MAXIMUM LIKELIHOOD ESTIMATION OF FRACTIONALLY INTEGRATED TIME SERIES MODELS†

Fallaw Sowell

Carnegie Mellon University

October 1987

Revised March 1989

ABSTRACT

To estimate the parameters of a stationary fractionally integrated vector time series with multiple differencing parameters, the unconditional exact likelihood function is derived. This allows the simultaneous estimation of all the parameters of the model by maximum likelihood. Attention is given to the recursive formulas that allow efficient evaluation of the likelihood function. Limitations and problems of previous estimation procedures for fractionally integrated models are also presented.

Fallaw Sowell

GSIA

Carnegie Mellon University

Pittsburgh, PA 15213

I. INTRODUCTION

Many economic time series appear to satisfy conditions which are inconsistent with the assumptions of standard econometric models. This has been the conclusion of recent papers by Gourieroux, Maurel, and Monfort (1987) (multivariate) and Sowell (1989) (univariate). Both papers consider a linear regression model where the underlying series are fractionally integrated. The results show that familiar asymptotic distributions are not achieved when series are fractionally integrated. These asymptotic results underscore the importance of being able to estimate fractionally integrated time series models. The current paper derives the likelihood function for fractionally integrated vector time series models. This allows the estimation of the parameters of the model by unconditional maximum likelihood. Previous estimation procedures for fractionally integrated models are reviewed and their limitations and deficiencies are noted.

Support for the idea that aggregate macroeconomic data are generated from fractionally integrated models exists. On a theoretical level, Granger (1980) showed that the aggregate, of a particular class of individual AR(1) models, follows a fractionally integrated process. There is also a large collection of applied studies that present evidence that many economic time series follow fractionally integrated models including: consumer price index, whole sale price index (Geweke and Porter-Hudak (1982)); stock market returns (Greene and Fielitz (1977)); exchange rates (Booth et al (1982)); daily futures prices for soybeans, soybean oil and soybean meal (Helms et al (1984)); yield on short term treasury bills (Shea (1987)). These studies were able to estimate a fractional differencing parameter for the given time series. Unfortunately, the other parameters of the model were not accurately estimated. The inability to accurately estimate all the parameters in a fractionally

integrated model is a problem common to several estimation techniques, which is alleviated by maximum likelihood.

The next section is an introduction to fractionally integrated models. Section three derives the likelihood function for a fractionally integrated vector time series when the errors are normal. Special case of the general model are presented in the fourth section. The fifth section outlines the maximization of the likelihood function to obtain maximum likelihood estimates of all the parameters of the model. Particular attention is given to efficient and recursive procedures. The final section contains a summary and notes directions for future research.

II. AN INTRODUCTION TO FRACTIONAL MODELS

The general differencing operator can be defined for $d \geq -1/2$ by

$$(1-L)^d \equiv (1-L)^{d-[d+1/2]}(1-L)^{[d+1/2]} \tag{1}$$

where $[\xi]$ is the greatest integer function. In the general definition of $(1-L)^d$, the first exponent is between -1/2 and 1/2, while the second exponent is an integer. For $d \in [-1/2,1/2]$, the fractional difference operator $(1-L)^d$ is defined by its Maclaurin series (Binomial Theorem) to be

$$(1-L)^d \equiv \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j L^j \tag{2}$$

where $\binom{d}{j}(-1)^j = \frac{\Gamma(d+1)(-1)^j}{\Gamma(d-j+1)\Gamma(j+1)} = \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)}$.

Because $1/\Gamma(\xi)$ is bounded and has roots at the nonpositive integers (see Gradshteyn and Ryzhik (1980) notes to equation (8.310.2)) the sum defining $(1-L)^d$ has a finite number of nonzero terms for $d \in [-1/2,1/2]$ and $d \neq 0$.

Extending the standard notation $z_t \sim I(d)$ could also be written $z_t \sim ARIMA(p,d,q)$ where p and q are finite nonnegative integers and the roots of the AR and MA polynomials are outside the unit circle. The characteristics of this univariate model are presented in Granger and Joyeux (1980), Hosking (1981) and Sowell (1986).

The multivariate generalization would be z_t a kx1 vector time series such that

$$\Theta(L)D(L)Vz_t=\Phi(L)\varepsilon_t \tag{3}$$

where $\Theta(L)$ and $\Phi(L)$ are kxk matrix polynomials in the lag operator L. It will be assumed that $D(L)=DIAG[(1-L)^{d_1}, (1-L)^{d_2},\ldots,(1-L)^{d_k}]$, $\Theta(L)$ is of order p, $\Phi(L)$ is of order q, $\Theta(0)=\Phi(0)=I_k$, the roots of $|\Theta(\xi)|$ and $|\Phi(\xi)|$ are outside the unit circle and $\varepsilon_t \sim IIDN_k(0,\Sigma)$. The constant kxk matrix V is nonsingular. The simple form of the differencing matrix, D(L), means that the characteristics of the fractional z_t vector series stated below can be obtained by the univariate proofs applied element by element. In particular:

a) z_t is stationary if $d_i<1/2$ for $i=1,2,\ldots,k$.

b) z_t possess an invertible moving average representation if $d_i>-1/2$.

c) If the spectral density of z_t is denoted $f_z(\lambda)$ then as $\lambda \to 0$ $f_z(\lambda) \sim [\kappa_{ij}\lambda^{-(d_i+d_j)}]$ where each κ_{ij} is a constant and is independent of d_i and d_j.

d) If the autocovariances of z_t are denoted $\gamma_z(s)=E[x_t x_{t-s}']$, then as $s \to \infty$, $\gamma_z(s) \sim [h_{ij}s^{d_i+d_j-1}]$, where each h_{ij} is a constant and depend on d_i and d_j.

The explicit forms of the spectral density and the autocovariances function for a stationary fractionally integrated vector time series are presented below.

The distinguishing characteristic of a fractionally integrated time series is its persistence or antipersistence. These characteristics can be seen from the shape of the diagonal elements (i.e. i=j) of the spectral density and the diagonal elements of the autocovariance functions. If $0<d_i<1/2$ the process is characterized by strong positive dependency between distant observations. This is noted in the frequency domain by the spectral density increasing to an infinite value at the zero frequency. In the time domain the persistence is indicated by the slow decline of the autocovariances as the distance between observations increases.

If $-1/2<d_i<0$ the process is characterized by quick dampening and negative dependency between observations. In the frequency domain, this antipersistence is indicated by the rapid decline of the spectral density to zero, as the frequency approaches zero. The time domain indicates the antipersistence by the rapid decline of the autocovariances.

