European University Institute
ECONOMICS DEPARTMENT

EUI Working Paper **ECO** No. 97/16

Likelihood Analysis of Seasonal Cointegration

Søren JOHANSEN
European University Institute, Florence
and
Ernst SCHAUMBURG
Princeton University

*

========

Printed in Italy in May 1997
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

Likelihood Analysis of Seasonal Cointegration

Søren Johansen
Department of Economics, European University Institute, Florence
and
Ernst Schaumburg
Department of Economics, Princeton University

March 14, 1997

Abstract

The error correction model for seasonal cointegration is analyzed. Conditions are found under which the process is integrated of order 1 and cointegrated at seasonal frequency, and a representation theorem is given. The likelihood function is analyzed and the numerical calculation of the maximum likelihood estimators is discussed. The asymptotic distribution of the likelihood ratio test for cointegrating rank is given. It is shown that the estimated cointegrating vectors are asymptotically mixed Gaussian. The results resemble the results for cointegration at zero frequency when expressed in terms of a complex Brownian motion. Tables are provided for asymptotic inference.

Key words: Autoregressive process; Granger's theorem; Error correction model; Complex Brownian motion

JEL *classification:* C32

1 Introduction

This paper contains a systematic treatment of the statistical analysis of seasonal cointegration in the vector autoregressive model. The theory started with the paper by Hylleberg, Engle, Granger, and Yoo (1990) which gave the main results on the representation and the single equation tests for cointegration at complex frequencies. An analysis of seasonal cointegration of Japanese consumption was given in Engle, Granger, Hylleberg, and Lee (1993). Seasonal cointegration analysis is prompted by the empirical finding that the vector autoregressive model often describe macro data quite well. The occurrence of unit roots in the fitted process implies that it is non-stationary, and the occurrence of unit roots at seasonal frequency implies a non-stationary seasonal variation. This again implies the possibility of seasonal cointegration, and the phenomenon that the seasonality drifts, such that "summer becomes winter".

The paper on maximum likelihood inference by Lee (1992) sets the stage for the analysis of multivariate systems. Unfortunately it does not treat all aspects of asymptotic inference, and the test for cointegration rank is only partially correct. The two papers by Gregoir (1993a,b) deal with a very general situation of unit roots allowing for processes to be integrated of order greater than 1 at any frequency, but do not treat likelihood inference.

The purpose of this paper is therefore to improve on the previous analysis and discuss maximum likelihood estimation, calculation of test statistics, and derivation of asymptotic distributions in the context of the vector autoregressive model. In the process of doing so it is natural to give the mathematical theory of the Granger representation, the error correction model, the role of the constant term, and seasonal dummies. The basic new trick is the introduction of the complex Brownian motion, which makes many calculations more natural and greatly simplifies formulae for limit distributions. We focus mainly on complex roots, since the case with roots at 1 is well known from the literature, see Johansen (1996), and the situation with a root at -1 can be dealt with using the same methods, see Lee (1992).

We consider the autoregressive model defined for an $n-$dimensional process X_t by the equations:

$$X_t = \sum_{j=1}^{l} \Pi_j X_{t-j} + \Phi D_t + \varepsilon_t, \tag{1}$$

where we assume that the initial values $X_0, \ldots, X_{-l+1}$ are fixed. When deriving estimators and test statistics we also assume that ε_t are i.i.d. $N_n(0, \Omega)$, while the asymptotic results are proved under the assumption that the errors are i.i.d. with finite variance and mean zero. This assumption can be further relaxed, see Chan and Wei (1988), and the comments in Section 4. The deterministic terms D_t may contain a constant, a linear term, or seasonal dummies. Various models defined by restrictions on the deterministic terms will be considered. The properties of the process generated by (1) are as usual expressed in terms of the characteristic polynomial

$$A(z) = I_n - \sum_{j=1}^{l} \Pi_j z^j,$$

with determinant $|A(z)|$, and where I_n is an $n \times n$ identity matrix.

The paper is organized as follows: in Section 2 the error correction model for seasonal cointegration of processes that are integrated of order 1 at seasonal frequency is discussed. The equations are solved in the form of a Granger representation theorem, applying a general result about inversion of matrix polynomials. This is applied to analyze the role of constant, linear term, and seasonal dummies, see Franses and Kunst (1995). In Section 3 the Gaussian likelihood analysis and calculation of maximum likelihood estimators in the model with unrestricted deterministic terms, as well as in some models, defined by restrictions of deterministic terms, is discussed. In Section 4 some technical asymptotic results on the behavior of the process and product moments are given. Section 5 contains asymptotic results for the maximum likelihood estimator of the cointegrating vectors, and the likelihood ratio test for cointegration rank at seasonal frequency.

In Appendix A a brief description of the (real) matrix representation of complex matrices is given along with proofs of the technical results in Section 4. Finally, Appendix B contains tables of limit distributions of the likelihood ratio tests for cointegrating rank for various models defined by restrictions on the deterministic terms.

2 The representation theorem and the error correction model

This section contains the necessary analytic results from the theory of real polynomials $A(z)$ with values in the set of $n \times n$ matrices. Theorem 1 gives Lagrange's expansion for a polynomial around arbitrary points and it is shown in Corollary 2 how this contains the formulation of an error correction model. The basic result, however, is Theorem 3, which gives a necessary and sufficient condition for the inverse matrix polynomial to have poles of order 1. In Theorem 4 this result is interpreted as a representation of the solution of the autoregressive equations allowing for integrated processes and cointegration at seasonal frequency, thereby generalizing Granger's theorem for $I(1)$ processes. The new Granger representation theorem is applied to discuss the role of constant, linear term, and in particular seasonal dummies. This section is concluded with some examples of models for annual, semi-annual, and quarterly data.

2.1 Some notation

We are concerned with roots of the equation $|A(z)| = 0$, in particular unit roots, for which $|z| = 1$. For a complex number $z = e^{i\theta}$, the complex conjugate is the same as the inverse, $\bar{z} = z^{-1} = e^{-i\theta}$. We use the notation $\beta^* = \beta_R' - i\beta_I'$ for the adjoint matrix, where β_R and β_I are real matrices.

Let z_m be a root of $|A(z)| = 0$, such that $A(z_m)$ has reduced rank. Then there exists two complex matrices α_m and β_m such that $A(z_m) = \alpha_m\beta_m^*$. Note that since the roots may be complex, the matrices α_m and β_m may be complex. However, since the coefficients of $A(z_m)$ are real, the roots and the corresponding matrices α_m and β_m come in complex conjugate pairs.

Corresponding to s distinct complex numbers $z_1, \ldots, z_s$, we introduce the polynomials

$$\begin{aligned} p(z) &= \textstyle\prod_{m=1}^s (1 - \bar{z}_m z), \\ p_j(z) &= \textstyle\prod_{m\neq j}^s (1 - \bar{z}_m z) = \frac{p(z)}{1-\bar{z}_j z},\ z \neq z_j, \\ p_{kj}(z) &= \textstyle\prod_{m\neq k,j}^s (1 - \bar{z}_m z) = \frac{p(z)}{(1-\bar{z}_k z)(1-\bar{z}_j z)},\ z \neq z_j, z_k. \end{aligned}$$

2.2 The error correction model

The error correction formulation is a simple consequence of Lagrange's expansion of $A(z)$ around the $s + 1$ points $z = 0, z_1, \ldots, z_s$.

Theorem 1 *The polynomial $A(z)$ can be expanded around the points $0, z_1, \ldots, z_s$ as follows*

$$A(z) = p(z)I_n + \sum_{m=1}^{s} A(z_m)\frac{p_m(z)z}{p_m(z_m)z_m} + p(z)zA_0(z),$$

where $A_0(z)$ is a matrix polynomial.

Proof. The matrix polynomial

$$A(z) - \sum_{m=1}^{s} A(z_m) \frac{p_m(z)z}{p_m(z_m)z_m} - p(z)I_n$$

is zero for $z = 0, z_1, \ldots, z_s$, and hence each of the entries can be factorized into $p(z)z$ times a polynomial. It follows that the difference can be written as $p(z)zA_0(z)$ for some matrix polynomial $A_0(z)$. ■

An immediate consequence of this is the error correction formulation, see Hylleberg, Engle, Granger, and Yoo (1990).

Corollary 2 *Let $z_1, \ldots, z_s$ be the unit roots of $|A(z)| = 0$, such that the matrices $A(z_m)$ are of reduced rank: $A(z_m) = -\alpha_m \beta_m^*$, with α_m and β_m complex matrices of dimension $n \times r_m$ and rank r_m. If $A(L)X_t = \Phi D_t + \varepsilon_t$, then X_t satisfies an error correction model:*

$$\begin{aligned} p(L)X_t &= \textstyle\sum_{m=1}^{s} \alpha_m \beta_m^* \frac{p_m(L)L}{p_m(z_m)z_m} X_t - p(L)A_0(L)LX_t + \Phi D_t + \varepsilon_t \\ &= \textstyle\sum_{m=1}^{s} \alpha_m \beta_m^* X_t^{(m)} - A_0(L)p(L)X_{t-1} + \Phi D_t + \varepsilon_t, \end{aligned} \tag{2}$$

where we have introduced

$$X_t^{(m)} = \frac{p_m(L)L}{p_m(z_m)z_m} X_t.$$

The idea behind this formulation (see Theorem 4, and condition (4)), is that X_t is a non-stationary process and the $p(L)X_t$ is stationary, that is, for any $h = 1, 2, \ldots$, the distribution of $(p(L)X_t, p(L)X_{t+1}, \ldots, p(L)X_{t+h})$ does not depend on t. The processes $X_t^{(m)} (m = 1, \ldots, s)$ are non-stationary but, as we shall see below, the components of $X_t^{(m)}$ have the same common non-stationary trends which are removed by the linear combinations β_m^* and $\beta_m^* X_t^{(m)}$ is stationary. Thus, the stationary "differences" $p(L)X_t$ react to equilibrium errors given by $\beta_m^* X_t^{(m)}$ through the adjustment coefficients α_m. The result can therefore be formulated as cointegration of the (complex) processes $X_t^{(m)}$, or if expressed in terms of X_t, we find that the stationarity of $\beta^* X_t^{(m)} = \beta^* \frac{p_m(L)L}{p_m(z_m)z_m} X_t$ can be expressed as a complicated cointegration relation between the process X_t and its lags (see section 2.5). This has been called polynomial cointegration by Engle and Yoo (1991). Note also that a different set of roots gives rise to a different error correction formulation.

2.3 Granger's representation theorem

We define the derivative $\dot{A}(z_m)$ of $A(z)$ at $z = z_m$. If the polynomial $|A(z)|$ has a root at $z = z_0$ then $A(z_0)$ is not invertible. We say that $A(z)^{-1}$ has a pole of order k $(k = 0, 1, \ldots)$ at $z = z_0$ if

$$\lim_{z \to z_0} (1 - \frac{z}{z_0})^k A^{-1}(z)$$

exists and is non-zero. We next prove a result that gives a necessary and sufficient condition for the inverse function to have a pole of order 1 at the point z_0. This condition clearly requires $A(z_0)$ to be singular, but we also need a condition on the derivative of $A(z)$ at z_0, which restricts the behavior of $A(z)$ in a neighborhood of z_0. For any (complex) matrix c of dimension $n \times r$ we define $c_\perp$ as a full rank (complex) matrix of dimension $n \times (n - r)$, such that $c' c_\perp = 0$. Note that $(c^*)_\perp = (c_\perp)^*$.

Theorem 3 *Assume that the roots of $|A(z)| = 0$ satisfy $|z| > 1 + \delta$ or $z \in (z_1, \ldots, z_s)$ with $|z_m| = 1$ for some $\delta > 0$, and that $A(z_m) = -\alpha_m \beta_m^*$. Then the matrix polynomial $A(z)$ is invertible on the disk $|z| \leq 1 + \delta$, except at the points $(z_1, \ldots, z_s)$ where $A^{-1}(z)$ has a pole. A necessary and sufficient condition for the pole to be of order 1 is that*

$$|\alpha_{m\perp}^* \dot{A}(z_m) \beta_{m\perp}| \neq 0, \, m = 1, \ldots, s. \tag{3}$$

In this case, we get an expansion of $A^{-1}(z)$ of the form

$$A^{-1}(z) = \sum_{m=1}^{s} C_m \frac{1}{1 - \bar{z}_m z} + C_0(z), \, z \neq (z_1, \ldots, z_s),$$

where

$$\lim_{z \to z_m} (1 - \bar{z}_m z) A(z)^{-1} = C_m = -\bar{z}_m \beta_{m\perp} (\alpha_{m\perp}^* \dot{A}(z_m) \beta_{m\perp})^{-1} \alpha_{m\perp}^*,$$

and where $C_0(z)$ has a convergent power series for $|z| \leq 1 + \delta$. That is, $C_0(z)$ has no poles and $A^{-1}(z)$ has poles of order 1. Moreover it holds that

$$\frac{p_m(z) z}{p_m(z_m) z_m} A^{-1}(z) = C_m \frac{1}{(1 - \bar{z}_m z)} + C_m(z), \, z \neq (z_1, \ldots, z_s),$$

for some power series $C_m(z)$ convergent for $|z| \leq 1 + \delta$.

Proof. The usual expression for the inverse of a matrix

$$A^{-1}(z) = \frac{Adj(A(z))}{|A(z)|},\ z \neq (z_1, \ldots, z_s)$$

shows that $A(z)^{-1}$ has poles at the roots $z = z_1,\ \ldots,\ z_s$ since $|A(z_m)| = 0,\ m = 1, \ldots, s$. We want to show that the pole at $z = z_m$ is of order 1 and has the form

$$C_m \frac{1}{(1 - \bar{z}_m z)},$$

such that the continuous function $C_0(z)$ defined by

$$C_0(z) = A(z)^{-1} - \sum_{m=1}^{s} C_m \frac{1}{(1 - \bar{z}_m z)},\ z \neq (z_1, \ldots, z_s),$$

has no poles in the disk $|z| \leq 1 + \delta$, for some $\delta > 0$.

This is proved by investigating the functions in a neighborhood of each of the poles and showing that, by subtracting the poles given in the sum, the poles in $A^{-1}(z)$ are eliminated. Thus, we first focus on the root $z = z_1$ where $A(z_1) = -\alpha\beta^*$, and we have left out the subscript to simplify the notation.

