Financial Econometrics
Lecture 7: Long Memory and Fractional Integration

Richard G. Pierse

1 Introduction

One observed property of many financial data series is that they appear to have *long memory*, either in mean or in variance. This means that the effect of shocks on financial time series take a very long time to disappear. One way to model such behaviour is through *fractionally integrated* time series processes that lie conceptually somewhere between *I(0)* and *I(1)* processes. Fractionally integrated models may be either stationary or non-stationary. Even when they are weakly stationary however, fractionally integrated processes have autocorrelation functions that decay very slowly and so exhibit long memory. Fractionally integrated processes have been applied both to *ARMA* models leading to *ARFIMA* models and to models of conditional volatility to lead to fractionally integrated *GARCH* and fractionally integrated stochastic volatility models. Long memory processes are reviewed in Robinson (1994) and Baillie (1996).

2 Fractional Integration

Consider the model

$$\Delta^d y_t = (1-L)^d y_t = u_t \tag{2.1}$$

where u_t is an independent identically distributed (iid) disturbance with $\mathrm{E}(u_t) = 0$ and $\mathrm{var}(u_t) = \sigma^2$. When $d = 0$, then y_t is simply white noise and is *weakly stationary*. When $d = 1$ then y_t is a random walk process and is *non-stationary*. In this case we say that y_t is integrated of order 1, and write $y_t \sim I(1)$. We now want to consider the properties of (2.1) for *fractional* values of d lying between 0 and 1.

2.1 The Binomial Theorem

What is the meaning of fractional values of d? The general version of Newton's *Binomial Theorem* allows the expression $(1-L)^d$ to be expanded as an infinite series for any *real* value of $d > -1$. The expansion is given by

$$\Delta^d = (1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k \tag{2.2}$$

where the binomial coefficients $\binom{d}{k}$ are defined by

$$\binom{d}{k} = \frac{d(d-1)(d-2)\cdots(d-k+1)}{k!}$$

and $k!$ (read as k *factorial* or k *shriek*) is

$$k! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot (k-1) \cdot k$$

with $0! = 1$.

Using these definitions, we can rewrite (2.2) as

$$(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \cdots \, . \tag{2.3}$$

For integral d, the coefficients will be zero for $k > d$ and so the infinite series is truncated at order d. Thus, for example, we have

$$(1-L)^2 = 1 - 2L + L^2$$

and

$$\begin{aligned}(1-L)^5 &= 1 - 5L + \frac{20}{2}L^2 - \frac{60}{6}L^3 + \frac{120}{24}L^4 - \frac{120}{120}L^5 \\ &= 1 - 5L + 10L^2 - 10L^3 + 5L^4 - L^5.\end{aligned}$$

For fractional d, $0 < d < 1$, the infinite series will not be truncated. For example

$$(1-L)^{0.5} = 1 - 0.5L + \frac{-0.25}{2}L^2 - \frac{.375}{6}L^3 + \frac{-0.9375}{24}L^4 - \frac{3.28125}{120}L^5 + \cdots$$

which is an infinite lag expansion. Depending on the value of d, the fractionally differenced series $(1-L)^d y_t$ may be stationary or non-stationary. It turns out that the condition for stationarity is that $d < 0.5$. Thus for the example case $d = 0.5$, the sum of the infinite series of coefficients, 1, -0.5, -0.125, -0.0625, -0.0390625, -0.02734375 $, \cdots$, does not converge fast enough.

2.2 Stationary Fractionally Integrated Processes

It can be shown that, for $d < 0.5$, the process (2.1) is *weakly stationary* and has an *infinite order MA (Wold)* representation given by

$$y_t = (1 - L)^{-d} u_t = \sum_{i=0}^{\infty} \theta_i u_{t-i} \tag{2.4}$$

where

$$\theta_k = \frac{d(1+d)(2+d)\cdots(k-1+d)}{k!}.$$

If $d > -0.5$ then the *MA* process (2.4) is *invertible* and y_t has the alternative infinite order *AR* representation

$$y_t = \sum_{i=1}^{\infty} \phi_i y_{t-i} + u_t \tag{2.5}$$

where

$$\phi_k = \frac{-d(1-d)(2-d)\cdots(k-1-d)}{k!}.$$

A comparison between the autocorrelation function of a stationary fractionally integrated model and that of a simple *short-memory AR(1)* process can be illustrated by an example.