III. PREVIOUS ESTIMATION TECHNIQUES

Several estimation procedure for fractionally integrated time series have been proposed. Unfortunately two general problems have plagued these estimation techniques, first is the inability to handle multivariate data and the second problem is that the AR and MA parameters are typically not accurately recovered. These two problems are alleviated in the current estimation procedure by deriving the unconditional likelihood function. Before deriving the likelihood function, the previous estimation techniques will be reviewed and their inability to estimate the general model considered

in this paper is noted. Readers not interested in the limitations and problems of previous estimation procedures, can skip the remainder to this section.

All previous estimation procedures allow for only one differencing parameter and all but one are univariate procedures. These procedures will be considered in two groups the two-step procedures and the one-step procedures. The first estimation procedures were the two-step procedures. In the first step the differencing parameter is estimated and in the second step the other parameters of the model were estimated. These steps may be repeated until the parameters converge. The two-step procedures only differ in the first step. All the two-step procedures then use the same procedure to transform the observed series to obtain a series that (hopefully) follows an ARMA(p,q) model. Once this transformed series is obtained the remaining parameters of the model are estimated by standard time series procedures.

The first step of estimating the differencing parameter can be carried out in at least four ways. The first procedure uses the estimated rescaled range exponent, H (see McLeod and Hipel (1978) for a survey of this statistic). The differencing parameter can then be identified as $\hat{d}=\hat{H}-(1/2)$. This procedure exploits the fact that fractionally integrated time series are asymptotically self similar. The next two procedures exploit the special shape of the spectral density of a fractionally integrated time series. In Janacek (1982) the differencing parameter is written as a function of the Fourier coefficients of the log spectral density. The differencing parameter estimate is obtained by estimating these Fourier coefficients by numerical integration of the log periodogram. An alternative procedure, presented in Geweke and Porter-Hudak (1983), exploits the behavior of the spectral density around 0. Using frequencies near zero, a univariate regression of the log

periodogram on the log of the frequencies is performed. The slope estimate from this simple univariate regression is an estimate of the differencing parameter. The fourth procedure presented is Shea (1987) is maximum likelihood estimation of the regression model used in Geweke and Porter-Hudak.

The above procedures give an estimate of the single differencing parameter. The other parameters of the model must still be recovered. The estimated differencing parameter is used to transform the observed series into a series that follows an ARMA(p,q) model. If the model of the observed series z_t, is $\theta(L)(1-L)^d z_t=\Phi(L)\varepsilon_t$ the estimated differencing parameter $\hat{d}$ is used to obtain the series $u_t=(1-L)^{\hat{d}} z_t$ which will follow the model $\theta(L)u_t=\Phi(L)\varepsilon_t$. A problem is that $(1-L)^{\hat{d}}$ is defined as an infinite lag polynomial, hence the series u_t can only be obtained when there exists and infinite realization of z_t. The result of this problem is that the transformation cannot be completely performed. Attempts to transform the series to the ARMA(p,q) model are only approximations.

Hosking (1981) and Geweke and Porter-Hudak (1983) both outline procedures to transform the observed series to leave only the parameters of the ARMA(p,q) process. The procedure presented in Hosking (1981) is a time domain transformation and the procedure presented in Geweke and Porter-Hudak is a frequency domain transformation. The two procedures appear to differ but are actually equivalent (shown below).

In the time domain the AR representation is used to transform the observed series. If u_t~ARMA and $(1-L)^d z_t = u_t$ then by expanding $(1-L)^d$ it is possible to write

$$u_t = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} z_{t-j}$$

This suggests after obtaining $\hat{d}$ an approximation to an ARMA can be obtained by using the truncated AR model, i.e use $\hat{d}$ for d in the above model and set $z_{t-j}=0$ for t-j outside of the sample.

An alternative to the time domain transformation is the frequency domain approach. This approach is rather simple; calculate the Fourier transform of the observed series, multiply by the Fourier transform of the fractional differencing operator based on $\hat{d}$, and then obtain the transformed series by calculating the inverse Fourier transform. It is hoped that the effect of this transformation is a series that follows the ARMA(p,q) model of the u_t process. The actual effect of the transformation is identical to the time domain truncated AR transformation.

For any onesided operator these two approaches are equivalent. To see this simply calculate the inverse Fourier transform of the product of the Fourier transform of the operator $\sum_{j=0}^{\infty} \phi_j L^j$ and the Fourier transform of the sample $x_1, x_2, \ldots, x_T$:

$$\tilde{u}_s = \frac{1}{2\pi} \int_0^{2\pi} e^{is\lambda} \sum_{j=0}^{\infty} \phi_j e^{-ij\lambda} \sum_{t=1}^{T} x_t e^{-it\lambda} d\lambda$$

$$= \frac{1}{2\pi} \int_0^{2\pi} e^{is\lambda} \sum_{k=1}^{\infty} \sum_{t=1}^{\min[T,k]} \phi_{k-T} x_t e^{-ik\lambda} d\lambda$$

$$= \begin{cases} 0, & \text{if } s \leq 0 \\ \sum_{t=1}^{\min[T,s]} \phi_{s-t} x_t, & \text{if } s > 1 \end{cases}$$

For the sample $s = 1,2,\ldots,T$ this is simply the truncated polynomial

$$\tilde{u}_s = \sum_{t=1}^{s} \phi_{s-t} x_t = \sum_{j=0}^{s-1} \phi_j x_{s-j}.$$

The problem is for any finite realization: (ignoring any bias in $\hat{d}$ which only further complicates the estimation of the ARMA parameters)

$$\tilde{u}_t = u_t - \sum_{j=t}^{\infty} \frac{\Gamma(-\hat{d}+j)}{\Gamma(-\hat{d})\Gamma(j+1)} z_{t-j}$$

$$= u_t - \sum_{k=0}^{\infty} \frac{\Gamma(-\hat{d}+t+k)}{\Gamma(-\hat{d})\Gamma(t+k+1)} z_{-k}.$$

After applying the truncated AR, each transformed variable is the sum of an element from an ARMA process and a linear combination of an infinite number of unobserved terms. The transformed series will not have the ARMA model of the u_t series.

As long as d is consistently estimated in the first step then the two-step estimates are all consistent. For all finite realizations these procedures cause residual autocorrelation.

Two alternatives to the two-step procedures have been proposed, one in Li and McLeod (1986) and the other in Fox and Taqqu (1986). In Li and Mcleod (1986) the procedure suggested is to truncate the infinite sum that defines $(1-L)^{-d}$ and use typical ARMA estimation procedures. Truncating the infinite series is equivalent to the two-step procedure proposed by Hosking (1981) which as argued above leads to misleading estimates.

The procedure proposed in Fox and Taqqu (1986) is an approximation to the maximization of the Gaussian likelihood function and simultaneously estimates all the parameters of the univariate model. The estimates are consistent, asymptotically normal and converge at the familiar rate of root T. However, as noted in Dahlhaus (1988) the small sample behavior of this type of estimator may be poor. Dahlhaus (1988) concludes that exact maximum likelihood is optimal.

