Indirect Estimation of Long Memory Volatility Models

Nigel Wilkins
School of Finance and Economics
University of Technology, Sydney
PO Box 123
Broadway NSW 2007
Australia
Ph: 61 2 95147784
Fax: 61 2 95147711
nigel.wilkins@uts.edu.au

January 2004

Abstract

An indirect estimator is proposed for two long memory volatility models; the fractionally integrated generalised autoregressive conditional heteroskedasticity (FIGARCH) model and the long memory stochastic volatility (LMSV) model. The small sample properties of the indirect estimator are compared to the small sample properties of conventional maximum likelihood estimators. It is found that the indirect estimator has the potential to perform favourably with respect to maximum likelihood for higher order parameterised FIGARCH and LMSV models.

Keywords: Fractional Integration, Persistence, Simulation.

JEL Classifications: C13, C15.

1 Introduction

Recent empirical research examining the time dependent conditional variances of financial variables has found that long memory is a relevant factor to be taken into consideration; see for example, Ding, Granger and Engle (1993), Ding and Granger (1996), Bollerslev and Mikkelsen (1996), and also Engle and Bollerslev (1986) for the original contribution on infinitely long memory conditional variances. This evidence of long range behaviour has lead to the development of integrated models for conditional variance. Under the restrictive integer framework, a stationary volatility process for a financial asset (where the level of integration is $d = 0$) exhibits exponential decay in response to an exogenous shock, while a nonstationary volatility process (where $d = 1$) displays no tendency to revert to its unconditional volatility, or variance, after an exogenous shock and actually retains some permanent component. In a manner similar to that for the conditional mean, it has been found that this $d = 0, 1$ dichotomy imposes too heavy a restriction on the allowable behaviour of the conditional second moment of a time series. Rather than restricting attention to integer levels of integration, it seems sensible to allow a compromise between the stationary models of conditional variance and nonstationary or integrated models. For the conditional mean, this has lead to the development and popularisation of the autoregressive fractionally integrated moving average (ARFIMA) model due to Granger and Joyeux (1980) and Hosking (1981); see also Baille (1996) for a survey of developments in ARFIMA modelling.

A natural framework on which to build a fractional model for the conditional variance is the autoregressive conditional heteroskedasticity (ARCH) model due to Engle (1982) or the generalised ARCH (GARCH) model due to Bollerslev (1986). Both models have received extensive application in modelling the time varying properties of the conditional variance of financial time series; see Bollerslev, Chou and Kroner (1992) or Bollerslev, Engle and Nelson (1994) for a review of ARCH models. Baille, Bollerslev and Mikkelsen (1996) propose the fractionally integrated GARCH (FIGARCH) model and a time domain maximum likelihood (TDML) estimator for the parameters in this model. Alternatively, the stochastic volatility (SV) class of models, see Taylor (1994) for a survey, can be utilised as a framework on which to build a fractional model. Harvey (1993) was the first to consider this by specifying a fractional noise process to drive the long memory component of the SV model. Breidt, Crato and deLima (1998) propose the long memory stochastic volatility (LMSV) model and a frequency domain based maximum likelihood (FDML) estimator for the parameters in this model. Unfortunately, the maximum likelihood estimators for the FIGARCH and LMSV models have potential problems in their application. The TDML estimator for the FIGARCH model has truncation and starting value problems, while the FDML estimator for the LMSV model can be prone to sizeable finite sample bias problems. This paper overcomes these estimation problems by proposing an indirect estimator for the two models based on the Gorieroux, Monfort and Renault (1993) version of the indirect estimator. The indirect estimator has already received application to fractional models for the conditional mean of a process; see Martin and Wilkins (1999) for details.

The outline of the paper is as follows. Section 2 briefly reviews models for conditional variance. Section 3 develops an indirect estimation procedure for FIGARCH and LMSV models by discussing alternative auxiliary models that may be implemented, computationally efficient simulation procedures and an estimation algorithm for empirical application. Section 4 reports the results of a set of simulation experiments to compare the small sample properties of the maximum likelihood and indirect estimators. Section 5 concludes.

2 Models for Conditional Variance

Consider some time series R_t with conditional mean equation

$$R_t = f(x_t) + \sigma_t \varepsilon_t \tag{1}$$

where, in very general terms, $f(x_t)$ denotes some functional form of the explanatory variables x_t, σ_t is the time varying conditional standard deviation (variance) and ε_t is iid$(0,1)$. Assuming the specification $f(x_t) = \mu_t$ where μ_t is the conditional mean of R_t, then (1) can be expressed as the (conditional) mean corrected series

$$r_t = R_t - \mu_t = \sigma_t \varepsilon_t \tag{2}$$

where the time series r_t is now a white noise (WN) process.

2.1 Autoregressive Conditional Heteroskedasticity

The process r_t exhibits ARCH(q) effects if σ_t can be written

$$\sigma_t^2 = \omega + \alpha(L) r_t^2 \tag{3}$$

where ω is a positive constant and $\alpha(L) = \sum_{i=1}^{q} \alpha_i L^i$ is a lag operator polynomial in L with all α_i non-negative to ensure σ_t^2 is positive for all possible realisations of r_t. When the appropriate specification for (3) includes lagged σ_t^2, the model becomes the generalised ARCH(p,q), or GARCH(p,q) model due to Bollerslev (1986)

$$\sigma_t^2 = \omega + \alpha(L) r_t^2 + \beta(L) \sigma_t^2 \tag{4}$$

where $\beta(L) = \sum_{i=1}^{p} \beta_i L^i$ is a lag operator polynomial in L with, once again, all β_i non-negative to ensure σ_t^2 is always positive. Bollerslev (1986) shows that (4) can be rearranged to give

$$[1 - \alpha(L) - \beta(L)]\, r_t^2 = \omega + [1 - \beta(L)]\, v_t \tag{5}$$

where $v_t = r_t^2 - \sigma_t^2$ is a random shock, or error component in conditional variance and (5) is interpreted as being an ARMA(q^*, p) model for r_t^2 where $q^* = \max\{p, q\}$.

