Discussion Papers

Department of Economics
University of Copenhagen

No. 10-28

An Extension of Cointegration to Fractional Autoregressive Processes

Søren Johansen

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

An extension of cointegration to fractional autoregressive processes

Søren Johansen*
University of Copenhagen
and CREATES
Presented at
The Finnish Statistical Society
May 2010

October 15, 2010

Abstract

This paper contains an overview of some recent results on the statistical analysis of cofractional processes, see Johansen and Nielsen (2010b). We first give an brief summary of the analysis of cointegration in the vector autoregressive model and then show how this can be extended to fractional processes. The model allows the process X_t to be fractional of order d and cofractional of order $d-b \geq 0$; that is, there exist vectors β for which $\beta' X_t$ is fractional of order $d-b$. We analyse the Gaussian likelihood function to derive estimators and test statistics. The asymptotic properties are derived without the Gaussian assumption, under suitable moment conditions. We assume that the initial values are bounded and show that they do not influence the asymptotic analysis

The estimator of β is asymptotically mixed Gaussian and estimators of the remaining parameters are asymptotically Gaussian. The asymptotic distribution of the likelihood ratio test for cointegration rank is a functional of fractional Brownian motion.

Keywords: Cofractional processes, cointegration rank, fractional cointegration, likelihood inference, vector autoregressive model.

JEL Classification: C32.

*The author gratefully acknowledges support from Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation.
Address: Department of Economics, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen K. Denmark, Email: Soren.Johansen@econ.ku.dk

1 Introduction

Granger (1983) defined the notion of cointegration as a formulation of the phenomenon that nonstationary processes can have linear combinations that are stationary. It was his investigations of the relation between cointegration and error correction that brought modeling of vector autoregressions with unit roots and cointegration to the center of attention in applied and theoretical econometrics; see Engle and Granger (1987). We begin with a brief account of the properties and analysis of the cointegrated vector autoregressive model, CVAR.

This serves as background for the main topic, which is the generalization of this analysis to a class of fractional processes.

2 The I(1) cointegration model

2.1 Examples of cointegration from economics and climate research

One of the first examples of the statistical analysis of cointegration was the paper by Campbell and Shiller (1987). They considered a present value model for the price of a stock Y_t at the end of period t and the dividend y_t paid during period t, see Figure 1. The expectations hypothesis is expressed as

$$Y_t = \theta(1-\delta)\sum_{i=0}^{\infty}\delta^j E_t y_{t+i} + c,$$

where c and θ are positive constants and the discount factor δ is between 0 and 1. The notation $E_t y_{t+i}$ means model based conditional expectations of y_{t+i} given information in the data at the end of period t. By subtracting θy_t, the model is written as

$$Y_t - \theta y_t = \theta(1-\delta)\sum_{i=0}^{\infty}\delta^j E_t(y_{t+i} - y_t) + c.$$

If the processes y_t and Y_t are nonstationary and Δy_t and ΔY_t are stationary, the present value model implies that the right hand side and hence the left hand side are stationary. Thus, there is cointegration between Y_t and y_t with a cointegration vector $\beta' = (1, -\theta)$.

Another example is an analysis of measurements of mean annual temperature and height of sea level taken from Hansen et al. (2001). The variables are clearly trending and probably nonstationary. They are analysed from the point of view of cointegration in Johansen (2010) and Schmith et *al.* (2010), see Figure 2.

2.2 Integration and cointegration

We call a $p-$dimensional process X_t integrated of order 1, $I(1)$, if ΔX_t is stationary, and

$$\Delta X_t - E(\Delta X_t) = \sum_{n=0}^{\infty} C_n \varepsilon_{t-n},$$

where ε_t is i.i.d. $(0, \Omega)$, is a linear process with coefficients satisfying $\sum_{n=0}^{\infty}|C_n|^2 < \infty$ for which $\sum_{n=0}^{\infty} C_n \neq 0$, that is, a so-called $I(0)$ process. The expansion

$$C(z) = \sum_{n=0}^{\infty} C_n z^n = C + (1-z)\sum_{n=0}^{\infty} C_n^* z^n,$$

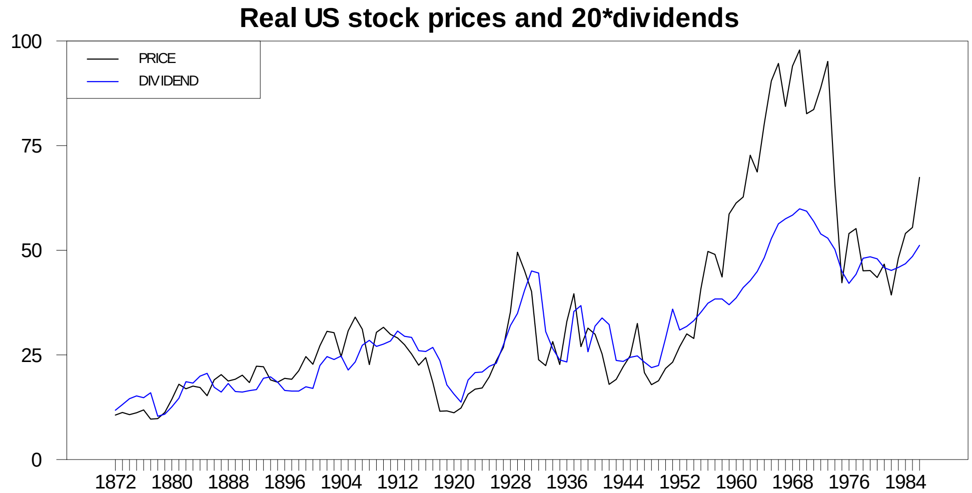

Figure 1: Data fra Campbell and Shiller (1987) of real US Stock prices and dividends (scaled). The data is annual from 1871 to 1986.