IV. THE LIKELIHOOD FUNCTION

Consider a kx1 stationary fractionally integrated vector time series z_t with differencing operators satisfy $|d_i|<1/2$. Let Z_T be a kTx1 sample of T observations such that $Z_T=[z_1' \; z_2' \; \ldots \; z_T']'$ and $Z_T \sim N_{kT}(0,\sum)$, with probability density function

$$f(Z_T,\sum) = (2\pi)^{-(kT)/2}|\sum|^{-1/2}\exp\{-(1/2)Z_T'\sum^{-1}Z_T\}.$$

This implies a block Toeplitz form for the covariance matrix,

$$\underset{kTxkT}{\sum} = [\; \Sigma(i-j)) \;] \text{ for } i,j=1,2,\ldots,T$$

where $\Sigma(r)=E[z_t z_{t+r}']$ are the autocovariances of z_t. The likelihood function, for a given sample and model parameterization, can be evaluated once the covariance matrix is written in terms of the parameters of the model.

The following approach will be used to achieve this parameterization. First, the model for z_t is written in its moving average representation. Second, the spectral density function of z_t, $f_z(\lambda)$, is calculated. Finally, the autocovariance function, $\Sigma(s)$, is calculated by

$$\Sigma(s) = \frac{1}{2\pi}\int_0^{2\pi} f_z(\lambda)e^{i\lambda s}d\lambda. \qquad (4)$$

IV.1 The General Model

The general model (3) admits different orders of integration of the different elements in the random vector z_t (i.e. $d_i \neq d_j$ for $i \neq j$). By introducing the V matrix the model allows some linear combination(s) of the elements of z_t to be of a different (lower) order of integration than the

element of z_t. When this occurs the time series z_t is said to be cointegrated. If z_t is cointegrated the parameters of the V matrix are the coefficients of these linear combinations. In the typical case where the z_t vector is not cointegrated $V=I_k$. Define the process u_t by $u_t=D(L)Vz_t$. The assumptions about z_t imply the vector time series u_t is generated by the ARMA(p,q) process $\Theta(L)u_t=\Phi(L)\varepsilon_t$.

IV.2 The Spectral Density of u_t

The Wold representation of u_t is

$$u_t = \Theta(L)^{-1}\Phi(L)\varepsilon_t = \frac{B(L)}{A(L)}\varepsilon_t.$$

$B(L)$ is a matrix polynomial of order $M \leq ((k-1)p) + q$; and $A(L)$ is a scalar polynomial of order $H \leq kp$. The roots of $A(\xi)$ are outside the unit circle, so $A(\xi)$ can be written

$$A(\xi) = \prod_{n=1}^{H} (1-\rho_n\xi)^{-1}, \quad \text{where} \quad |\rho_n| < 1 \quad \text{for } n=1,2,\ldots,H.$$

The spectral density of u_t, denoted $f_u(\lambda)$, is

$$f_u(\lambda) = \frac{B(\omega)\Sigma B(\omega^{-1})'}{A(\omega)A(\omega^{-1})} = B(\omega)\Sigma B(\omega^{-1})' \prod_{n=1}^{H} (1-\rho_n\omega)^{-1}(1-\rho_n\omega^{-1})^{-1}$$

where $\omega = e^{-i\lambda}$. Assuming that the roots of $A(\xi)$ are all unique it is possible to use the partial fraction decomposition of the product, to write

$$f_u(\lambda) = B(\omega)\Sigma B(\omega^{-1})' \sum_{j=1}^{H} \omega^H \zeta_j \left[\frac{\rho_j^{2H}}{(1-\rho_j\omega)} - \frac{1}{(1-\rho_j^{-1}\omega)} \right]$$

where $\zeta_j = \dfrac{1}{\rho_j \prod_{i=1}^{H} (1-\rho_i\rho_j) \prod_{\substack{m=1 \\ m\neq j}}^{H} (\rho_j-\rho_m)}$.

Note $B(\omega)$ is of finite order therefore each element of $B(\omega)\Sigma B(\omega^{-1})'$ can be written as a finite polynomial in ω, with both positive and negative values of the exponents. So the spectral density of u_t can be written with representative (i,j) element notation as $f_u(\lambda) = [f_u(\lambda)_{ij}]$ where

$$f_u(\lambda)_{i,j} = \sum_{-M}^{M} \psi_{i,j}(\ell)\omega^{\ell} \sum_{m=1}^{H} \omega^{H} \zeta_m \left[\frac{\rho_m^{2H}}{(1-\rho_m\omega)} - \frac{1}{(1-\rho_m^{-1}\omega)} \right] \tag{5}$$

If the (i,j) element of $B(\omega)$ is denoted

$$B_{i,j}(\omega) = \sum_{n=0}^{M} B_{i,j}(n)\ \omega^{n}$$

and the (i,j) element of Σ is σ_{ij} then $\psi_{i,j}(\ell)$ has the form

$$\psi_{i,j}(\ell) = \sum_{h=1}^{k} \sum_{t=1}^{k} \sum_{s=\max[0,\ell]}^{\min[M,M-\ell]} \sigma_{ht}\ B_{i,h}(s)\ B_{j,t}(s-\ell).$$

The analysis thus far applies to general vector ARMA(p.q) models, i.e. the spectral density of every wide sense stationary vector ARMA(p,q) model can be written in the form of equation (5).

IV.3 The Spectral Density of z_t

The moving average representation of Vz_t is: $z_t = V^{-1}D(L)^{-1}u_t$. This means that the spectral density of z_t can be written

$$f_z(\lambda) = V^{-1}D(\omega)^{-1}f_u(\lambda)D(\omega^{-1})^{-1}V^{-1}'.$$

Expanding and using the spectral density of u_t, the spectral density of z_t can be written with representative (i,j) element notation as $f_z(\lambda) = [f_z(\lambda)_{i,j}]$ where

$$f_z(\lambda)_{i,j} = \sum_{n=1}^{k}\sum_{r=1}^{k} v^{in}v^{jr} f_u(\lambda)_{n,r}(1-\omega)^{-d_n}(1-\omega^{-1})^{-d_r}$$

where the (i,j) element of V^{-1} is denoted v^{ij}. This rich structure is only a result of having the general matrix V in the model. In the case of $V=I_k$ the spectral density of z_t with a representative element (i,j) is simply

$$f_z(\lambda)_{i,j} = f_u(\lambda)_{i,j}(1-\omega)^{-d_i}(1-\omega^{-1})^{-d_j}.$$

Upon substituting in the general functional form for $f_u(\lambda)_{i,j}$ the (i,j) element of $f_z(\lambda)$ becomes

$$\sum_{\ell=-M}^{M}\sum_{m=1}^{H}\sum_{n=1}^{k}\sum_{r=1}^{k} v^{in}v^{jr}\psi_{n,r}(\ell)\zeta_m\left[\frac{\rho_m^{2H}}{(1-\rho_m\omega)} - \frac{1}{(1-\rho_m^{-1}\omega)}\right](1-\omega)^{-d_n}(1-\omega^{-1})^{-d_r}\omega^{H+\ell}. \tag{6}$$

This is the general form of an (i,j) element of the spectral density for a kx1 vector time series, which is generated by a fractional ARMA(p,q) model.