Engle and Bollerslev (1986) extend (5) by allowing $1 - \alpha(L) - \beta(L)$ to contain a unit root, giving the integrated GARCH(q^*, p), IGARCH(q^*, p) model

$$\Pi(L)(1-L)\varepsilon_t^2 = \omega + [1 - \beta(L)]\, v_t \tag{6}$$

where $\Pi(L) = (1-L)^{-1}[1 - \alpha(L) - \beta(L)] = (1-L)^{-1}[1 - \Phi(L)]$ is of order $q^* - 1$; see also Nelson (1990). In practical application, $d = 1$ has been found to be a suitable parameterisation for this class of model. The two GARCH extremes represented in (4) and (6) might be considered too rigid to approximate the sensitive dynamics in the data generating process (DGP) of a financial time series accurately. This suggests the relevance of more general, fractionally integrated, versions of these models. This is advantageous because incorporating fractional integration into the conditional second moment of a time series allows the conditional variance function to display persistent but still mean reverting behaviour. This appears to be more consistent with actual financial returns data than that behaviour captured in the $d = 0$ or $d = 1$ models.

The fractionally integrated GARCH, or FIGARCH model due to Baille, Bollerslev and Mikkelsen (1996) overcomes the restrictions of these models. The FIGARCH model is obtained by using the fractional differencing operator in (6) rather than the integer differencing operator. The model now becomes

$$\Pi(L)(1-L)^d r_t^2 = \omega + [1 - \beta(L)]\, v_t \tag{7}$$

which is in its ARMA(q^*, p) representation for r_t^2 and the fractional differencing filter is as conventionally defined

$$(1-L)^d = \frac{(j-d-1)!}{j!(-d-1)!} L^j \tag{8}$$

for $j = 1, 2, \ldots, \infty$. Rearranging (7) results in an alternative expression for the FIGARCH(p, d, q^*) model

$$\sigma_t^2 = \omega + \beta(L)\sigma_t^2 + \left[1 - \beta(L) - \Pi(L)(1-L)^d\right] r_t^2 \tag{9}$$

which shows how σ_t^2 evolves over time. In an analogous fashion to the exponential GARCH model of Nelson (1991), Bollerslev and Mikkelsen (1996) extend (7) to the fractionally integrated EGARCH(p, d, q^*), or FIEGARCH(p, d, q^*) model. The FIGARCH model has been applied to stock market returns data by Bollerslev and Mikkelsen (1996) and Psaradakis and Sola (1995)

As an illustration, the specific formulation for the FIGARCH$(1, d, 1)$ model is

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \left[1 - \beta_1 L - (1 - \pi_1 L)(1-L)^d\right] r_t^2 \tag{10}$$

where $\pi_1 = \alpha_1 + \beta_1$ is as defined in reference to (6). Equation (10) can be expressed as the observationally equivalent infinite order ARCH representation

$$\sigma_t^2 = \frac{\omega}{1-\beta_1} + \left[1 - \frac{(1-\pi_1 L)(1-L)^d}{(1-\beta_1 L)}\right] r_t^2 = \frac{\omega}{1-\beta_1} + \lambda_1(L) r_t^2 \tag{11}$$

where $\lambda_1(L) = \sum_{i=1}^{\infty} \lambda_{1,i} L^i$. Similarly, the FIGARCH$(1, d, 0)$ model is obtained by setting $\pi_1 = 0$ in (10)

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \left[1 - \beta_1 L - (1-L)^d\right] r_t^2 \tag{12}$$

and this has the observationally equivalent infinite order ARCH representation

$$\sigma_t^2 = \frac{\omega}{1-\beta_1} + \left[1 - \frac{(1-L)^d}{(1-\beta_1 L)}\right] r_t^2 = \frac{\omega}{1-\beta_1} + \lambda_2(L) r_t^2 \tag{13}$$

where $\lambda_2(L) = \sum_{i=1}^{\infty} \lambda_{2,i} L^i$. Following Bollerslev and Mikkelsen (1996), the ARCH coefficients in the lag operator polynomial $\lambda_1(L)$ have the recursive form

$$\lambda_{1,i} = \pi_1 - \beta_1 + d \tag{14}$$

for $i = 1$ and

$$\lambda_{1,i} = \beta_1 \lambda_{1,i-1} + \left[\frac{(i-1-d)}{i} - \pi_1\right] \delta_{i-1} \tag{15}$$

for $i = 2, 3, \ldots$ and where $\delta_i = \delta_{i-1}(i-1-d)/i$ is a recursive function for the coefficients in the expression for $(1-L)^d$ in (8). The coefficients in the lag operator polynomial $\lambda_2(L)$ are found by setting $\pi_1 = 0$ in (14) and (15). The two FIGARCH models in (10) and (12) are used as alternative DGPs for the Monte Carlo experiments in Section 4.

Estimation of the conventional GARCH model is nontrivial, although straightforward, because of the restrictions on the model coefficients to ensure conditional variance is always positive. In the case of the FIGARCH model, this complexity is increased because negative coefficients in the differencing filter in (8) cannot be avoided. This implies the simplest class of model for long memory conditional variances, the FIGARCH$(0, d, 0)$ model is not defined because the resulting negative coefficients in (9) cannot be avoided. For the more general FIGARCH(p, d, q^*) model, a general expression for the required parameter restrictions is not yet available, but, as Bollerslev and Mikkelsen (1996) note, the restrictions necessary for specific FIGARCH(p, d, q^*) models can be obtained on a case by case basis. For the FIGARCH$(1, d, 1)$ model, these are

$$\beta_1 - d \leq \pi_1 \leq \frac{(2-d)}{3} \tag{16}$$

and

$$d\left[\pi_1 - \frac{(1-d)}{2}\right] \leq \beta_1(\pi_1 - \beta_1 + d) \tag{17}$$

which simplify to

$$0 < \beta_1 < d < 1 \tag{18}$$

for the FIGARCH$(1, d, 0)$ model.

