Maximum likelihood estimation of the multivariate fractional cointegrating model

Johan Lyhagen*

Working Paper Series in Economics and Finance No. 233.

Abstract

Departures from an economic equilibrium should be mean reverting. The deviations are often assumed to be integrated of order zero but this is too restrictive. It is sufficient that the shocks are integrated of an order less than one, i.e. they may be fractionally integrated. A fractionally cointegrated system is developed. Further, estimation and testing are discussed, analytically and by Monte Carlo simulations. The Monte Carlo simulations shows that it is much more severe to ignore fractional cointegration than incorporating it when it is not present.

Key words: Fractional integration; Granger representation theorem; Likelihood ratio test; Monte Carlo.

JEL: C12, C13, C32.

1 Introduction

The standard technique for estimation of vector autoregressive models with cointegrating restrictions was developed in a series of papers by Søren Johansen (see e.g. Johansen, 1988, Johansen, 1991). The procedure is based on the reduced rank regression technique presented by Anderson (1951) and used by Velu and Reinsel (1987) in the case of stationary processes. The estimation method assumes Gaussian errors and the estimation of the cointegrating vectors is given by the solution to a certain eigenvalue problem.

*Department of Economic Statistics, Stockholm School of Economics, P.O. Box 6501, S-113 83 Stockholm. E-mail: STJL@HHS.SE. Financial support from the Tore Browaldhs Foundation for Scientific Research and Education is gratefully acknowledged. We thank M.K. Andersson and M.P Gredenhoff for comments. The usual disclaimer applies.

Another common assumption is that the variable x_t is integrated of order one, $I(1)$, and that the cointegrating vector, β, reduces the cointegrating order to zero, i.e. $\beta^T x_t$ is $I(0)$. The relationship $\beta^T x_t$ is often interpreted as a long run equilibrium. Deviations from the equilibrium should be mean reverting, and Cheung and Lai (1993) show that this is the case for fractionally integrated process of order $d < 1$. Thus, the restriction that $\beta^T x_t$ is $I(0)$ may be to strong. The concept of fractional integration were introduced by Granger and Joyeux (1980) and Hosking (1981). Beran (1994) provide a survey. Andersson and Gredenhoff (1998) shows that using the standard maximum likelihood approach on a fractional cointegrated system gives severe bias and large root mean squared errors for the "impact" matrix Π. Hence, a fractionally cointegrated model has, at least, two justifications: to generalize an unnecessarily restrictive model and to achieve better estimates of the parameters of interest.

The idea of fractional cointegration is not new, for example Granger (1981) points out that fractional cointegration is possible when discussing cointegration. Most of the paper in the area of fractional cointegration deals with the Engle-Granger approach, that is estimating a univariate model and then test if the residuals are $I(1)$ against integrated of an order less than one, see e.g. Cheung and Lai (1993), Baillie and Bollerslev (1994) and Booth and Tse (1995). Sowell (1986) considers a multivariate cointegration model, where it is assumed that x_t is integrated of an order less than 0.5, which may be very restrictive assumption.

This paper presents a general multivariate cointegration model, where the equation errors is allowed to follow a fractional specification, with Gaussian errors and derives the maximum likelihood estimator. Further, test of the cointegrating rank and distribution of the test statistic are discussed.

The paper is organized as follows. Section 2 describes specification and estimation of the fractionally cointegrated model. The Granger representation Theorem is the issue of Section 3, while estimation of the number of cointegrating restrictions is dealt with in Section 4. In Section 5 small sample properties of the estimator and the test procedure are analyzed while a conclusion ends the paper.

2 A fractionally cointegrated model

Consider the model

$$A(L)\, x_t = e_t, \tag{1}$$

where x_t is a random vector of order $k \times 1$, $A(L)$ is an infinite matrix polynomial in the lag operator L and e_t $k \times 1$ vector of Gaussian errors. Define

the following matrix polynomial

$$D(L) = \begin{bmatrix} (1-L)^{1-d_1} - (1-L) & 0 & \\ 0 & (1-L)^{1-d_2} - (1-L) & \\ & & \ddots \end{bmatrix} \tag{2}$$

and assume that $A(L)$ is expandable as $A(L) = A^1(L)(1-L) + \alpha D(L)\beta'$ where α and β are $k \times r$ matrices, and all $d_i > 0$. Further let r_i denote the number of cointegrating relations that share d_i and r^* the number of different values on d. Then an error correction representation exists and can be written as

$$A^1(L)(1-L)x_t = \alpha D(L)\beta' x_t + e_t. \tag{3}$$

The proof is simple and omitted. Note that since the polynomial $(1-L)^{1-d_i} - (1-L)$ has a zero as the first coefficient, the right hand side only contains lagged values of x_t. If all d_i equal one the filter is L and the standard cointegration model arises. The model

$$A^1(L)\tilde{D}(L)x_t = \alpha D(L)\beta' x_t + e_t \tag{4}$$

where

$$\tilde{D}(L) = \begin{bmatrix} (1-L)^{\tilde{d}_1} & 0 & \\ 0 & (1-L)^{\tilde{d}_2} & \\ & & \ddots \end{bmatrix} \tag{5}$$

could be considered, but in this paper we make the restriction that $\tilde{d}_1 = \tilde{d}_2 = ... = \tilde{d}_k = 1$. This is not a very restrictive assumption since most economic variables have been found to be $I(1)$.