IV.4 The Autocovariance of z_t

The s^{th} autocovariance of z_t will be calculated by using (4). From the structure of the (i,j) element of the spectral density function this reduces to the evaluation of the integral[1] (recall $\omega=e^{-i\lambda}$)

$$\frac{1}{2\pi}\int_0^{2\pi}\left[\frac{\rho^{2H}}{(1-\rho e^{-i\lambda})} - \frac{1}{(1-\rho^{-1}e^{-i\lambda})}\right](1-e^{-i\lambda})^{-w}(1-e^{i\lambda})^{-v}e^{-i\lambda h}d\lambda$$

This integral will be denoted $C(w,v,h,\rho)$. The (i,j) element of the s^{th} autocovariance can be written

$$\sum_{\ell=-M}^{M}\sum_{m=1}^{H}\sum_{n=1}^{k}\sum_{r=1}^{k} v^{in}v^{jr}\psi_{n,r}(\ell)\,\zeta_m C(d_n,d_r,H+\ell-s,\rho_m) \tag{7}$$

Because i and j only enter in the V parameters, each element of the s^{th} autocovariance is simply a linear combination of the same values of $C(d_n,d_r,h,\rho_m)$. This allows much quicker evaluation of the autocovariances, then just calculating each element term by term.

If p=0, the partial fraction decomposition need not be considered and the autocovariances have the form

$$\sum_{\ell=-q}^{q}\sum_{n=1}^{k}\sum_{r=1}^{k} v^{in}v^{jr}\psi_{n,r}(\ell)\,\frac{\Gamma(1-d_n-d_r)\Gamma(d_r+s-\ell)}{\Gamma(d_r)\Gamma(1-d_r)\Gamma(1-d_n-s+\ell)}.$$

which can be easily evaluated to obtain the autocovariances.

IV.5 Nonintegral Form of $C(w,v,h,\rho)$

The general autocovariances ($p \neq 0$) have been written in terms of the function $C(w,v,h,\rho)$, and this function has been defined as an integral. To calculate the autocovariances an alternative formula for the function $C(w,v,h,r)$, which is easier to evaluate will be derived. Note that the variable h can take on any integer value, ρ can be any complex number in the unit circle, while w and v are each restricted to real values on the open interval (-1/2,1/2). Using geometric series expansions $C(w,v,h,\rho)$ can be written

$$\rho^{2H}\sum_{m=0}^{\infty}\rho^{m}\frac{1}{2\pi}\int_0^{2\pi}(1-e^{-i\lambda})^{-w}(1-e^{i\lambda})^{-v}e^{-i\lambda(h+m)}d\lambda + \sum_{n=1}^{\infty}\rho^{n}\frac{1}{2\pi}\int_0^{2\pi}(1-e^{-i\lambda})^{-w}(1-e^{i\lambda})^{-v}e^{-i\lambda(h-n)}d\lambda.$$

The interchange of the sum and integral is justified in APPENDIX I. The problem is reduced to the following evaluation

$$\frac{1}{2\pi}\int_0^{2\pi}(1-e^{-i\lambda})^{-w}(1-e^{i\lambda})^{-v}e^{i\lambda h}d\lambda = \frac{\Gamma(1-w-v)\,(-1)^{h}}{\Gamma(1-w-h)\Gamma(1-v+h)}$$

which is shown in APPENDIX II. With this

$$C(w,v,h,\rho) = \Gamma(1-w-v)\left[\rho^{2H}\sum_{m=0}^{\infty}\frac{\rho^{m}(-1)^{h+m}}{\Gamma(1-w+h+m)\Gamma(1-v-h-m)} + \sum_{n=1}^{\infty}\frac{\rho^{n}(-1)^{h-n}}{\Gamma(1-w+h-n)\Gamma(1-v-h+n)}\right]. \tag{8}$$

The advantage of two infinite sums rather than an integral may not be immediately apparent. The benefit is that the infinite sums are well defined hypergeometric functions (see APPENDIX III), which can be quickly and accurately approximated to any desired accuracy on a (personal) computer. In some case the sums reduce allowing even simpler evaluation of $C(w,v,h,\rho)$. All the different forms of $C(w,v,h,\rho)$ are presented in APPENDIX III.

V. SPECIFIC MODELS

V.1 Cointegration and the Identification of V

Let δ_i be the order of differencing that needs to be applied to z_{it}, the i^{th} element of the observed series Z_t, so that $(1-L)^{\delta_i} z_{it}$ follows an ARMA(p,q) process where p and q are finite. If some of the δ_i's are equal then it is possible that linear combinations of the corresponding z_{it}'s possess lower orders of integration. This is referred to as cointegration and the coefficients of the linear combinations are called cointegrating vectors. Previous models of cointegration have typically restricted the observed series to I(1) and the linear combinations to I(0). Fractional integration extends this model by allowing for several different orders of integration. The different orders of integration requires identifying the cointegrating vectors for each orders of integration; this will implicitly identify the matrix V in equation (3). The identification of the cointegrating vectors requires knowing

1. the number of different orders of integration, denoted s;
2. the number of linearly independent series for each order of integration, denoted n_i for i=1,2,...,s; and
3. the number of cointegrating vectors for each of the orders of integration, denoted r_i for i=1,2,...,s-1.

It is possible that a cointegrating linear combination of series, which are $I(d_i)$, is $I(d_{i+\ell})$ [where $d_{i+\ell} < d_i$] and is cointegrated with other series that are $I(d_{i+\ell})$. This implies that the sum of the n_i may be greater than k. Assume the elements of VZ_t in equation (3) are ordered so that

$$D(L) = \text{blockdiagonal } [(1-L)^{a_1} I_{n_1-r_1}, (1-L)^{a_2} I_{n_2-r_2}, \ldots, (1-L)^{a_s} I_{n_s}]$$

where $a_1 > a_2 > \ldots > a_s$, $n_s + \Sigma_{i=1}^{s-1}(n_i-r_i) = k$ and $d_1 = \ldots = d_{n_1-r_1} = a_1$, $d_{n_1-r_1+1} = \ldots = d_{n_1-r_1+n_2-r_2} = a_2$, etc. If $r_i = 0$, there is no cointegration in the n_i series which are integrated of order a_i.