2.1.1 Time Domain Maximum Likelihood Estimation

By assuming conditional normality, it is possible to estimate (9) by maximising the natural logarithm of the conventional ARCH likelihood function

$$\ln L\{\Theta_1; r_t^2\} = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left[\ln\sigma_t^2 + \frac{r_t^2}{\sigma_t^2}\right] \tag{19}$$

where $\Theta_1 = \{d, \omega, \beta_1, \beta_2, \ldots, \beta_p, \pi_1, \pi_2, \ldots, \pi_{q^*}\}$ is the FIGARCH parameter vector, r_t^2 is obtained using the conditional mean equation and σ_t^2 is obtained approximately according to (9) by using the fractional differencing filter truncated at some point beyond which the contribution of extra terms in the summation is negligible. Estimation of the FIGARCH model using (19) is referred to as time domain maximum likelihood (TDML) estimation.

Three points should be noted about this TDML estimator. First, the truncation of the differencing filter means that maximum likelihood estimation is approximate only, although the extent of the approximation can be reduced by increasing the truncation point. This procedure is therefore not an exact estimator like the TDML estimator for the parameters of the ARFIMA(p, d, q) model for the conditional mean; see Sowell (1992a). Second, because of the truncation, estimation using (19) can be expensive in small samples. Although this might not be too much of a concern when using financial data sets, it still involves the loss of sample data and therefore sample information. Third, the truncation means the actual series being used is not the true fractionally differenced series and therefore there is a bias being introduced into the estimation procedure. This can be justified using an argument similar to that in Sowell (1992b) when discussing the estimation of the parameters in the conditional mean equation with an approximate TDML estimator.

Estimation using (19) is often referred to as quasi maximum likelihood (QML) estimation because the assumption of a normally distributed error process does not always hold in financial data sets. Using a robustified estimator of the covariance matrix of the parameter estimates is asymptotically valid however, see Weiss (1986) and Bollerslev and Wooldridge (1992). A heteroskedasticity consistent covariance matrix can be obtained using

$$H(\widehat{\Theta}_1)^{-1}G(\widehat{\Theta}_1)H(\widehat{\Theta}_1)^{-1} \tag{20}$$

where $G(\widehat{\Theta}_1)$ is the outer product of the gradients and $H(\widehat{\Theta}_1)$ is the hessian, both of which are evaluated at $\widehat{\Theta}_1$.

2.1.2 Frequency Domain Maximum Likelihood Estimation

The ARMA representation of the FIGARCH model in (7) allows the model to be estimated using a frequency domain based likelihood function. The interpretation of (7) as an ARFIMA(q^*, d, p) model for the squared errors allows standard time series results to be used to obtain the spectrum for r_t^2. In its most general form, the spectrum is

$$I_{r^2}(\lambda) = \frac{\sigma_\nu^2}{2\pi} \frac{|\Pi(e^{-i\lambda})|^2}{|\beta(e^{-i\lambda})|^2} \left|1 - e^{-i\lambda}\right|^{-2d} \tag{21}$$

where $|\Pi(e^{-i\lambda})|$ and $|\beta(e^{-i\lambda})|$ are the spectral generating functions corresponding to the lag operator polynomials $\Pi(L)$ and $\beta(L)$, and λ is angular frequency.

Equation (21) now allows specification of the spectral likelihood function, see Fox and Taqqu (1986),

$$\ln L\{\Theta_1; p(\lambda_j)\} = -\frac{1}{2\pi} \sum_{j=1}^{T/2} \frac{p(\lambda_j)}{I_{r^2}(\lambda_j)} \tag{22}$$

where Θ_1 is the parameter vector defined in reference to (19), $\lambda_j = 2\pi j/T$ for $j = 1, 2, \ldots, T/2$ is the jth Fourier frequency and $p(\lambda_j)$ is the estimated periodogram for the squared data. Once estimates for the parameters in the lag operator polynomials in the ARFIMA representation in (7) have been obtained, they can be used to generate an estimate of the constant ω, and all parameters can then be substituted into (9) to obtain the more familiar ARCH specification of the model. Estimation using (22) is advantageous because it avoids the restrictions inherent in the TDML procedure.

2.2 Stochastic Volatility

The ARCH class of models explain the conditional variance as a function of (squared) past errors, thus imposing dependence between r_t and σ_t^2. An alternative to this is the SV model, which specifies a DGP for σ_t^2 that is independent of r_t. The SV model can be written

$$\sigma_t^2 = \sigma^2 \exp(\frac{\nu_t}{2})^2 \tag{23}$$

where ν_t is some process that imposes the heteroskedasticity in σ_t^2 and that is also independent of r_t. Most simply, ν_t can be taken to be an AR(1) process, $\nu_t = \phi_1 \nu_{t-1} + \eta_t$, with the normal requirement that $|\phi_1| < 1$ for stationarity to hold. An alternative fractionally integrated DGP for the conditional variance is considered by Harvey (1993) and extended by Breidt, Crato and de Lima (1998). The equation explaining the evolution of the conditional variance is given the specification

$$\sigma_t = \sigma \exp(\frac{\nu_t}{2}) \tag{24}$$

where now ν_t follows the ARFIMA(p,d,q) process

$$\Phi(L)(1-L)^d \nu_t = \delta + \Theta(L)\varsigma_t \tag{25}$$

where $\Phi(L) = \sum_{i=0}^{p} \phi_i L^i$, $\Theta(L) = \sum_{i=0}^{q} \theta_i L^i$ and ς_t is WN$(0, \sigma_\varsigma^2)$. The long memory stochastic volatility, or LMSV(p,d,q) model is defined by the specification of the ARFIMA model in (25).