To better understand the fractional cointegration model, we take a closer look at

$$(1-L)x_t = \alpha D(L)\beta' x_t + e_t, \tag{6}$$

which may be rewritten as

$$(1-L)x_t - e_t = \alpha D(L)\beta' x_t. \tag{7}$$

The left hand of (7)is $I(0)$, hence $\alpha D(L)\beta' x_t$ must also be $I(0)$. The long run equilibrium is $\beta' x_t$ and the matrix filter $D(L)$ makes it exactly $I(0)$. d is interpreted as how much less integrated the linear combination is relatively the original variables, e.g. when $d = 0$ the linear combination is integrated of the same order as x_t (that is of order one).

When there is only one cointegrating vector, or when all cointegrating relations share the same d, then D and β "commute" so it is possible to

estimate the fractionally cointegrating model by using the standard reduced rank regression suggested by Johansen (1988), i.e. estimate

$$(1-L)\,x_t = \alpha\beta' x_t^d + e_t \tag{8}$$

where $x_t^d = \left((1-L)^{1-d} - (1-L)\right) x_t$. In the more complicated case of more than one cointegrating vector, and not the same d for all the cointegrating vectors, equation (6) may be rewritten,

$$\begin{aligned}
(1-L)\,x_t &= \alpha D(L)\,\beta' x_t + e_t \\
&= \alpha D(L) \begin{bmatrix} \beta_1' x_t \\ \beta_2' x_t \\ \vdots \\ \beta_r' x_t \end{bmatrix} + e_t \\
&= \alpha \begin{bmatrix} \beta_1' x_t^{d_1} \\ \beta_2' x_t^{d_2} \\ \vdots \\ \beta_r' x_t^{d_r} \end{bmatrix} + e_t \\
&= \alpha\tilde{\beta}'\tilde{x}_t + e_t
\end{aligned} \tag{9}$$

where $\tilde{x}_t' = \left(x_t^{d_1}, x_t^{d_2}, \ldots, x_t^{d_{r^*}}\right)'$ and

$$\tilde{\beta} = \begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{r^*} \end{bmatrix}$$

Note that if every d is unique then $r^* = r$. This matrix can in turn be written in terms of individual restrictions on the cointegrating vectors:

$$\tilde{\beta} = \left(H_1\beta_1, H_2\beta_2, \ldots, H_{r^*}\beta_{r^*}\right), \tag{10}$$

where the matrix H_i has an identity matrix at block i and zeros elsewhere. Estimation of models with this kind of restrictions are dealt with by Johansen (1992), pages. 62-65 and Johansen (1995), pages. 110-112. The estimation algorithm is simple: first get initial estimates of $\beta_2, \ldots, \beta_{r^*}$ and estimate β_1 with the concentrated likelihood conditioning on a starting value for d. That is, solve the eigenvalue problem

$$\frac{\left|S_{00|\beta_2,\ldots,\beta_{r^*}}\right| \left|\beta_1'\left(S_{11|\beta_2,\ldots,\beta_{r^*}} - S_{10|\beta_2,\ldots,\beta_{r^*}} S_{00|\beta_2,\ldots,\beta_{r^*}}^{-1} S_{01|\beta_2,\ldots,\beta_{r^*}}\right)\beta_1\right|}{\left|\beta_1'\left(S_{11|\beta_2,\ldots,\beta_{r^*}}\right)\beta_1\right|} \tag{11}$$

where $S_{00|\beta_{2,},\ldots,\beta_{r^*}}$ is the concentrated product moment matrix based on $(1-L)\,x_t$ and $S_{11|\beta_{2,},\ldots,\beta_{r^*}}$ is based on $x_t^{d_1}$. Then, concentrate with respect to $\beta_1' x_t^{d_1}, \beta_3' x_t^{d_3}, \ldots, \beta_{r^*}' x_t^{d_{r^*}}$ and estimate β_2. This is continued until β_{r^*} is estimated by use of the concentrated likelihood and the likelihood value is calculated by the standard formula. These steps are repeated until the increase in likelihood is sufficiently small. This gives the likelihood value for given d. Minimize numerically over d to find the value of d that maximizes the likelihood.

Johansen (1995) proposed that the starting values should be chosen by first estimating the unrestricted β, then solving the eigenvalue problem

$$\left|\lambda \hat{\beta}'\hat{\beta} - \hat{\beta}' H_i \left(H_i' H_i\right)^{-1} H_i' \hat{\beta}\right| = 0 \tag{12}$$

for the eigenvalues and eigenvectors, v. The starting value is $\beta_i = \hat{\beta}\left(v_1, \ldots, v_{r_i}\right)$ and note that r_i is the number of cointegrating vectors sharing the parameter d_i.

3 Granger Representation Theorem

This section closely follows the version of the Granger Representation Theorem presented in Banerjee et al. (1993) but adjusted for the fractionally cointegrated model.

As assumed in the previous section the process may be written as

$$A\left(L\right) x_t = A^1\left(L\right)\left(1-L\right) x_t + \alpha D\left(L\right) \beta' x_t = e_t, \tag{13}$$

Then there are four assumptions

A1. The characteristic polynomial $A\left(z\right)$ only has roots equal to or strictly greater than one, i.e. $\left|A\left(z\right)\right| = 0$ implies $z = 1$ or $\left|z\right| > 1$.

