

# **Distinguishing short and long memory volatility specifications**

Shiuyan Pong, Mark B. Shackleton\* and Stephen J. Taylor

Department of Accounting and Finance,  
Lancaster University, LA1 4YX, U.K.

\* +44 1524 594131, [m.shackleton@lancaster.ac.uk](mailto:m.shackleton@lancaster.ac.uk)

Forthcoming in the Econometrics Journal

March 2008

# **Distinguishing short and long memory volatility specifications**

March 2008

## **Abstract**

Asset price volatility appears to be more persistent than can be captured by individual, short memory, autoregressive or moving average components. Fractional integration offers a very parsimonious and tempting formulation of this long memory property of volatility but other explanations such as structural models (aggregates of several autoregressive components) are possible. Given the ability of the latter to mimic the former, we investigate the extent to which it is possible to distinguish short from long memory volatility specifications. For a likelihood ratio test in the spectral domain, we investigate size and power characteristics by Monte Carlo simulation. Finally applying the same test to Sterling/Dollar returns, we draw conclusions about the minimum number of structural factors that must be present to mimic the long memory volatility properties that are empirically observed.

Keywords: long memory, volatility, spectral test, power, size.

# 1 Introduction

The persistent nature of asset price volatility and power transforms of asset returns has been well documented in numerous studies, commencing with Taylor (1986). Ding, Granger and Engle (1993) observed hyperbolic decay in the autocorrelations of powers of daily absolute returns obtained from U.S. stock indices, while Andersen et al (2001a, 2001b) found the same phenomenon in realized volatilities. Fractionally integrated, long memory models have thus received considerable interest because of their ability to capture the slowly decaying autocorrelations of volatility. Long memory models are characterized by a hyperbolic decay rate in autocorrelation, which is consistent with findings from empirical data.

Baillie, Bollerslev and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996) introduced long memory processes in the context of conditional variance by extending the GARCH models of Bollerslev (1986) and the exponential ARCH models of Nelson (1991). Breidt, Crato and de Lima (1998) proposed a long memory stochastic volatility model by incorporating a fractionally integrated process in a standard volatility scheme. Andersen et al (2001a, 2001b) suggest that this ARFIMA model is well suited for realized volatility that is constructed from high-frequency intraday returns.

Although the evidence for long memory effects in volatility is at first sight compelling, there is an alternative explanation for these effects. Gallant, Hsu and Tauchen (1999) show that the sum of two (short memory) AR(1) processes will appear to have long memory features when the parameters are selected appropriately. A similar observation is made by Barndorff-Nielsen and Shephard (2001). Alizadeh, Brandt and Diebold (2002) show that the sum of two AR(1) processes describes FX volatility better than one and Pong, Shackleton, Taylor and Xu (2004), find that the sum of two AR(1) processes performs as well as a fractionally integrated process for forecasting the realization of exchange rate volatility up to three months ahead.

Given the interest in volatility persistence, it would seem important to resolve the properties of its memory. This issue is particularly important for option traders when they value options that expire several months into the future. Taylor (2005, page 394) compares the implied volatility term structures for short and long memory ARCH models, and shows that the valuation of S & P 100 options that expire after one year can be very different for short and long memory assumptions. Ohanissian, Russell and Tsay (2007) obtain the same conclusion for S & P 500 options when volatility is modeled by a diffusion process. Thus popular pricing models, based upon the affine jump-diffusion framework of Duffie, Pan and Singleton (2000), may provide inappropriate theoretical prices.

Granger (1980) showed that under certain conditions a linear combination of an infinite set of AR(1) processes is a fractionally integrated process, which may therefore explain why the sum of two AR(1) processes can mimic long memory if the parameters are chosen judiciously; however, mimicking behavior will break down as the sample length increases. Given the close relation between the sum of two AR(1) processes and a fractionally integrated process, it is natural to ask a fundamental question: given a certain dataset, can these two alternative data generating processes be distinguished statistically? Likewise, for a given source of data, how much more is needed to distinguish between a true long memory process and one that merely mimics long memory within that data length?

To answer these questions, in this paper we design and investigate a statistical test operating in the frequency domain based on likelihood ratios. Through a Monte Carlo study, we make use of our proposed statistical test to find out the probability of identifying the correct process given that the data was generated either by a fractionally integrated process or by a sum of two AR(1) processes designed to mimic long memory.

By generating simulated series of different lengths we examine the effect of sample size on the possibility of distinguishing these two processes. We find that the

probability of identifying the correct process can be relatively low, even for series of 2,000 observations. It therefore appears that traders of long-lived options will often have to cope with uncertainty about the memory characteristics of volatility when they price options. We also document recommended series lengths for differentiating between a fractionally integrated process and a sum of AR(1) processes.

Another test procedure has recently been developed by Ohanissian, Russell and Tsay (2007). They test the null hypothesis that a process has a long memory by comparing estimates of the long memory parameter  $d$  across data frequencies.

Section 2 describes the definitions and features of long memory and short memory processes. Long memory and short memory processes (fractionally integrated and aggregated factor processes respectively) are introduced and it is shown how the spectral densities of these two models can be very similar when the parameters are chosen appropriately.

Section 3 describes our proposed statistical test to distinguish long memory from short memory and also the Monte Carlo design. We then present our results based on simulated series of four different lengths in Section 4. The findings clearly highlight the effect of series length on the probability of identifying the correct type of memory.

Section 5 provides estimation results for Sterling/Dollar (GBP/USD) exchange rate volatility. The Monte Carlo simulation evidence is then used to interpret the empirical test results. With less than 3,000 observations, it cannot be concluded if long memory or short memory characterizes the realized volatility series. However, with the number of observations multiplied by six, we show that the mimicking process can be rejected. Further conclusions are provided in Section 6.

## 2. Long, short and mimicking memory processes

There are several definitions that categorize stochastic processes as having either a long memory or a short memory. Relevant discussions are provided by McLeod and Hipel (1978), Brockwell and Davis (1991), Baillie (1996) and Granger and Ding (1996). Consider a discrete-time stochastic process  $\{y_t\}$ , with autocorrelation function  $\rho_\tau$  and spectral density  $f(\omega)$ . A covariance stationary process possesses a short memory if

$\sum_{\tau=1}^n \rho_\tau$  converges as  $n$  tends to infinity, otherwise it is said to have a long memory.

The autocorrelation structure  $\rho_\tau$  of a short memory process is geometrically bounded, while its spectral density  $f(\omega)$  is bounded for all frequencies. In contrast, the autocorrelations of a long memory process have a hyperbolic decay and a spectral density that is unbounded at low frequencies. Representing the degree of fractional integration by  $d$  (for a value less than 0.5) and dividing the autocorrelations  $\rho_\tau$  and spectral density  $f(\omega)$  by  $\tau^{2d-1}$  and  $\omega^{-2d}$  respectively, the resulting values will converge to positive constants when  $\tau \rightarrow \infty$  and  $\omega \rightarrow 0$ . Table 1 summarizes the limiting properties of the autocorrelations and the spectral density for both short and long memory processes.

### 2.1 Fractionally integrated models

Granger and Joyeux (1980) and Hosking (1981) introduced a flexible class of long memory processes, called autoregressive, fractionally integrated, moving average or ARFIMA models. For an excellent survey of long memory processes, including applications to financial data, see Baillie (1996). The ARFIMA( $p, d, q$ ) model for a process  $\{y_t\}$  is given, using the lag operator  $L$ , by:

$$\phi(L)(1-L)^d(y_t - \mu) = \theta(L)\varepsilon_t \quad (1)$$

where  $d$  again represents the order of fractional integration,  $\phi(L) = 1 - \phi_1 L + \dots - \phi_p L^p$  represents the lag polynomial for the autoregressive component,  $\theta(L) = 1 - \theta_1 L + \dots - \theta_q L^q$  represents the lag polynomial for the moving average component and  $\mu$  is the expectation of  $y_t$ . The respective degrees of the polynomials  $\phi(L)$  and  $\theta(L)$  are  $p$  and  $q$ . The roots of  $\phi(L)$  and  $\theta(L)$  lie outside the unit circle and  $\{\varepsilon_t\}$  is a zero-mean white noise process with variance  $\xi^2$ . When  $0 < d < 1$ , the process has a long memory property and it is covariance stationary when  $d < 0.5$ .