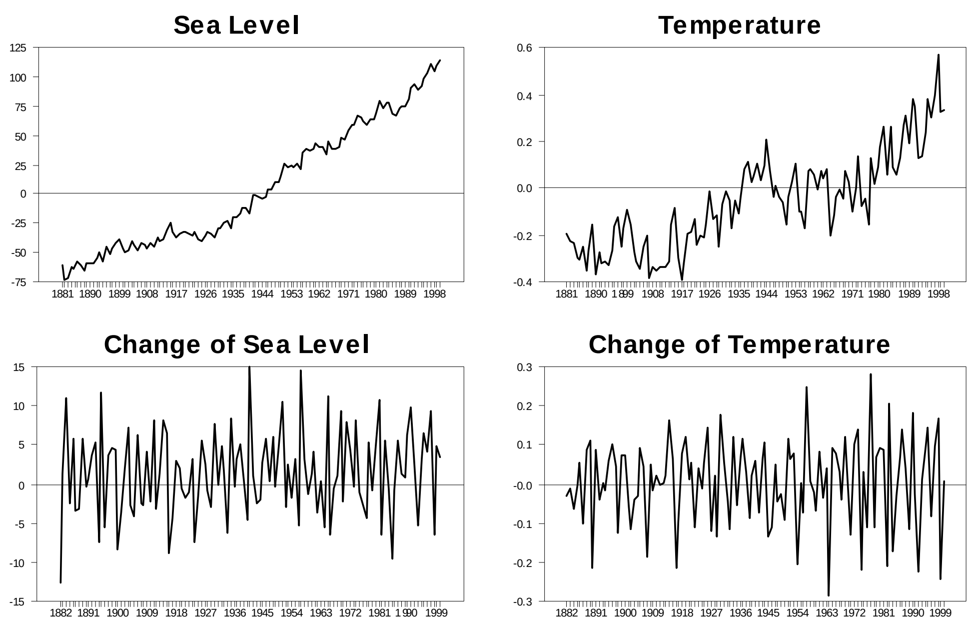

Figure 2: Plot of annual data of sea level and temperature anomalies in levels and differences from 1881 to 1995. Note the clear nonstationarity in the levels, which could be due to a stochastic trend or possibly a deterministic trend. The differences, however, behave like stationary processes.

where $C = C(1)$, shows that

$$X_t = \sum_{i=1}^{t} E(X_i) + C \sum_{i=1}^{t} \varepsilon_i + \sum_{n=0}^{\infty} C_n^* \varepsilon_{t-n} + X_0,$$

so that X_t is nonstationary because $C \neq 0$. We call a vector β a cointegrating vector if $\beta' X_t$ is stationary and the number of linearly independent cointegrating vectors is the cointegration rank.

The cointegrated vector autoregressive model, CVAR(k), for the $p-$dimensional process X_t is given by the equations

$$\mathcal{H}_r : \Delta X_t = \alpha\,(\beta' X_{t-1} + \rho' t) + \sum_{i=1}^{k-1} \Gamma_i\, \Delta X_{t-i} + \mu + \varepsilon_t, \tag{1}$$

where ε_t are i.i.d. with mean zero and variance Ω. The matrices α and β are $p \times r$ where $0 \leq r \leq p$.

There are many surveys of the theory of cointegration; see for instance Watson (1994) or Johansen (2006). The topic has become part of most textbooks in econometrics; see among others, Hamilton (1994), Lütkepohl (2006) or Johansen (1996). For a general account of the methodology of the cointegrated vector autoregressive model and its application to economic data; see Juselius (2006).

The process X_t is uniquely defined by (1) as a function of initial values, parameters and innovations $\varepsilon_1, \ldots, \varepsilon_t$. The properties of the solution of these equations are studied by means of the characteristic (matrix) polynomial

$$\Psi(z) = (1-z) I_p - \Pi z - (1-z) \sum_{i=1}^{k-1} \Gamma_i z^i. \tag{2}$$

The solution is given by the coefficients in the expansion of $C(z) = \Psi(z)^{-1}$. This has a pole at z if $\det \Psi(z) = 0$, and the position of the poles determines the stochastic properties of the solution of (1).

Example 1. A bivariate process is given for $t = 1, \ldots, T$ by the equations

$$\begin{aligned} \Delta X_{1t} &= \alpha_1 (X_{1t-1} - X_{2t-1}) + \varepsilon_{1t}, \\ \Delta X_{2t} &= \alpha_2 (X_{1t-1} - X_{2t-1}) + \varepsilon_{2t}. \end{aligned}$$

Subtracting the equations, it is seen that $y_t = X_{1t} - X_{2t}$ is autoregressive with one lag, and stationary if $|1 + \alpha_1 - \alpha_2| < 1$. Similarly we find that $S_t = \alpha_2 X_{1t} - \alpha_1 X_{2t}$ is a random walk, and that

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \frac{1}{\alpha_2 - \alpha_1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} S_t - \frac{1}{\alpha_2 - \alpha_1} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} y_t.$$

This shows, that when $|1+\alpha_1-\alpha_2| < 1$, X_t is $I(1)$, $X_{1t}-X_{2t}$ is stationary, and $\alpha_2 X_{1t} - \alpha_1 X_{2t}$ is a random walk, so that X_t is a cointegrated $I(1)$ process with cointegrating vector $\beta' = (1, -1)$. We call S_t a common stochastic trend and α the adjustment coefficients. ■

2.3 The Granger Representation Theorem

If the characteristic polynomial $\Psi(z)$ defined in (2) has a unit root, then $\Psi(1) = -\Pi$ is singular, of rank $r < p$, and the solution of (1) is not stationary. We denote by $\alpha_\perp$ a $p \times (p-r)$ matrix for which $\alpha'\alpha_\perp = 0$ and $(\alpha, \alpha_\perp)$ has full rank.