Consider therefore a value of z such that $0 < |z - z_1| \leq \varepsilon$. From the expansion

$$A(z) = A(z_1) + (z - z_1)\dot{A}(z_1) + (z - z_1)^2 A_1(z),$$

where $A_1(z)$ is a remainder polynomial, it follows by multiplying by $(\alpha, \alpha_\perp)^*$ and $(\beta, 1/(1 - \bar{z}_1 z)\beta_\perp)$ that, since $A(z_1) = -\alpha\beta^*$,

$$\begin{aligned} & \tilde{A}(z) \\ = \ & (\alpha, \alpha_\perp)^* A(z) \left(\beta, \beta_\perp \tfrac{1}{1 - \bar{z}_1 z}\right) \\ = \ & \begin{pmatrix} -\alpha^*\alpha\beta^*\beta & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \alpha^*\dot{A}(z_1)\beta(z - z_1) & -z_1\alpha^*\dot{A}(z_1)\beta_\perp \\ \alpha_\perp^*\dot{A}(z_1)\beta(z - z_1) & -z_1\alpha_\perp^*\dot{A}(z_1)\beta_\perp \end{pmatrix} \text{ for some} \\ & +(z - z_1)A_2(z), \end{aligned}$$

remainder polynomial $A_2(z)$. Here and in the following we often use such a notation for a remainder term, when we expand a polynomial or a power series. The function $\tilde{A}(z)$ is a matrix polynomial and therefore has no poles. Further

$$\tilde{A}(z_1) = -\begin{pmatrix} \alpha^*\alpha\beta^*\beta & z_1\alpha^*\dot{A}(z_1)\beta_\perp \\ 0 & z_1\alpha_\perp^*\dot{A}(z_1)\beta_\perp \end{pmatrix}$$

has full rank if and only if assumption (3) holds since

$$|\tilde{A}(z_1)| = (-1)^n |\alpha^*\alpha||\beta^*\beta||z_1\alpha_\perp^*\dot{A}(z_1)\beta_\perp| \neq 0.$$

In this case, $\tilde{A}(z)$ is invertible for $|z-z_1| \leq \varepsilon$ for some ε, with the property that $0 < \varepsilon < \min_{k \neq l} |z_k - z_l|$, and we find by the expansion

$$\tilde{A}^{-1}(z) = \tilde{A}^{-1}(z_1) + (z - z_1)M_2(z),$$

that for $z \neq z_1$

$$\begin{aligned} A^{-1}(z) &= \left(\beta, \beta_{\perp}\tfrac{1}{1-\bar{z}_1 z}\right) \tilde{A}^{-1}(z)(\alpha, \alpha_{\perp})^* \\ &= -\tfrac{\bar{z}_1}{1-\bar{z}_1 z}\beta_{\perp}(\alpha_{\perp}^* \dot{A}(z_1)\beta_{\perp})^{-1}\alpha_{\perp}^* + M_3(z) \\ &= C_1 \tfrac{1}{1-\bar{z}_1 z} + M_3(z), \end{aligned}$$

where C_1 is given in Theorem 3 and

$$M_3(z) = ((z - z_1)\beta, -\beta_{\perp} z_1)M_2(z)(\alpha, \alpha_{\perp})^*.$$

Here $M_2(z)$is a convergent power series which is a notation for the remainder term in the expansion of $\tilde{A}^{-1}(z)$. Hence $A^{-1}(z) - C_1\frac{1}{1-\bar{z}_1 z} = M_3(z)$ has no pole at $z = z_1$ and extends by continuity to the point $z = z_1$.

The same argument can be used to remove the other poles from $A^{-1}(z)$ and the theorem has been proved. ■

The next result is a representation of the solution of the error correction model (2). The equations (2) determine the process X_t as a function of the errors ε_i $(i = 1, \ldots, t)$ and the initial values of the process. We give the result for the model without deterministic terms, and later formulate and apply the result for the general case.

Theorem 4 *Let X_t satisfy $A(L)X_t = \varepsilon_t$, and let the equation $|A(z)| = 0$ have roots outside the unit disk and at $z_1, \ldots, z_s$ with absolute value 1, such that X_t is non-stationary. Then*

$$A(z_m) = -\alpha_m\beta_m^*, \; m = 1, \; \ldots, \; s,$$

where α_m and β_m are $n \times r_m$ matrices of rank $r_m < n$.

The processes $p(L)X_t$ and $p_m(L)\beta_m^* X_t$ can be made stationary by a suitable choice of initial values, if and only if

$$|\alpha_{m\perp}^* \dot{A}(z_m)\beta_{m\perp}| \neq 0. \tag{4}$$

In this case, X_t and $X_t^{(m)}$ can be given the representation

$$X_t = \sum_{m=1}^{s} C_m \bar{z}_m^t S_t^{(m)} + \sum_{m=1}^{s} \bar{z}_m^t A_m + Y_t, \tag{5}$$

and

$$X_t^{(m)} = \frac{p_m(L)L}{p_m(z_m)z_m} X_t = C_m \bar{z}_m^t S_{t-1}^{(m)} + \bar{z}_m^t A_m + Y_t^{(m)}, \tag{6}$$

where $S_t^{(m)}$ *is given by*

$$S_t^{(m)} = \sum_{j=0}^{t} z_m^j \varepsilon_j,$$

and the random variables A_m *depend on initial conditions such that* $\beta_m^* A_m = 0$, *and finally* Y_t *and* $Y_t^{(m)}$ *are stationary processes.*

Thus, the non-stationary process X_t can be made stationary by the difference operator $p(L)$ and, since $\beta_m^* X_t^{(m)}$ is stationary, we call X_t seasonally cointegrated at $z_m = e^{i\theta_m}$, or at frequency θ_m, with cointegrating vectors β_m. Note that $S_t^{(m)}$ is not a random walk since $\Delta S_t^{(m)} = z_m^t \varepsilon_t$ are independent but not in general identically distributed. Note also the factor $\bar{z}_m^t$ in front of $S_t^{(m)}$ gives a type of non-stationarity that is different from the usual unit root non-stationarity. Finally note that since we allow for complex roots, the process $X_t^{(m)}$ and the coefficients α_m and β_m are in general complex. Since, however, the data and the coefficients in $A(z)$ are real, the roots come in complex conjugate pairs. Hence a reduced rank condition at a complex root automatically implies a reduced rank condition at the complex conjugate root. This will complicate the statistical analysis below.

The difference between the results in Theorem 3 and Theorem 4 is that in order to interpret Theorem 3 for stochastic processes, care has to be taken of initial values in the representation (5) and (6) in order to translate the results about the power series into results about the lag operator.

Proof. From Theorem 3 we find

$$p(z)I_n = [\sum_{m=1}^{s} C_m p_m(z) + p(z)C_0(z)]A(z).$$

Expressed in terms of the lag operator L, defined by $LX_t = X_{t-1}$, the relation applied to X_t is, since $A(L)X_t = \varepsilon_t$,

$$p(L)X_t = \sum_{m=1}^{s} C_m p_m(L)\varepsilon_t + p(L)Y_t,\ t = 1, \ldots, T, \tag{7}$$

where $Y_t = C_0(L)\varepsilon_t$. The right hand side is stationary, which shows that by choosing the initial values $X_0, \ldots, X_{-s+1}$, such that (7) is satisfied for $t = 0$, we see that $p(L)X_t$ becomes stationary.

We want to solve equation (7) for X_t by removing the polynomial $p(L)$ one factor at a time by summation. This will again involve the initial values.

Consider first the root $z = z_1$. The definition of $p(z)$ implies that

$$p(z) = (1 - \bar{z}_1 z)p_1(z),\ p_m(z) = (1 - \bar{z}_1 z)p_{m1}(z),\ m \neq 1,$$

and equation (7) is

$$(1 - \bar{z}_1 L)\left[p_1(L)(X_t - Y_t) - \sum_{m=2}^{s} C_m p_{m1}(L)\varepsilon_t\right] = C_1 p_1(L)\varepsilon_t,\ t = 0, 1, \ldots, T.$$

Solving these equations we find

$$p_1(L)(X_t - Y_t) - \sum_{m=2}^{s} C_m p_{m1}(L)\varepsilon_t = \bar{z}_1^{t+1} A + C_1 p_1(L)\bar{z}_1^t \sum_{j=0}^{t} z_1^j \varepsilon_j,\ t = 0, 1, \ldots, T. \tag{8}$$

Here $A = p_1(L)(X_{-1} - Y_{-1}) - \sum_{m=1}^{s} C_m p_{m1}(L)\varepsilon_{-1}$ is the initial value of the left hand side. Next notice that $(1 - \bar{z}_m L)\bar{z}_1^t = (1 - \bar{z}_m \bar{z}_1)\bar{z}_1^t$ implies

$$p_1(L)\bar{z}_1^t = p_1(\bar{z}_1)\bar{z}_1^t,$$

such that result of the above calculations can be expressed as:

$$p_1(L)(X_t - Y_t - C_1 \bar{z}_1^t S_t^{(1)} - \bar{z}_1^t A_1) = \sum_{m=2}^{s} C_m p_{m1}(L)\varepsilon_t,\ t = 0, 1, \ldots, T, \tag{9}$$

with $A_1 = \bar{z}_1 A / p_1(\bar{z}_1)$. We next choose the initial value of X_t to satisfy the further restriction

$$p_1(L)\beta_1^* X_{-1} = p_1(L)\beta_1^* Y_{-1} - \sum_{m=2}^{s} \beta_1^* C_m p_{m1}(L)\varepsilon_{-1},$$

such that $\beta_1^* A = 0$, and hence $\beta_1^* A_1 = 0$.

We then get from (8):

$$p_1(L)\beta_1^* X_t = p_1(L)\beta_1^* Y_t + \sum_{m=2}^{s} \beta_1^* C_m p_{m1}(L)\varepsilon_t,\ t = 0, 1, \ldots, T,$$

since $\beta_1^* C_1 = 0$. This shows that, with this choice of initial values, $p_1(L)\beta_1^* X_t$ is a stationary process.

Equation (9) has the same form as the one we started with in (7), except that the root $z = z_1$ has been removed. In the same way we successively eliminate the roots by summation of the equation, each time subtracting a term of the form $C_m \bar{z}_m^t S_t^{(m)} + \bar{z}_m^t A_m$ from the left hand side, and thereby prove the representation.

Conversely if the assumptions are satisfied and $p(L)X_t$ is stationary, it has the representation :

$$p(L)X_t = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}.$$

By expanding $C(z) = \sum_{i=0}^{\infty} C_i z^i$ around $z_1, \ldots, z_s$, we find

$$p(L)X_t = \sum_{m=1}^{s} C(z_m) \frac{p_m(L)}{p_m(z_m)} \varepsilon_t + p(L)C_0(L)\varepsilon_t,$$

for some convergent power series $C_0(z)$. Hence

$$p(L)\varepsilon_t = p(L)A(L)X_t = A(L)[\sum_{m=1}^{s} C(z_m) \frac{p_m(L)}{p_m(z_m)} + p(L)C_0(L)]\varepsilon_t,$$

which shows that

$$A^{-1}(z) = \sum_{m=1}^{s} C(z_m) \frac{1}{p_m(z_m)(1 - \bar{z}_m z)} + C_0(z).$$

Thus, all poles are of order one and by Theorem 3 and condition (4) holds. ■

The usual Granger representation of an $I(1)$ process as a random walk plus a stationary process is here replaced by a representation in terms of the processes $S_t^{(m)}$. The representation captures the phenomenon that the variance of an $I(1)$ process is increasing. In the present case, we find that the increasing part of the variance of $X_t^{(m)}$ is generated by $C_m \bar{z}_m^t S_t^{(m)}$, which contributes with a variance of the form:

$$C_m \bar{z}_m^t E[S_t^{(m)} S_t^{(m)*}] z_m^t C_m^* = C_m \bar{z}_m^t \sum_{i,j=0}^{t} z_m^i E[\varepsilon_i \varepsilon_j'] \bar{z}_m^j z_m^t C_m^* = (t+1) C_m \Omega C_m^*.$$

The processes $X_t^{(m)}$ and $X_t^{(n)}$ are asymptotically uncorrelated in the sense that, for $n \neq m$, the covariances are:

$$\begin{aligned} & C_m E[\bar{z}_m^t S_t^{(m)} S_t^{(n)*} z_n^t] C_n^* \\ = \; & C_m \bar{z}_m^t z_n^t \sum_{i,j=0}^{t} z_m^i E[\varepsilon_i \varepsilon_j'] \bar{z}_n^j C_n^* \\ = \; & C_m \bar{z}_m^t z_n^t \sum_{i=0}^{t} z_m^i \Omega \bar{z}_n^i C_n^* = \bar{z}_m^t z_n^t \frac{1 - (z_m \bar{z}_n)^{t+1}}{1 - z_m \bar{z}_n} C_m \Omega C_n^*, \end{aligned}$$

which remains bounded as t tends to ∞, while the variances tend to ∞.

Thus, the process is composed of many different processes $S_t^{(m)}$ each of which has a variance that tends to infinity with t in the directions given by $\beta_{m\perp}$. In general one cannot expect that any given linear combination will eliminate all the non-stationarity created by the processes $S_t^{(m)}$. By applying the difference filter $p_m(L)$ one eliminates all non-stationarity except at the frequency θ_m, where a linear combination is needed to annihilate the matrix C_m.

Next Theorem 4 is applied to solve the autoregressive equations when they contain deterministic terms.