The autocorrelation function of the ARFIMA(0, $d$ ,0) process is written as:

$$\rho_\tau = \frac{\Gamma(1-d)\Gamma(\tau+d)}{\Gamma(d)\Gamma(\tau+1-d)} \quad (2)$$

where  $\Gamma(\cdot)$  represents the gamma function. The computation of the autocorrelations of the general ARFIMA( $p,d,q$ ) model is more complicated and detailed discussions about the calculations have been provided by Sowell (1992) and Chung (1994).

Using complex notation, the spectral density function of an ARFIMA( $p,d,q$ ) model is given (see Baillie, 1996) by:

$$f_{arfima}(\omega) = \frac{\xi^2}{2\pi} |\theta(e^{-i\omega})|^2 |\phi(e^{-i\omega})|^{-2} |1 - e^{-i\omega}|^{-2d}. \quad (3)$$

## 2.2 Structural AR-factor models

Granger and Newbold (1977) showed that when  $p$  distinct and independent AR(1) processes are aggregated, an ARMA( $p, p-1$ ) model is obtained which is categorized as a short memory structural model; the cumulative action of many factors on a system then determines its structure. Applications of structural factor models in finance can be found in Gallant, Hsu and Tauchen (1999), Alizadeh, Brandt and Diebold (2002) and Pong et al (2004), who all discuss and use the sum of two AR(1) processes, having an ARMA(2,1) characterization, to capture the dynamics of asset price volatility.

The ARMA( $p, p-1$ ) model for a process  $\{y_t\}$  is given by a restricted version of the more general ARFIMA model:

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t \quad (4)$$

with the respective degrees of the polynomials  $\phi(L)$  and  $\theta(L)$  equal to  $p$  and  $p-1$ . For the special case when  $p$  is 2, an ARMA(2,1) process is the sum of two AR(1) processes if and only if the roots of the quadratic equation  $z^2\phi(z^{-1}) = 0$  are both real; when these roots are real they equal the autoregressive parameters of the component AR(1) processes.

It is straightforward in theory, but algebraically time consuming, to derive the autocorrelation function of an ARMA( $p,q$ ) process. However, if the process is an ARMA( $p, p-1$ ) specification that can be rewritten as the sum of  $p$  independent AR(1) processes, then the calculations are made simpler. As the autocovariance function of the sum of independent processes is the sum of the autocovariance functions of the individual processes, the autocorrelations of the ARMA( $p, p-1$ ) model can be calculated as:

$$\rho_\tau = \frac{\sum_{k=1}^p \nu_k^2 \psi_k^\tau}{\sum_{i=k}^p \nu_k^2} \quad (5)$$

where  $\psi_k$  and  $\nu_k^2$  are respectively the autoregressive and variance parameters of the  $k^{th}$  AR(1) process. Furthermore, the spectral density function can be represented as the sum of the spectral densities of  $p$  AR(1) processes:

$$f_{sum}(\omega) = \frac{1}{2\pi} \sum_{k=1}^p \nu_k^2 (1 - \psi_k^2) \left| 1 - \psi_k e^{-i\omega} \right|^{-2}. \quad (6)$$



### 2.3 Comparison of ARFIMA and ARMA spectral densities

Gallant, Hsu and Tauchen (1999) point out that the sum of two AR(1) processes can mimic the appearance of long memory in volatility when an appropriate choice of parameters is made. For the commonly found value of  $d = 0.4$ , they used a least squares fit to find the two-factor process that best fits the spectral density of a fractionally integrated process. Specifically, an approximation to ARFIMA(0, $d$ ,0) can be made by an ARMA(2,1) obtained via the following spectral minimization:

$$\min_{\psi_1, \psi_2, \nu_1, \nu_2} \sum_{j=1}^{300} \left\{ \xi^2 |1 - e^{-i\omega_j}|^{-2d} - \sum_{k=1}^2 \nu_k^2 (1 - \psi_k^2) \left| 1 - \psi_k e^{-i\omega_j} \right|^{-2} \right\}^2 \quad (7)$$

where  $\omega_j = j \frac{\pi}{300}$ .

With  $d$  chosen as 0.4 and  $\xi$  as 0.2,  $(\psi_1, \nu_1, \psi_2, \nu_2) = (0.978, 0.136, 0.447, 0.216)$  minimizes the sum in equation (7). The values of the autoregressive parameters indicate that the first AR(1) process is highly persistent ( $\psi_1 = 0.978$ ) while the second has a transient nature ( $\psi_2 = 0.447$ ). The two variances are similar. Figure 1 shows the theoretical spectral densities of the ARFIMA(0, $d$ ,0) and ARMA(2,1) models. It can be seen that the spectral density of the long memory model is closely approximated by that of the short memory model over much of the frequency domain. The match is strikingly good and it may not be possible to distinguish the spectral densities with the naked eye.

For the realized volatility of USD/GBP, USD/DEM and USD/JPY exchange rates, spanning seven years, Pong et al (2004) use maximum likelihood estimation in the spectral domain to estimate ARFIMA and ARMA(2,1) models. They find that the likelihood function values of both models are very close, for each of the three exchange rates, and neither of the models possesses forecasting abilities that are statistically superior to the other. Their findings highlight the proximity of the fractionally integrated process and the two-factor AR process in terms of their persistence.

These results lead to two fundamental and linked questions. First, can two-factor and fractionally integrated processes be distinguished statistically? Second, what length of a financial time series is necessary to have a “good” chance of identifying the correct type of model? A Monte Carlo analysis can provide some answers to these questions. By generating simulated series from a specific process repeatedly, we can calculate the percentage probability of recovering the same process based on the Monte Carlo evidence. In addition, we can investigate this probability as a function of the length of the financial time series to determine the ability of a particular statistical test to distinguish long and short memory series.

### **3. Monte Carlo simulation**

A number of simulation studies have addressed the performances of different tests designed to distinguish long memory and short memory processes. Breidt, Crato and de Lima (1998) investigate the abilities of a spectral regression test, proposed by Geweke and Porter-Hudak (1983) and further developed by Robinson (1994), as well as the so-called R/S statistic test (see Beran, 1994) used to detect long memory in stock return volatility. They find that the tests are able to distinguish between ARFIMA(0, $d$ ,0) and ARFIMA(1,0,0) specifications incorporated in a stochastic volatility model. Smith, Taylor and Yadav (1997) use semi-parametric and maximum likelihood estimation methods to investigate the probability of selecting the correct specification given a simulated fractionally integrated process. They find that it is not always possible to identify the existence of long memory statistically for short simulated series with 256 observations.

Our research is similar in spirit but is different in four aspects. Firstly, we place emphasis specifically on two types of process in our simulation analysis, the fractionally integrated process and the two-factor process, because of their significance and popularity in volatility modeling and in particular the ability of the latter to mimic the former. Secondly, the typical time span of financial data employed in recent volatility modeling

research is in the neighborhood of ten years (see Gallant, Hsu and Tauchen (1999), Andersen et al, (2001a, 2001b) and Areal and Taylor (2002)). Therefore, the length of simulated series in our Monte Carlo study is chosen to match these features. Thirdly, previous studies concern the ability of statistical tests to identify a long memory feature when long memory actually exists in the data. Our paper provides a new dimension by looking at the possibility of mistaking a mimicking two-factor process for a long memory process. Finally, we investigate how the length of the time series affects the probability of identifying the correct type of process.

### 3.1 Simulation design

We design a statistical test to detect the correct type of memory in a Gaussian process which is composed of a pair of likelihood ratio tests. The Gaussian assumption is appropriate for the process of primary interest to us, namely the logarithm of realized volatility, because the empirical evidence shows that the unconditional distribution of log volatility is almost normal; this evidence commences with Andersen et al (2001a, 2001b) and has been confirmed by several subsequent studies, such as Areal and Taylor (2002) and Pong et al (2004).