Theorem 1 *(The Granger Representation Theorem) If $\Psi(z)$, defined by (2), has unit roots and $\alpha'_\perp \Gamma \beta_\perp$ has full rank for $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$, then $\Psi(z)$ as a pole of order one at $z = 1$, and*

$$(1-z)\Psi(z)^{-1} = C(z) = C + (1-z)C^*(z) \tag{3}$$

for $|z| \leq 1 + \delta$ for some $\delta > 0$, and

$$C = \beta_\perp(\alpha'_\perp \Gamma \beta_\perp)^{-1}\alpha'_\perp. \tag{4}$$

It follows that solution, X_t, of equation (1) has moving average representation

$$X_t = C\sum_{i=1}^{t} \varepsilon_i + C\mu t + \sum_{i=0}^{\infty} C_i^*(\varepsilon_{t-i} + \mu + \alpha\rho'(t-i)) + A, \tag{5}$$

where A depends on initial values, so that $\beta' A = 0$.

For a proof see Johansen (1996, 2008). This result implies that ΔX_t and $\beta' X_t$ are stationary, so that X_t is a cointegrated $I(1)$ process with r cointegration vectors β, because $\beta' C = 0$, and $p - r$ common stochastic trends $\alpha'_\perp \sum_{i=1}^t \varepsilon_i$.

This representation is useful for analysing the role of deterministic terms in the equation and for analysing the asymptotic properties of the process. Thus, the drift term μ is cumulated to the trend $C\mu t$, whereas $\alpha\rho'$ is not cumulated because $C\alpha = 0$. In the direction $C\mu$ the process is asymptotically dominated by the linear term, but orthogonal to that the random walk dominates.

2.4 Hypotheses on the rank

The models $\mathcal{H}_r$ are nested

$$\mathcal{H}_0 \subset \cdots \subset \mathcal{H}_r \subset \cdots \subset \mathcal{H}_p.$$

Here $\mathcal{H}_p$ is the unrestricted vector autoregressive model, so that α and β are unrestricted $p \times p$ matrices. The model $\mathcal{H}_0$ corresponds to the restriction $\alpha = \beta = 0$, which is the vector autoregressive model for the process in differences. Note that in order to have nested models, we allow in $\mathcal{H}_r$ all processes with rank less than or equal to r.

This formulation allows us to derive likelihood ratio tests for the hypothesis $\mathcal{H}_r$ in the unrestricted model $\mathcal{H}_p$. These tests can be applied to check if one's prior knowledge of the number of cointegration relations is consistent with the data, or alternatively to construct an estimator of the cointegration rank.

Note that when the cointegration rank is r, the number of common trends is $p - r$. Thus if one can interpret the presence of r cointegration relations, one should also interpret the presence of $p - r$ independent stochastic trends or $p - r$ driving forces in the data.

2.5 Hypotheses on long-run coefficients

The purpose of modeling economic data is to test hypotheses on the coefficients, thereby investigating whether the data supports an economic hypothesis or rejects it. If X_t consists of the log of a price index in US and Australia, and the log exchange rate is e_t, then the law of one price, $p^{au} - p^{us} + e = 0$, is formulated in the model as the hypothesis that $(1, -1, 1)$ is a cointegrating relation or $p_t^{au} - p_t^{us} + e_t$ is stationary. Similarly, the hypothesis of price homogeneity $p^{au} = p^{us}$ is formulated as $(1, -1, 0)$ is a cointegrating vector or $p_t^{au} - p_t^{us} + e_t$ is stationary.

2.6 Hypotheses on adjustment coefficients

The coefficients in α measure how the process adjusts to disequilibrium errors. Of particular interest is the hypothesis of weak exogeneity, which is the hypothesis that some rows of α are zero; see Engle, Hendry and Richard (1983).

The process X_t is decomposed as $X_t = (X_{1t}', X_{2t}')'$, and the matrices are decomposed similarly so that the model equations without deterministic terms become

$$\begin{aligned} \Delta X_{1t} &= \alpha_1 \beta' X_{t-1} + \textstyle\sum_{i=1}^{k-1} \Gamma_{1i} \Delta X_{t-i} + \varepsilon_{1t}, \\ \Delta X_{2t} &= \alpha_2 \beta' X_{t-1} + \textstyle\sum_{i=1}^{k-1} \Gamma_{2i} \Delta X_{t-i} + \varepsilon_{2t}. \end{aligned}$$

The conditional model for ΔX_{1t} given ΔX_{2t} and the past is

$$\Delta X_{1t} = \omega \Delta X_{2t} + (\alpha_1 - \omega \alpha_2) \beta' X_{t-1} + \sum_{i=1}^{k-1} (\Gamma_{1i} - \omega \Gamma_{2i}) \Delta X_{t-i} + \varepsilon_{1t} - \omega \varepsilon_{2t}, \tag{6}$$

where $\omega = \Omega_{12} \Omega_{22}^{-1}$. If $\alpha_2 = 0$, there is no levels feedback from $\beta' X_{t-1}$ to ΔX_{2t}, and if the errors are Gaussian, X_{2t} is called weakly exogenous for α_1 and β. In this case only the conditional model (6) need to be analysed, because the error term $\varepsilon_{1t} - \omega \varepsilon_{2t}$ is independent of the error term ε_{2t}.