Thus, assume that X_t is the solution to the equations:

$$A(L)X_t = \Phi D_t + \varepsilon_t,$$

where $A(z)$ satisfies the conditions set out in Theorem 4. In this case:

$$p(L)X_t = \sum_{m=1}^{s} C_m p_m(L)(\varepsilon_t + \Phi D_t) + p(L)C_0(L)(\varepsilon_t + \Phi D_t),$$

with solution:

$$X_t = \sum_{m=1}^{s} C_m \bar{z}_m^t S_t^{(m)} + \sum_{m=1}^{s} C_m \bar{z}_m^t \Phi \sum_{j=0}^{t} z_m^j D_j + \sum_{m=1}^{s} \bar{z}_m^t \tilde{A}_m + \tilde{Y}_t,$$

where $\tilde{Y}_t - E(\tilde{Y}_t)$ is stationary, and $\tilde{A}_m$ depends on initial values such that $\beta_m^* \tilde{A}_m = 0$. It is seen, that if $D_j = 1$ then the deterministic term gives rise to the trend $\bar{z}_m^t \sum_{j=0}^t z_m^j = \bar{z}_m^t(1 - z_m^{t+1})/(1 - z_m)$ which remains bounded, unless $z_m = 1$ in which case we get a linear trend in the process. Similarly, if $D_j = j$ we find

$$\bar{z}_m^t \sum_{j=0}^{t} j z_m^j = \frac{t+1}{1-\bar{z}_m} - \frac{(1-\bar{z}_m^{t+1})}{(1-\bar{z}_m)^2},$$

which is a linear trend if $z_m \neq 1$ and a quadratic trend if $z_m = 1$. Thus, if $\Phi D_t = \Phi_0 + \Phi_1 t$ we get a quadratic trend from $z_m = 1$, which vanishes if we choose $\alpha_{m\perp}' \Phi_1 = 0$. If $\Phi_1 = 0$, we get a linear trend which vanishes when $\alpha_{m\perp}' \Phi_0 = 0$. The next sub-section discusses similar result for seasonal dummies.

2.4 The role of seasonal dummies

If we have data measured at frequency $\tilde{s}$ per year, we often include seasonal dummies to model the effect of a seasonally changing mean. We here consider data with unit roots at (some of) the seasonal frequencies $\theta_m = 2\pi im/\tilde{s}$, $(m = 1, \ldots, \tilde{s})$.

This is the situation if for instance we have quarterly data and have unit roots in the process at $z = \pm i$ and $z = 1$. We denote the roots of unity z_m $(m = 1, \ldots, \tilde{s})$, and assume for simplicity here that $z_1 = 1$, $z_2, \ldots, z_s$ are roots of the process, $s \leq \tilde{s}$. The results are easily modified if this is not the case.

The seasonal dummies are defined by the $\tilde{s}$ vector D_t, with the property that $D_t = D_{t+\tilde{s}}$. This is a difference equation with characteristic roots equal to the roots of unity, and the solution is:

$$D_t = \sum_{m=1}^{\tilde{s}} \bar{z}_m^t d_m,$$

for some linearly independent $\tilde{s}-$vectors d_m, which can be determined by the initial $\tilde{s}$ values of D_t

$$d_m = \frac{1}{\tilde{s}} \sum_{j=1}^{\tilde{s}} z_m^j D_j.$$

With this notation we find:

$$\sum_{j=0}^{t} z_m^j D_j = \sum_{j=0}^{t} z_m^j \sum_{n=1}^{\tilde{s}} \bar{z}_n^j d_n = (t+1)d_m + \sum_{n \neq m} \frac{1 - z_m^{t+1}\bar{z}_n^{t+1}}{1 - z_m \bar{z}_n} d_n,\ m = 1, \ldots, s.$$

Thus, the seasonal dummy generates a trend $C_m \Phi d_m \bar{z}_m^t t$ in X_t at the unit root z_m $(m = 1, \ldots, s)$. For $z_m \neq 1$ this trend oscillates due to the factor $\bar{z}_m^t$, which is probably unwanted in the description of data. The trend is removed by assuming $C_m \Phi d_m = 0$, or $\alpha_{m\perp}^* \Phi d_m = 0$, or $\Phi d_m = \alpha_m \rho_m^*$, for some ρ_m $(1 \times r_m)$. This result was first proved by Franses and Kunst (1995).

We reparametrize the model by introducing the parameters $\Phi_m = \Phi d_m$. The vectors d_m come in complex conjugate pairs, which also holds for the new parameters. The deterministic terms in the equation become:

$$\Phi D_t = \sum_{m=1}^{\tilde{s}} \Phi_m \bar{z}_m^t.$$

We now restrict Φ_m $(m = 2, \ldots, s)$, by $\Phi_m = \alpha_m \rho_m^*$. In this way the oscillating trends are avoided, while leaving open the possibility of a linear trend generated by the unit root $z = 1$. If we also want to restrict this, we further assume that $\alpha_{1\perp}' \Phi_1 = 0$.

We conclude this section by a few illustrative examples.

2.5 Examples

Some examples of models for annual, semi-annual and quarterly data are given.

2.5.1 Annual data

If $z = 1$ is the only unit root in the process, then $p(z) = 1 - z$ and $X_t^{(1)} = X_{t-1}$. Model (2) is the usual error correction model for $I(1)$ variables:

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t,$$

where further dynamics and deterministic terms are left out.

2.5.2 Semi-annual data

If the unit roots are $z = \pm 1$, then $p(z) = (1 - z)(1 + z) = 1 - z^2$, and we find:

$$\begin{aligned} X_t^{(1)} &= \frac{(1+L)L}{2} X_t = \frac{1}{2}(X_{t-1} + X_{t-2}), \\ X_t^{(2)} &= \frac{(1-L)L}{2} X_t = \frac{1}{2}(X_{t-1} - X_{t-2}), \end{aligned}$$

such that (2) becomes

$$X_t - X_{t-2} = \frac{1}{2}\alpha_1\beta_1'(X_{t-1} + X_{t-2}) + \frac{1}{2}\alpha_2\beta_2'(X_{t-1} - X_{t-2}) + \varepsilon_t.$$

The reason for considering this case is that interpretation is somewhat easier than for quarterly models.

Consider for instance a process consisting of semi-annual income and consumption. In this case, $X_t^{(1)}$ is just the annual average, and the model specifies that this process is a non-stationary $I(1)$ process which cointegrates, such that annual consumption follows annual income in a stationary way through the cointegrating coefficients β_1. The process $X_t^{(2)}$, however, measures the variation within a year and the seasonal unit root at $z = -1$ implies that this process has a seasonal non-stationarity. This means that when averaged within a year it becomes stationary. The cointegrating vector β_2 gives the linear combination of annual variation of consumption and income which cointegrates.

Thus, not only the non-stationary yearly average but also the non-stationary variation within a year have to move together according to the model.

In order to understand the type of non-stationarity induced by a unit root at $z = -1$, consider the process

$$X_t = (-1)^t S_t^{(2)} = (-1)^t \sum_{j=0}^{t} (-1)^j \varepsilon_j = \sum_{j=0}^{t} (-1)^j \varepsilon_{t-j},$$

which enters the representation theorem.

Since the normal distribution is symmetric, the process $S_t^{(2)}$ is a random walk, and the factor $(-1)^t$ changes the sign of every second term, which give rise to the oscillating behavior that we see in seasonally varying processes. It is obvious that differencing such a process will not give stationarity, whereas one can obtain a stationary process by smoothing using a moving average $\frac{1}{2}(X_t + X_{t-1})$.

Note that when the random walk $S_t^{(2)}$ is positive for an interval, then X_t oscillates systematically between positive and negative values, but when $S_t^{(2)}$ gets too close to zero, or a large draw of ε_t occurs, then it can change sign with the result that the peaks of X_t are shifted one period, such that "summer becomes winter". This is a characteristic property of processes generated by the seasonal error correction model. It is easy in the example to check the role of constant, linear term, and seasonal dummies on the behavior of the process.

2.5.3 Quarterly data

Next consider quarterly data and unit roots at $z = \pm 1, \pm i$. In this case:

$$p(z) = (1 - z)(1 + z)(1 + iz)(1 - iz) = 1 - z^4.$$

The processes that are needed in the error correction model are

$$\begin{aligned}
X_t^{(1)} &= \tfrac{1}{4}(X_{t-1} + X_{t-2} + X_{t-3} + X_{t-4}), \\
X_t^{(2)} &= \tfrac{1}{4}(X_{t-1} - X_{t-2} + X_{t-3} - X_{t-4}), \\
X_t^{(3)} &= \tfrac{1}{4i}(X_{t-1} + iX_{t-2} - X_{t-3} - iX_{t-4}), \\
X_t^{(4)} &= -\tfrac{1}{4i}(X_{t-1} - iX_{t-2} - X_{t-3} + iX_{t-4}).
\end{aligned}$$

The error correction model contains 4 terms, and we express them using real variables, see also (15). If we let $X_t^{(3)} = X_{Rt}^{(3)} + iX_{It}^{(3)}$ and $X_t^{(4)} = \bar{X}_t^{(3)} = X_{Rt}^{(3)} - iX_{It}^{(3)}$ we find

$$\begin{aligned}
X_{Rt}^{(3)} &= \frac{1}{4}(X_{t-2} - X_{t-4}), \\
X_{It}^{(3)} &= -\frac{1}{4}(X_{t-1} - X_{t-3}).
\end{aligned}$$

The error correction model becomes

$$\begin{aligned} X_t - X_{t-4} &= \alpha_1 \beta_1' X_t^{(1)} + \alpha_2 \beta_2' X_t^{(2)} \\ &\quad + (\alpha_R \beta_R' + \alpha_I \beta_I')(X_{t-2} - X_{t-4}) \\ &\quad + (\alpha_R \beta_I' - \alpha_I \beta_R')(X_{t-1} - X_{t-3}) + \varepsilon_t, \end{aligned}$$

where we have absorbed the factor $\frac{1}{4}$ into the coefficients, and for ease of notation we let $\alpha_3 = \alpha_R + i\alpha_I$, $\alpha_4 = \bar{\alpha}_3$, $\beta_3 = \beta_R + i\beta_I$, $\beta_4 = \bar{\beta}_3$.

Note that the coefficient matrix to $X_{t-2} - X_{t-4}$ is rather complicated. It need not even have reduced rank. The same parameters appear in the coefficient to $X_{t-1} - X_{t-3}$. Thus, the type of polynomial cointegration obtained here is difficult to interpret. It has been suggested to assume that $\alpha_R \beta_I' - \alpha_I \beta_R' = 0$, in order to simplify the equations (see Lee 1992), but it is seen that this is a peculiar restriction on all coefficients, which is hard to interpret. If instead $\beta_I = 0$ then the equations contain the term

$$\alpha_R \beta_R' (X_{t-2} - X_{t-4}) - \alpha_I \beta_R' (X_{t-1} - X_{t-3}) = (\alpha_R L - \alpha_I) \beta_R' (X_{t-1} - X_{t-3}).$$

This shows that X_t is polynomially cointegrated. This has the advantage that only one set of linear combinations of $(1-L^2)X_t$ appears, and the interpretation is that $\beta_R'(X_{t-1} - X_{t-3})$ is either stationary or cointegrates with its own lag. If also $\alpha_I = 0$, the equation contains a term of the form $\alpha_R \beta_R'(X_{t-2} - X_{t-4})$ with the interpretation that $X_{t-2} - X_{t-4}$ cointegrates with cointegrating vector β_R.

The error correction model is different if the process only has roots at $z = 1, \pm i$, since then $p(z) = (1-z)(1+iz)(1-iz) = (1-z)(1+z^2)$. In this case,

$$\begin{aligned} X_t^{(1)} &= \frac{(1+L^2)}{2} L X_t = \frac{1}{2}(X_{t-1} + X_{t-3}), \\ X_t^{(2)} &= \frac{(1-L)(1-iL)}{2(1-i)} L X_t = \frac{1}{2(1-i)} (\Delta X_{t-1} - i \Delta X_{t-2}), \\ X_t^{(3)} &= \frac{(1-L)(1+iL)}{2(1+i)} L X_t = \frac{1}{2(1+i)} (\Delta X_{t-1} + i \Delta X_{t-2}), \end{aligned}$$

where $\Delta = 1 - L$. We find the real and imaginary part as follows:

$$X_{Rt}^{(2)} = \frac{1}{4} \Delta_2 X_{t-1},\ X_{It}^{(2)} = \frac{1}{4} \Delta^2 X_{t-1},$$

where $\Delta_2 = 1 - L^2$. The error correction model is:

$$\begin{aligned} & X_t - X_{t-1} + X_{t-2} - X_{t-3} \\ = \ & \alpha_1 \beta_1' (X_{t-1} + X_{t-3}) + (\alpha_R \beta_R' + \alpha_I \beta_I') \Delta_2 X_{t-1} + (\alpha_R \beta_I' - \alpha_I \beta_R') \Delta^2 X_{t-1} + \varepsilon_t. \end{aligned}$$

In this section we have given a general version of Granger's representation theorem which clarifies when we get a seasonally cointegrated solution to autoregressive equations and when the solution is integrated of order 1 at seasonal frequency. We also gave a discussion of trends generated by constant, linear term, and seasonal dummies, and the restrictions needed to avoid them. In the next section these results are applied to define the statistical models we want to analyze.

3 The models for seasonal cointegration and their statistical analysis

In this section the statistical model for autoregressive processes integrated of order 1 at seasonal frequency which allows for seasonal cointegration is defined. Various models defined by restrictions on the deterministic terms are given. We discuss Gaussian maximum likelihood estimation and the formulation of some hypotheses on the cointegrating ranks, the cointegrating vectors, and the adjustment coefficients.

3.1 Statistical models defined by restrictions on deterministic terms

The $n-$dimensional vector autoregressive model for seasonal cointegration is defined by the equations:

$$p(L)X_t = \sum_{m=1}^{s} \alpha_m \beta_m^* X_t^{(m)} + \sum_{j=1}^{k} \Gamma_j p(L) X_{t-j} + \Phi D_t + \varepsilon_t, \; t = 1, \ldots, T. \quad (10)$$

Here ε_t are i.i.d. $N_n(0, \Omega)$, and the parameters α_m, β_m $(m = 1, \ldots, s)$, Γ_j $(j = 1, \ldots, k)$, Φ, and Ω are freely varying, except that the α_m and β_m come in complex conjugate pairs. We assume that D_t consists of deterministic terms. Note that the lag length is $l = k + s$ since $p(L)$ is an s'th order lag polynomial. The dimension of α_m and β_m is $n \times r_m$, and the initial values are fixed in the analysis of the likelihood function.

If the roots of the process are also roots of unity, corresponding to a given frequency $\tilde{s}$ of the data, we can introduce seasonal dummies D_t in the model.

As seen in Section 2 they give rise to trends in the process and it was shown how these can be avoided by restriction of the parameters. We decompose the parameter Φ as $\Phi D_t = \sum_{m=1}^{\tilde{s}} \Phi_m \bar{z}_m^t$, and assume that $\Phi_m = \alpha_m \rho_m^*$, for some matrix ρ_m of dimension $1 \times r_m$, $(m = 1, \ldots, s)$.