A2. α and β are $n \times r$ matrices of full rank.

A3. $\alpha_\perp' A^1\left(1\right) \beta_\perp$ has full rank $n - r$.

A4. $0 < d \leq 1$.

Assumption A1 and A3 ensures that x_t is stationary after differencing while A2 make sure that x_t is not stationary.. The last assumption make sure that the part of the filtering polynomial that depends on d is not dominated by $\left(1-z\right)$, which would imply that the value of the polynomial evaluated at $z = 1$ is zero. With these assumptions we may prove that:

R1. Δx_t is stationary,

R2. $\beta' x_t$ is mean reverting,

R3. $D(L)\beta' x_t$ is integrated of order zero,

R4. Δx_t has a moving average representation given by

$$\Delta x_t = C(L) e_t,$$

R5. $C(1) = \beta_\perp \left(\alpha'_\perp A^1(1) \beta_\perp\right)^{-1} \alpha'_\perp$ has rank $n - r$,

R6. $\beta' C(1) = 0$,

R7. $C(1)\alpha = 0$,

R8. $x_t = x_0 + C(1) \sum_{i=1}^{t} e_i + S_t$,

R9. $x_t = x_0 + F\tau_t + S_t$,

where $C(L) = C(1) + (1 - L) C_1(1)$, $S_t = C_1(L) e_t$, F is of rank $k - r$ and τ_t is a linear combination of $e_1, ..., e_t$. The proofs are in the Appendix except for R9 where the proof is identical to the one in Hylleberg and Mizon (1989) and omitted. It is seen from R2, R3 and R8 that even though x_t is integrated of order one, $\beta' x_t$ is integrated of an order less than one and $D(L)\beta' x_t$ is integrated of order zero. As in the standard case R5 ensures that the process is as most integrated of order one. R4, R6, R7 and R8 gives the MA and the AR representation and the duality between them. The last one is the common trends representation.

4 Test for the cointegrating rank

In order to discuss the distributions of the trace and the maximum eigenvalue tests we need some distributional results. The following theorem is simply a multivariate form of Theorem 2 in Sowell (1990). Define the sum of vectors $S_N = \sum_{i=1}^{N} \varepsilon_t$ and let $W_d(t)$ denote a vector fractional Brownian motion. Fractional Brownian motions were first introduced by Mandelbrot and Van Ness (1968).

Theorem 1 *If ε_t is a k dimensional vector which is $I(d^*)$, $-0.5 < d^* < 0.5$ and if $(1 - L)^{d^*} \varepsilon_t = u_t$ have zero mean. are IID, and $E|u_t|^r < \infty$ for $r \geq max\left[4, -8d^*/(1 + 2d^*)\right]$, then $Z_N(t) = \sigma_N^{-1} S_{[Nt]} \Rightarrow W_d(t)$*

The integration properties of a sum of two series is dominated by the series with the largest integration order. This implies that $\left((1-L)^{1-d}-(1-L)\right)x_t$ behaves asymptotically equivalently to the series $(1-L)^{1-d}x_t$ when $d>0$. We have some other useful results, when $-0.5<d<0.5$,

$$S_{00} \xrightarrow{a.s.} \Sigma_{00} \tag{14}$$
$$\beta_i' S_{11}^{ii} \beta_i \xrightarrow{a.s.} \Sigma_{\beta_i\beta_i} \tag{15}$$
$$\beta_i' S_{10}^{i} \xrightarrow{a.s.} \Sigma_{\beta_i 0} \tag{16}$$
$$T^{-1} S_{11}^{ii} \xrightarrow{w} \int_0^1 W_{d_i} W_{d_i}' du \tag{17}$$
$$S_{10}^{i} \xrightarrow{w} \int_0^1 W_{d_i} dW' \tag{18}$$

The first result is a standard one. The following two results are due to the fact that $\beta_i'\left((1-L)^{1-d_i}-(1-L)\right)x_t$ is $I(0)$ and the last two are derived from Theorem 1 above. The formal proofs would just differ marginally from the proofs in Johansen (1995) pages. 146-147, and are thus omitted. If $0.5\leq d\leq 1$ then the following result replaces the last two

$$S_{11}^{ii} \xrightarrow{a.s.} \Sigma_{d_i d_i} \tag{19}$$
$$S_{10}^{i} \xrightarrow{a.s.} \Sigma_{d_i 0} \tag{20}$$

where $\Sigma_{d_i d_i}$ denoted the covariance matrix of fractionally integrated variables. Beran (1994) pages. 73-77 and references therein can be used to demonstrate how to normalize and under which conditions these moments converge to a distribution.