Following Breidt, Crato and de Lima (1998), the natural logarithm of the square of (2) with (24) substituted in can be rearranged to give

$$\ln(r_t)^2 = E[\ln \varepsilon_t^2] + \nu_t + \epsilon_t \tag{26}$$

where $\epsilon_t = \ln r_t^2 - E[\ln r_t^2]$ is an iid$(0, \sigma_\epsilon^2)$ error process. Using (26), the persistence in conditional variance nature of the series can now be examined using conventional ARFIMA methods. As a direct extension of (21), the spectrum of (26) is

$$I_{r^2}(\lambda) = \frac{\sigma_\varsigma^2}{2\pi} \frac{|\Theta(e^{-i\lambda})|^2}{|\Phi(e^{-i\lambda})|^2} \left|1 - e^{-i\lambda}\right|^{-2d} + \frac{\sigma_\epsilon^2}{2\pi} \tag{27}$$

which can be found directly by invoking the linearity property of the spectrum. As an example, the simplest model that can be considered is the LMSV$(0,d,0)$ model

$$r_t = \sigma \exp\left\{ \frac{(1-L)^{-d}\varsigma_t}{2} \right\} \varepsilon_t \tag{28}$$

while a more general specification is the LMSV$(1,d,1)$ model

$$r_t = \sigma \exp\left\{ \frac{(1-\phi L)^{-1}(1-L)^{-d}(1+\theta L)\varsigma_t}{2} \right\} \varepsilon_t \tag{29}$$

Both of these models are used in Section 4 as alternative DGPs for the Monte Carlo experiments. An advantage of the LMSV model is that the restrictions on the parameters in the FIGARCH model are not present, see (16) to (18), and the simplest fractional model for conditional variances with $p, q = 0$ but $d \neq 0$ can now be considered.

2.2.1 Frequency Domain Maximum Likelihood Estimation

Estimation of the SV model using TDML procedures requires numerical integration at each observation; see for example Gourieroux, Monfort and Renault (1993). A frequency domain based likelihood function specified using (27) avoids this problem. From Breidt, Crato and de Lima (1998), the spectral representation of the likelihood function for the transformation of the LMSV model in (26) is

$$\ln L\{\Theta_2; p(\lambda_j)\} = -\frac{2\pi}{T} \sum_{j=1}^{T/2} \left[\ln I_{r^2}(\lambda_j) + \frac{p(\lambda_j)}{I_{r^2}(\lambda_j)} \right] \tag{30}$$

where $\Theta_2 = \{d, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma_\varsigma^2, \sigma_\epsilon^2\}$ is the LMSV parameter vector and $p(\lambda_j)$ is now the estimated periodogram for the natural logarithm of the squared mean corrected data. Estimation of the LMSV model using (30) is referred to as frequency domain maximum likelihood (FDML) estimation. Breidt, Crato and de Lima (1998) show the strong consistency of the estimates obtained using (30). The convenient spectral representation of fractional processes and the truncation involved in maximising (19), implies that maximisation of (30) has the potential to be computationally easier and more accurate than (19).

3 Indirect Estimation

3.1 The Indirect Estimator

At its most elementary level, the indirect estimator obtains parameter estimates for one model by estimating the parameters of another model. The estimation procedure has two basic requirements. The first is that the direct model of interest, that is the model the empirical researcher wishes to estimate, must be easy and computationally efficient to simulate. This model of interest is never actually "directly" estimated, only simulated. The second requirement is that there exists some indirect, or auxiliary model that is straightforward to estimate and that can be considered to be representative of the direct model of interest. The parameter estimates for this auxiliary model are of no interest in their own right. Indirect estimation yields consistent parameter estimates for the direct model by comparing the estimates of the auxiliary model obtained using the actual data to those obtained using artificial data simulated from the direct model. The direct model is then repeatedly simulated within an iterative routine until the parameter estimates for the auxiliary model obtained using the actual data and the simulated data are equal, or sufficiently close. The set of parameters for the direct model used to generate the last simulated series of numbers are the indirect parameter estimates for the direct model.

Martin and Wilkins (1999) consider estimating the parameters in the general ARFIMA(p, d, q) model using simple AR(p) processes as auxiliary models. As a specific example, denote the autoregressive parameter vector in an AR auxiliary model as ρ, and the estimate of ρ obtained using the actual data as $\rho(a)$. Now denote the parameter vector for the direct model in (25) as $\Psi = \{d, \delta, \Phi(1), \Theta(1), \sigma^2\}$. Indirect estimation involves simulating (25), for a particular set of parameter values, say Ψ_1, and then estimating the AR auxiliary model with this simulated series. Denote the estimate of the auxiliary model parameter using this simulated series as $\rho_1(s)$. Then, intuitively, the indirect estimator of Ψ is that Ψ_i corresponding to the ith simulated series where

$$\rho_i(s) \rightarrow \rho(a) \tag{31}$$

with a satisfactory tolerance level prespecified. When the dimension of the parameter vectors Ψ and ρ are the same, the criterion simplifies to $\rho_i(s) = \rho(a)$. This will be the case, for example, when estimating a zero mean ARFIMA$(0, d, 0)$ model with $\varepsilon_t \sim (0, 1)$ using an AR(1) process as an auxiliary model.

Rather than simulating the model only once for every Ψ_i, increased precision for the estimator may be obtained by simulating $h = 1, 2, \ldots, H$ series or paths, where each path consists of an independent drawing of random numbers with which the direct model is simulated. For H such paths, $\rho_i(s)$ is replaced by an average of the parameter estimates from all H paths, $1/H \sum_{h=1}^{H} \rho_{i,h}(s)$, and the criterion in (31) becomes

$$\frac{1}{H}\sum_{h=1}^{H} \rho_{i,h}(s) \to \rho(a) \tag{32}$$

This general principle, of course, carries over to fractionally integrated models for the conditional variance. Now denote the parameter vector of interest as $\mathbf{\Theta}$ which represents the full set of parameters for the FIGARCH model (that is, $\mathbf{\Theta}_1$) or the LMSV model (that is, $\mathbf{\Theta}_2$). The GMR indirect estimator is formally given by

$$\widehat{\mathbf{\Theta}} = \underset{\mathbf{\Theta}}{\text{Argmin}} \left(\mathbf{\Pi}(a) - \frac{1}{H}\sum_{h=1}^{H} \mathbf{\Pi}_h(s)\right)' \mathbf{\Omega} \left(\mathbf{\Pi}(a) - \frac{1}{H}\sum_{h=1}^{H} \mathbf{\Pi}_h(s)\right) \tag{33}$$

where $\mathbf{\Pi}(a)$ and $\mathbf{\Pi}_h(s)$ are vectors containing the parameter estimates for the auxiliary model using the actual and simulated data sets respectively, $\mathbf{\Omega}$ is a weighting matrix defined in Gourieroux, Monfort and Renault (1993) and the subscript i denoting simulated series has been dropped. The weighting matrix $\mathbf{\Omega}$ is of importance when the dimension of $\mathbf{\Pi}$ is greater than the dimension of $\mathbf{\Theta}$, otherwise the criterion in (33) simplifies to

$$\frac{1}{H}\sum_{h=1}^{H} \mathbf{\Pi}_h(s) = \mathbf{\Pi}(a) \tag{34}$$

when the dimension of the direct and auxiliary models are the same.