We first consider a series simulated by an ARFIMA(0, $d$ ,0) model. The objective of the test is to choose between ARFIMA(0, $d$ ,0) and ARMA(2,1) models. Since these models are not nested, they cannot be compared directly. However, they can be tested indirectly by using the encompassing ARFIMA(2, $d$ ,1) specification as a bridge. A likelihood ratio test can be used to choose between ARFIMA(0, $d$ ,0) and ARFIMA(2, $d$ ,1) by testing the null hypothesis that  $(\phi_1, \phi_2, \theta_1) = (0, 0, 0)$ . The test should (asymptotically) nearly always accept the null hypothesis if the true process is ARFIMA(0, $d$ ,0) and the significance level is small. We can simultaneously use another likelihood ratio test to choose between the ARMA(2,1) and ARFIMA(2, $d$ ,1) models by testing the null hypothesis that  $d = 0$ , when the null hypothesis should be rejected (asymptotically). There are four combinations of results from the pair of LR tests. Firstly, the combined LR test can accept ARFIMA(0, $d$ ,0)

and reject ARMA(2,1). Secondly, it can accept ARMA(2,1) and reject ARFIMA(0, $d$ ,0). Thirdly, it can accept both the ARFIMA(0, $d$ ,0) and ARMA(2,1) models. Finally, it can reject both models and accept ARFIMA(2, $d$ ,1). The left panel of Figure 2 provides a graphic presentation of the test design.

All our estimates of ARMA(2,1) models do not restrict the AR and MA parameters, so we do not explicitly assume that these models are the sum of two AR(1) processes. However, we always find that these estimated models do correspond to AR(1) sums, i.e. the quadratic equation  $z^2\phi(z^{-1}) = 0$  always has two real roots.

By replicating the experiment and documenting the percentage of times each of these four conclusions is obtained we find the probability of identifying the existence of long memory using the combined LR test. Asymptotically the tests should nearly always accept ARFIMA(0, $d$ ,0) and reject ARMA(2,1); however, bearing in mind their proximity it is expected that the test results will be less than perfect given a series of finite length. Our Monte Carlo results, which are shown in Section 4, do indeed show that long and short memories are not always distinguishable.

On the other hand, if the series is truly generated by an ARMA(2,1) model for our simulation experiment, the percentage of detections of short memory can be obtained using the same procedure, except that one modification has been made. One leg of the LR tests is the same as before, in which the ARMA(2,1) and ARFIMA(2, $d$ ,1) models are compared, while the design of the second leg is changed by comparing ARFIMA(1, $d$ ,1) instead of ARFIMA(0, $d$ ,0) with ARFIMA(2, $d$ ,1) in order to keep the same degrees of freedom for the two LR tests. The test design is shown in the right panel of Figure 2.

To carry out the LR test, we perform the maximum likelihood estimation in the frequency domain. This estimation methodology is computationally efficient, particularly for long time series, because the Fast Fourier Transform is applicable. Whittle (1951) provides the expression for the log-likelihood function of a process in the frequency

domain, under the assumption of normality of the residuals. The log-likelihood function for  $N$  observations is

$$\log L = -\frac{1}{2} \sum_{j=1}^{N-1} \log[2\pi f(\omega_j)] - \frac{1}{2} \sum_{j=1}^{N-1} [\hat{f}(\omega_j)/f(\omega_j)]. \quad (8)$$

The theoretical spectral density,  $f(\omega_j)$ , at frequency  $\omega_j = 2\pi j/N$ , is always computed from equation (3), while  $\hat{f}(\omega_j)$  is the value of the periodogram at frequency  $\omega_j$ .

Discussion of maximum likelihood estimation in the frequency domain has been provided by Fox and Taqqu (1986), Cheung and Diebold (1994) and Baillie (1996) in the context of long memory. The maximum likelihood estimation can also be performed by full maximum likelihood estimation (MLE) in the time domain; however, it is computationally intensive especially for long time series. Cheung and Diebold (1994) find that spectral-likelihood estimation is preferable to full MLE when the mean of the process is not required. Breidt, Crato and de Lima (1998) show that strong consistency can be obtained by MLE in the frequency domain. These findings explain our choice of MLE in the frequency domain for our simulation exercise.

Davies and Harte (1987) suggest an algorithm for simulating ARFIMA models which makes use of a two-step Fourier transform to simulate a series which matches the autocovariance structure of the fractionally-integrated process. Beran (1994) provides a review of standard approaches for simulating ARFIMA models, including the Cholesky decomposition of the covariance matrix of the time series and the Davies and Harte algorithm. Our Monte Carlo analysis relies on the Davies and Harte algorithm to simulate the ARFIMA(0, $d$ ,0) process. The simulation of the ARMA(2,1) process is relatively simple, which can be done by summing two independent first-order autoregressive Gaussian processes.

We set the minimum length of simulated series as 2,000. Assuming 250 trading days in a year, such a length is equivalent to a time span of eight years for a daily sampling frequency and it represents the typical length of financial time series employed in volatility modeling. In order to investigate the effect of series length on identification of the true processes, we also include simulated series with lengths of 4,000, 8,000 and 12,000 in our study. While it is predicted that the power of the test increases when the simulated series is lengthened, we aim to identify a length which gives us a sufficient level of confidence to identify the correct type of memory in a process. Finally, our analysis is based on 1,000 replications of each combined LR test.

## 4. Results

The upper part of Table 2 shows the results of parameter estimation when the simulated series is generated from an ARFIMA(0, $d$ ,0) model with  $d = 0.4$  and  $\xi = 0.2$ . This value of  $d$  represents the typical degree of fractional integration found in recent volatility studies (e.g. Andersen et al (2001a, 2001b), Areal and Taylor (2002)). The first two rows of Table 2 show the mean value of the estimated  $\hat{d}$  for different simulated lengths when we estimate the correct specification. The values of  $\hat{d}$  are estimated using the maximum likelihood estimation method as well as the semi-parametric method proposed by Geweke and Porter-Hudak (1983). It is found that the mean estimates are close to the true value of  $d$ . However, what deserves more attention are the results found in the next four rows which show the mean estimates of the parameters when we mistakenly suppose the two-factor model is the correct model. The values of the estimates indicate that the typical mistaken process is composed of a highly persistent process and a transient process.

The lower part of Table 2 shows the estimation results when the simulated series is generated by the two-factor model with  $(\psi_1, \nu_1, \psi_2, \nu_2) = (0.978, 0.136, 0.447, 0.216)$ . These input values are obtained using the least squares approximation stated in equation

(7), to match a fractionally integrated process with  $d = 0.4$ . The values of the estimates for the two-factor model are found to be close to the input values as shown in the table. The figures in the next two rows show the average value of  $\hat{d}$  when we mistake the correct specification as ARFIMA(0, $d$ ,0). It is found that the average values of  $\hat{d}$  are significantly greater than zero and are in the neighborhood of 0.3.

The above results carry important implications. They suggest that it is often easy to claim a series has a long memory by performing some of the popular statistical tests for detecting long memory but without considering the short memory alternative specifications (and also vice versa).

#### 4.1 Identifying long memory

Using the Monte Carlo design suggested in Section 3.1, we investigate the probability of identifying long memory given a simulated series generated from an ARFIMA(0, $d$ ,0) model. We define  $\lambda_1$  as the likelihood ratio test statistic obtained from the LR test that compares ARFIMA(0, $d$ ,0) with ARFIMA(2, $d$ ,1) and  $\lambda_2$  as that obtained from comparing ARMA(2,1) with ARFIMA(2, $d$ ,1). Figures 3a-d show plots of  $\lambda_2$  against  $\lambda_1$ , with both variables plotted on a logarithmic scale. The dotted lines on the figures represent the cut-off boundaries of both tests at the 5% significance level. For this significance level, a test statistic greater than 7.81 leads to a rejection of the ARFIMA(0, $d$ ,0) model in the first LR test and a ratio greater than 3.84 leads to a rejection of the ARMA(2,1) model in the second LR test.

We first look at the results when the simulated series consists of 2,000 observations, as shown in Figure 3a. It can be easily noticed that there are many points located in the region for accepting the ARMA(2,1) model. This indicates that there is a significant probability of failing to reject the wrong specification. Figures 3b-d show the plots when the lengths of simulated series are 4,000, 8,000 and 12,000 respectively. The values of  $\lambda_2$  generally shift upwards as the simulated length increases and as a result fewer points

are found within the acceptance region of the second leg of the test. It shows that, as expected, the ability of the test to detect long memory improves as the simulated length increases.

The chance of obtaining each of the four different conclusions is also displayed on Figure 3. The numbers in the upper-left quadrants represent the fraction of simulations for which the correct specification is obtained, i.e. the ARFIMA(0, $d$ ,0) model is accepted while the ARMA(2,1) model is rejected. This value equals 72% when the simulated length is 2,000. The probability of identifying the ARFIMA(0, $d$ ,0) model increases substantially to 91% when the length of the simulated series is extended to 4,000; the probability approaches 95% (the asymptotic value of the test) when the length reaches 8,000.