In this case:

$$
\begin{aligned}
p(L)X_t &= \sum_{m=1}^{s} \alpha_m \begin{pmatrix} \beta_m \\ \rho_m \end{pmatrix}^* \begin{pmatrix} X_t^{(m)} \\ \bar{z}_m^t \end{pmatrix} \\
&\quad + \sum_{j=1}^{k} \Gamma_j p(L) X_{t-j} + \sum_{m=s+1}^{\tilde{s}} \Phi_m \bar{z}_m^t + \varepsilon_t,
\end{aligned} \tag{11}
$$

where $\{z_{s+1}, \ldots, z_{\tilde{s}}\}$ are the roots of unity which are not unit roots of the process.

If instead we allow for a linear trend, we do not restrict at zero frequency but use the model

$$
\begin{aligned}
p(L)X_t &= \alpha_1 \beta_1' X_t^{(1)} + \sum_{m=2}^{s} \alpha_m \begin{pmatrix} \beta_m \\ \rho_m \end{pmatrix}^* \begin{pmatrix} X_t^{(m)} \\ \bar{z}_m^t \end{pmatrix} \\
&\quad + \sum_{j=1}^{k} \Gamma_j p(L) X_{t-j} + \Phi_1 + \sum_{m=s+1}^{\tilde{s}} \Phi_m \bar{z}_m^t + \varepsilon_t.
\end{aligned} \tag{12}
$$

In these two last models the parameters specified are varying freely, with the only restriction that the parameters α_m, β_m, and Φ_m come in complex conjugate pairs. Note how the roots of unity that do not correspond to unit roots in the process enter with unrestricted coefficients Φ_m $(s < m \leq \tilde{s})$, and do not give rise to trends in the process. Thus, consider for instance the terms with $m = s+1$ and $m = s+2$, in (12). Let $z_{s+1} = \bar{z}_{s+2} = e^{i\theta}$ and $\Phi_{s+1} = \bar{\Phi}_{s+2} = \Phi_R + i\Phi_I$, say. The corresponding terms in the regression are:

$$
\Phi_{s+1} \bar{z}_{s+1}^t + \Phi_{s+2} \bar{z}_{s+2}^t = 2(\Phi_R \cos(t\theta) + \Phi_I \sin(t\theta)).
$$

Thus, the regression will include the extra regressors $\cos(t\theta)$ and $\sin(t\theta)$, with unrestricted coefficients. This term will not generate a trend since

$$
\sum_{i=0}^{t} z_m^i \bar{z}_{s+1}^i
$$

is bounded, whenever the index m corresponds to the roots in the characteristic polynomial, that is, $m \leq s < s+1$.

3.2 Some algorithms for estimation

The statistical analysis of (10) leads to a non-linear regression problem since the coefficients α_m and β_m enter through their product. We here discuss estimation of the model without restrictions on deterministic terms and mention in the end of this subsection how to modify the algorithm if the deterministic terms are restricted, as in models (11) and (12).

Since the cointegration model (10) does not restrict the matrices Γ_j and Φ, we can concentrate the likelihood function with respect to these and define residuals R_{0t}, $R_{1t}^{(m)}$, and $R_{\varepsilon t}$ by regression of $p(L)X_t$, $X_t^{(m)}$, and ε_t on D_t and lagged values of $p(L)X_t$. Thus, we get the equation:

$$R_{0t} = \sum_{m=1}^{s} \alpha_m \beta_m^* R_{1t}^{(m)} + R_{\varepsilon t}. \tag{13}$$

An algorithm for estimating this model (see Boswijk 1995), is the following: for fixed β coefficients the model is a linear regression model that determines the $\alpha's$ and Ω by regression of n variables R_{0t} on $\sum_{m=1}^{s} r_m$ variables $\beta_1^* R_{1t}^{(1)}, \ldots, \beta_m^* R_{1t}^{(m)}$. For fixed values of α and Ω, however, we have a linear regression model in the β coefficients which can be estimated by generalized least squares. This determines a switching algorithm, which in each step increases the likelihood function, but the second step involves vectorizing β_m so we need a total of $n \sum_{m=1}^{s} r_m$ regressors.

Another algorithm can be based on first and second derivatives of the likelihood function and an application of the Gauss Newton algorithm. This algorithm also involves a large number of variables in general.

Finally, we describe an algorithm which is slightly simpler, and which can be proved to give estimators which are asymptotically equivalent to the maximum likelihood estimators, since the regressors $X_t^{(m)}$ are asymptotically uncorrelated, in the sense that

$$T^{-2} \sum_{t=1}^{T} X_t^{(m)} X_t^{(n)*} \xrightarrow{P} 0,\ z_n \neq z_m,$$

see Corollary 7.

The idea of the algorithm is that when focussing on one frequency we can concentrate out the other regressors by ignoring the constraint of reduced rank at these other frequencies, see Lee (1992). It is an interesting consequence that if one performs a usual cointegration analysis for the unit root $z = 1$ we get consistent estimators, even though there may be seasonal cointegration. Furthermore we can make valid asymptotic inference using the results given later. Thus we can for instance test hypotheses on the coefficient of the cointegrating vector with asymptotic χ^2 tests. What breaks down completely of course is the interpretation, since $\beta' X_t$ is not stationary.

We illustrate the situation of a complex root, since the real roots 1 and -1 are easily handled in a similar way.

Consider therefore the situation where, say, $z_1 = e^{i\theta}$ and $z_2 = e^{-i\theta}$ are two complex roots with $0 < \theta < \pi$. Note that $A(e^{i\theta}) = \bar{A}(e^{-i\theta})$ and $X_t^{(2)} = \bar{X}_1^{(1)}$. For notational reasons we use α and β without subscripts now and let $\alpha_1 = \alpha, \beta_1 = \beta$ which in turn implies $\alpha_2 = \bar{\alpha}$, , and $\beta_2 = \bar{\beta}$. Thus we write the model equation (13) as

$$R_{0t} = \alpha\beta^* R_{1t}^{(1)} + \bar{\alpha}\bar{\beta}^* \bar{R}_{1t}^{(1)} + \sum_{m=3}^{s} \alpha_m \beta_m^* R_{1t}^{(m)} + R_{\varepsilon t}. \tag{14}$$

We concentrate with respect to $R_{1t}^{(m)}$, where $m \neq (1, 2)$, that is, we remove the restriction of reduced rank at $z_3, \ldots, z_s$. This gives residuals U_{0t}, U_{1t}, and $U_{\varepsilon t}$ and we find the equations

$$\begin{aligned} U_{0t} &= \alpha\beta^* U_{1t} + \bar{\alpha}\bar{\beta}^* \bar{U}_{1t} + U_{\varepsilon t} \\ &= 2Real[(\alpha_R + i\alpha_I)(\beta_R - i\beta_I)'(U_{Rt} + iU_{It})] + U_{\varepsilon t} \\ &= 2[(\alpha_R \beta_R' + \alpha_I \beta_I')U_{Rt} + (\alpha_R \beta_I' - \alpha_I \beta_R')U_{It}] + U_{\varepsilon t} \\ &= 2(\alpha_R, -\alpha_I) \begin{pmatrix} \beta_R & -\beta_I \\ \beta_I & \beta_R \end{pmatrix}' \begin{pmatrix} U_{Rt} \\ U_{It} \end{pmatrix} + U_{\varepsilon t} \\ &= \breve{\alpha} \boldsymbol{\beta}' \begin{pmatrix} U_{Rt} \\ U_{It} \end{pmatrix} + U_{\varepsilon t}, \end{aligned} \tag{15}$$

where $\breve{\alpha} = 2(\alpha_R, -\alpha_I)$, and we use the matrix notation $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ for matrices with complex structure:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_R & -\alpha_I \\ \alpha_I & \alpha_R \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_R & -\beta_I \\ \beta_I & \beta_R \end{pmatrix}.$$

In Appendix A the matrix representation of complex numbers and matrices is explained. This representation is noted throughout by boldface. Since the roots come in complex pairs the sum $\sum_{m=3}^{s} \alpha_m \beta_m^* X_t^{(m)}$ is real, such that both U_{0t} and $U_{\varepsilon t}$ are real. The statistical problem appears to be a reduced rank regression problem, at least if $n > 2r_1$, but the matrix $\boldsymbol{\beta}$ is not unrestricted since, by construction, it must have complex structure.

In order to express the partially maximized likelihood function we introduce the product moments

$$
\begin{aligned}
S_{11} &= T^{-1}\sum_{t=1}^{T}\begin{pmatrix} U_{Rt} \\ U_{It}\end{pmatrix}\begin{pmatrix} U_{Rt} \\ U_{It}\end{pmatrix}', \quad (2n \times 2n)\,, \\
S_{10} &= T^{-1}\sum_{t=1}^{T}\begin{pmatrix} U_{Rt} \\ U_{It}\end{pmatrix}U_{0t}', \quad (2n \times n)\,, \\
S_{00} &= T^{-1}\sum_{t=1}^{T}U_{0t}U_{0t}', \quad (\, n \times n)\,, \\
S_{\varepsilon 1} &= T^{-1}\sum_{t=1}^{T}U_{\varepsilon t}\begin{pmatrix} U_{Rt} \\ U_{It}\end{pmatrix}', \quad (n \times 2n)\,.
\end{aligned}
$$

Finally, we define $S_{11.0} = S_{11} - S_{10}S_{00}^{-1}S_{01}$.

For fixed value of $\boldsymbol{\beta}$ we can concentrate the likelihood function with respect to the parameters $\breve{\alpha} = 2(\alpha_R, -\alpha_I)$ and Ω and find, apart from a constant factor,

$$
L_{\max}^{-\frac{2}{T}}(\boldsymbol{\beta}) = |\hat{\Omega}| = |S_{00} - S_{01}\boldsymbol{\beta}(\boldsymbol{\beta}'S_{11}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'S_{10}| = |S_{00}|\frac{|\boldsymbol{\beta}'S_{11.0}\boldsymbol{\beta}|}{|\boldsymbol{\beta}'S_{11}\boldsymbol{\beta}|}. \tag{16}
$$

This minimization cannot be solved as an eigenvalue problem since the $2n \times 2r$ matrix $\boldsymbol{\beta}$ has complex structure while S_{11} and $S_{11.0}$ do not have complex structure.

We can minimize (16) by an iterative procedure using the Gauss Newton algorithm or we can use the idea of switching between $(\breve{\alpha}, \Omega)$ and $\boldsymbol{\beta}$ as in (15). Applying the switching algorithm here only involves $2nr_1$ parameters from the cointegrating relations and a similar number from the adjustment coefficients.

Finally, the maximum likelihood estimator can be calculated iteratively as follows. For fixed values of $\beta_2, \ldots, \beta_s$ we concentrate the likelihood function with respect to $\alpha_2, \ldots, \alpha_s$. Then the equations have the form (15) and we apply the switching algorithm to determine α_1 and β_1. Next fix $\beta_1, \beta_3, \ldots, \beta_s$ and repeat the procedure as above until convergence. In this way one can, by focussing on one frequency at a time, reduce the dimension of the matrices involved in the regressions.

If instead we consider the problem of reduced rank at $\theta = 0$ or π then we get the product moments as before but now with $X_t^{(1)}$, say, corrected for all the other components. In this case all residuals are real and the matrices S_{11}, S_{01}, and S_{00} are all of dimension $n \times n$, and the problem can then be solved by reduced rank regression, see Lee (1992).

Finally we can use the same ideas to estimate models (11) and (12) with the various restrictions on deterministic terms. The coefficients Φ_j with $j > s$, and possibly $j = 1$, can be concentrated out in the preliminary regression, and in the reduced rank regressions we just replace $X_t^{(1)}$ by the extended variables $X_{et}^{(1)} = (X_t^{(1)\prime}, \bar{z}_1^t)'$, see (11) and (12), such that the residuals U in (15) are based upon the variable

$$(X_{Rt}^{(1)\prime}, \cos(\theta t), X_{It}^{(1)\prime}, -\sin(\theta t))',$$

and the cointegrating coefficient is

$$\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\rho} \end{pmatrix} = \begin{pmatrix} \beta_R & -\beta_I \\ \rho_R & -\rho_I \\ \beta_I & \beta_R \\ \rho_I & \rho_R \end{pmatrix}.$$

In summary the analysis is: in model (11), we concentrate out the coefficients Φ_j, $j > s$, using regression. The switching algorithm is then applied to the residuals from the extended variables $X_{et}^{(m)} = (X_t^{(m)\prime}, \bar{z}_m^t)$. In model (12) we eliminate Φ_1 by regression and do not extend the variable $X_t^{(1)}$.

The algorithm has been programmed in Gauss, see Schaumburg (1996) and RATS, see Dahl Pedersen (1996).

3.3 Hypotheses of interest

The first hypothesis of interest is the test for reduced rank at complex frequency. This requires maximization of the likelihood function under model $H(r)$, that is, the assumption of reduced rank r at the complex frequency θ as discussed in the previous subsection. We then compare the maximum with the maximum obtained from the unrestricted VAR, which corresponds to $r = n$. Thus, the test statistic is

$$-2\log Q(H(r)|H(n)) = T\log\left(\frac{|\boldsymbol{\beta}' S_{11.0}\boldsymbol{\beta}||S_{11}|}{|\boldsymbol{\beta}' S_{11}\boldsymbol{\beta}||S_{11.0}|}\right).$$

Other hypotheses of interest are hypotheses on the cointegrating coefficients $\boldsymbol{\beta}$. The most interesting perhaps is the hypothesis that $\boldsymbol{\beta}$ is real, since without this simple structure the interpretation becomes rather tedious. This hypothesis is formulated by Lee (1992), and in the present notation becomes the restriction

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_R & 0 \\ 0 & \beta_R \end{pmatrix}. \tag{17}$$

Note that due to the non-identification of $\boldsymbol{\beta}$ we can equivalently formulate this hypothesis as $\beta_R = \beta_I$. Finally, we consider the assumption that

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_R & 0 \\ 0 & \alpha_R \end{pmatrix}, \tag{18}$$

which allows for a simple interpretation, see the examples in sub-section 2.4. Maximization of the concentrated likelihood function (16) under any of these restrictions again requires an iterative algorithm. By comparing the obtained maxima with and without restrictions (17) and (18) we obtain the likelihood ratio test statistic. Prior hypotheses about the structure of the cointegrating relations are tested by the likelihood ratio test, by suitably modifying the maximization algorithm.