3.2 Auxiliary Models and Simulation Procedures

3.2.1 FIGARCH Auxiliary Models

With respect to choosing an auxiliary model and in reference to the special cases in (11) and (13), note that the general FIGARCH(p, d, q) model can be expressed as the infinite order ARCH model

$$\sigma_t^2 = \frac{\omega}{1 - \beta(1)} + \lambda(L) r_t^2 \tag{35}$$

where $\lambda(L) = 1 - [1 - \beta(L)]^{-1}\Pi(L)(1-L)^d$. This suggests that a natural choice of auxiliary model for the FIGARCH model is given by an ARCH(k)

$$\sigma_t^2 = \lambda_0 + \sum_{j=1}^{k} \lambda_j r_{t-j}^2 \tag{36}$$

where k is chosen to be greater than or equal to the parameter dimension of the FIGARCH model. In practice, (36) is equivalent to a simple AR(k) for the squared errors

$$r_t^2 = \lambda_0 + \sum_{j=1}^{k} \lambda_j r_{t-j}^2 + \zeta_t \tag{37}$$

where ζ_t is WN, since, by definition, $r_t^2 = E_{t-1}(r_t^2) + \zeta_t$, $E_{t-1}(r_t^2) = \sigma_t^2$ and $E_{t-1}(\zeta_t) = E(\zeta_t) = 0$, where E_{t-1} denotes the expectation conditional on the information set up to time $t-1$. This implies that $\sigma_t^2 = r_t^2 - \zeta_t$, thus allowing (37) to be used. Using (37) negates the need to use maximum likelihood estimation for the auxiliary model, thus reducing the computational burden of the indirect procedure.

Another auxiliary model contender, that possibly has a stronger theoretical justification than (37), is the ARMA(q^*, p) model for r_t^2 that is obtained from (5)

$$r_t^2 = \pi_0 + \sum_{j=1}^{q^*} \pi_i r_{t-i}^2 + \nu_t + \sum_{j=1}^{p} \beta_i \nu_{t-i} \tag{38}$$

where $\pi_i = \alpha_i + \beta_i$. However, this latter auxiliary model is not considered because of the increased computational burden it contains in estimating the moving average parameters when using the GMR (1993) indirect estimator. The Gallant and Tauchen (GT) (1996) indirect estimator can avoid this problem.

3.2.2 LMSV Auxiliary Models

The infinite order ARCH representation of the EGARCH(p, q) model due to Nelson (1991) suggests one possible auxiliary model that can be considered for the LMSV model is

$$\ln \sigma_t^2 = \omega + \sum_{j=1}^{k} \varphi_j \ln \sigma_{t-j}^2 + z_t \tag{39}$$

From this, an appropriate choice of auxiliary model might simply be an AR(k) for the natural logarithm of the squared errors

$$\ln r_t^2 = \omega + \sum_{j=1}^{k} \varphi_j \ln r_{t-j}^2 + \zeta_t \tag{40}$$

where ζ_t is once again a WN error process. The EGARCH model can be viewed as a discrete time version of the SV model and so would appear to be an appropriate choice as an auxiliary model.

3.2.3 Simulation Procedures

The simulation model can be written $\sigma_t = f(r_t)$ for the FIGARCH model and $\sigma_t = f(\nu_t)$ for the LMSV model. For the FIGARCH model, σ_t can be simulated by expanding the lag operator polynomial for r_t^2 in (9) to obtain the moving average error process

$$\tilde{r}_t^2 = \left[1 - (\beta_1 L + \ldots + \beta_p L^p) - (\pi_1 L + \ldots + \pi_{q^*} L^{q^*}) \sum_{j=1}^{n} \frac{(j-d-1)!}{j!(-d-1)!} L^j \right] r_t^2 \tag{41}$$

which will be of order $s = \max\{p, n + q^*\}$ and where (8) has been truncated at n. This particular truncation should not cause concern because n is not dependent on the size of the sample, and so can be made arbitrarily large. The error process $\tilde{r}_t^2$ can then be used to generate an AR(p) process for σ_t^2 using $\beta(1)$ as the autoregressive weights. The LMSV model is more straightforward to simulate as already established random number simulators can be used to generate the ARFIMA DGP, see Martin and Wilkins (1999) for a discussion of alternative procedures that may be implemented. Once the ν_t have been obtained they can be substituted into (2) using (24).

3.3 Estimation Algorithm

Consider the AR(k) for the squared errors in (37) as an auxiliary model for the FIGARCH model and the AR(k) for the natural logarithm of the squared errors in (40) as an auxiliary model for the LMSV model, then the estimation algorithm follows that in Martin and Wilkins (1999) and can be expressed as follows:

1. Estimate the auxiliary models (37) and (40) for a given lag length $k = k^*$, using the actual data, $r_t(a)$, and compute $\mathbf{\Pi}(a)$.

2. Choose an initial set of parameter estimates for the FIGARCH(p, d, q) model

$$\mathbf{\Theta}_1^{(0)} = \{d^{(0)}, \omega^{(0)}, \beta_i^{(0)}, i = 1, 2, \ldots, p, \pi_i^{(0)}, i = 1, 2, \ldots, q^*, \sigma_\varepsilon^{2(0)}\} \tag{42}$$

   or for the LMSV(p, d, q) model

$$\mathbf{\Theta}_2^{(0)} = \{d^{(0)}, \phi_i^{(0)}, i = 1, 2, \ldots, p, \theta_i^{(0)}, i = 1, 2, \ldots, q, \sigma_\varsigma^{2(0)}, \sigma_\varepsilon^{2(0)}\} \tag{43}$$

3. Draw a set of random numbers w_t, from a $N(0,1)$ distribution. For the LMSV model, this set of random numbers must be partitioned such that $w_t = \{w_{1,t}, w_{2,t}\}$ in order to obtain the two sets of independently distributed processes, ε_t and ς_t.