In addition to having a snapshot of the results at the 5% significance level, it is interesting to see how the performance of the LR tests changes with the significance level. The plot of the probability of rejecting the ARMA(2,1) model against the probability of rejecting the ARFIMA(0, $d$ ,0) model for different length series is given by Figure 4a for different lengths of the simulated series. The former and latter probabilities are labeled and can be regarded as the power and the size of the test respectively. As expected, the power increases with the size of the test. However, the convex shapes of the curves show that the rate of change in power decreases when the size becomes greater. Given any size, the longer the simulated series the higher is the power. When the length reaches 8,000, the power is close to 100% even for a very small test size.

## 4.2 Identifying short memory

We now investigate the ability of the combined LR test to identify memory correctly given that the underlying series is generated from an ARMA(2,1) model. The first LR test compares the ARFIMA(1, $d$ ,1) and the ARFIMA(2, $d$ ,1) models and the second compares the ARMA(2,1) and the ARFIMA(2, $d$ ,1) models. The parameter values of the ARMA(2,1) model are chosen as shown in Section 2.3.



The scatter diagrams of the likelihood-ratio test statistics  $\lambda_1$  and  $\lambda_2$  can be found in Figures 5a-d, where a similar pattern to that of Figure 3 is seen. The boundaries of the acceptance region for both tests (shown as dotted lines) are 3.84 for a 5% significance level. Several points can be found in the acceptance region of the first test when the length is 2,000 and it implies a significant probability of failing to reject the wrong specification. As the length of the simulated series increases, the whole pattern shifts towards the right and it indicates that it is easier to identify the true ARMA(2,1) process.

The probabilities of obtaining the various conclusions for different simulated lengths at the 5% significance level are also shown on the same figure. Given that the simulated series are generated from an ARMA(2,1) model we find that the probability of failing to reject the ARFIMA(1, $d$ ,1) specification is only substantial for a series length of 2,000. On 43% of occasions the long memory hypothesis is not rejected. This probability can be further broken down into two components: for 38% of the series the ARFIMA(1, $d$ ,1) and ARMA(2,1) model cannot be differentiated and 5% of the time accepting ARFIMA(1, $d$ ,1) is the sole conclusion. The probability of choosing the correct ARMA(2,1) specification is 55%, which can be found in the lower-right quadrant. The performance of the test improves as the length of the series increases. The probability of identifying short memory rises to 82% when the length is 4,000 and the probability of choosing the correct specification is close to 95% when the length of the simulated series is either 8,000 or 12,000.

Figure 4b shows the size-power curve given the simulated series is from an ARMA(2,1) process. The pattern of the graph is similar to that of Figure 4a. In particular, it is again found that the power of the test increases as the length of the simulated series becomes longer given any test size. In addition, the rate of change of power is higher when the test size is smaller. Finally, the power is close to 100% even for a small size when the length of simulation reaches 8,000.

### 4.3 Sensitivity analysis

Sensitivity analysis is now performed with regards to changes in the parameter values of ARFIMA and ARMA models. We investigate the effect of a reduction in the persistence level of a process on the probability of identifying the correct specification. In particular, to simulate the long memory series, we suppose  $d$  is reduced from 0.4 to one of 0.3, 0.2 and 0.1. To generate short memory series, we choose the three following parameter sets:  $(\psi_1, \nu_1, \psi_2, \nu_2) = (0.967, 0.095, 0.304, 0.204), (0.942, 0.069, 0.165, 0.196), (0.894, 0.048, 0.063, 0.196)$ , which correspond to solving equation (7) while using  $d = 0.3, 0.2$  and 0.1 and  $\xi = 0.2$  as inputs.

Tables 3A and 3B display the results of the sensitivity analysis when the series are generated by ARFIMA(0, $d$ ,0) and ARMA(2,1) models respectively and all tests have a 5% significance level. We find that the performance of the combined LR test generally improves as the simulated series is lengthened for all persistence levels. However, the probability of identifying the correct type of memory deteriorates as the persistence level decreases for both cases. This finding is expected since a long memory process behaves more like its short memory counterpart as the degree of fractional integration diminishes. The probability of identifying the correct type of memory is found to be relatively low for small values of  $d$  or its matching AR parameters even for a very long time series. For example, we can see from Table 3A that when the length is 12,000 the probability of correctly identifying the ARFIMA(0, $d$ ,0) specification is 64% when  $d = 0.1$  and such a probability is much less than the 95% obtained when  $d = 0.4$ . Similar findings are obtained when the underlying process is ARMA(2,1), which are shown in Table 3B. They indicate that when the level of persistence is low, a very large number of observations is required in order to have acceptable power to identify the correct specification.

#### **4.4. Implications for volatility modeling**

Our simulation study illustrates that it is possible to mistake a fractionally integrated series for a two-factor process, irrespective of the amount of data used (and vice versa). This result carries significant implications for volatility modeling. Claiming that a volatility series has a fractionally integrated structure while not considering the factor specification represents a bold assertion. It is therefore wise to consider both specifications before any conclusion on the type of memory present in the volatility series is made. Our proposed test procedures serve this purpose and they identify the probability of incorrect conclusions.

With limited data, our study shows that it is not always possible to distinguish a long memory process and a short memory process statistically. In volatility modeling and forecasting research, the typical time span of financial data employed is in the neighborhood of ten years, which is equivalent to 2,500 daily observations. Our results illustrate that based on this kind of history it is difficult to tell whether the long or short memory model describes the volatility process better. However, the encouraging news is that the ability to distinguish long memory and short memory processes improves if a more substantial data history is investigated.

### **5. Empirical evidence**

In this section, we show how the Monte Carlo results can be applied in an empirical study to distinguish between long and short memory in a volatility process. The data used in our empirical study are GBP/USD spot exchange rates provided by Olsen and Associates. The raw data is in the form of five-minute exchange rate differences computed using mid-quote exchange rates. The sample period is from July 2, 1987 to December 31, 1998. Although the foreign exchange market operates twenty-four hours a day, seven days a week, trading volume and volatility of the market diminish

significantly during weekends and holidays, as depicted for example by Bollerslev and Domowitz (1993). To eliminate the weekend effect, we exclude the period from 21:00 GMT Friday to 21:00 GMT Sunday from our study. We also exclude several holidays: bank holidays in May and August for the U.K. and Memorial Day, July Fourth, Labour Day and Thanksgiving and the day after for the U.S. In addition, the returns during Christmas (24/12 to 26/12), Good Friday, Easter Monday and New Year (31/12 to 01/01) are also removed. We are left with a volatility series spanning 2,867 days.

Our analysis is carried out based on the realized volatility series of the GBP/USD exchange rate. Realized volatility is constructed by summing the squared intraday returns sampled at a particular frequency. Andersen et al (2001a, 2001b) have shown that when certain regularity conditions apply, as sampling becomes more frequent, the realized volatility is an increasingly accurate measure of integrated return volatility. The realized variance for day  $t$  is defined by the following equation:

$$\sigma_t^2 = \sum_{j=1}^n r_{t,j}^2 \quad (9)$$

where  $r_{t,j}$  is the return in interval  $j$  on day  $t$  and  $n$  represents the total number of intervals in a day. To construct the realized volatility series we adopt a five minute sampling frequency in order to lessen the effects of market microstructure. As a result, 288 five-minute returns are used to construct one daily observation of realized volatility. Following Andersen et al (2001a, 2001b), we model  $\log(\sigma_t^2)$  because its distribution is closer to Gaussian.

The upper panel of Table 4A shows the results of the parameter estimation from MLE in the frequency domain. The value of  $\hat{d}$  in the ARFIMA(1, $d$ ,1) model is 0.348 while  $\hat{\phi}_1$  and  $\hat{\theta}_1$  are 0.943 and 0.899 respectively. The estimated values of the parameters of the ARMA(2,1) model are  $(\hat{\phi}_1, \hat{\phi}_2, \hat{\theta}_1) = (1.201, -0.221, 0.797)$ . The autoregressive parameters in the two AR(1) components that aggregate to the

ARMA(2,1) model are equal to 0.973 and 0.228. These results are consistent with the estimates in Pong et al (2004), who use a shorter time span for the same exchange rate.