Consider the rank zero hypothesis against that of full rank using the trace statistic $\left(-T\sum_{i=r+1}^{k} ln\left(1-\lambda_i\right)\right)$. Note that (see Johansen (1995) p.152)

$$-T\sum_{i=1}^{k}\left(1-\lambda_i\right) = -T\sum_{i=1}^{k}\lambda_i + o_p\left(1\right) \tag{21}$$

The equation

$$\left|\lambda S_{11}^{ii} - S_{10}^{i} S_{00}^{-1} S_{01}^{i}\right| = 0 \tag{22}$$

or equivalently

$$\left|\lambda I - \left(S_{11}^{ii}\right)^{-1} S_{10}^{i} S_{00}^{-1} S_{01}^{i}\right| = 0 \tag{23}$$

may be used to show that

$$T\sum_{i=1}^{p}\lambda_i = Ttr\left(\left(S_{11}^{ii}\right)^{-1} S_{10}^{i} S_{00}^{-1} S_{01}^{i}\right) \tag{24}$$

Table 1: Asymptotic distribution trace statistic 95%

			d		
$p-r$	1	0.9	0.8	0.7	0.6
1	4.039	3.926	3.783	3.768	3.658
2	12.300	11.726	11.204	10.740	10.295
3	24.252	22.866	21.463	20.144	18.803
4	40.205	37.433	34.587	31.993	29.897
5	59.567	55.153	50.585	46.365	43.054
6	83.317	76.730	70.443	64.464	59.257
7	111.054	101.802	92.932	84.561	77.623
8	142.459	130.475	118.701	107.703	98.323
9	178.084	162.551	147.442	133.333	121.599
10	217.120	198.663	179.706	162.345	147.355

Using the results above, the quantity $\left(S_{11}^{ii}\right)^{-1} S_{10}^{i} S_{00}^{-1} S_{01}^{i}$ goes in distribution towards $\left(\int_0^1 W_{d_i} W_{d_i}' du\right)^{-1} \int_0^1 W_{d_i} dW' \, \Sigma_{00}^{-1} \int_0^1 W_{d_i} dW'$. Defining the standard fractional Brownian motion $B = \Sigma_{00}^{-1/2} W_{d_i}$ the asymptotic distribution of the trace test is

$$tr\left(\int_0^1 B_{d_i} dB' \left(\int_0^1 B_{d_i} B_{d_i}' du\right)^{-1} \int_0^1 B_{d_i} dB'\right). \tag{25}$$

Once again, the proof is a simple modification of the one in Johansen (1995), pages. 158-160, and hence omitted. The distribution of the maximum eigenvalue test is derived in the same way.

The fractional testing procedure is more complicated compared to the case when the variables are not fractionally cointegrated. We suggest to estimate with one cointegrating relationship in a first step and test rank zero against rank one. If the test suggests that the rank null is rejected, we carry on by estimating with rank two and test for rank one against rank two. This is continued until null is not rejected. The maximum of the likelihood is proportional to the determinant of the residual covariance matrix, Σ_ε, which may be written

$$\begin{aligned} |\Sigma_\varepsilon| &= \left|S_{00|\beta_2,\ldots,\beta_{r^*}}\right| \prod_{i=1}^{r_1} \left(1-\lambda_i^1\right) \\ &= \left|S_{00|\beta_1,,\beta_3,\ldots,\beta_{r^*}}\right| \prod_{i=1}^{r_2} \left(1-\lambda_i^2\right) \\ &\vdots \end{aligned}$$

$$= \left|S_{00|\beta_1,\ldots,\beta_{r^*-1}}\right| \prod_{i=1}^{r_{r^*}} \left(1-\lambda_i^{r^*}\right)$$

where λ_i^j is the ith root of the solution to the eigenvalue problem where d_j has been used to filter the series. A problem is that there are r^* possible tests for rank $r-1$ against rank r. The likelihood ratio test is the maximum eigenvalue test. It is also possible to use the trace test. The value of the likelihood under the alternative is then proportional to the value gained when the multiplicative summation index is k. This does not correspond to a model of interest as then it would imply that the rank may be larger than the number of variables. The proposed solution to the choice of which of the r^* possible tests to use is to use the test which is most far from the null, i.e. choose the test which has the smallest $p-$value as is done in the standard framework.. Table (1) shows the asymptotic distribution of the trace statistic for different values of d.

5 Small sample properties

The properties of the model are evaluated by means of Monte Carlo simulation. The simulation has two aims: namely to analyze the variability in the cointegrating relation and the properties of the tests.

A two equation model (see Engle and Granger, 1987, or Banerjee et al., 1993, p.137) is generated by

$$\begin{aligned} x_t + \beta y_t &= u_t && (26)\\ x_t + \alpha y_t &= e_t && (27)\\ (1-L)^{1-d_u} u_t &= \varepsilon_{1t} && (28)\\ (1-L)^{1-d_e} e_t &= \varepsilon_{2t} && (29) \end{aligned}$$

where ε_{1t} and ε_{2t} both are independently and standard bivariate normally distributed with expectation zero, variances equal to σ_1^2 and σ_2^2 respectively and a covariance denoted by σ_{12}. This model has many interesting special cases. When u_t and e_t are not mean reverting, a system in differences is the correct model to analyze.. If $1-d_u = 1-d_e = 0$ a system in levels emerge, if one of u_t and e_t is mean reverting and the other is not then the system is cointegrated, e.g. if $1-d_u = 0$ and $1-d_e = 1$ then $[1,\beta]'$ is a cointegrating vector. A fractional cointegrating model arises when either of d_u and d_e is not integer valued and the other is zero. After some simple manipulation

(see Banerjee et al., 1993, for the case when e_t is an AR(1) process),

$$\begin{bmatrix} \Delta x_t \\ \Delta y_t \end{bmatrix} = (\beta-\alpha)^{-1} \begin{bmatrix} \alpha \\ -1 \end{bmatrix} \left[(1-L)^{1-d_u} - (1-L)\right] \begin{bmatrix} 1 & \beta \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \begin{bmatrix} \zeta_{1t} \\ \zeta_{2t} \end{bmatrix} \tag{30}$$

where ζ_{it} is a linear combination of ε_{1t} and ε_{2t}. This model has been analyzed by Cheung and Lai (1993) when analyzing the power of the GPH test for fractional cointegration. They put $\alpha = 1$ and $\beta = 2$. Estimation by the Johansen method when there is fractional cointegration is analyzed by Andersson and Gredenhoff (1998) using the same model and parameters. They found that the consequences of ignoring fractional cointegration may be severe for the estimates of the matrix $\Pi = \alpha\beta'$, using familiar notation.