The asymptotic results that allow these procedures to be applied in practice are given in the next sections.

4 Asymptotic results

This section deals with some technical results on asymptotic behavior of various processes and product moments. The proofs are given in Appendix A. We assume throughout that the processes are generated by autoregressive equations without deterministic terms and that the ε are i.i.d. with mean zero and variance Ω. We start with the sums $S_t^{(m)}$ and then find the limiting behavior of $X_t^{(m)}$ and finally investigate S_{11}, S_{10}, S_{00}, and $S_{\varepsilon 1}$ which are based on residuals from the regression (15).

The limit distribution of the $S_t^{(m)}$ is found in Chan and Wei (1988) who show the following result:

Lemma 5 *If $S_t^{(m)} = \sum_{j=0}^t z_m^j \varepsilon_j$ and $z_m = \exp(i\theta_m)$, then*

$$T^{-\frac{1}{2}} S_{[Tu]}^{(m)} \xrightarrow{w} \delta_m \left(W_R^{(m)}(u) + iW_I^{(m)}(u) \right) = \delta_m W_m, \quad \delta_m = \begin{cases} \frac{1}{\sqrt{2}}, & 0 < \theta_m < \pi \\ 1, & \theta_m = 0, \pi \end{cases}, \tag{19}$$

where $W_R^{(m)}$ and $W_I^{(m)}$ are independent Brownian motions with variance matrix Ω, if $\theta_m \neq (0, \pi)$. For $\theta_m \in \{0, \pi\}$, W_m is a Brownian motion with variance matrix Ω. Moreover these Brownian motions are independent for different values of θ_m.

Another result that follows from their calculations is the following.

Theorem 6 *For* $T \to \infty$ *:*

$$T^{-2}\sum_{t=1}^{T} S_t^{(m)} S_t^{(n)*} \xrightarrow{w} \delta_m \delta_n \int_0^1 W_m W_n^* du, \tag{20}$$

$$T^{-1}\sum_{t=1}^{T} S_{t-1}^{(m)} \bar{z}_n^t \varepsilon_t' \xrightarrow{w} \delta_m \delta_n \int_0^1 W_m (dW_n^*). \tag{21}$$

If further $f(t)$ *is a complex function such that* $F(t) = \sum_{i=1}^{t} f(i)$ *is bounded, then*

$$T^{-2}\sum_{t=1}^{T} S_t^{(m)} f(t) S_t^{(n)*} \xrightarrow{P} 0. \tag{22}$$

■

The non-stationary component of $X_t^{(m)}$ is $\bar{z}_m^t S_t^{(m)}$ and the results of Lemma 5 and Theorem 6 translate into results about product moments involving $X_t^{(m)}$.

Corollary 7 *The asymptotic properties of the process and product moment matrices are given by*

$$T^{-\frac{1}{2}} \bar{z}_m^{[Tu]} X_{[Tu]}^{(m)} \xrightarrow{w} \delta_m C_m W_m(u) \tag{23}$$

$$T^{-2}\sum_{t=1}^{T} X_t^{(m)} X_t^{(m)*} \xrightarrow{w} \delta_m^2 C_m \int_0^1 W_m W_m^* du C_m^*, \tag{24}$$

$$T^{-2}\sum_{t=1}^{T} X_t^{(m)} X_t^{(n)*} \xrightarrow{w} 0, \tag{25}$$

$$T^{-1}\sum_{t=1}^{T} X_t^{(m)} \varepsilon_t' \xrightarrow{w} \delta_m^2 C_m \int_0^1 W_m (dW_m)^*, \tag{26}$$

where δ_m *and* W_m *are given in Lemma 5, and* C_m *in Theorem 3.*

Next we find the asymptotic properties of the product moment matrices S_{00}, S_{10}, S_{11}, and $S_{1\varepsilon}$. These are defined in terms of residuals U_{0t}, U_{1t}, and $U_{\varepsilon t}$ which in turn are defined in terms of $X_{Rt}^{(1)}$, $X_{It}^{(1)}$, and $p(L)X_t$ corrected for $X_t^{(m)}$, $m \neq (1, 2)$.

In the following we let $z_1 = e^{i\theta}$, $0 < \theta < \pi$, and use the notation $S_t^{(1)} = S_{Rt}^{(1)} + iS_{It}^{(1)}$, $X_t^{(1)} = X_{Rt}^{(1)} + iX_{It}^{(1)}$ and $C_1 = C_R^{(1)} + iC_I^{(1)}$. With the matrix representation of the complex processes, see Appendix A, we find:

$$\mathbf{X}_t^{(1)} = \begin{pmatrix} X_{Rt}^{(1)} & -X_{It}^{(1)} \\ X_{It}^{(1)} & X_{Rt}^{(1)} \end{pmatrix}, \mathbf{C}_1 = \begin{pmatrix} C_R^{(1)} & -C_I^{(1)} \\ C_I^{(1)} & C_R^{(1)} \end{pmatrix}, \mathbf{S}_t^{(1)} = \begin{pmatrix} S_{Rt}^{(1)} & -S_{It}^{(1)} \\ S_{It}^{(1)} & S_{Rt}^{(1)} \end{pmatrix}, \tag{27}$$

such that the complex representation

$$X_t^{(1)} = C_1 S_{t-1}^{(1)} \bar{z}_1^t + o_P(T^{\frac{1}{2}}),$$

in matrix notation becomes

$$\mathbf{X}_t^{(1)} = \begin{pmatrix} C_R^{(1)} & -C_I^{(1)} \\ C_I^{(1)} & C_R^{(1)} \end{pmatrix} \begin{pmatrix} S_{Rt}^{(1)} & -S_{It}^{(1)} \\ S_{It}^{(1)} & S_{Rt}^{(1)} \end{pmatrix} \begin{pmatrix} \cos(t\theta) & \sin(t\theta) \\ -\sin(t\theta) & \cos(t\theta) \end{pmatrix} + o_P(T^{\frac{1}{2}}).$$

From this we find by multiplying from the right by $(1, 0)'$, that

$$\begin{pmatrix} X_{Rt}^{(1)} \\ X_{It}^{(1)} \end{pmatrix} = \mathbf{C}_1 \mathbf{S}_{t-1}^{(1)} \begin{pmatrix} \cos(t\theta) \\ -\sin(t\theta) \end{pmatrix} + o_P(T^{\frac{1}{2}}).$$

Define the $\sigma-$field $\mathcal{F}_t$ as

$$\mathcal{F}_t = \sigma \left\{ p(L)X_{t-1}, \ldots, p(L)X_{t-k}, \beta_m^* X_t^{(m)}, \, m \neq 1, 2 \right\},$$

that is, the σ-field generated by the stationary processes in the model except those that are derived from $X_{Rt}^{(1)}$ and $X_{It}^{(1)}$. Note that $\mathcal{F}_t$ is generated by variables before time t, since $X_t^{(m)}$ depends on lagged X_t.

We define the variances and covariances

$$Var \left\{ \begin{array}{c} p(L)X_t \\ \boldsymbol{\beta}' \begin{pmatrix} X_{Rt}^{(1)} \\ X_{It}^{(1)} \end{pmatrix} \end{array} \middle| \mathcal{F}_t \right\} = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}.$$

Lemma 8 *The following identities hold*

$$\Sigma_{0\beta} = \breve{\alpha} \Sigma_{\beta\beta}, \tag{28}$$

$$\Sigma_{00} = \breve{\alpha}' \Sigma_{\beta\beta} \breve{\alpha} + \Omega, \tag{29}$$

$$\Sigma_{\beta\beta.0}^{-1} \Sigma_{\beta 0} \Sigma_{00}^{-1} = \breve{\alpha}' \Omega^{-1}, \tag{30}$$

$$\Sigma_{\beta\beta}^{-1} - \Sigma_{\beta\beta.0}^{-1} + \Sigma_{\beta\beta.0}^{-1} \Sigma_{\beta 0} \Sigma_{00}^{-1} \breve{\alpha} = 0. \tag{31}$$

Theorem 9 *Asymptotic properties of product moment matrices defined from* $X_t^{(1)}$, *corrected for* $X_t^{(m)}$ $(m \neq 1, 2)$, *and lagged values of* $p(L)X_t$ *are given by*

$$T^{-1}S_{11} \xrightarrow{w} \frac{1}{4}\mathbf{C}_1 \int_0^1 \mathbf{W}_1\mathbf{W}_1' du \mathbf{C}_1', \; 0 < \theta < \pi,$$

$$S_{1\varepsilon} \xrightarrow{w} \frac{1}{2}\mathbf{C}_1 \int_0^1 \mathbf{W}_1(d\mathbf{W}_1)' \begin{pmatrix} I_n \\ 0 \end{pmatrix}, \; 0 < \theta < \pi,$$

where

$$\mathbf{W}_1 = \begin{pmatrix} W_R^{(1)} & -W_I^{(1)} \\ W_I^{(1)} & W_R^{(1)} \end{pmatrix}.$$

Note that S_{11} is $2n \times 2n$ but does not have complex structure. The limit of $T^{-1}S_{11}$, however, has complex structure as do the matrices $\mathbf{C}_1$ and $\mathbf{W}_1$. Note also that the $\mathbf{W}_1$ process appearing in Theorem 9 is the complex valued Brownian motion W_1 from Lemma 5 in the matrix representation of complex processes.

5 Asymptotic inference on rank and cointegrating relations

The main result about the estimator $\hat{\beta}$ is that it is asymptotically mixed Gaussian such that asymptotic inference on its coefficients can be conducted in the χ^2 distribution. The test statistic for hypotheses on the rank at seasonal frequency has a limit distribution, which is similar to the usual one at frequency zero, when expressed in terms of the complex Brownian motion.

5.1 The asymptotic distribution of $\hat{\boldsymbol{\beta}}$

Although it is necessary to apply numerical algorithms for calculating $\hat{\boldsymbol{\beta}}$, we can use the derived expression for the likelihood function (16) to obtain the asymptotic distribution of the maximum likelihood estimator. This is done by exploiting the fact that $\hat{\boldsymbol{\beta}}$ must be a solution to a set of first order conditions for maximizing (16).

The parameter $\boldsymbol{\beta}$ is not identified unless normalized in some way. This normalization can be accomplished by defining $\boldsymbol{\beta}_b = \boldsymbol{\beta}(\mathbf{b}'\boldsymbol{\beta})^{-1}$ for some $\mathbf{b}$ $(2n \times$

$2r$) of complex structure with the property that $\boldsymbol{\beta}'\mathbf{b}$ has full rank, and let $\boldsymbol{\alpha}_b = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{b}$, such that $\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\alpha}_b\boldsymbol{\beta}_b'$. For the analysis in the following it is convenient first to normalize the estimator on the true value $\boldsymbol{\beta}$ and choose $\mathbf{b} = \bar{\boldsymbol{\beta}} = \boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}$. We thus define $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\bar{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}})^{-1}$, and note that

$$\boldsymbol{\beta}'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = 0.$$

Thus, we only have to investigate the limit of $T\bar{\boldsymbol{\beta}}_\perp'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Results are given for the model without deterministic terms and later it is mentioned how they are modified for models (11) and (12). The result is given for the case of a complex frequency, since the result for the case $\theta = \pi$ can be proved exactly as for the case $\theta = 0$, which is well known in the literature, see Johansen (1996).

Theorem 10 *In the model with no deterministic term the asymptotic distribution of the estimator $\tilde{\boldsymbol{\beta}}$ is consistent and asymptotically mixed Gaussian:*

$$T\bar{\boldsymbol{\beta}}_\perp'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{w}{\rightarrow} [\int_0^1 \mathbf{F}\mathbf{F}'du]^{-1} \int_0^1 \mathbf{F}(d\mathbf{V})',$$

where

$$\mathbf{F} = \boldsymbol{\beta}_\perp'\mathbf{C}_1\mathbf{W}_1,$$

and

$$\mathbf{V} = (\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\mathbf{W}_1.$$

We have here used the complex matrix notation:

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}, \boldsymbol{\alpha} = \begin{pmatrix} \alpha_R & -\alpha_I \\ \alpha_I & \alpha_R \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_R & -\beta_I \\ \beta_I & \beta_R \end{pmatrix}.$$

Since the upper left hand corner and the lower right hand corner of the matrix $T\bar{\boldsymbol{\beta}}_\perp'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ are identical, there is some redundancy built into the notation. Note that $\mathbf{F}$ and $\mathbf{V}$ are independent. With the present notation the results appear as the usual ones for the case of a unit root $z = 1$, see Johansen (1996).