4. Using (2), simulate the FIGARCH model

$$r_t(s) = \left\{ \frac{\omega^{(0)}}{1-\beta_1^{(0)}-\ldots-\beta_p^{(0)}} + \left[1 - \frac{(\pi_1^{(0)}L+\ldots+\pi_{q^*}^{(0)}L^{q^*})(1-L)^{d^{(0)}}}{(\beta_1^{(0)}L+\ldots+\beta_p^{(0)}L^p)} \right] \varepsilon_t^2 \right\}^{1/2} \varepsilon_t \tag{44}$$

where $\varepsilon_t = \sigma_\varepsilon^{(0)} w_t$, or the LMSV model

$$r_t(s) = \sigma_\varepsilon^{(0)} \exp\left\{ \frac{(1-L)^{-d^{(0)}} \left(1+\theta_1^{(0)}L+\cdots+\theta_q^{(0)}L^q\right) \varsigma_t}{2\left(1-\phi_1^{(0)}L-\cdots-\phi_p^{(0)}L^p\right)} \right\} \varepsilon_t \tag{45}$$

where $\varepsilon_t = \sigma_\varepsilon^{(0)} w_{1,t}$ and $\varsigma_t = \sigma_\varsigma^{(0)} w_{2,t}$.

5. Estimate the auxiliary models (37) and (40) for the lag length k^* using the simulated data $r_t(s)$, and calculate $\mathbf{\Pi}(s)$.

6. Repeat steps 4 and 5, $h = 1, 2, \ldots, H$ times.

7. Calibrate the parameter vector $\mathbf{\Theta}_1^{(0)}$ for the FIGARCH model or $\mathbf{\Theta}_2^{(0)}$ for the LMSV model to satisfy the GMR indirect estimation criterion in (33).

4 Monte Carlo Experiments

The asymptotic properties of consistency and normality of the indirect estimator follow directly from the results in Gourieroux, Monfort and Renault (1993). The performance of the indirect estimator for the long memory volatility models in finite samples is however unknown.

4.1 Simulation Design

The sample size for the Monte Carlo experiments is set at $T = 1000$ observations, a sample size not uncommon in applied financial economics and $R = 1000$ replications are performed. In simulating each series, $T + l = 6000$ observations are generated and the first $l = 5000$ observations are then truncated to ensure that there are no starting value dependencies in the simulated series. The indirect estimator utilises $H = 2$ and $H = 10$ simulation paths. In all cases, the dimension of the parameter vector for the auxiliary model is the same as the dimension of the parameter vector for the direct model, allowing $\mathbf{\Omega} = \mathbf{I}$ to be used in (33).

Two first order FIGARCH models are considered. The DGP for the FIGARCH$(1, d, 1)$ model is (10) and the DGP for the FIGARCH$(1, d, 0)$ model is (12). The indirect results for the FIGARCH$(1, d, 0)$ model are based on an AR(2) auxiliary model for the squared errors according to that in (37) while the indirect results for the FIGARCH$(1, d, 1)$ model are based on an AR(3) auxiliary model for the squared errors. Both auxiliary models contain a constant term. The maximum likelihood procedure uses a truncation parameter of $T/5 = 200$ observations; the sensitivity of the TDML estimator on this truncation parameter is examined later. The restrictions in (16) to (18) are imposed in the estimation routines for both the TDML and indirect estimators. The DGP parameter values for each FIGARCH model are reported in Tables 1 and 2. For each DGP, six parameter sets are investigated.

Two first order LMSV models are considered. The DGP for the LMSV$(0, d, 0)$ model is (28) and the DGP for the LMSV$(1, d, 1)$ model is (29). The indirect results for the LMSV$(0, d, 0)$ model are based on an AR(1) auxiliary model for the natural logarithm of the squared errors according to that in (40) while the indirect results for the LMSV$(1, d, 1)$ model are based on an AR(3) auxiliary model. The FDML estimator uses the likelihood function in (30) specified using (27). The DGP parameter values for each LMSV model are reported in Tables 4 and 5.

4.2 Comparison of Alternative Estimators: The FIGARCH Model

The bias and root mean square error (RMSE) of the TDML and indirect estimators for the FIGARCH DGP are reported in Tables 1 and 2. Across the two tables and the 12 DGP specifications, the bias levels for the estimators are mostly comparable; the TDML estimator is consistently the more efficient however. Despite this, the TDML estimator does struggle a few times, for example the bias levels for the $d = 0.9, \omega = 0.4, \beta_1 = 0.3, \pi_1 = 0.2$ specification are high, but the indirect estimator appears to struggle more. The general performance of the indirect estimator seems to improve as H increases from 2 to 10. However, for the indirect estimator the results in these tables are not strong. The TDML estimator dominates in terms of bias and RMSE for nearly every specification. Further investigation into auxiliary models and simulation paths is necessary to establish the indirect estimator's optimal properties.

Table 3 reports the impact of the truncation parameter on the TDML estimator for the $T = 1000$ sample. Four truncation parameters are examined, ranging from $m = \sqrt{T} \approx 32$ to $m = T/4 \approx 250$, and three DGPs are utilized from Table 1. The results suggest the different truncation parameters used in the TDML estimator do not uniformly affect the finite sample properties of the estimator. It might be expected that as m increases bias may fall and RMSE rise, since the smaller estimation sample will reduce efficiency but avoid any starting value problems biasing the estimates. Table 3 suggests, at least for the sample size being considered, that the TDML estimator is reasonably robust to the values of m chosen.

4.3 Comparison of Alternative Estimators: The LMSV Model

The small sample properties of the FDML and indirect estimators for the LMSV model are reported in Tables 4 and 5. Five values of d are considered for the LMSV$(0, d, 0)$ DGP in Table 4. The FDML estimator has superior finite sample properties; with RMSEs that are particularly low relative to the indirect estimator. The performance of the indirect estimator does appear to improve as H goes from 2 to 10. Table 5 suggests the performance of the indirect estimator improves relative to the FDML estimator for the slightly more complicated LMSV$(1, d, 1)$ DGP. Several of the RMSEs for the two estimators are reasonably close and the indirect estimator has a better bias estimate for a couple of the specifications; although this would not appear to be a strong result. Again, the performance of the indirect estimator improves as H goes from 2 to 10. Overall, it would appear the indirect estimator performs relatively better for the LMSV DGP than for the FIGARCH DGP.