The same parameter values are used in our Monte Carlo simulation. The estimated parameters are standardized such that one should compare with 0.5 and -1 for the estimate of β and α respectively. The cointegrating rank is assumed to be one and the fractional parameter used is 0.25, 0.45, 0.75, 0.85, 0.95 and 1. The sample sizes are 50, 100, 200 and 400 and the number of replicates is 1000. The difference parameter d is treated in three ways, 1) $d = 1$, as in usual maximum likelihood method, 2) $d = d_{DGP}$, i.e. the parameter is assumed known, and 3) estimated. The number of replicates when calculating small sample critical values is 50000 and 10000 when analyzing power. For the power comparison the following additional values of d are used, 0.05, 0.15 and 0.35.

In Tables (2) to (4) , the results concerning α, β and d are displayed for sample size 100. Tables for the other sample sizes are given in Appendix B. Most of the results are expected, for instance $\sqrt{MSE}$ decreases with increasing d. The same is true for the bias, except for α where the picture is unclear. In larger sample both the bias and $\sqrt{MSE}$ decrease. This last conclusion is somewhat contradictive to the one of Andersson and Gredenhoff (1998) that claims that $\Pi = \alpha\beta'$ has an increasing bias with increasing sample size. Comparing the estimators, the standard method have both larger bias and $\sqrt{MSE}$ compared both to knowing and estimating d for values of d up to 0.85. For the largest values of d the difference is small between methods, both the standard method and estimating d have small bias and $\sqrt{MSE}$. Overall, the Monte Carlo indicates that estimating d gives results that are more close to the results of knowing d than using the standard method. The estimator of d has larger bias and $\sqrt{MSE}$ than the ones reported by Sowell (1992).

When comparing the power it is standard practice to size adjust the tests. As seen above, the distribution of the test statistic depends on the estimated value of d under the alternative, hence the nominal size also depends on

Table 2: Bias and rmse of β, n=100, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	.11827	3.17360	-.01310	.12159	-.03067	.59866
0.45	-.00902	.09198	-.00200	.03357	-.00331	.04130
0.75	-.00065	.01352	-.00031	.01170	-.00041	.01247
0.85	-.00039	00944	-.00023	.00901	-.00032	.00932
0.95	-.00022	.00679	-.00018	.00676	-.00024	.00691
1.00	-.00016	.00582	-.00016	.00582	-.00020	.00594

Table 3: Bias and rmse of α, n=100, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	.42798	19.92663	-.02948	4.87964	-.12445	6.95811
0.45	-.01023	4.49964	.09079	.42642	.10117	.54120
0.75	.02518	.18036	.02399	.16914	.02519	.17234
0.85	.01598	.13795	.01620	.13541	.01672	.13689
0.95	.11663	.01219	.11627	.01232	.11702	.01260
1.00	.01106	.10945	.01106	.10945	.01129	.11000

d. This makes it very difficult to size adjust the power when estimating d. A power comparison is made between the standard method and when d is assumed known, bearing in mind that the previous results indicate that the model is, in some sense, closer to the estimation method when d is known than the standard estimation method. The null is no cointegrating relations and the alternative is one cointegrating relation and the result is presented in Table (5). The power increases with d and with sample size for both methods but the increase in power differs substantially where knowing d has the larger power. The same conclusions are made if the maximum eigenvalue test were used instead. It is interesting to note that the maximum eigenvalue test has larger power than the trace test for *all* cases considered, although the difference is small.

6 Conclusion

In this paper a multivariate cointegrated model is presented. The main advantage of this model compared to previous models is that departures from equilibrium are not restricted to be ruled by a process which is integrated

Table 4: Bias and rmse of d, n=100, 1000 replicates

	$d = \hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$
0.25	.15006	.79755
0.45	.05902	.37054
0.75	.09101	.22518
0.85	.09523	.20593
0.95	.09147	.18998
1.00	.08818	.18203

Table 5: Size adjusted power of trace test based on 10000 replicates, 5% test

	$n = 50$		$n = 100$		$n = 200$	
d_{DGP}	$d = 1$	$d = d_{DGP}$	$d = 1$	$d = d_{DGP}$	$d = 1$	$d = d_{DGP}$
0.05	.05640	.05860	.05750	.06920	.06280	.08660
0.15	.08820	.11270	.12570	.21720	.17920	.46270
0.25	.15630	.23680	.28660	.54620	.45380	.93210
0.35	.27540	.42280	.53340	.86610	.75510	.99920
0.45	.44990	.64490	.75850	.98140	.91490	1.000
0.75	.98930	.99550	1.000	1.000	1.000	1.000
0.85	.99950	.99940	1.000	1.000		
0.95	.99990	.99990	1.000	1.000		

of order zero but may be fractionally integrated. Hence, a much broader class of processes may be approximated by this model. Estimation of the involved parameters and testing for the number of cointegrating vectors are discussed. Further, the Granger representation theorem is extended to include the proposed fractional cointegration model. A Monte Carlo simulation study is conducted to evaluate the estimation and testing procedure. The result suggest that the efficiency loss by assuming a fractional cointegration model instead of the standard when the standard is true is minor while the consequences of ignoring fractional cointegration may be severe.

