</td>
<td>−1.379</td>
<td>(0.262) (0.006) (0.020)</td>
<td>0.588</td>
<td>0.650</td>
<td>0.474</td>
<td>0.002</td>
<td>(0.209)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>1.001</td>
<td>0.530</td>
<td>−0.761</td>
<td>(0.015) (0.011) (0.014)</td>
<td>0.647</td>
<td>0.699</td>
<td>0.477</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>1.001</td>
<td>0.778</td>
<td>−0.495</td>
<td>(0.010) (0.015) (0.014)</td>
<td>0.693</td>
<td>0.740</td>
<td>0.495</td>
<td>0.049</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>St(3)</td>
<td>1.068</td>
<td>0.442</td>
<td>−0.769</td>
<td>(0.857) (0.010) (0.019)</td>
<td>0.575</td>
<td>0.809</td>
<td>0.693</td>
<td>0.008</td>
<td>(0.060)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>0.997</td>
<td>0.768</td>
<td>−0.298</td>
<td>(0.012) (0.012) (0.016)</td>
<td>0.627</td>
<td>0.841</td>
<td>0.680</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>0.999</td>
<td>0.927</td>
<td>−0.158</td>
<td>(0.008) (0.009) (0.009)</td>
<td>0.671</td>
<td>0.866</td>
<td>0.642</td>
<td>0.189</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
<td>St(3)</td>
<td>0.975</td>
<td>0.693</td>
<td>−0.290</td>
<td>(0.098) (0.018) (0.031)</td>
<td>0.553</td>
<td>0.953</td>
<td>0.900</td>
<td>0.279</td>
<td>(0.073)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>0.998</td>
<td>0.954</td>
<td>−0.049</td>
<td>(0.015) (0.016) (0.013)</td>
<td>0.594</td>
<td>0.965</td>
<td>0.873</td>
<td>0.615</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>0.999</td>
<td>0.992</td>
<td>−0.022</td>
<td>(0.008) (0.008) (0.010)</td>
<td>0.629</td>
<td>0.971</td>
<td>0.844</td>
<td>0.729</td>
</tr>
<tr>
<td>1.2</td>
<td>0.1</td>
<td>St(3)</td>
<td>1.209</td>
<td>0.319</td>
<td>−1.176</td>
<td>(1.180) (0.007) (0.019)</td>
<td>0.607</td>
<td>0.669</td>
<td>0.480</td>
<td>0.004</td>
<td>(0.069)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>1.224</td>
<td>0.648</td>
<td>−0.565</td>
<td>(0.016) (0.011) (0.015)</td>
<td>0.685</td>
<td>0.739</td>
<td>0.495</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>1.223</td>
<td>0.951</td>
<td>−0.294</td>
<td>(0.012) (0.018) (0.013)</td>
<td>0.760</td>
<td>0.807</td>
<td>0.517</td>
<td>0.050</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>St(3)</td>
<td>1.396</td>
<td>0.619</td>
<td>−0.428</td>
<td>(0.188) (0.013) (0.024)</td>
<td>0.617</td>
<td>0.843</td>
<td>0.708</td>
<td>0.054</td>
<td>(0.068)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>1.400</td>
<td>1.080</td>
<td>0.037</td>
<td>(0.020) (0.018) (0.015)</td>
<td>0.840</td>
<td>0.952</td>
<td>0.714</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>1.399</td>
<td>1.299</td>
<td>0.178</td>
<td>(0.010) (0.013) (0.009)</td>
<td>0.994</td>
<td>0.998</td>
<td>0.779</td>
<td>0.999</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1</td>
<td>St(3)</td>
<td>3.025</td>
<td>2.900</td>
<td>0.810</td>
<td>(0.381) (0.055) (0.034)</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>0.959</td>
<td>(0.099)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>2.999</td>
<td>2.867</td>
<td>1.045</td>
<td>(0.044) (0.048) (0.016)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>U</td>
<td>2.998</td>
<td>2.976</td>
<td>1.078</td>
<td>(0.026) (0.028) (0.009)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.019</td>
</tr>
</tbody>
</table>