5 Conclusion

The development, estimation and testing for long memory volatility models is a rapidly growing research area for econometricians and applied financial economists. Fractionally integrated models of conditional variance are important because of their empirical relevance and the flexibility they offer over the relatively rigid nature of integer restricted GARCH, IGARCH and SV models. Baille (1996) remarks that comparison of FIGARCH and LMSV models is a promising area for future research. However, as with fractionally integrated models for the conditional mean, the difficulties associated with estimation threaten to inhibit their widespread application. This paper has addressed this issue by extending the indirect estimator for ARFIMA models in Martin and Wilkins (1999) to fractionally integrated GARCH and fractionally integrated SV models. Simulation results comparing maximum likelihood and indirect estimators suggests there is scope for further investigation into alternative indirect estimator specifications.

References

[1] Baillie, R.T. (1996) Long memory processes and fractional integration in econometrics. **Journal of Econometrics**, 73, 5-59.

[2] Baillie, R.T., T. Bollerslev and H.O. Mikkelson (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. **Journal of Econometrics**, 74, 3-30.

[3] Bollerslev, T. (1986) Generalised autoregressive conditional heteroskedasticity. **Journal of Econometrics**, 31, 307-327.

[4] Bollerslev, T., R.Y. Chou and K.F. Kroner (1992) ARCH modelling in finance. **Journal of Econometrics**, 52, 5-59.

[5] Bollerslev, T. and R.F. Engle (1993) Common persistence in conditional variances. **Econometrica**, 61(1), 167-186.

[6] Bollerslev, T., R.F. Engle and D.B. Nelson (1994) ARCH models. In R.F. Engle and D. McFadden, **The Handbook of Econometrics 4**, North-Holland.

[7] Bollerslev, T. and H.O. Mikkelson (1996) Modeling and pricing long memory in stock market volatility. **Journal of Econometrics**, 73, 151-184.

[8] Bollerslev, T. and J.M. Wooldridge (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying conditional variances. **Econometric Reviews**, 11, 143-172.

[9] Breidt, F.J., N. Crato and P.J.F. de Lima (1998) The detection and estimation of long memory in stochastic volatility. **Journal of Econometrics**, 83, 325-348.

[10] Ding, Z. and C.W.J. Granger (1996) Modeling volatility persistence of speculative returns: A new approach. **Journal of Econometrics**, 73, 185-215.

[11] Ding, Z., C.W.J. Granger and R.F. Engle (1993) A long memory property of stock market returns and a new model. **Journal of Empirical Finance**, 1, 83-106.

[12] Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. **Econometrica**, 50, 987-1008.

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

[14] 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.

[15] Gallant, R. and G. Tauchen (1996) Which moments to match? **Econometric Theory**, 12, 657-681.

[16] 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.

[17] Gourieroux, C.A., A. Monfort and E. Renault (1993) Indirect inference. **Journal of Applied Econometrics**, 8, s85-s118.

[18] 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(1), 15-29.

[19] Harvey, A.C. (1993) Long memory in stochastic volatility. Working Paper.

[20] Hosking, J. (1981) Fractional differencing. **Biometrika**, 68, 165-176.

[21] Martin, V.L. and N.P. Wilkins (1999) Indirect estimation of ARFIMA and VARFIMA models. **Journal of Econometrics**, 93, 149-175.

[22] Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. **Econometrica**, 59, 347-370.

[23] Psaradakis, Z. and M. Sola (1995) Modelling long memory in stock market volatility: A fractionally integrated generalised ARCH approach. **Discussion Paper No. DP 12-95, London Business School**.

[24] Sowell, F. (1992a) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. **Journal of Econometrics**, 53, 165-188.

[25] Sowell, F. (1992b) Modeling long-run behaviour with the fractional ARIMA model. **Journal of Monetary Economics**, 29, 277-302.

[26] Taylor, S. (1994) Modelling stochastic volatility: A review and comparative study. **Mathematical Finance**, 4, 183-204.

[27] Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing. **Econometric Theory**, 2, 107-131.

Table 1: Small Sample Properties of the FIGARCH$(1, d, 0)$ Estimators

DGP	TDML *Bias*	TDML *RMSE*	Indirect: $H = 2$ *Bias*	Indirect: $H = 2$ *RMSE*	Indirect: $H = 10$ *Bias*	Indirect: $H = 10$ *RMSE*
$d = 0.5$	0.063	0.177	0.118	0.250	-0.045	0.152
$\omega = 0.4$	-0.119	0.160	-0.052	0.289	0.145	0.292
$\beta = 0.3$	0.051	0.178	0.153	0.273	-0.005	0.161
$d = 0.7$	0.050	0.128	-0.017	0.231	-0.051	0.209
$\omega = 0.4$	-0.051	0.118	0.028	0.278	0.080	0.266
$\beta = 0.3$	0.043	0.150	0.119	0.262	0.009	0.279
$d = 0.7$	0.065	0.180	0.010	0.167	-0.099	0.187
$\omega = 0.4$	-0.086	0.157	0.037	0.254	0.151	0.251
$\beta = 0.5$	0.057	0.165	0.052	0.141	-0.070	0.192
$d = 0.9$	0.000	0.067	-0.205	0.305	-0.202	0.298
$\omega = 0.4$	-0.001	0.089	0.146	0.299	0.063	0.258
$\beta = 0.3$	-0.004	0.098	-0.045	0.271	-0.007	0.276
$d = 0.9$	0.003	0.091	-0.195	0.266	-0.201	0.228
$\omega = 0.4$	-0.004	0.116	0.192	0.306	0.034	0.165
$\beta = 0.5$	-0.002	0.113	-0.142	0.279	0.005	0.117
$d = 0.9$	-0.002	0.116	-0.203	0.260	-0.153	0.177
$\omega = 0.4$	-0.007	0.140	0.244	0.277	0.194	0.225
$\beta = 0.7$	-0.008	0.110	-0.145	0.219	-0.122	0.150

Details of the DGP are given in Section 4.1.