References

Andersson, M.K. and Gredenhoff, M.P. (1998) Power and bias of likelihood based inference in the cointegration model under fractional cointegration, *Working Paper Series in Economics and Finance*, 221, Stockholm School of Economics.

Anderson, T.W. (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics*, 22, 327-351.

Baillie, R. and Bollerslev, T. (1994) Cointegration, fractional cointegration and exchange rate dynamics, *The Journal of Finance*, 49, 737-745.

Booth, G.G. and Tse, Y. (1995) Long memory in interest rate futures markets: A fractional cointegration analysis, *The Journal of Futures Markets*, 15, 573-584.

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F. (1993) *Co-integration, Error Correction and the Econometric Analysis of Non-stationary Data*, Oxford University Press, Oxford.

Beran, J. (1994) *Statistics for long-memory processes*, Chapman & Hall, New York.

Cheung, Y-W. and Lai, K. (1993) A fractional cointegration analysis of purchasing parity, *Journal of Business & Economic Statistics*, 11, 103-112.

Efron, B. and Tibshirani, R.J. *An introduction to the Bootstrap*, Chapman & Hall, New York.

Engel, R.F. and Granger, C.W.J. (1987) ointegration and error correction: Representation, estimation and testing, *Econometrica*, 55, 251-276.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification, *Journal of Econometrics*, 16, 121-130.

Granger, C.W.J. and Joyeux, R. (1980) An introduction to long memory time series and fractional differencing, *Journal of Time series Analysis*, 1, 15-39.

Hosking, J.R.M. (1981) Fractional differencing, *Biometrika*, 73, 217-221.

Hylleberg, S. and Mizon, G.E. (1989) Cointegration and error correction mechanisms, *The Economic Journal*, (supplement) 99, 113-125.

Johansen, S. (1988) Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59, 1551-1580.

Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Auto-regressive Models*, Oxford University Press, Oxford.

Johansen, S. and Juselius, K. (1995) Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK, *Journal of Econometrics*, 52, 211-244.

Li, H. and Maddala, G.S. (1997) Bootstrapping cointegrating regressions, *Journal of Econometrics*, 80, 297-318.

Mandelbrot, B.B. and van Ness, J.W. (1968) Fractional Brownian motions, fractional noise and applications, *SIAM rev.*, 10, 422-437.

Sowell, F. (1986) Fractionally integrated vector time series, Ph.D. Dissertation, Duke University.

Sowell, F. (1990) The fractional unit root distribution, *Econometrica*, 58, 495-505.

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

Velu, R.P., and Reinsel, G.C. (1990) Reduced rank regression with autoregressive errors, *Journal of Economcetrics*, 35, 317-335.

Appendix A: Proof of Granger representation Theorem

Proof of Granger Representation Theorem.

Multiply (13) by α' and $\alpha_{\perp}^{'}$ respectively

$$\begin{aligned} \alpha' A^1(L)(1-L)x_t + \alpha'\alpha D(L)\beta' x_t &= \alpha^{'} e_t \quad (31) \\ \alpha_{\perp}^{'} A^1(L)(1-L)x_t &= \alpha_{\perp}^{'} e_t \quad (32) \end{aligned}$$

Define the new variables $w_t = (\beta'\beta)^{-1}\beta' x_t$ and $v_t = (\beta_{\perp}'\beta_{\perp})^{-1}\beta_{\perp}'\Delta x_t$. Further define $\bar{\beta} = \beta(\beta'\beta)^{-1}$ and $\bar{\beta}_{\perp} = \beta_{\perp}(\beta_{\perp}'\beta_{\perp})^{-1}$. Let $R = (\beta, \beta_{\perp})$ then is has full rank and $R'(R'R)R = I$ which implies $(\beta'\bar{\alpha} + \beta_{\perp}'\bar{\alpha}_{\perp}) = I$, then

$$\Delta x_t = \left(\beta'\bar{\beta} + \beta_{\perp}'\bar{\beta}_{\perp}\right)\Delta x_t = \beta\Delta w_t + \beta_{\perp} v_t \quad (33)$$