Identification of V, or equivalently the identification of the cointegrating vectors, can be achieved by the following procedure. Multiply Z_t by P_1, a known nonsingular kxk matrix of k(k-1) zeros and k ones, so the elements of P_1Z_t are arranged with the first n_1 elements $I(a_1)$. The r_1 linearly independent cointegrating vectors associated with a_1 span an r_1 dimensional subspace and any vector in that subspace is a cointegrating vector. Hence, only the subspace span by the cointegrating vectors is identifiable. This subspace is identified by the free parameters in the kxk matrix

$$v_1 = \text{blockdiagonal } [I_{n_1-r_1}, v_1, I_{k-n_1}]$$

where v_1 is an $r_1 x r_1$ matrix with diagonal elements restricted to one. The off diagonal element of v_1 are the identifiable parameters of the cointegrating vectors associated with a_1. Multiply $v_1 P_1 Z_t$ by P_2, a known nonsingular kxk matrix of k(k-1) zeros and k ones, so the elements of $P_2 v_1 P_1 Z_t$ which are $I(a_2)$ are in positions (n_1-r_1+1) to (n_1+n_2). The subspace span by the r_2 cointegrating vector associated with a_2 is identified by the free parameters in the kxk matrix

$$v_2 = \text{blockdiagonal}\ [I_{n_1-r_1}, I_{n_2-r_2}, v_2, I_{k-(n_1-r_1)-n_2}]$$

where v_2 is an $r_2 x r_2$ matrix with diagonal elements restricted to one. The off diagonal element of v_2 are the identifiable parameters of the cointegrating vectors associated with a_2. Applying the same procedure for each orders of integration with cointegrating vectors identifies the elements of V by

$$V = v_{s-1} P_{s-1} \cdots v_2 P_2 v_1 P_1$$

where P_i are known matrices,

$$v_i = \text{blockdiagonal}[I_{(n_1-r_1)+\ldots+(n_{i-1}-r_{i-1})}, I_{n_i-r_i}, v_i, I_{k-(n_1-t_1)-\ldots-(n_{i-1}-r_{i-1})-n_i}]$$

and each v_i is an $r_i x r_i$ matrix with diagonal elements restricted to one. The off diagonal element of each v_i represent the identifiable parameters of V.

V.2 The Bivariate Model

The simplest model of cointegration is the bivariate model. Assume $Z_t=[x_t\ y_t]'$, where x_t and y_t are each integrated of order d, but there exist $[\theta\ 1]'\in R^2$ such that $[\theta\ 1]Z_t \sim I(d-b)$ where $b>0$. The model for Z_t can be written

$$\begin{bmatrix} (1-L)^d & 0 \\ 0 & (1-L)^{d-b} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \theta & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

The error process $u_t=[u_{1t}\ u_{2t}]'$ is assumed to be generated by an ARMA(p,q) process that is stationary and possesses an invertible MA representation, i.e. u_t can be written $\Phi(L)u_t = \Theta(L)\varepsilon_t$ where, the roots of the polynomials $|\Phi(\xi)|$ and $|\Theta(\xi)|$ are outside the unit circle, $\Phi(0)=\Theta(0)=I_2$ and

$$\varepsilon_t \sim IID\ N_2 \left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{bmatrix} \right].$$

If the Wold representation of u_t is written $u_t = \frac{B(L)}{A(L)}\varepsilon_t$, then the 2x2 s^{th} autocovariance can be written

$$\Sigma(s) = \sum_{j=1}^{H} \zeta_j \sum_{k=1}^{2} \sum_{h=1}^{2} \sigma_{kh}^2 \sum_{n=0}^{M} \sum_{m=0}^{M}$$

$$\begin{aligned} &[[1\ -\theta]'[1\ -\theta]B_{1,k}(n)B_{1,h}(m)\ C(d,d,2p+n-m-s,\rho_j) \\ &\quad + [1\ -\theta]'[-1\ 0]B_{1,k}(n)B_{2,h}(m)\ C(d,d-b,2p+n-m-s,\rho_j) \\ &\quad\quad + [-1\ 0]'[1\ -\theta]B_{2,k}(n)B_{1,h}(m)\ C(d-b,d,2p+n-m-s,\rho_j) \\ &\quad\quad\quad + [-1\ 0]'[-1\ 0]B_{2,k}(n)B_{2,h}(m)\ C(d-b,d-b,2p+n-m-s,\rho_j)]. \end{aligned}$$

If p=0 the autocovariances will have the form

$$\Sigma(s) = \sum_{k=1}^{2}\sum_{h=1}^{2}\sigma_{kh}^{2}\sum_{n=0}^{q}\sum_{m=0}^{q}$$

$$\Big[[1\ -\theta]'[1\ -\theta]B_{1,k}(n)B_{1,h}(m)\ \frac{\Gamma(1-2d)\Gamma(d-n+m+s)}{\Gamma(1-d)\Gamma(d)\Gamma(1-d+n-m-s)}$$

$$+\ [1\ -\theta]'[-1\ 0]B_{1,k}(n)B_{2,h}(m)\ \frac{\Gamma(1-2d+b)\Gamma(d-b-n+m+s)}{\Gamma(1-d+b)\Gamma(d-b)\Gamma(1-d+n-m-s)}$$

$$+\ [-1\ 0]'[1\ -\theta]B_{2,k}(n)B_{1,h}(m)\ \frac{\Gamma(1-2d+b)\Gamma(d-n+m+s)}{\Gamma(1-d)\Gamma(d)\Gamma(1-d+b+n-m-s)}$$

$$+\ [-1\ 0]'[-1\ 0]B_{2,k}(n)B_{2,h}(m)\ \frac{\Gamma(1-2d+2b)\Gamma(d-b-n+m+s)}{\Gamma(1-d+b)\Gamma(d-b)\Gamma(1-d+b+n-m-s)}\Big].$$

VI. MAXIMIZATION OF THE LIKELIHOOD FUNCTION

The superiority of maximum likelihood estimates for time series models has been noted in Hillmer and Tiao (1979), Ansley and Newbold (1980) and Dahlhaus (1988). The consistency and asymptotic normality of maximum likelihood estimates for the fractional univariate model are proved in Li and McLeod (1986).

Using the formulas presented above the likelihood function is easily evaluated on a computer. The main parts of the program are:

1. Calculate the Wold representation of the model as a matrix of ratios of polynomials.
2. Factor the common denominator polynomial of the Wold representation.
3. Calculate the ζ_i's.
4. Calculate the different $C(w,v,h,\rho)$ values.
5. Evaluate the covariance matrix.
6. Calculate the Cholesky decomposition and determinant of the inverse.
7. Calculate the log likelihood function value.