2.7 Likelihood analysis of the $I(1)$ model

The model equations are nonlinear in α and β. Nevertheless the algorithm of reduced rank regression, see Anderson (1951), allows one to calculate the maximum likelihood estimators explicitly, by an eigenvalue routine. In model (1) we stack X_{t-1} and t and find the equation

$$\mathcal{H}_r : \Delta X_t = \alpha \begin{pmatrix} \beta \\ \rho \end{pmatrix}' \begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \sum_{i=1}^{k-1} \Gamma_i \, \Delta X_{t-i} + \mu + \varepsilon_t. \tag{7}$$

The maximum likelihood estimator of $(\beta', \rho')'$ is given by reduced rank of ΔX_t on $(X_{t-1}', t)'$ corrected for a constant and $\Delta X_{t-i}, i = 1, \ldots, k-1$. This shows that $(\beta', \rho')'$ are the r canonical variates that are most correlated with ΔX_t corrected for a constant and lagged differences.

The test statistic is simply expressed in terms of the eigenvalues, which are the squared canonical correlations. Such a test statistic was already considered by Bartlett (1948).

2.8 Asymptotic distribution of the rank test

The asymptotic distribution of the likelihood ratio test for rank in model (1), involves Brownian motion which appears as the limit

$$T^{-1/2}\sum_{i=1}^{[Tu]}\varepsilon_i \Longrightarrow W(u) \text{ on } D^p[0,1];$$

see Billingsley (1968).

Theorem 2 *Let ε_t be i.i.d. $(0,\Omega)$ in model (1). Under the assumptions that the cointegration rank is r, the asymptotic distribution of the likelihood ratio test statistic for rank r is given by*

$$-2logLR(\mathcal{H}_r|\mathcal{H}_p) \xrightarrow{d} tr\{\int_0^1 (dB)F'(\int_0^1 FF'du)^{-1}\int_0^1 F(dB)'\}, \tag{8}$$

where F is defined by

$$F(u) = \left(\begin{array}{c} B(u) \\ u \end{array}\middle| 1\right),$$

and $B(u)$ is the $p-r$ dimensional standard Brownian motion.

Note that the expression for F reflects that of $(X_{t-1}', t)'$ corrected for a constant, but the lagged differences have no influence in the limit. The limit distribution is tabulated by simulation, as it is analytically quite intractable. Note that the limit distribution does not depend on the parameters, but only on $p-r$, the number of common trends, and the type of deterministic term.

In the model without deterministics the same result holds, but with $F(u) = B(u)$. A special case of this, for $p = 1$, is the Dickey-Fuller test and the distributions (8) are called the Dickey–Fuller distributions with $p-r$ degrees of freedom; see Dickey and Fuller (1981).

2.9 Asymptotic distribution of the estimators

The estimator $(\hat{\beta}, \hat{\rho})$ suitably normalized, converges to a mixed Gaussian distribution, even when estimated under continuously differentiable restrictions, see Johansen (1991). This result implies that likelihood ratio tests for hypotheses on (β, ρ) are asymptotically χ^2 distributed. Furthermore the estimators of the adjustment parameters α and the short-run parameters Γ_i are asymptotically Gaussian and asymptotically independent of the estimator for (β, ρ).

It is therefore possible to scale the deviations $\hat{\beta} - \beta$ in order to obtain an asymptotic Gaussian distribution. Note that the scaling matrix is not an estimate of the asymptotic variance of $\hat{\beta}$, but an estimate of the asymptotic *conditional* variance given the information in the data. It is therefore not the asymptotic distribution of $\hat{\beta}$ that is used for inference, but the *conditional* distribution given the information; see Basawa and Scott (1983) or Johansen (1995) for a discussion.

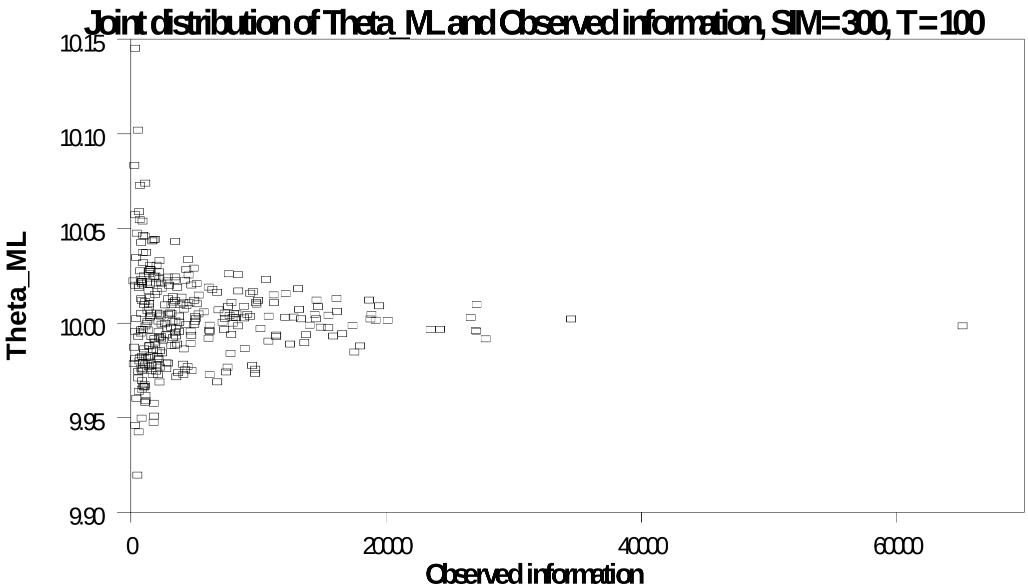

Figure 3: The joint distribution of $\hat{\theta}$ and the observed information $(\sum_{i=1}^{T} x_{2t-1}^2/\hat{\sigma}^2)$ in the model $x_{1t} = \theta x_{2t-1} + \varepsilon_{1t}$ and $\Delta x_{2t} = \varepsilon_{2t}$. Note that the larger the information, the smaller is the uncertainty in the estimate $\hat{\theta}$.