Proof. The proof that the maximum likelihood estimator is consistent can be given along the same lines as the proof of consistency in Johansen (1997). It is here pointed out that since the cointegration model is a sub-model of a Gaussian regression model, it is possible to find an upper bound of the likelihood function outside a neighborhood of the true value. This is then applied to prove consistency. In the following we assume that $\tilde{\boldsymbol{\beta}}$ exists and is consistent. It follows that $\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}} \xrightarrow{P} \Sigma_{\beta\beta}$, $\tilde{\boldsymbol{\beta}}' S_{11.0}\tilde{\boldsymbol{\beta}} \xrightarrow{P} \Sigma_{\beta\beta.0}$ and $S_{00} \xrightarrow{P} \Sigma_{00}$. From

The concentrated likelihood function is given by

$$-2\log L(\boldsymbol{\beta}) = T\log\frac{|\boldsymbol{\beta}' S_{11.0}\boldsymbol{\beta}|}{|\boldsymbol{\beta}' S_{11}\boldsymbol{\beta}|} + T\log|S_{00}|.$$

We expand the likelihood function around $\tilde{\boldsymbol{\beta}}$ applying the result

$$\log|(x+h)'A(x+h)| = \log|x'Ax| + 2tr\{(x'Ax)^{-1}x'Ah\} + O(|h|^2). \qquad (32)$$

This gives the first order condition

$$tr\{[(\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}} S_{11} - (\tilde{\boldsymbol{\beta}}' S_{11.0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{11.0}]\mathbf{h}\} = 0,$$

for all $\mathbf{h}$ of complex structure. Hence

$$[(\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{11} - (\tilde{\boldsymbol{\beta}}' S_{11.0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{11.0}]^c = 0, \qquad (33)$$

where $[\ldots]^c$ denotes the complexified matrix, see Appendix A. We first find the weak limit for the matrix in (33) before it is complexified. Multiplying from the right by $\boldsymbol{\beta}_\perp$, which has complex structure, we find

$$\begin{aligned}
&(\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{11}\boldsymbol{\beta}_\perp - (\tilde{\boldsymbol{\beta}}' S_{11.0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{11.0}\boldsymbol{\beta}_\perp \\
&= [(\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}})^{-1} - (\tilde{\boldsymbol{\beta}}' S_{11.0}\tilde{\boldsymbol{\beta}})^{-1}]\tilde{\boldsymbol{\beta}}' S_{11}\boldsymbol{\beta}_\perp + (\tilde{\boldsymbol{\beta}}' S_{11.0}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{10}S_{00}^{-1}S_{01}\boldsymbol{\beta}_\perp \\
&= [\Sigma_{\beta\beta}^{-1} - \Sigma_{\beta\beta.0}^{-1}]\tilde{\boldsymbol{\beta}}' S_{11}\boldsymbol{\beta}_\perp + \Sigma_{\beta\beta.0}^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1}S_{01}\boldsymbol{\beta}_\perp + o_P(1).
\end{aligned}$$

From (15) we find

$$S_{01}\boldsymbol{\beta}_\perp = \breve{\alpha}\boldsymbol{\beta}' S_{11}\boldsymbol{\beta}_\perp + S_{\varepsilon 1}\boldsymbol{\beta}_\perp = \breve{\alpha}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})' S_{11}\boldsymbol{\beta}_\perp + \breve{\alpha}\tilde{\boldsymbol{\beta}}' S_{11}\boldsymbol{\beta}_\perp + S_{\varepsilon 1}\boldsymbol{\beta}_\perp.$$

Inserting this above we find

$$\begin{aligned}
&[\Sigma_{\beta\beta}^{-1} - \Sigma_{\beta\beta.0}^{-1} + \Sigma_{\beta\beta.0}^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1}\breve{\alpha}]\tilde{\boldsymbol{\beta}}' S_{11}\boldsymbol{\beta}_\perp \\
&+\Sigma_{\beta\beta.0}^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1}[\breve{\alpha}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'\bar{\boldsymbol{\beta}}_\perp\boldsymbol{\beta}_\perp' S_{11}\boldsymbol{\beta}_\perp + S_{\varepsilon 1}\boldsymbol{\beta}_\perp] \\
&= -\breve{\alpha}'\Omega^{-1}[\breve{\alpha}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\bar{\boldsymbol{\beta}}_\perp\boldsymbol{\beta}_\perp' S_{11}\boldsymbol{\beta}_\perp + S_{\varepsilon 1}\boldsymbol{\beta}_\perp],
\end{aligned}$$

since the first term is zero by (31), and the coefficient simplifies by (30). The weak limit of this is

$$-\breve{\alpha}'\Omega^{-1}\left[\breve{\alpha}\mathbf{B}_\infty'\frac{1}{4}\int_0^1 \mathbf{F}\mathbf{F}'du + \frac{1}{2}(I_n, 0)\int_0^1 (d\mathbf{W}_1)\mathbf{F}'\right],$$

where $\mathbf{F} = \boldsymbol{\beta}'_{\perp}\mathbf{C}_1\mathbf{W}_1$ and $\mathbf{B}_\infty$ is the weak limit of $T\overline{\boldsymbol{\beta}}'_{\perp}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Thus, the limit of (33) becomes

$$\left[\breve{\alpha}'\Omega^{-1}\breve{\alpha}\mathbf{B}'_\infty\frac{1}{4}\int_0^1 \mathbf{F}\mathbf{F}'du - \frac{1}{2}\breve{\alpha}'\Omega^{-1}(I_n, 0)\int_0^1 (d\mathbf{W}_1)\mathbf{F}'\right]^c = 0.$$

We still have to simplify this result before we can solve the equation for $\mathbf{B}_\infty$. Since $\mathbf{B}'_\infty \int_0^1 \mathbf{F}\mathbf{F}'du$ and $\int_0^1 (d\mathbf{W}_1)\mathbf{F}'$ have complex structure, the first order condition (33) is equivalent to

$$\frac{1}{4}\left[\breve{\alpha}'\Omega^{-1}\breve{\alpha}\right]^c \mathbf{B}'_\infty \int_0^1 \mathbf{F}\mathbf{F}'du - \frac{1}{2}\left[\breve{\alpha}'\Omega^{-1}(I_n, 0)\right]^c \int_0^1 (d\mathbf{W}_1)\mathbf{F}' = 0,$$

where $[\breve{\alpha}'\Omega^{-1}\breve{\alpha}]^c = 2\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}$ and $[\breve{\alpha}'\Omega^{-1}(I_n, 0)]^c = \boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}$, and

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix}.$$

This shows that

$$\mathbf{B}_\infty = \left(\int_0^1 \mathbf{F}\mathbf{F}'du\right)^{-1} \int_0^1 \mathbf{F}(d\mathbf{V}),$$

where $\mathbf{V} = \mathbf{W}_1{}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}$. ■

Next we give a result for the estimator of $\boldsymbol{\beta}$ normalized on a matrix $\mathbf{b}$, that is, $\hat{\boldsymbol{\beta}}_b = \tilde{\boldsymbol{\beta}}(\mathbf{b}'\tilde{\boldsymbol{\beta}})^{-1}$.

Theorem 11 *Let $\boldsymbol{\beta}$ be the cointegrating vector at seasonal frequency, normalized by $\boldsymbol{\beta}'\mathbf{b} = I_{2r}$. In the model with no deterministic terms, $\hat{\boldsymbol{\beta}}_b$ is consistent and asymptotically mixed Gaussian:*

$$T(\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}) \xrightarrow{w} (I_{2n} - \boldsymbol{\beta}\mathbf{b}')\boldsymbol{\beta}_{\perp}[\int_0^1 \mathbf{F}\mathbf{F}'du]^{-1} \int \mathbf{F}(d\mathbf{V})', \tag{34}$$

where

$$\begin{aligned} \mathbf{F} &= \boldsymbol{\beta}'_{\perp}\mathbf{C}_1\mathbf{W}_1, \\ \mathbf{V} &= (\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\mathbf{W}_1. \end{aligned}$$

The asymptotic conditional variance matrix is

$$(I_{2n} - \boldsymbol{\beta}\mathbf{b}')\boldsymbol{\beta}_{\perp}[\int_0^1 \mathbf{F}\mathbf{F}'du]^{-1}\boldsymbol{\beta}'_{\perp}(I_{2n} - \mathbf{b}\boldsymbol{\beta}') \otimes (\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}, \tag{35}$$

which by Theorem 9 is estimated consistently by

$$T(I_{2n} - \hat{\boldsymbol{\beta}}_b\mathbf{b}')\hat{\boldsymbol{\beta}}_{b\perp}[4\hat{\boldsymbol{\beta}}'_{b\perp}S_{11}\hat{\boldsymbol{\beta}}_{b\perp}]^{-1}\hat{\boldsymbol{\beta}}'_{b\perp}(I_{2n} - \mathbf{b}\hat{\boldsymbol{\beta}}'_b) \otimes (\hat{\boldsymbol{\alpha}}'_b\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{\alpha}}_b)^{-1}. \tag{36}$$

Thus linear and non-linear overidentifying hypotheses on the coefficients of the just identified vector $\boldsymbol{\beta}_b$ can be tested asymptotically by construction of $t-$ratios using (36) as variance matrix.

Proof. The proof of (34) follows from Theorem 10 by the expansion

$$\hat{\boldsymbol{\beta}}_b = (I_{2n} - \boldsymbol{\beta}(\mathbf{b}'\boldsymbol{\beta})^{-1}\mathbf{b}')(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\mathbf{b}'\boldsymbol{\beta})^{-1} + O_P(|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}|^2).$$

The proof that (35) is a consistent estimator follows from Theorem 9. ■

If instead we consider the models (11) or (12) we get much the same results. A detailed study will show that the estimated cointegrating vectors $\hat{\beta}_m$ are T consistent but their extension $\hat{\rho}_m$ is only $T^{\frac{1}{2}}$ consistent. This gives some difficulties in the formulation, but the end result is that one can treat the full extended vector as asymptotically Gaussian with a variance matrix given by (36), see Harbo *et al.* (1998) for the details in the case of zero frequency.

5.2 Test for cointegrating rank

This section contains a test to determine the rank r of β at the seasonal frequency $z_1 = e^{i\theta}$. We here concentrate on deriving the result for testing at strictly complex frequencies, which yields a result similar to the usual test but involving complex Brownian motions. The results for $\theta = \pi$ can be found in Lee (1992). We focus on the model without deterministic terms and give the results for the other cases without proof.

Theorem 12 *In the model with no deterministic terms we assume that the cointegrating rank at complex seasonal frequency is r. The asymptotic distribution of the likelihood ratio test statistic for the hypothesis of $r < n$ cointegrating relations is asymptotically distributed as*

$$\frac{1}{2}tr\{\int_0^1 (d\mathbf{B})\mathbf{B}'[\int_0^1 \mathbf{B}\mathbf{B}'du]^{-1}\int_0^1 \mathbf{B}(d\mathbf{B}')\}, \tag{37}$$

where $\mathbf{B}$ is standard complex Brownian motion of dimension $2(n-r)$

$$\mathbf{B} = \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix}.$$

The distribution is tabulated by simulation in Table 1.

Proof. From (16) the maximized likelihood function for $n = r$ is

$$L_{\max}^{-\frac{2}{T}} = |S_{00}| \frac{|S_{11.0}|}{|S_{11}|}. \tag{38}$$

The likelihood ratio test statistic is then

$$Q^{-\frac{2}{T}}(H(r)|H(n)) = \frac{|S_{11}||\tilde{\boldsymbol{\beta}}' S_{11.0} \tilde{\boldsymbol{\beta}}|}{|S_{11.0}||\tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}}|}.$$

Now choose $\tilde{\boldsymbol{\beta}}_\perp$ orthogonal to $\tilde{\boldsymbol{\beta}}$ and use the identities

$$\begin{aligned}
&|(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}_\perp)'||S_{11}||(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}_\perp)'| \\
&= |(\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}_\perp)' S_{11} (\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}_\perp)'| \\
&= \left| \begin{array}{cc} \tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}} & \tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}}_\perp \\ \tilde{\boldsymbol{\beta}}_\perp' S_{11} \tilde{\boldsymbol{\beta}} & \tilde{\boldsymbol{\beta}}_\perp' S_{11} \tilde{\boldsymbol{\beta}}_\perp \end{array} \right| \\
&= \left| \tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}} \right| \left| \tilde{\boldsymbol{\beta}}_\perp' S_{11} \tilde{\boldsymbol{\beta}}_\perp - \tilde{\boldsymbol{\beta}}_\perp' S_{11} \tilde{\boldsymbol{\beta}} \left(\tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}} \right)^{-1} \tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}}_\perp \right| \\
&= \left| \tilde{\boldsymbol{\beta}}' S_{11} \tilde{\boldsymbol{\beta}} \right| \left| \tilde{\boldsymbol{\beta}}_\perp' S_{11.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp \right|,
\end{aligned}$$

and a similar one for the matrix $S_{11.0}$ to prove the expression

$$-2 \log Q(H(r)|H(n)) = -T \log \frac{\left| \tilde{\boldsymbol{\beta}}_\perp' S_{11.0\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp \right|}{\left| \tilde{\boldsymbol{\beta}}_\perp' S_{11.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp \right|}. \tag{39}$$

The idea of the proof is to derive the asymptotic distribution of (39) by noting that it is a function of $\hat{\boldsymbol{\beta}}$, for which the distribution is derived in Theorem 11.

From the consistency of $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}_\perp'$, it follows from Theorem 9 that

$$T^{-1} \tilde{\boldsymbol{\beta}}_\perp' S_{11.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp \xrightarrow{w} \frac{1}{4} \boldsymbol{\beta}_\perp' \mathbf{C}_1 \int_0^1 \mathbf{W}_1 \mathbf{W}_1' du \mathbf{C}_1' \boldsymbol{\beta}_\perp = \frac{1}{4} \int_0^1 \mathbf{F}\mathbf{F}' du,$$

and the same result holds for $\tilde{\boldsymbol{\beta}}_\perp' S_{11.0\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp$. Thus, the ratio in (39) tends to 1, and from (32),

$$\begin{aligned}
&-2 \log Q(H(r)|H(n)) \\
&= -T \log |I_{2n} - (\tilde{\boldsymbol{\beta}}_\perp' S_{11.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp)^{-1} \tilde{\boldsymbol{\beta}}_\perp' S_{10.\hat{\beta}} S_{00.\hat{\beta}}^{-1} S_{01.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp| \\
&= tr \left\{ (T^{-1} \tilde{\boldsymbol{\beta}}_\perp' S_{11.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp)^{-1} \tilde{\boldsymbol{\beta}}_\perp' S_{10.\hat{\beta}} S_{00.\hat{\beta}}^{-1} S_{01.\hat{\beta}} \tilde{\boldsymbol{\beta}}_\perp \right\} + o_P(1).
\end{aligned}$$

In order to determine the limit of this quantity we first consider

$$S_{00.\hat{\beta}} = S_{00} - S_{01}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}'_{\perp} S_{10}$$
$$\stackrel{P}{\to} \Sigma_{00} - \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0} = \Omega + \breve{\alpha}\Sigma_{\beta\beta}\breve{\alpha}' - \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0} = \Omega.$$

Next consider

$$\tilde{\boldsymbol{\beta}}'_{\perp} S_{10.\hat{\beta}} = \tilde{\boldsymbol{\beta}}'_{\perp} S_{10} - \tilde{\boldsymbol{\beta}}'_{\perp} S_{11}\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}' S_{11}\tilde{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{\beta}}' S_{10}$$
$$= \tilde{\boldsymbol{\beta}}'_{\perp} S_{1\varepsilon} + \tilde{\boldsymbol{\beta}}'_{\perp} S_{11}\boldsymbol{\beta}\breve{\alpha}' - \tilde{\boldsymbol{\beta}}'_{\perp} S_{11}\tilde{\boldsymbol{\beta}}'\Sigma_{\beta\beta}^{-1}\Sigma_{\beta 0} + o_P(1)$$
$$= \boldsymbol{\beta}'_{\perp} S_{1\varepsilon} - \boldsymbol{\beta}'_{\perp} S_{11}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\breve{\alpha}' + o_P(1)$$
$$= \boldsymbol{\beta}'_{\perp} S_{1\varepsilon} - \boldsymbol{\beta}'_{\perp} S_{11}\boldsymbol{\beta}_{\perp}\bar{\boldsymbol{\beta}}'_{\perp}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})\breve{\alpha}' + o_P(1).$$