Not knowing the true volatility process, we now use LR tests to choose the most suitable model for the log realized volatility. Two test values and the corresponding significance levels are shown in the lower panel of Table 4A. We first consider the LR test statistic for comparing the ARFIMA(1, $d$ ,1) and ARFIMA(2, $d$ ,1) models. The results show that the null hypothesis  $\phi_1 = 0$  cannot be rejected at the 5% significance level, as the p-value of this test is 23%; consequently this test can accept ARFIMA(1, $d$ ,1) as the correct specification. We then compare the log-likelihoods of the ARMA(2,1) and ARFIMA(2, $d$ ,1) specifications. The null hypothesis  $d = 0$  cannot be rejected at a 5% significance level, the p-value being 11%. The combined result of the two tests indicates that we cannot statistically conclude whether the ARFIMA(1, $d$ ,1) or the ARMA(2,1) model is the better model for the log volatility when the length of our volatility series is “only” 2,867 daily observations.

Our Monte Carlo evidence indicates that a longer series is required to improve the statistical performance of the combined LR test. There are two ways to increase the amount of data for our estimation. The first method is to lengthen the sample period so that more daily observations of realized volatilities are obtained. However this method depends on data availability. The second method is to shorten the sampling interval so that more observations are obtained given a fixed time span. We will adopt the latter approach for our empirical analysis and estimate four-hour realized volatilities by summing the squares of 48, five-minute returns. As a result, 17,202 observations are produced, which is six times the number of observations obtained on a daily basis.

We may suppose that four-hour realized volatilities (RVs) are the product of latent volatility, a cyclical term and a residual factor that averages unity. The logarithms of the 4-hour RVs will have the same memory characteristics (long or short) as the log

of the latent 4-hour volatility process. Furthermore, as the 24-hour RV is the sum of six 4-hour RV terms we expect the latent 4-hour and the latent 24-hour series to have identical memory characteristics (Andersen and Bollerslev (1997)): either both have a short memory or both have a long memory with the same value of  $d$ .

Figure 6 shows the logarithm of the periodogram of the four-hour realized volatility series from January 1 1993 to December 31 1994. Periodicity can be found at the daily frequency (and its harmonics) which can be attributed to intraday volatility patterns that reflect the opening and closing of various financial centers in the world, their attendant macroeconomic news schedules, etc. The shape at the left end of the spectrum (near  $\omega = 0$ ) indicates a high volatility persistence. In order not to bias the estimation of the long memory spectrum when performing maximum likelihood estimation, we need to get rid of the seasonal component in the volatility series. We apply a relatively simple but efficient frequency-domain method that removes the seasonal effects, as recommended in Shackleton (1998) which is similar to the method applied in Deo, Hurvich and Lu (2006). By omitting observations at the daily frequency and its harmonics (and also the ten adjacent ordinates) on the periodogram, we obtain a deseasonalised periodogram. The omission of the ordinates is not expected to distort the pure stochastic components of the volatility process, since only a small fraction of ordinates in the periodogram are deleted (the percentage deleted is less than 1%).

Table 4B shows the results of maximum likelihood estimation for the ARFIMA(1, $d$ ,1), ARMA(2,1) and ARFIMA(2, $d$ ,1) models using all 17,202 observations. When modeled as ARFIMA(1, $d$ ,1), the value of  $d$  equals 0.470 and  $\phi_1$  and  $\theta_1$  equal 0.507 and 0.719 respectively. When modeled as ARMA(2,1),  $\phi_1$ ,  $\phi_2$  and  $\theta_1$  equal 1.195,  $-0.204$  and 0.922 respectively. The autoregressive parameters in the two AR(1) components that aggregate to the ARMA(2,1) model are equal to 0.990 and 0.206 respectively. Finally, when modeled as ARFIMA(2, $d$ ,1), the value of  $d$  equals 0.428 while  $\phi_1$ ,  $\phi_2$  and  $\theta_1$  equal 0.438,  $-0.050$  and 0.598 respectively.

Table 4B also shows the log-likelihood values for the different models which are used to perform the likelihood-ratio tests. When we compare ARFIMA(1, $d$ ,1) and ARFIMA(2, $d$ ,1), the LR test results in the lower panel of Table 4B show that the ARFIMA(1, $d$ ,1) hypothesis is rejected at significance levels below 1%. When ARMA(2,1) and ARFIMA(2, $d$ ,1) are compared, it is found that the ARMA(2,1) hypothesis is also rejected at levels below 1%. The statistical results conclude that the ARFIMA(2, $d$ ,1) model is the best of those evaluated for the logarithm of four-hour realized volatility.

The above results only apply to the situation when the number of parameters is small. Models such as ARFIMA(3, $d$ ,2) and ARMA(3,2) are not included in our study because it is difficult to estimate the required number of parameters. Andersen and Bollerslev (1997, 1998) suggest that the log volatility process can be expressed as the sum of many AR(1) processes. Following Granger (1980) they show that under certain assumptions the autocorrelation structure of the log volatility process approaches that of a fractional integrated process as the number of AR(1) processes becomes large. Consequently the fractionally integrated dynamics can be considered as a reduced form of a high-order multi-factor volatility process. Our empirical results show that the two-factor model is as good statistically as the fractionally integrated model when the history of the volatility process is relatively short. However, as the number of observations increases, the long memory representation is better than the two-factor structure in explaining the volatility process, but potentially still inferior to structural component models of yet higher order. Consequently, our results cannot exclude the possibility that the real volatility process is defined by the sum of many AR(1) components. However, the number of components must be more than two.

## 6. Conclusions

Evidence of long memory in the volatility process has been well documented in many studies for both the equity and foreign exchange markets. However this should more precisely be explained as evidence for long memory *effects* since there are short memory processes that can have similar abilities to capture and mimic the persistent nature of asset price volatility (the sum of two AR(1) processes is an example).

In this paper, we show that it is easy to claim that a series has a long memory by performing popular statistical tests for detecting long memory but without considering short memory alternative specifications. However, it is not always possible to identify the correct memory specification when tests are evaluated for both types. Option traders, in particular, must therefore often face an important degree of uncertainty about the correct specification of volatility processes.

We find that the length of the financial series and therefore the number of observations plays an important role in deciding if long or short memory is present. Through a Monte Carlo study, we examine if the long or short memory of a simulated series can be revealed statistically. Our results first show that it is indeed easy to mistake a long memory process as a short memory mimicking process and vice versa. We propose a test, which is composed of two likelihood ratio tests, to identify the correct specification using simulated series generated from ARFIMA(0, $d$ ,0) and ARMA(2,1) models with the parameters appropriately chosen. By generating simulated series of different length, we document the effect of series length on the ability to identify the correct type of memory. We find that it is often not possible to identify the correct type of memory when the simulated series is not long enough (for example 2,000 observations). However, the chance of such a mistake can be reduced by including more observations. For example, considering a persistence level equivalent to  $d$  equaling 0.4 and when the simulated series contains 12,000



observations, there is a 95% chance of identifying the correct nature of memory when the tests employ the 5% significance level.

Our Monte Carlo results are useful in identifying the correct type of memory in GBP/USD exchange rate volatility. Using approximately 3,000 observations of daily realized volatility, we cannot decide if long or short memory exists in the volatility series. However, the Monte Carlo results show that by lengthening the series there is a good chance to identify the correct specification. This is done by increasing the number of observations by raising the volatility sampling frequency from daily to four hours. It is then found that a long memory process is preferable to the sum of two AR processes in capturing the volatility dynamics although this could still be a parsimonious treatment of a yet higher-order, factor AR model.

## References

- Alizadeh, S., M.W. Brandt and F.X. Diebold, 2002, Range-based estimation of stochastic volatility models, *Journal of Finance* 57, 1047-1091.
- Andersen, T.G. and T. Bollerslev, 1997, Heterogenous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns, *Journal of Finance* 52, 975-1005.
- Andersen, T.G. and T. Bollerslev, 1998, Deutsche Mark-Dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies, *Journal of Finance* 53, 219-265.
- Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001a, The distribution of realized stock return volatility, *Journal of Financial Economics* 61, 43-76.
- Andersen, T.G., T. Bollerslev, F.X. Diebold and P. Labys, 2001b, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42-55.
- Areal, N.M.P.C. and S. J. Taylor, 2002, The realized volatility of FTSE-100 futures prices, *Journal of Futures Markets* 22, 627-648.
- Baillie, R.T., 1996, Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5-59.
- Baillie, R.T., T. Bollerslev and H.O. Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 74, 3-30.
- Barndorff-Nielsen, O.E. and N. Shephard, 2001, Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics, *Journal of the Royal Statistical Society B* 63, 167-241.
- Beran, J., 1994, *Statistics for Long-Memory Processes*, Chapman and Hall, New York.
- Bollerslev, T., 1986, Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics* 31, 307-327.
- Bollerslev, T. and I. Domowitz, 1993, Trading patterns and prices in the interbank foreign exchange market, *Journal of Finance* 48, 1421-1443.
- Bollerslev, T. and H.O. Mikkelsen, 1996, Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* 73, 151-184.