Table 2: Small Sample Properties of the FIGARCH$(1, d, 1)$ Estimators

	TDML		Indirect: $H = 2$		Indirect: $H = 10$	
DGP	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
$d = 0.5$	0.092	0.193	-0.077	0.252	-0.085	0.171
$\omega = 0.4$	-0.081	0.154	0.118	0.270	0.114	0.274
$\beta_1 = 0.3$	0.128	0.228	0.077	0.246	0.143	0.280
$\pi_1 = 0.2$	-0.076	0.159	0.158	0.314	0.072	0.199
$d = 0.7$	0.041	0.123	-0.184	0.391	-0.197	0.289
$\omega = 0.4$	0.007	0.139	0.162	0.189	0.055	0.276
$\beta_1 = 0.3$	-0.062	0.208	0.196	0.351	0.177	0.299
$\pi_1 = 0.2$	-0.102	0.154	0.238	0.382	0.092	0.270
$d = 0.7$	0.097	0.170	-0.122	0.250	-0.205	0.229
$\omega = 0.4$	-0.057	0.170	0.178	0.269	0.222	0.255
$\beta_1 = 0.5$	0.040	0.195	-0.066	0.223	-0.021	0.165
$\pi_1 = 0.2$	-0.055	0.125	0.142	0.309	0.271	0.352
$d = 0.9$	-0.009	0.079	-0.316	0.367	-0.394	0.420
$\omega = 0.4$	0.056	0.119	0.033	0.275	0.023	0.220
$\beta_1 = 0.3$	-0.115	0.189	-0.053	0.278	0.027	0.263
$\pi_1 = 0.2$	-0.114	0.155	0.294	0.368	0.272	0.334
$d = 0.9$	0.002	0.091	-0.196	0.307	-0.232	0.261
$\omega = 0.4$	0.017	0.139	0.176	0.267	-0.041	0.141
$\beta_1 = 0.5$	-0.042	0.169	-0.087	0.281	-0.107	0.263
$\pi_1 = 0.2$	-0.060	0.121	0.134	0.286	0.352	0.430
$d = 0.9$	0.014	0.113	-0.171	0.238	-0.224	0.268
$\omega = 0.4$	-0.003	0.153	0.190	0.298	0.256	0.313
$\beta_1 = 0.7$	-0.010	0.127	-0.058	0.187	-0.057	0.193
$\pi_1 = 0.2$	-0.029	0.088	0.043	0.187	0.253	0.379

Details of the DGP are given in Section 4.1.

Table 3: Impact of Truncation on the TDML Estimator for the FIGARCH$(1, d, 0)$ DGP with $T = 1000$ Observations

	$m = \sqrt{T} \approx 32$		$m = T/10 = 100$		$m = T/5 = 200$		$m = T/4 = 250$	
DGP	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
$d = 0.5$	-0.009	0.115	0.042	0.144	0.063	0.177	0.058	0.182
$\omega = 0.4$	0.109	0.202	-0.056	0.138	-0.119	0.160	-0.132	0.170
$\beta = 0.3$	-0.005	0.121	0.038	0.145	0.051	0.178	0.050	0.186
$d = 0.7$	-0.011	0.129	0.044	0.150	0.065	0.180	0.089	0.173
$\omega = 0.4$	0.085	0.203	-0.036	0.156	-0.086	0.157	-0.112	0.170
$\beta = 0.5$	-0.010	0.133	0.033	0.150	0.057	0.165	0.079	0.167
$d = 0.9$	-0.027	0.115	-0.013	0.115	-0.002	0.116	-0.001	0.118
$\omega = 0.4$	0.057	0.194	0.019	0.155	-0.007	0.140	-0.013	0.151
$\beta = 0.7$	-0.022	0.107	-0.017	0.107	-0.008	0.110	-0.008	0.118

Details of the DGP are given in Section 4.1.

Table 4: Small Sample Properties of the LMSV$(0, d, 0)$ Estimators

	FDML		Indirect: $H = 2$		Indirect: $H = 10$	
DGP	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
$d = 0.1$	0.031	0.077	0.074	0.148	0.061	0.133
$d = 0.3$	-0.002	0.063	-0.007	0.139	-0.020	0.138
$d = 0.5$	0.009	0.047	-0.037	0.147	-0.047	0.148
$d = 0.7$	0.010	0.042	-0.015	0.085	-0.014	0.072
$d = 0.9$	0.017	0.036	-0.012	0.058	-0.007	0.046

Details of the DGP are given in Section 4.1.

Table 5: Small Sample Properties of the LMSV$(1, d, 1)$ Estimators

	FDML		Indirect: $H = 2$		Indirect: $H = 10$	
DGP	*Bias*	*RMSE*	*Bias*	*RMSE*	*Bias*	*RMSE*
$d = 0.2$	-0.028	0.090	-0.029	0.123	-0.031	0.103
$\phi_1 = 0.4$	0.050	0.119	0.048	0.132	0.047	0.121
$\theta_1 = 0.8$	-0.042	0.133	-0.070	0.202	-0.065	0.164
$d = 0.2$	-0.020	0.095	0.050	0.121	0.003	0.106
$\phi_1 = 0.6$	0.017	0.094	-0.012	0.135	-0.008	0.121
$\theta_1 = 0.8$	-0.007	0.118	-0.033	0.163	-0.031	0.136
$d = 0.4$	-0.021	0.092	-0.067	0.155	-0.059	0.144
$\phi_1 = 0.4$	0.046	0.129	0.095	0.199	0.090	0.183
$\theta_1 = 0.8$	-0.054	0.149	-0.084	0.201	-0.093	0.181
$d = 0.4$	-0.019	0.109	-0.023	0.108	-0.024	0.115
$\phi_1 = 0.6$	0.016	0.106	0.036	0.140	0.035	0.139
$\theta_1 = 0.8$	-0.008	0.116	-0.054	0.170	-0.058	0.163

Details of the DGP are given in Section 4.1.