From Theorems 9 and 10 we find that this converges towards

$$\tfrac{1}{2}\int_0^1 \mathbf{F}(d\mathbf{W}_1')\begin{pmatrix} I_n \\ 0 \end{pmatrix} - \tfrac{1}{4}\int_0^1 \mathbf{F}(d\mathbf{W}_1')\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}\breve{\alpha}'$$
$$= \tfrac{1}{2}\int_0^1 \mathbf{F}(d\mathbf{W}_1')(I_{2n} - \boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}')\begin{pmatrix} I_n \\ 0 \end{pmatrix}$$
$$= \tfrac{1}{2}\int_0^1 \mathbf{F}(d\mathbf{W}_1')\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\begin{pmatrix} I_n \\ 0 \end{pmatrix}.$$

Thus,

$$\tilde{\boldsymbol{\beta}}'_{\perp} S_{10.\hat{\beta}} S_{00.\hat{\beta}}^{-1} S_{01.\hat{\beta}}\tilde{\boldsymbol{\beta}}_{\perp} \stackrel{w}{\to} \frac{1}{2}\int_0^1 \mathbf{F}(d\mathbf{W}_1')M\int_0^1 (d\mathbf{W}_1)\mathbf{F}'\frac{1}{2},$$

where M is given by

$$M = \boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\begin{pmatrix} I_n \\ 0 \end{pmatrix}\Omega^{-1}\begin{pmatrix} I_n \\ 0 \end{pmatrix}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp},$$

such that

$$M^c = \frac{1}{2}\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp}$$
$$= \frac{1}{2}\boldsymbol{\alpha}_{\perp}(\boldsymbol{\alpha}'_{\perp}\boldsymbol{\Omega}\boldsymbol{\alpha}_{\perp})^{-1}\boldsymbol{\alpha}'_{\perp},$$

since

$$\left[\begin{pmatrix} I_n \\ 0 \end{pmatrix}\Omega^{-1}\begin{pmatrix} I_n \\ 0 \end{pmatrix}\right]^c = \frac{1}{2}\boldsymbol{\Omega}^{-1}.$$

The asymptotic distribution is then given by

$$-2\log Q(H(r)|H(n))$$
$$\stackrel{w}{\to} tr\{\left[\int_0^1 \mathbf{F}\mathbf{F}'du\right]^{-1}\int_{\mathbf{0}}^{\mathbf{1}} \mathbf{F}(d\mathbf{W}_{\mathbf{1}}')M\int_0^1 (d\mathbf{W}_1)\mathbf{F}'\}$$
$$= tr\{\left[\int_0^1 \mathbf{F}\mathbf{F}'\mathbf{du}\right]^{-1}\int_0^1 \mathbf{F}(d\mathbf{W}_{\mathbf{1}}')M^{\mathbf{c}}\int_{\mathbf{0}}^{\mathbf{1}} (d\mathbf{W}_{\mathbf{1}})\mathbf{F}'\},$$

since both the matrices $\int_0^1 \mathbf{F}\mathbf{F}'du$ and $\int_0^1 \mathbf{F}(d\mathbf{W}_1')$ have complex structure. Combining the results we find that

$$-2\log Q(H(r)|H(n)) \xrightarrow{w} \frac{1}{2}tr\{\int_0^1 (d\mathbf{B})\mathbf{B}') \left[\int_0^1 \mathbf{B}\mathbf{B}'du\right]^{-1} \int_0^1 \mathbf{B}(d\mathbf{B}')\},$$

where

$$\mathbf{B} = (\boldsymbol{\alpha}_\perp' \boldsymbol{\Omega} \boldsymbol{\alpha}_\perp)^{-\frac{1}{2}} \boldsymbol{\alpha}_\perp' \mathbf{W}_1.$$

■

By choosing to express the result in terms of the complex Brownian motion we find that, apart from the factor $\frac{1}{2}$, the result looks like the result for the real case, see Johansen (1996), for $z = 1$, and Lee (1992) for the case $z = -1$. The result given in (37) corresponds to formula (3.35) in Lee (1992). The calculations, in Lee (1992), of the likelihood ratio statistics (3.34), however, are not correct and there is an error in the proof giving the asymptotic properties. The choice of δ_q cannot be made as stated just below (A.42). The resulting formula for the limit distribution is, however, correct.

Finally consider the test for cointegration rank at complex frequency when there are deterministic terms in the model.

Theorem 13 *In model (11) the asymptotic distribution of the test statistic for the hypothesis of $r < n$ cointegrating relations at complex seasonal frequency is asymptotically distributed as*

$$\frac{1}{2}tr\{\int_0^1 (d\mathbf{B})\mathbf{H}'(\int_0^1 \mathbf{H}\mathbf{H}'du)^{-1} \int_0^1 \mathbf{H}(d\mathbf{B}')\},$$

where $\mathbf{B}$ is standard complex Brownian motion of dimension $2(n-r)$ and $\mathbf{H} = (\mathbf{B}', I_2)'$. The limit distribution is tabulated in Table 2.

Note how the properties of the extended process $(X_t^{(m)\prime}, \bar{z}_m^t)'$ are reflected in the extended Brownian motion $\mathbf{H}$.

Finally in model (12), which allows for a linear trend in the process, we find the same result but with the definition of $\mathbf{H}$ changed.

Theorem 14 *In model (12) the asymptotic distribution of the test statistic for the hypothesis of $r < n$ cointegrating relations at complex seasonal frequency is asymptotically distributed as*

$$\frac{1}{2}tr\{\int_0^1 (d\mathbf{B})\mathbf{H}'(\int_0^1 \mathbf{H}\mathbf{H}'du)^{-1} \int_0^1 \mathbf{H}(d\mathbf{B}')\},$$

where $\mathbf{B}$ *is standard complex Brownian motion of dimension* $2(n-r)$ *and* $\mathbf{H} = (\mathbf{B}' - \bar{\mathbf{B}}', I_2)'$.

Again the process $\mathbf{H}$ reflects the properties of the extended process, but this time $X_t^{(m)}$ is corrected for its average corresponding to fitting an unrestricted constant in the equations. Note that the average of z_m^t (for $z_m \neq 1$) tends to zero so that $B - \bar{B}$ is extended by I_2 as before. The limit distribution is tabulated by simulation in Table 3.

The proofs of Theorems 13 and 14 are similar to the proof of Theorem 12, and are not given here.

Appendix A

A 1. Complex matrices and real matrices with complex structure

Complex number $z = a + ib$ can be represented by the matrix

$$\mathbf{z} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

in the sense that this representation preserves linear operations and also complex multiplication, that is, if

$$(a + ib)(c + id) = e + if,$$

then

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} c & -d \\ d & c \end{pmatrix} = \begin{pmatrix} e & -f \\ f & e \end{pmatrix}.$$

We represent a complex $n \times q$ matrix $F = A + iB$ by the real $2n \times 2q$ matrix $\mathbf{F}$

$$\mathbf{F} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}.$$

We say that F is complex, but that $\mathbf{F}$ has complex structure. Throughout we use boldface to denote real matrices with this complex structure. Note that if $F^* = A' - iB'$ then F^* has representation

$$\mathbf{F}' = \begin{pmatrix} A' & B' \\ -B' & A' \end{pmatrix} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}'.$$

We consider the transformation of a $2n \times 2q$ matrix to a matrix of complex structure given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \rightarrow \frac{1}{2}\left[\begin{pmatrix} A & B \\ C & D \end{pmatrix} + \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} 0 & I_q \\ -I_q & 0 \end{pmatrix}\right]$$
$$= \frac{1}{2}\begin{pmatrix} A+D & B-C \\ C-B & A+D \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}^c.$$

We introduce the transformation

$$\mathcal{I}_{2n} = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix},$$

with the property that $\mathcal{I}_{2n}^2 = -I_{2n}$ and $\mathcal{I}_{2n}{}' = -\mathcal{I}_{2n}$. For a $2n \times 2q$ matrix M we then have

$$M^c = \frac{1}{2}\left(M + \mathcal{I}_{2n} M \mathcal{I}_{2q}{}'\right),$$

and if $\mathbf{M}$ has complex structure then $\mathcal{I}_{2n}\mathbf{M}\mathcal{I}_{2q}{}' = \mathbf{M}$, such that $\mathbf{M} = \mathbf{M}^c$. If M is any $2n \times 2q$ matrix and $\mathbf{h}$ $(2q \times 2n)$ has complex structure then

$$\begin{aligned}(M\mathbf{h})^c &= \frac{1}{2}\left(M\mathbf{h} + \mathcal{I}_{2n} M\mathbf{h}\mathcal{I}_{2n}'\right) = \frac{1}{2}(M\mathbf{h} + \mathcal{I}_{2n} M\mathcal{I}_{2q}'\mathcal{I}_{2q}\mathbf{h}\mathcal{I}_{2n}') \\ &= \frac{1}{2}(M + \mathcal{I}_{2n} M\mathcal{I}_{2q}')\mathbf{h} = M^{\mathbf{c}}\mathbf{h}.\end{aligned}$$

Finally, notice that if $tr\{M\mathbf{h}\} = 0$ for all $\mathbf{h}$ with complex structure, then $M^c = 0$, since

$$tr\{M\mathbf{h}\} = tr\{(M\mathbf{h})^c\} = tr\{M^c\mathbf{h}\} = 0,$$

for all $\mathbf{h}$ with complex structure implies that $M^c = 0$.

A 2. Asymptotics

This appendix contains brief proofs of some of the technical results stated in Section 4.

*Proof of Theorem 6***.** The first result (20) follows by the continuous mapping theorem and the second (21) by noting that $\Delta\bar{S}_t^{(n)} = \bar{z}_n^t \varepsilon_t$.

The third result (22) follows by a partial summation. Let $|A|^2 = tr\{A^*A\}$ for a complex matrix, and let $c = \sup_t |F(t)|$. Then

$$\begin{aligned}&T^{-2}\textstyle\sum_{t=1}^T S_t^{(m)} f(t) S_t^{(n)*} \\ &= T^{-2}\textstyle\sum_{t=1}^T S_t^{(m)}(F(t) - F(t-1))S_t^{(n)*} \\ &= T^{-2}\textstyle\sum_{t=1}^T S_t^{(m)} F(t) S_t^{(n)*} - T^{-2}\sum_{t=1}^T (S_{t-1}^{(m)} + \Delta S_t^{(m)})F(t-1)(S_{t-1}^{(n)} + \Delta S_t^{(n)})^* \\ &= T^{-2} S_T^{(m)} F(T) S_T^{(n)*} - T^{-2}\textstyle\sum_{t=1}^T S_{t-1}^{(m)} F(t-1)\Delta S_t^{(n)*} \\ &-T^{-2}\textstyle\sum_{t=1}^T \Delta S_t^{(m)} F(t-1) S_{t-1}^{(n)*} - T^{-2}\sum_{t=1}^T \Delta S_t^{(m)} F(t-1)\Delta S_t^{(n)*}.\end{aligned} \quad (40)$$

The first term is written as

$$(T^{-\frac{1}{2}}S_T^{(m)})T^{-1}F(T)(T^{-\frac{1}{2}}S_T^{(n)*}) \xrightarrow{P} 0,$$

since F is bounded and $T^{-\frac{1}{2}}S_T^{(m)}$ converges weakly. The second and third terms are evaluated as follows:

$$\begin{aligned} &E|T^{-2}\textstyle\sum_{t=1}^T S_{t-1}^{(m)}F(t-1)\Delta S_t^{(n)*}| \\ &\leq cT^{-2}\textstyle\sum_{t=1}^T E|S_{t-1}^{(m)}|E|\Delta S_t^{(n)*}| \\ &\leq c_1T^{-2}\textstyle\sum_{t=1}^T t^{\frac{1}{2}} \in O(T^{-\frac{1}{2}}). \end{aligned}$$

Thus, the second and third term tend to zero, and the last term is evaluated as

$$\begin{aligned} &T^{-2}E|\textstyle\sum_{t=1}^T \Delta S_t^{(m)}F(t-1)\Delta S_t^{(n)*}| \\ &\leq cT^{-2}\textstyle\sum_{t=1}^T E|\Delta S_t^{(m)}|E|\Delta S_t^{(n)*}| \in O(T^{-1}). \end{aligned}$$

■

Proof of Corollary 7. The relation (23) follows from (19). The relation

$$T^{-2}\sum_{t=1}^T X_t^{(m)}X_t^{(n)*} = T^{-2}\sum_{t=1}^T \bar{z}_m^t z_n^t C_m S_{t-1}^{(m)}S_{t-1}^{(n)*}C_n^* + o_P(1),$$

shows that the asymptotic behavior of the product moments depends on the boundedness of

$$\sum_{t=0}^T \bar{z}_m^t z_n^t = \frac{1-(\bar{z}_m z_n)^{T+1}}{1-\bar{z}_m z_n},$$

which remains bounded if $z_m \neq z_n$. Thus, for $z_m \neq z_n$ the product moment will converge to zero, whereas for $z_m = z_n$ we get the limit stated, which proves (24) and (25). The result (26) follows from (21). ■

Thus, the reason that the mixed moments tend to zero is not that $X_t^{(m)}$ and $X_t^{(n)}$ are asymptotically independent (which they are) but the factor $\bar{z}_m^t z_n^t$ which appears in the summation. The factor $\bar{z}_m^t$ comes from the representation of $X_t^{(m)}$ and also implies that the limit of $T^{-1}\sum_{t=1}^T X_t^{(m)}\varepsilon_t'$ does not involve the limit of $T^{-\frac{1}{2}}\sum_{t=1}^T \varepsilon_t$, but rather the limit of $T^{-\frac{1}{2}}\sum_{t=1}^T \bar{z}_m^t\varepsilon_t$.