Comparison of ACF of fractionally integrated (FI) and AR(1) model.
Source: Campbell, Lo and MacKinlay (1997)

Lag	*FI* ($d = \frac{1}{3}$)	*AR(1)* $\phi = 0.5$
1	0.500	0.500
2	0.400	0.250
3	0.350	0.125
4	0.318	0.063
5	0.295	0.031
10	0.235	0.001
25	0.173	2.98×10^{-8}
50	0.137	8.88×10^{-16}
100	0.109	7.89×10^{-31}

The table makes clear the big difference between the *geometric decay* of the *ACF* coefficients in the *AR(1)* model and the *hyperbolic decay* of the (stationary) *fractionally integrated* (*FI*) model. After *100* periods, the *FI* model still exhibits large autocorrelation while in the *AR(1)* model it is completely negligible.

3 *ARFIMA* models

ARIMA models can be generalised to incorporate fractional integration, leading to the *ARFIMA(p,d,q)* model:

$$\phi(L)\Delta^d y_t = \theta(L)u_t \tag{3.1}$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$

and

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

and where d now represents the order of fractional differencing.

The model can be rewritten as

$$\Delta y_t = \Delta^{1-d}\phi(L)^{-1}\theta(L)u_t = A(L)u_t$$

and, since $\Delta^{1-d} = 0$ for $d < 1$, any *ARFIMA* model is *trend-reverting* in the sense that shocks will have no permanent effect and $A(1) = 0$.

The *ACF* function of *ARFIMA* models will exhibit slow decay at a hyperbolic rate. For large s, the *ACF* can be approximated by

$$\rho_s \simeq \kappa s^{2d-1} \tag{3.2}$$

where κ is a constant. For $d < 0.5$, the exponent $2d - 1 < 0$ so the correlations eventually decay, but at a slow hyperbolic rate compared with the fast geometric decay in standard *ARMA* models.

3.1 Estimating d

It is possible to estimate the fractional differencing parameter d. One crude estimation method is based on the approximate formula for the *ACF* function (3.2). Taking logarithms of the absolute value of ρ_s,

$$\log|\rho_s| \simeq \log\kappa + (2d-1)\log s$$

for large s, and $2d - 1$ can be estimated by the coefficient β in the regression

$$\log|\widehat{\rho}_i| = \alpha + \beta\log i + u_i.$$

so that $\widehat{d} = (\widehat{\beta} + 1)/2$. This estimator is not very accurate however and its asymptotic properties are difficult to derive. They are discussed in Geweke and Porter-Hudak (1982).

A more satisfactory approach is to estimate d as part of the joint maximum likelihood estimation of the parameters of the *ARFIMA* model. The log-likelihood of the *ARFIMA* model is given by

$$\log L(\mathbf{y}; d, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\Sigma}) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathbf{y}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$$

where

$$\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{y})$$

is the variance covariance matrix of $\mathbf{y}$ which is a (complicated) function of the unknown model parameters d, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$.

This approach is developed by Sowell (1992) and is adopted in the estimation package *PcGive* by Doornik and Hendry (2001).

4 Long memory volatility models

Long memory in the conditional variance of financial data series is even more prevalent than long memory in the mean. This has led to the development of fractionally integrated versions of both the *GARCH* and the stochastic volatility model.

4.1 The *FIGARCH* model

Baillie, Bollerslev and Mikkelson (1996) incorporated fractional integration into the *GARCH* model to develop the *FIGARCH* model. Recall from lecture 2 that the *GARCH(1,1)* model

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

can be rewritten as the *ARMA* process

$$u_t^2 = \alpha_0 + (\alpha_1 + \beta_1)u_{t-1}^2 + \varepsilon_t - \beta_1 \varepsilon_{t-1}$$

where $\varepsilon_t = u_t^2 - \sigma_t^2$. For the case of the *FIGARCH(1,d,1)* model, by analogy, we have

$$\Delta^d u_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\Delta^d u_{t-1}^2 + \varepsilon_t - \beta_1 \varepsilon_{t-1},$$

which can be rewritten as

$$\sigma_t^2 = \alpha_0 + (1 - \Delta^d)u_t^2 - (\beta_1 - (\alpha_1 + \beta_1)\Delta^d)u_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{4.1}$$

Conditions that ensure that the conditional variance is always positive are that $\alpha_0 > 0$, $\alpha_1 + d \geq 0$, and $1 - 2(\alpha_1 + \beta_1) \geq d \geq 0$.