3 The CVAR$_{d,b}(k)$ model for fractional processes

The fractional process have been studied for many years, see for instance the monograph by Beran (1994), and have applications in for instance hydrology, cognitive science and finance. Such analyses are typically univariate but a statistical theory is developing for the multivariate processes; see for instance Marinucci and Robinson (2000) or Jeganathan (1999). The autoregressive models have turned out to be a useful tool in applied work, and what we want to survey here is a model and its asymptotic analysis, that combines the usefulness of the autoregressive model with the fractional processes; see Johansen and Nielsen (2010b) and Lasak (2008a,b) for a slight different approach to some of the results.

The fractional processes are linear processes generated by the fractional coefficients defined by the expansion

$$(1-z)^{-d} = \sum_{n=0}^{\infty} (-1)^n \binom{-d}{n} z^n = \sum_{n=0}^{\infty} \pi_n(d) z^n,$$

and the coefficients satisfy

$$\pi_n(d) = \frac{d(d+1)\dots(d+n-1)}{n!} = \frac{n^{d-1}}{\Gamma(d)}(1+o(1)).$$

The basic fractional process is generated by the equation

$$\Delta^d X_t = \varepsilon_t, \tag{9}$$

with solution

$$X_t = \Delta^{-d}\varepsilon_t = \sum_{n=0}^{\infty} \pi_n(d)\varepsilon_{t-n},\ d < 1/2.$$

If $d > 1/2$ the infinite sum does not exist and we define instead

$$X_t = \Delta_+^{-d}\varepsilon_t = \Delta^{-d}(\varepsilon_t 1_{\{t\geq 1\}}) = \sum_{n=0}^{t-1} \pi_n(d)\varepsilon_{t-n},\ d \geq 1/2,$$

which is the solution of equation (9) if $\varepsilon_t = 0, t \leq 0$. For $d = 1$, we get $\pi_n(1) = 1$, and $X_t = \Delta_+^{-1}\varepsilon_t = \sum_{i=1}^t \varepsilon_i$, an $I(1)$ process. For $d = 2$, we get $\pi_n(2) = n+1$, and $X_t = \Delta_+^{-2}\varepsilon_t = \sum_{i=1}^t \sum_{j=1}^i \varepsilon_j$, an $I(2)$ process.

In general we call X_t fractional of order d if $\Delta^d X_t$ is fractional of order zero, $I(0)$, and β a cofractional vector if $\beta' X_t$ is fractional of order $d-b$.

We want an autoregressive equation that generates fractional processes in order to be able to mimic the theory of the cointegrated VAR analysis to fractional processes.

Example 2 Assume X_t is fractional of order d and $\beta' X_t$ is fractional of order $d-b$, where b is the "cointegration gap." We formulate this as

$$\begin{aligned} \beta' X_t &= \Delta^{-(d-b)} u_{1t}, \\ \gamma' X_t &= \Delta^{-d} u_{2t}, \end{aligned}$$

for some γ for which (β, γ) is full rank and u_t is $I(0)$. It follows by solving the equation, using

$$\gamma_\perp(\beta'\gamma_\perp)^{-1}\beta' + \beta_\perp(\gamma'\beta_\perp)^{-1}\gamma' = I_p,$$

that we

$$\Delta^d X_t = -\gamma_\perp(\beta'\gamma_\perp)^{-1}\beta'(I-\Delta^b)\Delta^{d-b}X_t + \gamma_\perp(\beta'\gamma_\perp)^{-1}u_{1t} + \beta_\perp(\gamma'\beta_\perp)^{-1}u_{2t},$$

or, for $\alpha = -\gamma_\perp(\beta'\gamma_\perp)^{-1}$,

$$\Delta^d X_t = \alpha\beta'(I-\Delta^b)\Delta^{d-b}X_t + \tilde{u}_t.$$

This model is more or less the model suggested by Granger (1986), and justifies the model we now consider which allows for more lags. ■

The fractional vector autoregressive model, $\text{CVAR}_{d,b}(k)$, is defined by

$$\mathcal{H}_r : \Delta^d X_t = \alpha\beta'\Delta^{d-b}L_b X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t,\ t = 1, \ldots, T, \tag{10}$$

where ε_t are i.i.d. $(0, \Omega)$, $d \geq b$, and α and β are $p \times r$. The parameters are otherwise unrestricted, and $L_b = 1 - \Delta^b$ is the fractional lag operator. Note that the expansion of $L_b = 1 - \Delta^b$ has no term in $L^0 = 1$ and thus only lagged disequilibrium errors appear in (10).

This is a model for the observations $X_t,\ t = 1, \ldots, T$, but just as for the CVAR we need the initial values to calculate the differences. For model (10), we need infinitely many initial values in order to be able to calculate $\Delta^d X_t = \sum_{n=0}^{\infty} \pi_n(-d) X_{t-n}$. When $d > 0$ the coefficients are summable $\sum_{n=0}^{\infty} |\pi_n(-d)| \leq c \sum_{n=0}^{\infty} n^{-d-1} < \infty$, so $\Delta^d X_t$ exists if the initial values are bounded, which we assume from now on. The model is formulated so that the usual results from the $CVAR$ can be proved also for the new model. Using the polynomial $\Psi(z)$, see (2), the model can be formulated as $\Pi(L)X_t = \Psi(L_b)\Delta^{d-b}X_t = \varepsilon_t$.

That is, $\Delta^{d-b}X_t$ satisfies a vector autoregression in the lag operator L_b rather than the standard lag operator $L = L_1$. The CVAR model, analyzed in section 2, appears as the special case $d = b = 1$, and the interpretation of the model parameters is similar, i.e., the columns of β are the cofractional relations and α are the adjustment or loading coefficients.

Just as for the usual VAR model, the stochastic properties of X_t depend on the characteristic function $\Pi(z) = \Psi((1-(1-z)^b)(1-z)^{d-b}$ associated with (10).