- Breidt, F.J., N. Crato and P. de Lima, 1998, The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* 83, 325-348,
- Brockwell, P.J. and R.A. Davis, 1991, *Time Series: Theory and Methods*, second edition, Springer, New York.
- Cheung, Y.W. and F.X. Diebold, 1994, On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean, *Journal of Econometrics* 62, 301-316.
- Chung, C.-F., 1994, A note on calculating the autocovariances of fractionally integrated ARMA models, *Economics Letters* 45, 293-297.
- Davies, R.B. and D.S. Harte, 1987, Tests for Hurst effect, *Biometrika* 74, 95-102.
- Deo, R., C. Hurvich and Y. Lu, 2006, Forecasting realized volatility using a long-memory stochastic volatility model: estimation, prediction and seasonal adjustment, *Journal of Econometrics*, 131(1-2), 29-58.
- Ding, Z., C.W.J. Granger and R.F. Engle, 1993, A long memory property of stock markets and a new model, *Journal of Empirical Finance* 1, 83-106.
- Duffie, D., J. Pan and K.J. Singleton, 2000, Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* 68, 1343-1376.
- Fox, R. and M. S. Taqqu, 1986, Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics*, 14, 517-532.
- Gallant, A.R., C.-T. Hsu and G.E. Tauchen, 1999, Using daily range data to calibrate volatility diffusions and extract the forward integrated variance, *The Review of Economics and Statistics* 81, 617-631.
- Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221-238.
- Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, *Journal of Econometrics* 14, 227-238.
- Granger, C.W.J. and R. Joyeux, 1980, An introduction to long memory time series models and fractional differencing, *Journal of Time Series Analysis* 1, 15-39.
- Granger, C.W.J. and P. Newbold, 1977, *Forecasting Economic Time Series*, Academic Press. New York.
- Granger, C.W.J. and Z. Ding, 1996, Varieties of long memory models, *Journal of Econometrics* 73, 61-77.

- Hosking, J.R.M., 1981, Fractional differencing, *Biometrika* 68, 165-176.
- McLeod, A.I. and K.W. Hipel, 1978, Preservation of the rescale adjusted range 1. A reassessment of the Hurst phenomenon, *Water Resources Research* 14(3), 491-508.
- Nelson, D.B., 1991, Conditional heteroskedasticity in asset returns: a new approach, *Econometrica* 59, 347-370.
- Ohanissian, A., J.R. Russell and R.S. Tsay, 2008, True or spurious long memory? A new test, *Journal of Business and Economic Statistics*, 26(2), 161-175.
- Pong, S.Y., M.B. Shackleton, S.J. Taylor and X. Xu, 2004, Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models, *Journal of Banking and Finance* 28, 2541-2563.
- Robinson, P.M., 1994, Semiparametric analysis of long-memory time series, *Annals of Statistics* 22, 515-539.
- Shackleton, M.B., 1998, *Frequency domain and stochastic control theory applied to volatility and pricing in intraday financial data*, PhD Thesis, London Business School, University of London.
- Smith, J., N. Taylor and S. Yadav, 1997, Comparing the bias and misspecification in ARFIMA models, *Journal of Time Series Analysis* 18, 507-527.
- Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics* 53, 165-188.
- Taylor, S.J., 1986, *Modelling Financial Time Series*, John Wiley, Chichester, U.K.
- Taylor, S.J., 2005, *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, Princeton.
- Whittle, P., 1951, *Hypothesis Testing in Time Series Analysis*, Almquist and Wiksells, Uppsala.

Table 1. Limit properties of the autocorrelation function and the spectral density

Limiting Case	Short memory	Long memory
Autocorrelation $\tau \rightarrow \infty$	$\frac{ \rho_\tau }{\phi^\tau} \rightarrow C_1 > 0,$ $0 < \phi < 1$	$\frac{\rho_\tau}{\tau^{2d-1}} \rightarrow D_1 > 0,$ $d > 0$
Frequency $\omega \rightarrow 0$	$f(\omega) \rightarrow C_2 > 0$	$\frac{f(\omega)}{\omega^{-2d}} \rightarrow D_2 > 0,$ $d > 0$

The quantities  $C_1, C_2, D_1, D_2$  are positive constants.

Table 2. Averages of parameter estimates for simulated series

Simulated Process	Estimated Model	length= 2,000	length= 4,000	length= 8,000	length= 12,000	True values
ARFIMA (0,d,0)	ARFIMA(0,d,0) $\left\{ \begin{array}{l} \hat{d}_{MLE} \\ \hat{d}_{GPH} \end{array} \right.$	0.4022	0.4020	0.4016	0.4010	0.4000
		0.4027	0.4019	0.4018	0.4015	0.4000
	ARMA(2,1) $\left\{ \begin{array}{l} \psi_1 \\ \nu_1 \\ \psi_2 \\ \nu_2 \end{array} \right.$	0.9465	0.9541	0.9605	0.9632	
		0.1806	0.1836	0.1870	0.1884	
		0.2428	0.2527	0.2614	0.2643	
		0.1857	0.1870	0.1889	0.1895	
ARMA (2,1)	ARMA(2,1) $\left\{ \begin{array}{l} \psi_1 \\ \nu_1 \\ \psi_2 \\ \nu_2 \end{array} \right.$	0.9742	0.9763	0.9767	0.9774	0.9780
		0.1387	0.1378	0.1368	0.1369	0.1360
		0.4419	0.4447	0.4466	0.4472	0.4470
		0.2146	0.2150	0.2156	0.2156	0.2160
	ARFIMA(0,d,0) $\left\{ \begin{array}{l} \hat{d}_{MLE} \\ \hat{d}_{GPH} \end{array} \right.$	0.3030	0.3047	0.3029	0.3040	
		0.3317	0.3289	0.3245	0.3212	

The degree of fractional integration is estimated using the maximum likelihood estimation method and the GPH (1983) method modified by Robinson (1994). The parameters of the ARMA(2,1) model for simulation are chosen to mimic an ARFIMA(0,d,0) process with  $d = 0.4$ .

Table 3. Probabilities of correct memory choices as the level of persistence changes

A. Simulated series generated from the ARFIMA(0, $d$ ,0) process

Parameter/length	2,000	4,000	8,000	12,000
$d=0.4$	72.4%	90.6%	93.7%	94.6%
$d=0.3$	61.2%	87.9%	91.9%	94.8%
$d=0.2$	37.8%	66.8%	90.1%	94.3%
$d=0.1$	7.1%	19.5%	46.5%	64.3%

B. Simulated series generated from the two-factor process ARMA(2,1)

Parameters/length	2,000	4,000	8,000	12,000
$(\psi_1, v_1, \psi_2, v_2) = (0.97, 0.14, 0.44, 0.21)$	55.4%	81.9%	93.9%	95.7%
$(\psi_1, v_1, \psi_2, v_2) = (0.97, 0.10, 0.30, 0.20)$	46.3%	65.7%	91.9%	94.9%
$(\psi_1, v_1, \psi_2, v_2) = (0.94, 0.07, 0.17, 0.20)$	25.9%	48.1%	77.5%	90.1%
$(\psi_1, v_1, \psi_2, v_2) = (0.90, 0.48, 0.06, 0.20)$	9.6%	16.6%	30.4%	46.0%

Panels A and B show the percentage of simulations for which the correct process is identified using a significance level of 5%. The results in Panels A and B are obtained using simulated series generated by the ARFIMA(0, $d$ ,0) and ARMA(2,1) processes respectively. The parameter values of the ARMA(2,1) process are obtained by approximating the spectral density of an ARFIMA(0, $d$ ,0) process with  $d=0.4, 0.3, 0.2$  and 0.1 respectively.