Step 1 can be ignored if the model is parameterized as a Wold representation. Important savings in time are made by separating steps 4 and 5. Because the same $C(w,v,h,\rho)$ term appears in several autocovariances, all the $C(w,v,h,\rho)$ terms are calculated and saved in memory then recalled as needed in step 5. Step 4 is relatively quick because, as noted in APPENDIX III, the $C(w,v,h,\rho)$ terms can be calculated by a recursive formula. In step 6 a multivariate version of Levinson's algorithm (Sowell (1986)) is used to recursively calculate the terms of the Cholesky decomposition and determinant of the inverse. Because Levinson's algorithm is used to obtain the inverse of the covariance matrix as T grows, the number of calculations grows at the rate $(kT)^2$ instead of the typical $(kT)^3$. The likelihood function can be evaluated with the Cholesky decomposition of the inverse which saves $T(T-1)/2$ memory locations. If the log likelihood function is maximized by a procedure which requires derivatives then numerical derivatives must be used because the derivative of $\Gamma(x)$ does not possess a closed form.

To see how the maximum likelihood procedure works in practice, Monte Carlo simulation was used to compare the estimates from the maximum likelihood procedure and estimates from the procedures presented in Fox and Taqqu (1986) and Geweke and Porter-Hudak (GPH) (1982). A collection of various univariate models were considered because GPH's and Fox and Taqqu's procedures are only appropriate for univariate models. For each model considered, 100 samples of T observations were simulated. The differencing parameter of each simulated sample was estimated by exact maximum likelihood, the Fox & Taqqu procedure and the GPH procedure. The GPH procedure only estimates the d parameters while, both maximum likelihood and the Fox & Taqqu procedures estimated all the parameters of the model. The correct values of p and q were assumed known. Reported in Table 1 are the averages of the estimates of the

fractional differencing parameters for the 100 simulations. Also reported is the average time taken to perform the estimations.

* * * INSERT TABLE 1 HERE * * *

For the GPH procedure α was set equal to 1/2. The likelihood function and the negative of the objective function in Fox and Taqqu were maximized using the DFP procedure in GQOPT. For all the models the start values of the parameters were

$d = .01$

$\sigma_{\epsilon}^2 = 1.0$

$\phi_1 = .1$

$\phi_2 = .05$

$\theta_1 = \begin{cases} .1, \text{ if } p = 0 \\ .09, \text{ otherwise} \end{cases}$

$\theta_2 = .05$

All computations were performed in VAX FORTRAN on a VAX workstation 3200.

For models with p=q=0 all three procedures perform equally well in terms of bias. The most striking feature is the standard deviation of the GPH estimates are three to four times larger than for the maximum likelihood estimates. In general, the bias for the means of the maximum likelihood estimates are smaller than all the Fox & Taqqu means and all but one of the GPH means. As the number of parameters increases and/or as the roots of the AR and MA polynomials approach the unit circle the number of observations needed to apply asymptotics may increase dramatically. For the model $(1-L)^{.25}Z_t = (1-.5L-.36L^2)\epsilon_t$, all the estimation procedures give biased estimates in sample even with T=500. The smallest bias was associated with the maximum likelihood estimates for all the sample sizes.

MODEL	T	MAXIMUM		FOX & TAQQU		GPH	
		d	time	d	time	d	time
$(1-L)^{-.4}z_t = \epsilon_t$	100	-.409 (.092)	4.67	-.438 (.098)	1.20	-.394 (.317)	.18
$z_t = \epsilon_t$	100	-.006 (.075)	4.25	-.046 (.093)	1.22	-.026 (.266)	.19
$(1-L)^{.4}z_t = \epsilon_t$	100	.385 (.060)	4.61	.369 (.089)	1.14	.509 (.285)	.19
$(1+.7L)(1-L)^{.2}z_t = \epsilon_t$	100	.182 (.104)	9.32	.135 (.117)	1.51	.158 (.316)	.16
$(1-.8L)(1-L)^{.3}z_t = \epsilon_t$	200	.289 (.117)	43.79	.375 (.234)	5.01	.627 (.242)	.40
$(1-L)^{-.25}z_t = (1+.6L)\epsilon_t$	200	-.252 (.068)	33.03	-.278 (.07)	5.70	-.272 (.258)	.48
$(1+.6L+.09L^2)z_t = \epsilon_t$	200	-.022 (.133)	55.31	-.097 (.229)	6.97	-.012 (.240)	.48
$(1-1.0L+.16L^2)(1-L)^{-.35}z_t = \epsilon_t$	200	-.404 (.366)	73.38	-.841 (.413)	12.14	-.044 (.224)	.40
$(1-.6L)(1-L)^{-.2}z_t = (1-.3L)\epsilon_t$	300	-.270 (.265)	117.40	-.525 (.367)	13.03	-.121 (.201)	.70
$(1-L)^{.25}z_t = (1-.5L-.36L^2)\epsilon_t$	100	-.220 (.397)	15.06	-.608 (.235)	3.68	-.426 (.315)	.17
$(1-L)^{.25}z_t = (1-.5L-.36L^2)\epsilon_t$	200	-.076 (.324)	49.13	-.388 (.247)	7.53	-.301 (.221)	.40
$(1-L)^{.25}z_t = (1-.5L-.36L^2)\epsilon_t$	300	.016 (.295)	112.08	-.273 (.244)	13.54	-.201 (.162)	.72
$(1-L)^{.25}z_t = (1-.5L-.36L^2)\epsilon_t$	400	.080 (.265)	374.33	-.207 (.275)	25.58	-.159 (.173)	1.14
$(1-L)^{.25}z_t = (1-.5L-.36L^2)\epsilon_t$	500	.121 (.202)	1110.03	-.082 (.267)	30.99	-.096 (.181)	1.57

Table 1: Averge of estimated fractional differencing parameters and average time for the estimation in seconds for 100 simulations. Standard deviations reported in parenthesis. In the simulations $\epsilon_t \sim \text{IIDN}(0,1)$.

VII. SUMMARY AND EXTENSIONS

This has been a presentation of the unconditional Gaussian likelihood function for a fractionally integrated stationary vector time series model. With the likelihood function, it is now possible to obtain maximum likelihood estimates. This task is made easier by several recursive procedures which allow speedy evaluation of the autocovariances and inversion of the covariance matrix.

A related method of dealing with fractionally integrated time series would be to use a Bayesian approach. This has been advocated by Carlin et al (1987). Given a prior distribution on the parameters of the model and the likelihood function derived in this paper it would be straightforward to evaluate posterior densities.