We consider throughout the case $b \leq d$, so that $\Delta^{d-b}X_t$ can be calculated for bounded initial values, and for the asymptotic analysis we consider $b \geq 1/2$, which is the "strong cointegration" case in the terminology of Hualde and Robinson (2010). A consequence is that asymptotic inference for the rank involves fractional Brownian motion rather that the Brownian motion entering in (8).

3.1 Solution of the fractional autoregressive equations

We consider equation (10) written as $\Pi(L)X_t = \sum_{n=0}^{\infty} \Pi_n X_{t-n} = \varepsilon_t,\ t = 1, \ldots, T$. In order to derive a general expression for the solution in terms of initial values $X_{-n}, n = 0, 1, \ldots,$ and random shocks $\varepsilon_1, \ldots, \varepsilon_t$, we define two operators, see Johansen (2008),

$$\Pi_+(L)X_t = 1_{\{t\geq 1\}} \sum_{i=0}^{t-1} \Pi_i X_{t-i} \text{ and } \Pi_-(L)X_t = \sum_{i=t}^{\infty} \Pi_i X_{t-i}.$$

Here the operator $\Pi_+(L)$ is defined for any sequence and is invertible on sequences that are zero for $t \leq 0$. The coefficients of the inverse are found by expanding $\Pi(z)^{-1}$ around zero. The process $\Pi_-(L)X_t$ is defined, if we assume initial values of X_t fixed and bounded.

The solution of the equation $\Pi(L)X_t = \varepsilon_t$ is found by using $\Pi_+(L)$ and $\Pi_-(L)$. From

$$\varepsilon_t = \Pi(L)X_t = \Pi_+(L)X_t + \Pi_-(L)X_t,$$

we find, by applying $\Pi_+(L)^{-1}$ on both sides, that

$$X_t = \Pi_+(L)^{-1}\varepsilon_t - \Pi_+(L)^{-1}\Pi_-(L)X_t = \Pi_+(L)^{-1}\varepsilon_t + \mu_t,\ t = 1, 2, \ldots. \tag{11}$$

The first term is the stochastic component generated by $\varepsilon_1, \ldots, \varepsilon_t$, and the second a deterministic component generated by initial values. An example of this is the well known result that $X_t = \xi X_{t-1} + \varepsilon_t$ has the solution $X_t = \sum_{i=0}^{t-1} \xi^i \varepsilon_{t-i} + \xi^t X_0$ for any ξ.

The solution (11) of equation (10) is valid without any assumptions on the parameters. We next give results, see Johansen (2008, Theorem 8), which guarantee that the process is fractional of order d and cofractional from d to $d - b$. The conditions are given in terms of the roots of the polynomial $\det(\Psi(y))$ and the set $\mathbb{C}_b$, which is the image of the unit disk under the mapping $y = 1 - (1-z)^b$, see Figure 1.

The following result is Granger's Representation Theorem for the cofractional VAR model (10) generalizing Theorem 1 to fractional processes.

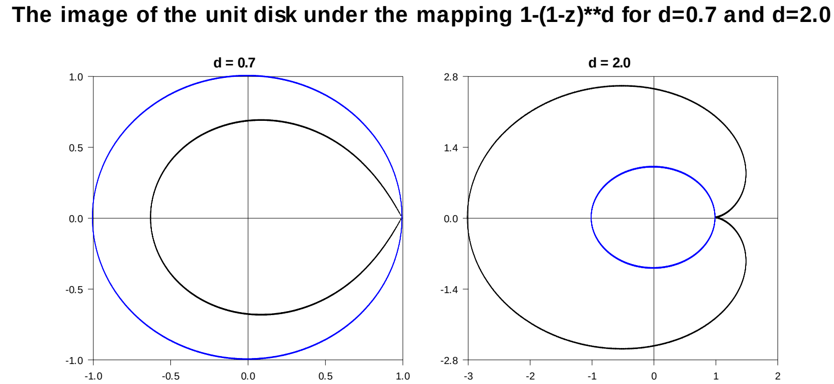

Figure 4: We illustrate the set $\mathcal{C}_d$ which is the image of the unit disk under the mapping $z \longmapsto 1-(1-z)^d$ which replaces the unit disk in the structure theory of fractional processes, see Theorem 3. Note that $\mathcal{C}_d$ is increasing in d, and for $d=1$ it is the unit disk.

Theorem 3 *Let* $\Pi(z) = (1-z)^{d-b}\Psi(1-(1-z)^b)$ *be given by* (**??**) *and let* $1/2 \leq b \leq d$. *Assume that* $\det(\Psi(y)) = 0$ *implies that either* $y = 1$ *or* $y \notin \mathbb{C}_b$ *and that* α *and* β *have rank* $r < p$. *Let* $\Gamma = I_p - \sum_{i=1}^k \Gamma_i$ *and assume that* $\det(\alpha_\perp' \Gamma \beta_\perp) \neq 0$, *so that* C *is defined by* ((5)). *Then*

$$(1-z)^d \Pi(z)^{-1} = C + (1-z)^b H(1-(1-z)^b),$$

where $H(1) \neq 0$ *and* $H(y)$ *is regular, see Phillips (1958), in a neighborhood of* $\mathbb{C}_b$. *It follows that the coefficient matrices* τ_n *defined by* $F(z) = H(1-(1-z)^b) = \sum_{n=0}^\infty \tau_n z^n$, $|z| < 1$, *satisfy* $\sum_{n=0}^\infty |\tau_n| < \infty$.