Table 4. Maximum likelihood estimation for log realized volatilities, for the £/\$ exchange rate

A. N = 2,867				B. N = 17,202			
I.				I.			
ARFIMA	(1,d,1)	(2,0,1)	(2,d,1)	ARFIMA	(1,d,1)	(2,0,1)	(2,d,1)
d	0.3479 (0.0476)		0.2614 (0.0742)	d	0.4701 (0.0322)		0.4279 (0.0390)
AR(1)	0.9430 (0.0355)	1.2008 (0.0397)	1.0142 (0.0307)	AR(1)	0.5075 (0.0313)	1.1951 (0.0123)	0.4384 (0.0302)
AR(2)		-0.2215 (0.0365)	-0.0621 (0.0313)	AR(2)		-0.2035 (0.0119)	-0.0495 (0.0272)
MA(1)	0.8988 (0.0595)	0.7966 (0.0306)	0.8756 (0.0513)	MA(1)	0.7190 (0.0278)	0.9223 (0.0063)	0.5983 (0.0248)
log likelihood	3027.72	3027.15	3028.45	log likelihood	12197.18	12171.80	12206.00
II.				II.			
<i>Model Comparison</i> (1,d,1) v (2,d,1) (2,0,1) v (2,d,1)				<i>Model Comparison</i> (1,d,1) v (2,d,1) (2,0,1) v (2,d,1)			
<i>Likelihood Ratio</i>				<i>Likelihood Ratio</i>			
<i>Test Statistic</i>				<i>Test Statistic</i>			
	1.47 (0.23)	2.60 (0.11)			17.64 (0.00)	68.39 (0.00)	

The estimation in Panel A is based on the logarithm of daily realized volatility series from July 1, 1987 to December 31, 1998 while that in Panel B is based on the logarithm of the 4-hour realized volatility series for the same period. Sub-panel I shows the estimates of the parameters of three different models using Maximum Likelihood Estimation. The values in parentheses represent the standard errors of the estimates. Sub-panel II shows the Likelihood Ratio Test Statistics when comparing two models. The values in parentheses are the levels of significance of the test statistics.



Figure 1: Spectral densities of ARMA(2,1) and ARFIMA(0,d,0) processes

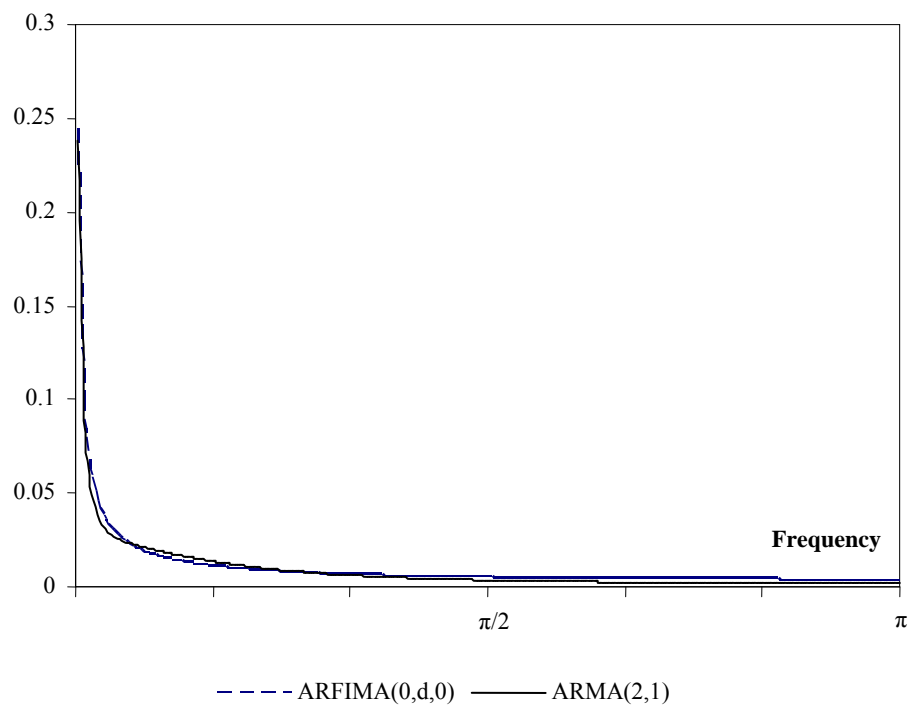


Figure 2. Design of the combined likelihood ratio test to detect the correct specification in simulated series.

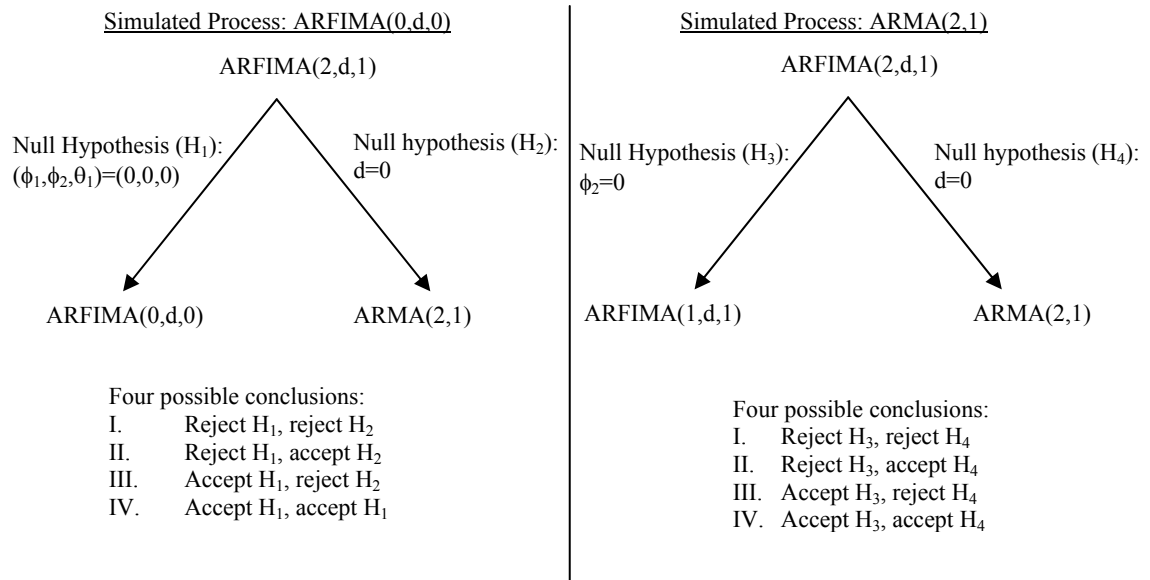
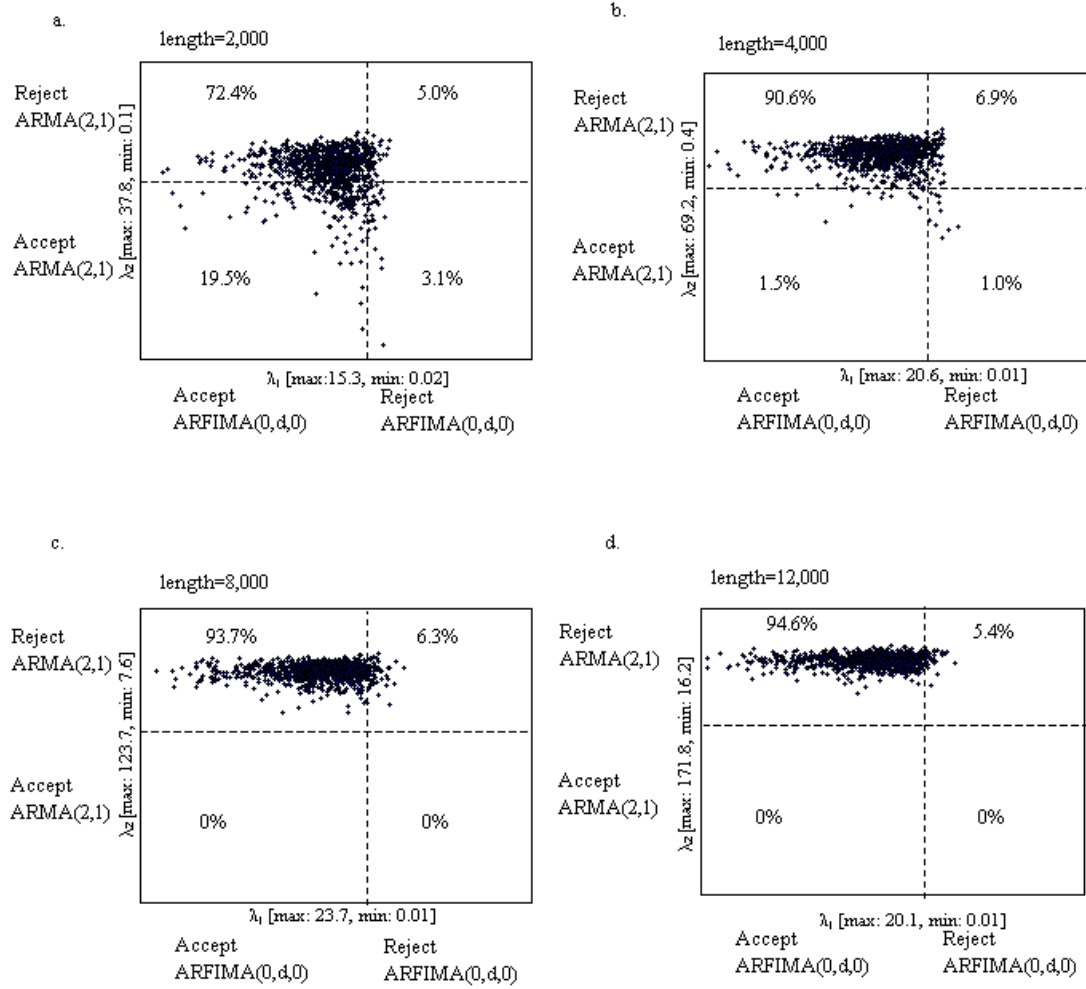