An immediate application of the estimation procedure presented is a new unit root test. Consider estimating a fractional ARIMA(p,d,q) model for a time series. If $\hat{d}$ is near 1/2 (within two asymptotic standard errors) or if a root of the AR polynomial is near 1 then first difference the series and again estimate the best ARIMA(p,d,q) model. This approach to unit root testing has a distinct advantage because it explicitly accounts for nuisance parameters in the model. Nuisance parameters, particularly MA parameters have been a problem for some unit root test as noted in Schwert (1987(a), 1987(b)).

APPENDIX I

JUSTIFICATION OF INTERCHANGE OF INTEGRALS

This is a demonstration of the integrability of

$$\int_0^{2\pi} \sum_{n=0}^{\infty} \rho^n (1-e^{-i\lambda})^{-w} (1-e^{i\lambda})^{-v} e^{i\lambda(h+n)} d\lambda$$

to allow interchanging the summation and the integral. The parameters satisfy the restrictions: w and v are on the open interval (-1/2,1/2) and $|\rho|<1$.

$$\int_0^{2\pi} \sum_{n=0}^{\infty} \left| \rho^n (1-e^{-i\lambda})^{-w} (1-e^{i\lambda})^{-v} e^{i\lambda(h+n)} \right| d\lambda$$

because $|e^{i\xi}|=1$

$$\leq \int_0^{2\pi} \sum_{n=0}^{\infty} |\rho|^n \left| (1-e^{-i\lambda})^{-w} (1-e^{i\lambda})^{-v} \right| d\lambda$$

$$= \frac{1}{1-|\rho|} \int_0^{2\pi} \left| (1-e^{-i\lambda})^{-w} (1-e^{i\lambda})^{-v} \right| d\lambda.$$

The integrand has poles at 0 and 2π. At both poles the integrand behaves like λ^{-v-w} near zero, which is integrable for $v + w < 1$.

APPENDIX II

EVALUATION OF THE FUNDAMENTAL INTEGRAL

This is the evaluation of the integral

$$\frac{1}{2\pi} \int_0^{2\pi} (1-e^{-i\lambda})^{-w}(1-e^{i\lambda})^{-v} e^{i\lambda h} d\lambda$$

$$= \frac{2^{-1-w-v}}{\pi} e^{i\pi(v-w)/2} \int_0^{2\pi} [\sin(\lambda/2)]^{-w-v} e^{i(2h+w-v)\lambda/2} d\lambda$$

because $(1-e^{i\xi}) = 2\sin(\xi/2)e^{i(\xi+3\pi)/2} = 2\sin(\xi/2)e^{i(\xi-\pi)/2}$

$$= \frac{2^{-w-v}}{\pi} e^{i\pi(v-w)/2} \int_0^{\pi} [\sin(z)]^{-w-v} e^{i(2h+w-v)z} dz$$

See Erdelyi et al. (1953) p.12 (1.5.29)

$$= \frac{\Gamma(1-w-v)\ (-1)^h}{\Gamma(1-w-h)\Gamma(1-v+h)}$$

APPENDIX III

ALTERNATIVE FORMS OF $C(w,v,h,\rho)$

Some of the autocovariances are written in terms of the function $C(w,v,h,\rho)$. The variable h can take on any integer value, ρ can be any complex number in the unit circle, while w and v are each restricted to real values on the open interval $(-1/2,1/2)$. As written in (7) the function is correct for all values of interest, however, because $1/\Gamma(\xi)$ has roots at the nonnegative integers some of the terms in the infinite sums are zero when w or v is zero. The different forms of the function will be derived using the following definition and equation. The hypergeometric function is defined by

$$F(a,b;c;\xi) \equiv \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+n)\Gamma(n+1)} \xi^n . \qquad \text{(IV.1)}$$

For all real values of x and integer values of n, as long as the arguments of the gamma functions are never nonpositive integers then

$$\frac{\Gamma(x+n)}{\Gamma(x)} = \frac{\Gamma(1-x)(-1)^n}{\Gamma(1-x-n)} . \qquad \text{(IV.2)}$$

This follows from equation (8.334.3) of Gradshteyn and Ryzhik (1980) and because $\sin(\pi(n+x)) = (-1)^n \sin(\pi x)$ for real x and integer n.

CASE I. $w \neq 0$ and $v \neq 0$

The terms in the first sum of (7) can be rewritten using (IV.2)

$$\frac{(-1)^{h+m}}{\Gamma(1-w+h+m)\Gamma(1-v-h-m)} = \frac{(-1)^h}{\Gamma(1-v-h)\Gamma(1-w-h)} \left[\frac{\Gamma(v+h+m)\Gamma(1-w+h)}{\Gamma(v+h)\Gamma(1-w+h+m)}\right].$$

Similarly, the terms in the second sum can be rewritten

$$\frac{(-1)^{h-n}}{\Gamma(1-w+h-n)\Gamma(1-v-h+n)} = \frac{(-1)^h}{\Gamma(1-v-h)\Gamma(1-w-h)} \left[\frac{\Gamma(w-h+n)\Gamma(1-v-h)}{\Gamma(w-h)\Gamma(1-v-h+n)}\right].$$

Using these relations equation (7) can be written

$$C(w,v,h,\rho) = \frac{\Gamma(1-w-v)(-1)^h}{\Gamma(1-v-h)\Gamma(1-w+h)} \left[\rho^{2H} \sum_{m=0}^{\infty} \frac{\rho^m \Gamma(v+h+m)\Gamma(1-w+h)}{\Gamma(v+h)\Gamma(1-w+h+m)} + \sum_{n=1}^{\infty} \frac{\rho^n \Gamma(w-h+n)\Gamma(1-v-h)}{\Gamma(w-h)\Gamma(1-v-h+n)} \right]$$

$$= \frac{\Gamma(1-w-v)\Gamma(v+h)}{\Gamma(1-w+h)\Gamma(1-v)\Gamma(v)} \left[\rho^{2H} F(v+h,1;1-w+h;\rho) + F(w-h,1;1-v-h;\rho) - 1 \right].$$

When calculating the likelihood function for a fractional model many different value of $C(w,v,h,r)$ must be evaluated. Fortunately, this does not imply multiple evaluation of the hypergeometric function. The w, v and ρ parameters remain fixed for all the autocovariances. Only h varies and it only varies by integer values. Using the recursive relationship, $x\Gamma(x)=\Gamma(1+x)$, and the definition of the hypergeometric function, it is straightforward to show

$$F(a,1;c;\rho) = \frac{c-1}{\rho(a-1)} \left[F(a-1,1;c-1;\rho) - 1 \right]$$

For each evaluation of the likelihood function only k^3p different hypergeometric functions need to be evaluated. The recursive calculation of the hypergeometric function allows quick evaluations of the $C(w,v,h,\rho)$ functions.