Equation ((10)) *is solved by*

$$X_t = C\Delta_+^{-d}\varepsilon_t + \Delta_+^{-(d-b)} Y_t^+ + \mu_t, \; t = 1, \ldots, T, \tag{12}$$

where $\mu_t = -\Pi_+(L)^{-1}\Pi_-(L)X_t$ *and* $Y_t^+ = \sum_{n=0}^{t-1} \tau_n \varepsilon_{t-n}$, *so that* Y_t *is fractional of order zero.*

Thus X_t *is fractional of order* d, *and because* $\beta' C = 0$, X_t *is cofractional since* $\beta' X_t = \Delta_+^{-(d-b)}\beta' Y_t^+ + \beta' \mu_t$ *is fractional of order* $d-b$.

The proof is given in Johansen (2008, Theorem 8); see also Johansen and Nielsen (2010a, Lemma 1) for the univariate case.

3.2 Assumptions for asymptotic analysis

For the asymptotic analysis we apply the result, e.g. Davydov (1970), that when $d > 1/2$ and $E|\varepsilon_t|^q < \infty$ for some $q > \max(2, 1/(d-1/2))$, then

$$T^{-d+1/2}\Delta_+^{-d}\varepsilon_{[Tu]} \Longrightarrow W_{d-1}(u) = \Gamma(d)^{-1}\int_0^u (u-s)^{d-1} dW(s) \text{ on } D^p[0,1], \tag{13}$$

where W denotes $p-$dimensional Brownian motion generated by ε_t, W_{d-1} is the corresponding fractional Brownian motion of type II. We also need a result for the product moments

$$T^{-d}\sum_{t=1}^{T}\Delta_{+}^{-d}\varepsilon_{t-1}\varepsilon_t' \xrightarrow{d} \int_0^1 W_{d-1}dW', d > 1/2 \tag{14}$$

see Jakubowski, Mémin, and Pages (1989), where $\xrightarrow{d}$ denotes convergence in distribution on $\mathbb{R}^{p\times p}$.

We next formulate the assumptions needed for the asymptotic results.

Assumption 1 *The process $X_t,\ t = 1,\ldots,T$, is given by ((10)) for some $k \geq 1$, for some value of the parameters satisfying the assumptions of Theorem 3, and the errors ε_t are i.i.d. $(0,\Omega)$. The initial values are bounded and for identification we assume $\Gamma_k \neq 0$. Finally*

$$1/2 \leq b \leq d \leq d_1. \tag{15}$$

The theory has been developed for observations $X_1,\ldots,X_T$ generated by (10) assuming that all initial values are observed and bounded, that is, conditional on $X_{-n},\ n = 0,1,\ldots$

This is standard in the literature on inference for nonstationary autoregressive processes, where the initial values are observed but not modeled, and inference is conditional on them. However, we do not set initial values equal to zero as is often done in the literature on fractional processes, but instead assume only that they are observed unmodelled bounded constants.

Of course in practice we have not observed infinitely many initial values, and we will have to set them to zero for $t \leq -N_0$, say. The asymptotic results do not depend on the choice of initial values, but there is obviously a finite sample problem, that need to be investigated.

3.3 Profile likelihood function and consistency of the MLE

For given (d,b) we calculate the maximum likelihood estimators by first performing a reduced rank regression of $\Delta^d X_t$ on $\Delta^{d-b}L_b X_t$ corrected for $\Delta^d L_b^i X_t, i = 1,\ldots,k$. This gives the likelihood profile function, which is then maximized as a function of just two parameters (d,b). By analysing the profile likelihood function as a continuous stochastic process in (d,b), using Kallenberg (2001), we can show that if $E(|\varepsilon_t|^q) < \infty$ for all q, then it converges uniformly in probability to a deterministic limit for $d-b \geq \delta_0 > 0$, and this again implies that with probability tending to one, the maximum likelihood estimator exists and is consistent.

We next find the limit distribution of the score function at the true value and show that the information matrix converges uniformly in a neighborhood of the true value. This implies that we can apply the usual expansion of the score function and find the asymptotic distribution of the maximum likelihood estimator.

Theorem 4 *If $1/2 < b < d < d_1$, and $E|\varepsilon_t|^q < \infty$ for some $q > \max(2,(b-1/2)^{-1})$, then the asymptotic distribution of the Gaussian maximum likelihood estimator $\hat{\phi} = (\hat{d},\hat{b},\hat{\alpha},\hat{\Gamma}_1,\ldots,\hat{\Gamma}_k)$ and $\hat{\beta}$ for model ((10)) is given by*

$$\begin{pmatrix} T^{1/2}(\hat{\phi}-\phi) \\ T^{b_0}\beta_\perp'(\hat{\beta}-\beta) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_{n_\phi}(0,\Sigma^{-1}) \\ (\int_0^1 FF'du)^{-1}\int_0^1 F(dV)' \end{pmatrix}, \tag{16}$$

where $F = (\beta_\perp'\beta_\perp)^{-1}\beta_\perp' C W_{b-1}$ *and* $V = (\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1}W$ *are independent. It follows that the asymptotic distribution of* $T^b\bar{\beta}_\perp'(\hat{\beta}-\beta)$ *is mixed Gaussian with conditional variance*

$$(\alpha_0'\Omega_0^{-1}\alpha_0)^{-1}\otimes(\int_0^1 F_0F_0'du)^{-1}. \tag{17}$$

This result is the same as for the standard cointegration model, except that Brownian motion is replaced by fractional Brownian motion.

3.4 Likelihood ratio test for cofractional rank

Using the same methods we can find the asymptotic distribution of the likelihood ratio statistic for cointegrating rank, $-2\log LR(\mathcal{H}_r|\mathcal{H}_p)$.

Theorem 5 *Under the assumptions of Theorem 4, the likelihood ratio statistic for rank has asymptotic distribution*

$$-2\log LR(\mathcal{H}_r|\mathcal{H}_p)\xrightarrow{d} tr\{\int_0^1(dB)B_{b-1}'(\int_0^1 B_{b-1}B_{b-1}'du)^{-1}\int_0^1 B_{b-1}(dB)'\}, \tag{18}$$

where B *is* $(p-r)$*-dimensional standard Brownian motion and* B_{b_0-1} *the corresponding fractional Brownian motion.*

Again this result mimics the usual result (8) where F is replaced by B_{b-1}.