Table 2: Series properties and persistence measures for the GARCH(1,1) model
<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha$</th>
<th>$d$</th>
<th>Dist'n</th>
<th>$\Psi_t$</th>
<th>$\xi_t^2$</th>
<th>$h_t$</th>
<th>$R_t$</th>
<th>$\xi_t^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>MeanLog</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$St(3)$</td>
<td>0.8</td>
<td>0.9</td>
<td>$N$</td>
<td>0.805</td>
<td>0.218</td>
<td>-1.418</td>
<td>0.571</td>
<td>0.634</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>0.801</td>
<td>0.622</td>
<td>-0.678</td>
<td>0.642</td>
<td>0.669</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>0.5</td>
<td>0.9</td>
<td>$N$</td>
<td>0.797</td>
<td>0.422</td>
<td>-0.738</td>
<td>0.556</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>0.798</td>
<td>0.710</td>
<td>-0.345</td>
<td>0.585</td>
<td>0.722</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>0.1</td>
<td>0.9</td>
<td>$N$</td>
<td>0.765</td>
<td>0.611</td>
<td>-0.400</td>
<td>0.523</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>0.779</td>
<td>0.763</td>
<td>-0.266</td>
<td>0.525</td>
<td>0.753</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>1</td>
<td>0.9</td>
<td>$N$</td>
<td>0.980</td>
<td>0.271</td>
<td>-1.199</td>
<td>0.588</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>1.002</td>
<td>0.777</td>
<td>-0.453</td>
<td>0.695</td>
<td>0.741</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>0.5</td>
<td>0.9</td>
<td>$N$</td>
<td>0.993</td>
<td>0.528</td>
<td>-0.507</td>
<td>0.575</td>
<td>0.818</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>0.999</td>
<td>0.887</td>
<td>-0.122</td>
<td>0.651</td>
<td>0.838</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>0.1</td>
<td>0.9</td>
<td>$N$</td>
<td>0.995</td>
<td>0.769</td>
<td>-0.172</td>
<td>0.536</td>
<td>0.893</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>0.976</td>
<td>0.914</td>
<td>-0.057</td>
<td>0.537</td>
<td>0.856</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>1.2</td>
<td>0.9</td>
<td>$N$</td>
<td>1.135</td>
<td>0.326</td>
<td>-1.013</td>
<td>0.606</td>
<td>0.682</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>1.199</td>
<td>0.617</td>
<td>-0.496</td>
<td>0.702</td>
<td>0.780</td>
</tr>
<tr>
<td>$St(3)$</td>
<td>0.5</td>
<td>0.9</td>
<td>$N$</td>
<td>1.211</td>
<td>0.631</td>
<td>-0.322</td>
<td>0.620</td>
<td>0.871</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$U$</td>
<td>1.193</td>
<td>1.062</td>
<td>0.063</td>
<td>0.977</td>
<td>0.992</td>
</tr>
</tbody>
</table>

Table 3: Series properties and persistence measures for the HY/FIGARCH model.
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 \( \xi^2_{T+j} \) for \( j > 0 \) replaced by its (assumed) conditional expectation.
Among many references describing this technique see for example Poon (2005) page 36, 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 \( \xi^2_t \) by \( \mathbb{E}_t \xi^2_t = h_t \), and so set\(^4\)

\[
\hat{h}_{t+1|t-1} = \omega + \theta_1 h_t + \sum_{j=2}^{\infty} \theta_j \xi^2_{t-j+1}. \tag{5.1}
\]

The volatility forecast error accordingly has the form

\[
f_{t+1|t-1} = h_{t+1} - \hat{h}_{t+1|t-1} = \theta_1 (\xi^2_t - h_t) = \theta_1 h_t (\xi^2_t - 1). \tag{5.2}
\]

In the general \( k \)-step ahead case,

\[
\hat{h}_{t+k|t-1} = \omega + \sum_{j=1}^{k-1} \theta_j \hat{h}_{t-j+k|t-1} + \theta_k h_t + \sum_{j=k}^{\infty} \theta_j \xi^2_{t-j+k}. \tag{5.3}
\]

and so

\[
f_{t+k|t-1} = \sum_{j=1}^{k-1} \theta_j f_{t-j+k|t-1} + \sum_{j=1}^{k} \theta_j h_{t-j+k} (\xi^2_{t-j+k} - 1) \tag{5.4}
\]

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

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

If \( \xi^2_t \) is replaced by \( \mathbb{E}_{t-1} \xi^2_t = 1 \) to construct the forecast, (5.2) reduces to

\[
f_{t+1|t-1} = (\delta - \beta) h_t (\xi^2_t - 1).
\]

The problem with this formulation, as the preceding analysis demonstrates, is that due to the
skewness of the distribution of \( \xi^2_t \), 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 (2.7) by using \( S \) as the forward
projection for unobserved \( \Psi_t \). \( S \) is not the value that \( \Psi_t \) is close to with highest probability,
and 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 “under-
shoots”. The forecast is unbiased in the sense \( \mathbb{E}(f_{t+k|t-1}) = 0 \) when this expectation is defined,
but this condition excludes the IGARCH and FIGARCH and other nonstationary cases. Even

\(^4\)We call this expression the two-step volatility forecast since \( h_t \) itself is of course the one-step forecast.
of central tendency of the shocks, denoted by $\hat{z}_t$ in the two-step forecasts constructed under different assumptions about the appropriate measure of central tendency of the shocks, denoted by $\hat{z}_t$.

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 $\hat{z}_t$ appropriate loss function.

if the mean squared forecast error is defined, in this context, it is not clear that the MSE is an appropriate loss function.

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. Denoting these by $\hat{z}_t$, this corresponds to computing the $k$-step volatility forecasts by the recursion

$$
\hat{h}_{t+k|t-1} = \omega + \frac{2}{3}\text{Median}(\hat{z}_t^2) \sum_{j=1}^{k} \theta_j \hat{h}_{t-j+k|t-1} + \sum_{j=k}^{\infty} \theta_j \hat{z}_t^2.
$$

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

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

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 is that future volatility is significantly overstated by conventional procedures.
Table 5: MAV 2-step forecast error in HY/FIGARCH, against $M$ (see (5.5))

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. The message for practitioners is that conventional forecasting methodologies, which are optimal under the assumption of symmetrically distributed shocks, may tend to overstate the degree of future volatility. This is, of course, an issue essentially of the preferred choice of loss function. Practitioners may elect to favour the unbiasedness and minimum MSE 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.

### References