Baillie *et al.* argue that the *FIGARCH* model is preferable to the *IGARCH* model that is often used in *GARCH* modelling with high frequency data. The *IGARCH* model implies the implausible and counterfactual result that shocks persist indefinitely. The *FIGARCH* model avoids this problem, while at the same time allowing for shocks to have long memory.

4.2 A long memory stochastic volatility model

Breidt, Crato and de Lima (1998) develop a simple stochastic volatility model with long memory. The model is defined by

$$y_t = \sigma_t u_t \tag{4.2}$$

where the volatility, σ_t, follows the process

$$\sigma_t^2 = \phi_1^2 \exp(\varepsilon_t) \tag{4.3}$$

and

$$(1-L)^d \varepsilon_t = \eta_t \tag{4.4}$$

where u_t and η_t are independent, normally distributed white noise processes with unit variance and $0 < d < 0.5$.

Squaring (4.2) and taking logarithms,

$$\begin{aligned} \log(y_t^2) &= \log \sigma_t^2 + \log u_t^2 \\ &= \log \phi_1^2 + \varepsilon_t + \log u_t^2 \\ &= (\log \phi_1^2 + \mathrm{E}(\log u_t^2)) + \varepsilon_t + (\log u_t^2 - \mathrm{E}(\log u_t^2)) \end{aligned}$$

or

$$\log(y_t^2) = \mu + \varepsilon_t + e_t \tag{4.5}$$

where $e_t = \log u_t^2 - \mathrm{E}(\log u_t^2)$ is a mean-zero disturbance term which is independent of η_t but which has a non-normal distribution. Equations (4.5) and (4.4) represent a model which can be estimated by quasi-maximum likelihood. Estimating this model on squared daily returns on companies in the S&P 500 index, Bollerslev and Jubinski (1999) found a median estimate of d of about 0.38.

5 Testing for fractional integration

5.1 The Range over Standard Deviation test

A test for fractional integration was originally proposed in the engineering literature by Hurst (1951) and applied to financial economics by Mandelbrot (1972).

This is called the '*range over standard deviation*' or *R/S* statistic and is defined by

$$R_0 = \widehat{\sigma}_0^{-1}\{\max_i \sum_{t=1}^{i}(y_t - \overline{y}) - \min_i \sum_{t=1}^{i}(y_t - \overline{y})\} \tag{5.1}$$

where

$$\widehat{\sigma}_0^2 = \frac{\sum_{t=1}^{T}(y_t - \overline{y})^2}{T}$$

is an estimate of the sample variance. The term in braces in (5.1) is called the range and is the difference between the largest and smallest partial sums of mean deviations of y_t. Since the sum of mean deviations over all observations is zero, this difference will always be non-negative.

One problem with the classic *R/S* statistic is that it is sensitive to the short-run dependence that may be present in any *I(0)* process, even in the absence of fractional integration. In view of this, Lo (1991) suggested a non-parametric correction to (5.1) leading to a *modified R/S* statistic

$$R_q = \widehat{\sigma}_q^{-1}\{\max_i \sum_{t=1}^{i}(y_t - \overline{y}) - \min_i \sum_{t=1}^{i}(y_t - \overline{y})\} \tag{5.2}$$

and

$$\widehat{\sigma}_q^2 = \widehat{\sigma}_0^2(1 + \frac{2}{T}\sum_{j=1}^{q} w_{qj}\widehat{\rho}_j)$$

where the w_{qj} are triangular weights defined by

$$w_{qj} = 1 - \frac{j}{q+1}$$

and $\widehat{\rho}_j$ is the sample autocorrelation coefficient defined by

$$\widehat{\rho}_j = \frac{\sum_{t=j+1}^{T}(y_t - \overline{y})(y_{t-j} - \overline{y})}{\sum_{t=1}^{T}(y_t - \overline{y})^2}.$$

The correction in (5.2) will pick up any short-run dependence in the series y_t so that the test will have power to distinguish fractionally integrated processes from other stationary processes. The appropriate choice of q is left to the user.