4 Conclusion

We have summarized the statistical theory of the CVAR

$$\Delta X_t = \alpha(\beta' X_{t-1} + \rho' t) + \sum_{i=1}^k \Gamma_i \Delta X_{t-i} + \mu + \varepsilon_t,$$

and indicated the asymptotic results needed for likelihood based inference.

We have then extended the model to

$$\Delta^d X_t = \alpha\beta'\Delta^{d-b}L_b X_t + \sum_{i=1}^k \Gamma_i\Delta^d L_b^i X_t + \varepsilon_t,\ 1/2\le b\le d,$$

by replacing Δ by Δ^b and applying the equation to $\Delta^{d-b}X_t$. This gives a model for fractional processes of order d which cointegrate to order $d-b$. Note, however, that we have not included any deterministic terms, Such models need to be formulated and analysed.

We have analyzed the conditional Gaussian likelihood given initial values, which are assumed bounded. We can show existence and consistency and derive the asymptotic distribution of the maximum likelihood estimator. In the asymptotic analysis we assumed i.i.d. errors with suitable moment conditions. We have derived the asymptotic distribution of the test for the rank of $\alpha\beta'$ and shown that it is expressed in terms of fractional Brownian motion, that inference on β is asymptotically mixed Gaussian, and finally that the estimators of the remaining parameters are asymptotically Gaussian.

References

1. Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327–351.
2. Bartlett, M. (1948). A note on the statistical estimation of the demand and supply relations from time series. *Econometrica* 16, 323–329
3. Basawa I.V. and D.J. Scott (1983). *Asymptotic Optimal Inference for Non-ergodic Models.* Springer, New York.
4. Beran, J. (1994). *Statistics for long-memory processes.* Chapman and Hall.
5. Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
6. Campbell, J. and R.J. Shiller (1987). Cointegration and tests of present value models. *Journal of Political Economy* 95, 1062–1088.
7. Davydov, Yu.A. (1970). The invariance principle for stationary processes. *Theory of Probability and Its Applications* 15, 487–498.
8. Engle, R.F., D.F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica* 51, 277–304.
9. Engle, R.F. and C.W.J. Granger, (1987). Co-integration and error correction: representation, estimating and testing. *Econometrica* 55, 251–276.
10. Granger, C.W.J. (1983). Cointegrated variables and error correction models. UCSD Discussion paper 83-13a.
11. Granger, C.W.J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economic and Statistics* 48, 213–228.
12. Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.
13. Hamilton J.D. (1994). *Time series analysis.* Princeton University Press, Princeton New Jersey.
14. Hansen, J.E., R. Ruedy, Mki. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl (2001). A closer look at United States and global surface temperature change. *Journal of Geophysical Research*, 106, 23947-23963.
15. Hualde, J. and P.M. Robinson (2010). Semiparametric inference in multivariate fractionally cointegrated systems. Forthcoming in *Journal of Econometrics.*
16. Jakubowski, A., Mémin, J., and G. Pages (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbb{D}^1$ de Skorokhod. *Probability Theory and Related Fields* 81, 111–137.
17. Jeganathan, P. (1999). On asymptotic inference in cointegrated time series with fractionally integrated errors. *Econometric Theory* 15, 583–621.
18. Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.
19. Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
20. Johansen, S. (1995). The role of ancillarity in inference for nonstationary variables. *Economic Journal* 13, 302–320.
21. Johansen, S. (1996). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford.

22. Johansen, S. (2006). Cointegration: a survey. In Mills, T.C. and Patterson, K. (eds) *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory.* Palgrave Macmillan, Basingstoke.
23. Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory* 24, 651–676.
24. Johansen, S. (2010). The analysis of nonstationary time series using regression, correlation and cointegration – with an application to annual mean temperature and sea level. Discussion paper, Department of Economics, University of Copenhagen.
25. Johansen, S. and M.Ø., Nielsen (2010a). Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics* 158, 51–66.
26. Johansen, S. and M. Ø., Nielsen (2010b). Likelihood inference for a vector autoregressive model which allows for fractional and cofractional processes (Discussion papers 10-15, Department of Economics, University of Copenhagen, 41 pp.)
27. Juselius, K. (2006). *The Cointegrated VAR Model: Econometric Methodology and Macroeconomic Applications.* Oxford University Press, Oxford.
28. Kallenberg, O. (2001). *Foundations of Modern Probability Theory.* 2nd edition, Springer, New York.
29. Lasak, K. (2008a). Likelihood based testing for no fractional cointegration. *Journal of Econometrics* 158, 67-77.
30. Lasak, K. (2008b). Maximum likelihood estimation of fractionally cointegrated systems. CREATES Research Paper 2008–53.
31. Lütkepohl H. (2006). I*ntroduction to Multiple Times Series Analysis.* Springer, New York.
32. Marinucci, D. and P.M. Robinson (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and their Applications* 86, 103–120.
33. Phillips, E.G. (1958). *Functions of a Complex Variable with Applications.* Oliver and Boyd.
34. Phillips P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
35. Robinson, P.M. and J. Hualde (2003). Cointegration in fractional systems with unkown integration orders. *Econometrica* 71, 1727–1766.
36. Schmith, T., Johansen, S., and P. Thejll (2010). Statistical analysis of global surface temperature and sea level using cointegration methods. Manuscript in preparation.
37. Watson, M. (1994). Vector autoregressions and cointegration. In: R. F. Engle and McFadden, D. (eds.,). *Handbook of Econometrics* Vol. 4, North Holland.