Proof of Lemma 8. From the model equations

$$p(L)X_t = \breve{\alpha}\boldsymbol{\beta}'\begin{pmatrix} X_{Rt}^{(1)} \\ X_{It}^{(1)} \end{pmatrix} + \sum_{m=3}^s \alpha_m\beta_m^* X_t^{(m)} + \sum_{j=1}^k \Gamma_j p(L)X_{t-j} + \varepsilon_t,$$

it follows by taking conditional variances and covariances given the lagged values of $p(L)X_t$ and the remaining linear combinations $\beta_m^* X_t^{(m)}$, that (28) and (29) hold.

In order to prove (30) we write it as

$$\Sigma_{\beta 0} = \Sigma_{\beta\beta.0}\breve{\alpha}'\Omega^{-1}\Sigma_{00},$$

and introduce the normalized vector

$$u = \Omega^{-\frac{1}{2}}\breve{\alpha}\Sigma_{\beta\beta}^{\frac{1}{2}}.$$

After some reductions, applying $\Sigma_{\beta 0} = \breve{\alpha}\Sigma_{\beta\beta}$, the relation (30) reduces to

$$u' = (I_r - u'(I_n + uu')^{-1}u)u'(I_n + uu'),$$

which follows from the identity

$$u'(I_n + uu')^{-1}u = (u'u)\,(I_r + u'u)^{-1}.$$

Next we multiply in (31) by $\Sigma_{\beta\beta.0}$ and $\Sigma_{\beta\beta}$ and find

$$\Sigma_{\beta\beta.0} - \Sigma_{\beta\beta} + \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} = 0,$$

which is zero by the definition of $\Sigma_{\beta\beta.0}$. ■

Proof of Theorem 9. We first give a result for product moments of $X_t^{(m)}$ and $X_t^{(n)}$ corresponding to complex roots z_m and z_n :

$$T^{-1}M_{11}^{(n,m)} = T^{-2}\sum_{t=1}^{T}\begin{pmatrix} X_{Rt}^{(m)} \\ X_{It}^{(m)} \end{pmatrix}\begin{pmatrix} X_{Rt}^{(n)} \\ X_{It}^{(n)} \end{pmatrix}'$$

$$= T^{-2}\sum_{t=1}^{T}\mathbf{C}_m\mathbf{S}_{t-1}^{m}\begin{pmatrix} \cos(t\theta_m)\cos(t\theta_n) & -\cos(t\theta_m)\sin(t\theta_n) \\ -\sin(t\theta_m)\cos(t\theta_n) & \sin(t\theta_m)\sin(t\theta_n) \end{pmatrix}\mathbf{S}_{t-1}^{n\prime}\mathbf{C}_n'+o_P(1).$$

The matrix in the middle is

$$\frac{1}{2}\begin{pmatrix} \cos((\theta_m-\theta_n)t)+\cos((\theta_m+\theta_n)t) & \sin((\theta_m-\theta_n)t)-\sin((\theta_m+\theta_n)t) \\ \sin((\theta_n-\theta_m)t)-\sin((\theta_m+\theta_n)t) & \cos((\theta_m-\theta_n)t)-\cos((\theta_m+\theta_n)t) \end{pmatrix},$$

which remain bounded when summed unless $\theta_m = \theta_n$, in which case the matrix equals

$$\frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \cos(2t\theta_m) & -\sin(2t\theta_m) \\ -\sin(2t\theta_m) & -\cos(2t\theta_m) \end{pmatrix},$$

where the last term is bounded when summed.

Hence $T^{-1}M_{11}^{(n,m)}$ converges to zero if $n \neq m$ and for $n = m$ has the same limit as

$$\frac{1}{2}T^{-2}\sum_{t=1}^{T}\mathbf{C}_m\mathbf{S}_{t-1}^{(m)}\mathbf{S}_{t-1}^{(m)\prime}\mathbf{C}_m' \xrightarrow{w} \frac{1}{4}\mathbf{C}_m\int_0^1\mathbf{W}_m\mathbf{W}_m'du\mathbf{C}_m'.$$

Similarly we find that for z_m complex:

$$\begin{aligned} M_{1\varepsilon}^{(m)} &= T^{-1}\sum_{t=1}^{T}\begin{pmatrix} X_{Rt}^{(m)} \\ X_{It}^{(m)} \end{pmatrix}\varepsilon_t' \\ &= \mathbf{C}_m(T^{-1}\sum_{t=1}^{T}\mathbf{S}_{t-1}^{(m)}\begin{pmatrix} \cos(\theta_m t) \\ -\sin(\theta_m t) \end{pmatrix}\varepsilon_t') + o_P(1) \\ &= \mathbf{C}_m T^{-1}\sum_{t=1}^{T}\mathbf{S}_{t-1}^{(m)}(\Delta\mathbf{S}_t^{(m)})'\begin{pmatrix} I_n \\ 0 \end{pmatrix} + o_P(1) \\ &\overset{w}{\to} \tfrac{1}{2}\mathbf{C}_m\int_0^1 \mathbf{W}_m(d\mathbf{W}_m)'\begin{pmatrix} I_n \\ 0 \end{pmatrix}. \end{aligned}$$

If either z_m or z_n are real, similar results can be proved. Finally, we want the results for the product moment matrices constructed from the residuals U_t. It is clear that the limit of the product moments of $X_{Rt}^{(m)}$ and $X_{It}^{(m)}$ are not influenced by the preliminary regression on the lagged values of $p(L)X_t$, since these are stationary. The matrix S_{11} is $M_{11}^{(1,1)}$ corrected for the other processes. Since the mixed moments $T^{-1}M_{11}^{(n,m)}$converge to zero, the limit of $T^{-1}S_{11}$ is the same as that of $T^{-1}M_{11}^{(1,1)}$. Similarly the limit of $M_{1\varepsilon}^{(1)}$ is the same as that of $S_{1\varepsilon}$. ■

Appendix B

Tables

In this Appendix the asymptotic distributions of the likelihood ratio test statistics for cointegrating rank at complex frequency are tabulated. The limit distributions all have the form

$$\frac{1}{2}tr\left\{\int_0^1 (d\mathbf{B})\mathbf{H}'\left[\int_0^1 \mathbf{H}\mathbf{H}'du\right]^{-1}\int_0^1 \mathbf{H}(d\mathbf{B})'\right\}, \tag{41}$$

where $\mathbf{B}$ is a $2(n-r)-$dimensional complex Brownian motion, and $\mathbf{H}$ is some process derived from $\mathbf{B}$ depending on the model for the deterministic terms.

The Brownian motion $\mathbf{B}$ is approximated by a 400 - step random walk and the statistic is calculated $100,000$ times.

The approximation formulae used are as follows. Let $B = (B_R', B_I')'$ denote a $2(n-r)$-dimensional Brownian motion, and let $(\varepsilon_t)_{t\geq 0}$ be a sequence of $2(n-r)$-dimensional i.i.d. $N_{2(n-r)}(0, I_{2(n-r)})$ variables, then

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T}\left(\sum_{k=1}^{t}\varepsilon_k\right)\overset{w}{\to}\int_0^1 Bdu,$$

$$\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{k=1}^{t-1}\varepsilon_k\right)\varepsilon_t' \xrightarrow{w} \int_0^1 B\,(dB)',$$
$$\frac{1}{T^2}\sum_{t=1}^{t}\left(\sum_{k=1}^{t}\varepsilon_k\right)\left(\sum_{k=1}^{t}\varepsilon_k\right)' \xrightarrow{w} \int_0^1 BB'du,$$

and

$$\mathbf{B} = \begin{pmatrix} B_R & -B_I \\ B_I & B_R \end{pmatrix}.$$

The extension of the process $X_t^{(m)}$ with $\bar{z}_m^t$ gives rise to a complex Brownian motion B extended by I_2

$$\begin{pmatrix} \mathbf{B} \\ I_2 \end{pmatrix} = \begin{pmatrix} B_R & -B_I \\ B_I & B_R \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If the model has an unrestricted constant the processes $X_{et} = (X_t^{(m)\prime}, \bar{z}_m^t)'$ is corrected for a constant. Note that $\frac{1}{T}\sum_{t=1}^{T}\cos(\theta_m t) \to 0$, such that limit becomes

$$\begin{pmatrix} \mathbf{B} - \bar{\mathbf{B}} \\ I_2 \end{pmatrix} = \begin{pmatrix} B_R - \bar{B}_R & -B_I + \bar{B}_I \\ B_I - \bar{B}_I & B_R - \bar{B}_R \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

References

Boswijk, H. P., 1995, Identifiability of cointegrated systems, Tinbergen Institute, Discussion paper, TI, 95-78.

Chan, N. H. and C. Z. Wei, 1988, Limiting distributions of least squares estimators of unstable autoregressive processes, The Annals of Statistics 16, 367-401.

Dahl Pedersen, I., 1996, A practical implementation of the seasonal cointegration theory, Masters Thesis, University of Copenhagen, 93 pp.

Engle R. F., Granger, C. W. J., Hylleberg, S., and H. S. Lee, 1993, Seasonal cointegration: The Japanese consumption function, Journal of Econometrics 55, 275-298.

Engle, R. F. and S. Yoo, 1991, Cointegrated economic times series: An overview with new results, in R. F. Engle and C. W. J. Granger, eds., Long Run Economic Relationships, Readings in Cointegration (Oxford University Press, Oxford), 237-266.

Franses, P. H. and R. M. Kunst, 1995, On the role of seasonal intercepts in seasonal cointegration, IHS working paper, Reihe Ökonomie no. 15.

Gregoir, S., 1993*b*, Multivariate time series with various hidden unit roots: Part I. Integral operator algebra and representation theorem, Discussion paper.

Gregoir, S., 1993*a*, Multivariate time series with various hidden unit roots: Part II: Estimation and Test, Discussion paper.

Harbo, I., Johansen S., Nielsen, B., and A. Rahbek, 1998, Asymptotic inference on cointegrating rank in partial systems, Journal of Business and Economic Statistics, forthcoming.

Hylleberg, S., Engle R. F., Granger, C. W. J., and S. B., Yoo, 1990, Seasonal integration and cointegration, Journal of Econometrics 44, 215-238.

Johansen, S., 1996, Likelihood-based Inference in Cointegrated Vector Autoregressive Models, 2.ed. (Oxford University Press, Oxford).

Johansen, S., 1997, Likelihood analysis of the $I(2)$ model, Scandinavian Journal of Statistics 24, 433-462.

Lee, H. S., 1992, Maximum likelihood inference on cointegration and seasonal cointegration, Journal of Econometrics 54, 1-47.

Schaumburg, E., 1996, Inference in seasonally integrated VAR models, Masters thesis, University of Copenhagen, 98 pp.

Table 1: Quantiles of the limit distribution (42) ($\mathbf{H} = \mathbf{B}$) for cointegrating rank at seasonal frequency for model (10) with no determinstic terms. The number of simulations is 100,000 and the random walk has 400 steps.

p-r	0.01	0.05	0.10	0.50	0.75	0.80	0.85	0.90	0.95	0.975	0.99
1	0.0228	0.114	0.234	1.50	2.95	3.41	3.99	4.80	6.20	7.57	9.45
2	4.21	5.74	6.73	11.4	14.6	15.5	16.6	18.1	20.4	22.6	25.3
3	16.3	19.4	21.3	29.2	34.1	35.5	37.0	39.1	42.3	45.3	48.9
4	36.3	41.1	43.8	54.8	61.5	63.2	65.3	67.9	72.0	75.7	80.3
5	64.1	70.5	74.2	88.3	96.6	98.7	101	105	110	114	119
6	99.6	108	112	129	139	142	145	149	155	160	166
7	143	153	158	178	190	193	196	201	207	213	220
8	194	205	211	235	248	251	255	260	268	274	282
9	252	265	272	299	313	317	322	327	336	343	352
10	318	333	341	370	387	391	396	402	411	419	429
11	391	408	417	449	467	472	477	484	494	503	513
12	472	490	500	535	555	560	566	573	584	594	605

Table 2: Quantiles of the limit distribution (42) ($\mathbf{H} = (\mathbf{B}', I_2)'$) for cointegrating rank at seasonal frequency for model (11) with restricted seasonal dummies and constant. The number of simulations is 100,000 and the random walk has 400 steps.

p-r	0.01	0.05	0.1	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	2.22	3.21	3.90	7.39	10.0	10.8	11.7	12.9	14.9	16.8	19.3
2	12.3	14.9	16.5	23.2	27.6	28.8	30.2	32.0	34.9	37.6	40.9
3	30.2	34.4	37.0	46.9	53.0	54.6	56.5	59.0	62.9	66.3	70.3
4	56.0	61.9	65.3	78.4	86.0	88.0	90.4	93.5	98.2	102	107
5	89.6	97.2	101	118	127	129	132	136	141	146	152
6	131	140	145	164	175	178	182	186	192	198	205
7	180	191	197	219	231	235	239	243	250	257	265
8	236	249	255	281	295	299	303	308	317	324	332
9	300	314	322	350	366	370	375	381	390	398	407
10	372	387	396	428	445	449	455	461	471	480	490
11	451	468	477	512	531	536	542	549	560	570	580
12	537	556	566	604	624	630	636	643	655	666	678

Table 3: Quantiles of the limit distribution (42) ($\mathbf{H} = (\mathbf{B}' - \bar{\mathbf{B}}', I_2)'$) for cointegrating rank at seasonal frequency for model (12) with restricted seasonal dummies and unrestricted constant. The number of simulations is 100,000 and the random walk has 400 steps.

p-r	0.01	0.05	0.1	0.5	0.75	0.8	0.85	0.9	0.95	0.975	0.99
1	2.06	3.77	5.10	13.9	23.7	27.0	31.4	37.5	48.4	59.4	74.6
2	15.0	19.9	23.3	43.7	62.7	68.4	75.8	86.3	104	123	146
3	39.1	48.1	54.3	88.8	118	127	138	153	179	204	237
4	74.4	88.9	98.6	149	190	201	216	237	270	304	348
5	122	143	156	225	278	293	312	338	381	421	475
6	181	209	227	316	383	401	425	457	508	556	622
7	254	289	312	422	503	526	554	591	651	712	790
8	338	383	410	544	639	666	699	743	813	881	969
9	435	489	524	680	792	822	860	910	992	1070	1170
10	546	609	648	832	960	995	1040	1100	1190	1280	1390
11	668	742	788	997	1140	1180	1230	1300	1400	1500	1630
12	804	888	941	1180	1340	1390	1440	1510	1630	1740	1880