In practice $C(w,v,h,\rho)$ needs only be evaluated (i.e. computer code need only be written) for the case w and v not equal to zero. This is because the subset of the parameter space defined by w=0 and/or v=0 has Lebesgue measure zero. However, the theoretical possibilities that w=0 and/or v=0 require deriving the special forms of $C(w,v,h,\rho)$ in these cases. There is nothing pathological about the likelihood function when w and/or v equal zero. The likelihood function is continuous and differentiable at these values. The

only difference is the existence of easier formula to evaluate the autocovariances.

CASE II. $w=0$ and $v\neq 0$.

A) $w=0$, $v\neq 0$ and $h \leq 0$.

$$c(0,v,h,\rho) = \rho^{2kp-h}F(v,1;1;\rho)$$

B) $w=0$, $v\neq 0$ and $h > 0$.

$$c(0,v,h,\rho) = \frac{\Gamma(v+h)}{\Gamma(v)\Gamma(2+h)}\left[\rho^{2kp}F(v+h,1;1+h,\rho) + \sum_{i=1}^{h}\rho^{i}\frac{\Gamma(v+h-i)\Gamma(1+h)}{\Gamma(v+h)\Gamma(1+h-i)}\right]$$

CASE III. $w\neq 0$ and $v=0$.

A) $w\neq 0$, $v=0$ and $h \leq 0$.

$$c(w,0,h,\rho) = \frac{\Gamma(w-h)}{\Gamma(w)\Gamma(1-h)}\left[\rho^{2kp}\sum_{\ell=0}^{-h}\rho^{\ell}\frac{\Gamma(w-h-\ell)\Gamma(1-h)}{\Gamma(w-h)\Gamma(1-h-\ell)} + F(w-h,1,1-h,\rho) - 1\right].$$

B) $w\neq 0$, $v=0$ and $h > 0$.

$$c(w,0,h,\rho) = \rho^{h}F(w,1;1;\rho)$$

CASE IV. $w=v=0$.

A) $w=v=0$ and $h \leq 0$.

$$c(0,0,h,\rho) = \rho^{2kp}\rho^{-h} = \rho^{2kp-h}$$

B) $w=v=0$ and $h > 0$.

$$c(0,0,h,\rho) = \rho^{h}\ .$$

This final case and the formulas in the paper gives an alternative procedure to derive the autocovariances for a multivariate ARMA(p,q) model.

ACKNOWLEDGEMENT

†This is a generalization of chapter 3 of the author's Ph.D. Thesis, which was supported by the Alfred P. Sloan Foundation through a Doctoral Dissertation Fellowship. The author thanks John Geweke and two anonymous referees for helpful suggestions, Ken Murphy for computing assistance and Nancy Reed for her careful preparation of the manuscript.

ENDNOTES

1 If p=0, (i.e., no AR terms) the term in brackets equals one. This reduces the special integral evaluated in APPENDIX II. Also, if p=0 the m indexed sum and the terms ζ_m will not be in the expression of the spectral density.

REFERENCES

Ansley, Craig F. and Newbold, Paul, "Finite Sample Properties of Estimators For Autoregressive Moving Average Models", Journal of Econometrics, 13, 198, p.159-183.

Booth, G. Geoffrey, Kaen, Fred R. and Koveos Peter E., "R/S Analysis of Foreign Exchange Rates Under Two International Monetary Regimes", Journal of Monetary Economics, 10, 1982, p.407-415.

Carlin, J. B. and A. P. Dempster, "Empirical Cases in the Bayesian Analysis of Seasonal Time Series." Department of Statistics, Harvard University, working paper August, 1987.

Dahlhaus, Rainer, "Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and A New Estimate," The Annals of Statistics, 16, 1988, p. 808-841.

Erdelyi, Arthur et al., Higher Transcendental Functions, Vol I. New: York: McGraw-Hill, 1953.

Fox, Robert and Murad S. Taqqu, "Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," The Annals of Statistics, 14, 1986, p. 517-532.

Geweke, John F., and Porter-Hudak, Susan. "The Estimation and Application of Long Memory Time Series Models." Journal of Time Series Analysis 4, 1982: 221-238.

Gourieroux, Christian, Francoise Maurel and Alain Monfort, "Regression and Non Stationarity", June 1987 (preliminary version).

Gradshteyn, I. S., and Ryzhik, I.M. Tables of Integrals, Series and Products. New York: Academic Press, 1980.

Granger, C. W. J. "Long Memory Relationships and the Aggregation of Dynamic Models." Journal of Econometrics 14, 1980: 227-238.

Granger, C. W. J. "Developments in the Study of Cointegrated Economic Variables." Oxford Bulletin of Economics and Statistics 48, 1986, p.213-228.

Granger, C. W. J., and Robert F. Engle, "Co-Integration and Error-Correction: Representation, Estimation and Testing." Econometrica 55, 1985, p.251-276.

Granger, C. W. J., and Joyeux, Roselyne. "An Introduction to Long-Memory Time Series Models and Fractional Differencing." Journal of Time Series Analysis 1, 1980, p.15-29.

Greene, Myron T. and Bruce D. Fielitz, "Long-Term Dependence in Common Stock Returns", Journal of Financial Economics 4, 1977, p.339-349.

Helms, Billy P., Kaen, Fred R. and Rosenman. Robert E., "Memory in Commodity Futures Contracts," The Journal of Futures Markets 4, 1984, p.559-567.

Hillmer, Steven C. and Tiao, George C., "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models", Journal of the American Statistical Association 74, 1979, p.652-660.

Hosking, J. R. M. "Fractional Differencing." Biometrica 68. 1981: 165-76.

Janacek, G. J. "Determining The Degree of Differencing for Time Series Via the Long Spectrum", Journal of Time Series Analysis 3, 1982, p.177-183.

Li, W. K. and A. I. McLeod, "Fractional Time Series Modelling", Biometrika 73, 1986, p.217-221.

McLeod, Angus Ian, and Keith William Hipel, "Preservation of the Rescaled Adjusted Range 1. A Reassessment of the Hurst Phenomenon." Water Resources Research 14, 1978, p.491-508.

Schwert, G. William "Tests for Unit Roots: A Monte Carlo Investigation." Center for Research in Government Policy and Business, University of Rochester, working paper GPB 87-01, 1987(a).

Schwert, G. William "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data." Journal of Monetary Economics 20, 1987(b), p.73-103.

Shea, Gary S. "Long-Memory Models of Interest Rates: Estimation, Forecasting, and Inference for Variance Bounds on the Interest Rate Term Structure", Paper from the Department of Finance, The Pennsylvania State University, June 1987.

Sowell, Fallaw, "Fractionally Integrated Vector Time Series," Ph.D. Thesis, Department of Economics Duke University, 1986.

Sowell, Fallaw, "The Fractional Unit Root Distribution," forthcoming in Econometrica, 1989.