On the null hypothesis that $d = 0$, the statistic $T^{-1/2}R_q$ converges asymptotically to a well defined random variable for which critical values are tabled in Lo (1991). Lo shows that this test is consistent against all alternatives in the class of long-memory *ARFIMA(p,d,q)* models where $-0.5 < d < 0.5$. However, if the process y_t exhibits *excess kurtosis* and so is fat-tailed, then there is evidence that the test will reject too often against alternatives with $d < 0$ and not often enough against the long-memory alternatives where $d > 0$.

5.2 A Lagrange Multiplier test

Agiakloglou and Newbold (1994) derive a Lagrange Multiplier test of the null hypothesis that $d = 0$ from the residuals $\widehat{u}_t$ from the estimated *ARIMA(p,0,q)* model

$$\widehat{\phi}(L)y_t = \widehat{\theta}(L)\widehat{u}_t. \tag{5.3}$$

The test is based on the *t-ratio* on the coefficient δ from the auxiliary regression

$$\widehat{u}_t = \sum_{i=1}^{p} \beta_i w_{t-i} + \sum_{i=1}^{q} \gamma_i v_{t-i} + \delta z_t + \eta_t$$

where

$$\widehat{\theta}(L)w_t = y_t$$
$$\widehat{\theta}(L)v_t = \widehat{u}_t$$

and

$$z_t = \sum_{j=1}^{m} \frac{1}{j}\widehat{u}_{t-j}.$$

Agiakloglou and Newbold find that, in finite samples, the power of their test decreases when p or q are greater than zero, or when the process (5.3) has a non-zero mean.

5.3 Tests against $d = 1$

It is important to be able to test between the integrated model with $d = 1$ and the fractionally integrated models with $0 < d < 1$. In fact, the standard *Augmented Dickey-Fuller* tests of the null hypothesis that $d = 1$ are consistent against the alternative of a fractional d. However, as might be expected, the tests will have less power against the alternative of a fractional d than against the more usual alternative of an *I(0)* model with autocorrelated errors.

References

[1] Agiakloglou, C. and P. Newbold (1994), 'Lagrange multiplier tests for fractional difference', *Journal of Time Series Analysis*, 15, 253–262.

[2] Baillie, R.T. (1996), 'Long memory processes and fractional integration in econometrics', *Journal of Econometrics*, 73, 5–59.

[3] Baillie, R.T., T. Bollerslev and H.O. Mikkelson (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics*, 74, 3–30.

[4] Bollerslev, T. and D. Jubinski (1999), 'Equality trading volume and volatility: latent information arrivals and common long-run dependencies', *Journal of Business and Economic Statistics*, 17, 9–21.

[5] Breidt, F.J., N. Crato and P. de Lima (1998), 'On the detection and estimation of long memory in stochastic volatility', Journal of Econometrics, 83, 325–348.

[6] Campbell, J.Y., A.W. Lo and A. C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

[7] Doornik, J. A. and D.F. Hendry (2001), *PcGive 10: An Interactive Econometric Modelling System*, Timberlake Publishing, West Wickham, UK.

[8] Geweke, J. and S. Porter-Hudak (1982), 'The estimation and application of long memory time series models', *Journal of Time Series Analysis*, 4, 221–238.

[9] 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, 15–29.

[10] Hosking, J.R.M. (1981), 'Fractional differencing', *Biometrika*, 68, 165–176.

[11] Hurst, H. (1951), 'Long term storage capacity of reservoirs', *Transactions of the American Society of Civil Engineers*, 116, 770–799.

[12] Lo, A.W. (1991), 'Long-term memory in stock market prices', *Econometrica*, 59, 1279–1313.

[13] Mandelbrot, B.B. (1972), 'Statistical methodology for non-periodic cycles: from the covariance to R/S analysis', *Annals of Economic and Social Measurement*, 1, 259–290.

[14] Robinson, P. (1994), 'Time series with strong dependence', in C.A. Sims (ed.) *Advances in Econometrics: Sixth World Congress, Vol 1*, Cambridge University Press, Cambridge, UK.

[15] Sowell, F.B. (1992), 'Maximum likelihood estimation of stationary univariate fractionally integrated time series models', *Journal of Econometrics*, 53, 165–188.