Substitute this into (31) and (32),

$$\begin{aligned} \alpha' A^1(L)\beta\Delta w_t + \alpha' A^1(L)\beta_{\perp} v_t + \alpha'\alpha D(L)\beta'\beta w_t &= \alpha^{'} e_t \quad (34) \\ \alpha_{\perp}^{'} A^1(L)\beta\Delta w_t + \alpha_{\perp}^{'} A^1(L)\beta_{\perp} v_t &= \alpha_{\perp}^{'} e_t \quad (35) \end{aligned}$$

and in matrix form

$$\tilde{A}(L)\begin{bmatrix} w_t \\ v_t \end{bmatrix} = \begin{bmatrix} \alpha^{'} \\ \alpha_{\perp}^{'} \end{bmatrix} e_t \quad (36)$$

where

$$\tilde{A}(z) = \begin{bmatrix} \alpha'\alpha D(z)\beta'\beta + \alpha' A^1(z)\beta(1-z) & \alpha' A^1(L)\beta_{\perp} \\ \alpha_{\perp}^{'} A^1(L)\beta(1-z) & \alpha_{\perp}^{'} A^1(L)\beta_{\perp} \end{bmatrix} \quad (37)$$

For $z = 1$ the determinant is

$$\left|\tilde{A}(1)\right| = |\alpha'\alpha|\,|D(1)|\,|\beta'\beta|\left|\alpha_{\perp}^{'} A^1(1)\beta_{\perp}\right| \quad (38)$$

which is nonzero by assumption and by the fact that $|D(1)| \neq 0$ when $0 < d \leq 1$. Substitute for $A^1(z)$ in $\tilde{A}(z)$ in terms of $\alpha D(L)\beta'$ and $A(L)$ we end up with, for $z \neq 1$,

$$\tilde{A}(z) = \begin{bmatrix} \alpha' \\ \alpha_{\perp}^{'} \end{bmatrix} A^1(z)\left[\beta, \beta_{\perp}(1-z)^{-1}\right] \quad (39)$$

with determinant

$$\left|\tilde{A}(z)\right| = |(\alpha, \alpha_{\perp})|\left|A^1(z)\right|\,|(\beta, \beta_{\perp})|\,(1-z)^{-(n-r)} \quad (40)$$

Excluding $z \neq 1$ the remaining roots is outside the unit circle and it is possible to invert $\tilde{A}(z)$. Hence, w_t and v_t may be given initial distributions such that they become mean reverting. As we have the relationship $\Delta x_t = \beta \Delta w_t + \beta_\perp v_t$ mean reverting of w_t and v_t implies mean reverting of Δx_t. Further $\beta' x_t = \beta' \beta w_t$ which is also mean reverting. This ends the proof of R1, R2 and R3.

Not that

$$\begin{aligned} \Delta x_t &= \beta \Delta w_t + \beta_\perp v_t \\ &= [\beta(1-L), \beta_\perp] \begin{bmatrix} w_t \\ v_t \end{bmatrix} \\ &= [\beta(1-L), \beta_\perp] \tilde{A}(z)^{-1} \begin{bmatrix} \alpha' \\ \alpha'_\perp \end{bmatrix} e_t \\ &= C(L) e_t \end{aligned} \tag{41}$$

which is the moving average representation and R4 is proved.

Evaluate the polynomial $C(L)$ in 1 then

$$\begin{aligned} C(1) &= [0, \beta_\perp] \tilde{A}(1)^{-1} \begin{bmatrix} \alpha' \\ \alpha'_\perp \end{bmatrix} \\ &= \beta_\perp \left(\alpha'_\perp A^1(1) \beta_\perp \right)^{-1} \alpha'_\perp \end{aligned} \tag{42}$$

which has rank $n - r$ by assumption and the definition of the orthogonal complements. Then R5 is proved..

R6 and R7 follows.

Above we had the expression

$$\Delta x_t = C(L) e_t \tag{43}$$

Use the decomposition $C(L) = C(1) + (1-L) C_1(1)$ then

$$\Delta x_t = C(1) e_t + (1-L) C_1(1) e_t \tag{44}$$

Integrating yields

$$x_t = x_0 + C(1) \sum_{i=1}^{t} e_i + S_t \tag{45}$$

where $S_t = C_1(L) e_t$, which ends the proofs.

Appendix B: Tables

Table 6: Asymptotic distribution max eigenvalue statistic 95%

			d		
$p-r$	1	0.9	0.8	0.7	0.6
1	4.039	3.926	3.783	3.768	3.658
2	11.230	10.707	10.139	9.651	9.293
3	17.795	16.799	15.947	15.057	14.336
4	24.367	22.952	21.726	20.554	19.426
5	30.379	28.570	27.038	25.448	24.175
6	36.265	34.228	32.275	30.409	28.896
7	42.465	40.174	37.554	35.383	33.682
8	48.451	45.630	42.894	40.565	38.606
9	54.617	51.589	48.600	45.609	43.471
10	60.738	57.295	54.057	50.907	48.094

Table 7: Asymptotic distribution trace statistic 90%

			d		
$p-r$	1	0.9	0.8	0.7	0.6
1	2.988	2.886	2.756	2.652	2.571
2	10.530	10.007	9.408	8.919	8.480
3	21.811	20.384	18.843	17.431	16.384
4	37.302	34.434	31.563	28.835	26.685
5	55.904	51.141	46.679	42.894	39.485
6	79.122	72.625	66.062	60.061	55.229
7	105.920	96.883	87.896	79.594	72.534
8	136.665	125.025	113.294	102.136	93.153
9	171.802	156.582	141.264	126.94	115.638
10	210.486	191.63	172.475	155.34	140.747