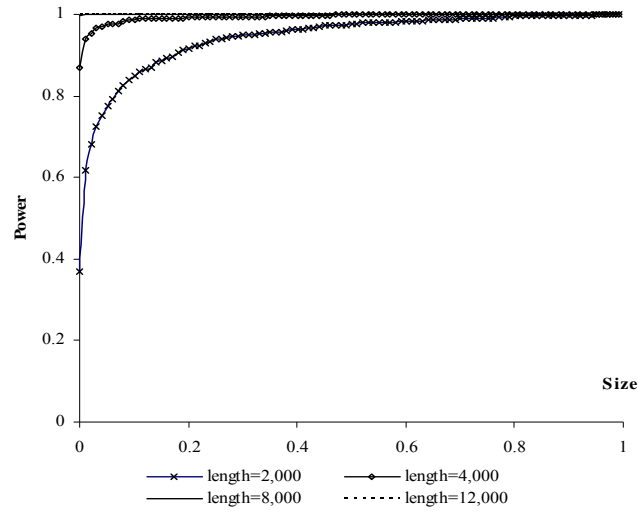
Figure 3: The log-likelihood ratios from combined likelihood ratio tests when simulated series are generated from the ARFIMA(0,d,0) process



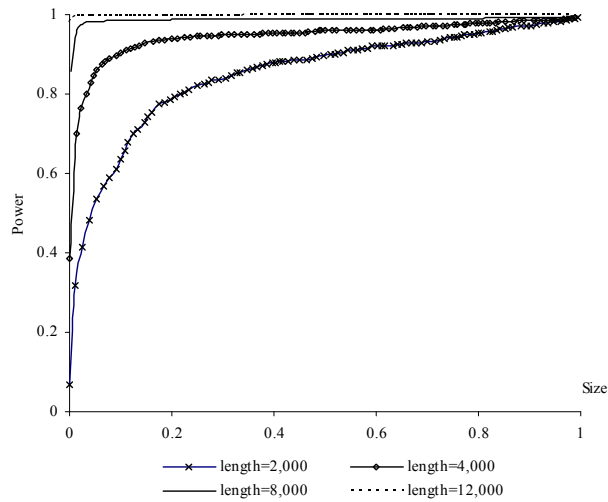
$\lambda_1$  is twice the log-likelihood ratio from the comparison of ARFIMA(0,d,0) against ARFIMA(2,d,1) while  $\lambda_2$  is twice the log-likelihood ratio from the comparison of ARMA(2,1) against ARFIMA(2,d,1). Both variables are plotted on a logarithmic scale. The dotted lines represent the cut-off boundaries (  $\lambda_1 = 7.81$  ,  $\lambda_2 = 3.84$  ) used to reject the null hypothesis (  $H_1 : \phi_1 = \phi_2 = \theta_1 = 0$  ,  $H_2 : d = 0$  ).

Figure 4: Comparisons of test size and power

a.

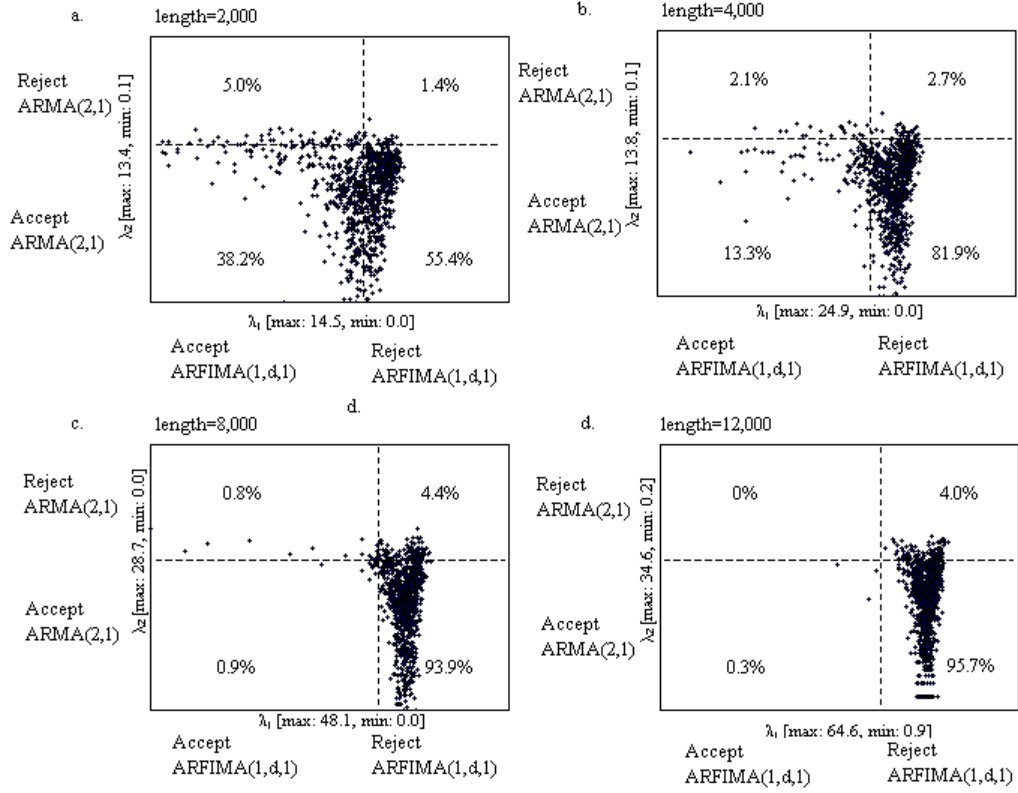


b.



Panel a shows the size-power curve for the combined likelihood test given the generating process is ARFIMA(0,d,0). The value of size represents the probability of rejecting ARFIMA(0,d,0), which is the correct specification, and that of power represents the probability of rejecting ARMA(2,1), which is the incorrect specification. Panel b shows the size-power curve for the combined likelihood test given the generating process is ARMA(2,1). The value of size represents the probability of rejecting ARMA(2,1), which is the correct specification, while that of power represents the probability of rejecting ARFIMA(1,d,1), which is the incorrect specification.

Figure 5: The log-likelihood ratios from combined likelihood ratio tests when simulated series are generated from the ARFIMA(2,1) process



$\lambda_1$  is twice the log-likelihood ratio from the comparison of ARFIMA(1,d,1) against ARFIMA(2,d,1) while  $\lambda_2$  is twice the log-likelihood ratio from the comparison of ARMA(2,1) against ARFIMA(2,d,1). Both variables are plotted on a logarithmic scale. The dotted lines represent the cut-off boundaries ( $\lambda_1 = 3.84$ ,  $\lambda_2 = 3.84$ ) used to reject the null hypothesis ( $H_1 : \phi_2 = 0$ ,  $H_2 : d = 0$ ).

Figure 6: Periodogram of the 4-hour log realized volatility for the USD/GBP exchange rate for the period from January 1 1993 to December 31 1994.