Table 8: Asymptotic distribution max eigenvalue statistic 90%

			d		
$p-r$	1	0.9	0.8	0.7	0.6
1	2.988	2.886	2.756	2.652	2.571
2	9.500	9.015	8.495	8.023	7.615
3	15.703	14.798	13.797	12.988	12.329
4	21.969	20.724	19.440	18.217	17.083
5	27.753	26.134	24.472	23.016	21.724
6	33.636	31.635	29.700	27.983	26.306
7	39.658	96.883	34.919	32.642	31.021
8	45.472	42.712	40.084	37.813	35.832
9	51.579	48.455	45.516	42.801	40.500
10	57.237	54.041	50.745	47.709	44.96

Table 9: Bias and rmse of β, n=50, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	.06106	2.71736	-.03136	.49571	-.06615	7.10189
0.45	-.00694	.91519	-.00481	.11133	.21099	6.76486
0.75	-.00088	.02437	-.00058	.02077	-.00100	.02335
0.85	-.00045	.01769	-.00034	.01658	-.00068	.01858
0.95	-.00023	.01331	-.00021	.01313	-.00049	.01445
1.00	-.00016	.01167	-.00016	.01167	-.00041	.01269

Table 10: Bias and rmse of α, n=50, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	-2.32403	99.80322	-.41883	6.87538	.05257	22.29025
0.45	-2.64326	89.24902	.00515	8.72210	1.01837	37.23937
0.75	.08890	.67998	.06718	.42663	.08831	.53026
0.85	.03922	.25018	.03839	.24384	.04535	.26704
0.95	.02741	.19420	.02743	.19375	.03064	.20422
1.00	.02406	.17861	.02406	.17861	.02630	.18630

Table 11: Bias and rmse of d, n=50, 1000 replicates

	$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$
0.25	.09820	2.21283
0.45	.11441	.75763
0.75	.16727	.38432
0.85	.34472	.17353
0.95	.31019	.16407
1.00	.15599	.29236

Table 12: Bias and rmse of β, n=200, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	.13025	5.30116	-.00867	.19107	-.00249	.08263
0.45	-.00738	.09967	.00005	.02139	-.00008	.02213
0.75	-.00012	.00819	.00006	.00683	.00005	.00687
0.85	-.00007	.00543	.00000	.00505	.00001	.00506
0.95	-.00004	.00370	-.00003	.00364	.00000	.00366
1.00	-.00003	.00308	-.00003	.00308	-.00001	.00309

Table 13: Bias and rmse of α, n=200, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	.56180	14.43707	.28929	5.29935	.13181	1.12982
0.45	5.05707	158.85227	.03502	.20994	.03803	.21636
0.75	.01160	.11674	.01037	.10912	.01090	.11052
0.85	.00709	.09134	.00676	.08906	.00704	.08994
0.95	.00518	.07764	.00495	.07702	.00499	.07744
1.00	.00436	.07262	.00436	.07262	.00457	.07314

Table 14: Bias and rmse of d, n=200, 1000 replicates

	$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$
0.25	.08598	.44248
0.45	.02936	.20673
0.75	.05577	.13922
0.85	.05782	.12936
0.95	.05570	.12014
1.00	.05383	.11588

Table 15: Bias and rmse of β, n=400, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	.02088	.58891	-.00182	.03367	-.00229	.03599
0.45	-.00416	.05196	-.00038	.01346	-.00025	.01365
0.75	-.00014	.00457	-.00011	.00387	-.00009	.00390
0.85	-.00010	.00285	-.00010	.00269	-.00009	.00269
0.95	-.00008	.00182	-.00008	.00180	-.00007	.00180
1.00	-.00007	.00146	-.00007	.00146	-.00006	.00147

Table 16: Bias and rmse of α, n=400, 1000 replicates

	$d=1$		$d=d_{DGP}$		$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$	$Bias$	$\sqrt{MSE}$
0.25	-.11705	7.49184	.04610	.27039	.05603	.31577
0.45	.05425	.37711	.01443	.13517	.01508	.13663
0.75	.00482	.08087	.00433	.07607	.00439	.07664
0.85	.00283	.06401	.00276	.06266	.00278	.06306
0.95	.00197	.05471	.00197	.05448	.00198	.05478
1.00	.00171	.05146	.00171	.05146	.00172	.05173

Table 17: Bias and rmse of d, n=400, 1000 replicates

	$d=\hat{d}$	
d_{DGP}	$Bias$	$\sqrt{MSE}$
0.25	.04880	.22907
0.45	.00650	.13129
0.75	.02896	.08810
0.85	.03078	.08154
0.95	.02994	.07594
1.00	.02901	.07344

Table 18: Size adjusted power of maximum eigenvalue test based on 10000 replicates, 5% test

	$n=50$		$n=100$		$n=200$	
d_{DGP}	$d=1$	$d=d_{DGP}$	$d=1$	$d=d_{DGP}$	$d=1$	$d=d_{DGP}$
0.05	.05840	.06090	.05980	.07000	.06340	.08350
0.15	.08940	.11390	.12640	.21720	.18860	.46530
0.25	.16100	.24090	.29200	.55990	.46500	.93930
0.35	.28280	.43660	.53940	.87810	.76060	.99930
0.45	.46280	.66260	.75950	.98500	.91240	1.000
0.75	.98950	.99630	1.000	1.000	1.000	1.000
0.85	.99940	.99960	1.000	1.000		
0.95	1.000	1.000	1.000	1.000		