Distribution approximations for cointegration tests with stationary exogenous regressors
Boswijk, H.P.; Doornik, J.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)

Download date: 05 Dec 2018
Distribution Approximations for Cointegration Tests with Stationary Exogenous Regressors

H. Peter Boswijk
Department of Quantitative Economics, Universiteit van Amsterdam*

Jurgen A. Doornik
Nuffield College, University of Oxford

10th February 1999

Abstract

The distribution of a functional of two correlated vector Brownian motions is approximated by a Gamma distribution. This functional represents the limiting distribution for cointegration tests with stationary exogenous regressors, but also for cointegration tests based on a non-Gaussian likelihood. The approximation is accurate, fast, and easy to use in comparison to both tabulated critical values and simulated $p$-values.

1 Introduction

Consider the vector error correction model (VECM)

$$\Delta X_t = \Pi^* X_{t-1}^* + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \Phi q_t + \varepsilon_t, \quad t = 1, \ldots, T, \quad (1)$$

where \( \{X_t\} \) is a \( p \)-vector time series, the starting values \( X_{1-k}, \ldots, X_0 \) are fixed, \( \{\varepsilon_t\} \) is i.i.d. \( N(0, \Omega) \), and \( X_{t-1}^* = (X'_{t-1}, d_t)' \), where \( d_t \) and \( q_t \) are deterministic regressors. The three deterministic specifications that are most commonly used, are, in the notation of Doornik et al. (1998) (also see Johansen, 1995, Section 5.7):

- \( H_z \) : both \( d_t \) and \( q_t \) are void; no deterministics;
- \( H_c \) : \( d_t = 1, q_t \) is void; restricted constant;
- \( H_l \) : \( d_t = t, q_t = 1 \); restricted linear trend.

*Correspondence to: H. Peter Boswijk, Department of Quantitative Economics, Universiteit van Amsterdam, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands. E-mail: peterb@fee.uva.nl.
When \( \text{rank} \Pi^* = r \leq p \) and some further restrictions on \((\Pi^*, \{\Gamma_j\})\) are satisfied, the model implies that \( X_t \) is cointegrated. The likelihood ratio (LR) statistic for the hypothesis \( H(r) : \text{rank} \Pi^* \leq r \) has been derived by Johansen (1988, 1995), and its limiting distribution (when \( \text{rank} \Pi^* = r \)) is characterized by the functional

\[
T = \text{tr} \left\{ \int_0^1 dB F' \left( \int_0^1 F F' du \right)^{-1} \int_0^1 F dB' \right\},
\]

where \( B(u) \) is a standard \((p-r)\)-vector Brownian motion, and where either \( F(u) = B(u) \) (model \( H_z \)), or \( F(u) = \{ B(u)', 1 \}' \) (model \( H_\omega \)), or \( F(u) = \{ |B(u) - \int_0^1 B du|', |u - \frac{1}{2}| \}' \) (model \( H_t \)).

Recently, Seo (1998) and Rahbek and Mosconi (1998) considered an extension of (1), where some exogenous stationary vector process \( Z_t \) is added to the regressors (together with some of its lags):

\[
\Delta X_t = \Pi^* X^*_t + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \Phi_\ell + \sum_{j=0}^m D_j Z_{t-j} + \varepsilon_t, \quad t = 1, \ldots, T. \tag{3}
\]

The limiting distribution under the null hypothesis of the LR statistic for \( H(r) \) in this case turns out to be characterized by

\[
Q = \text{tr} \left\{ \int_0^1 dWF' \left( \int_0^1 FF' du \right)^{-1} \int_0^1 F dW' \right\},
\]

where \( F(u) \) is the same as before, and \( W(u) \) is a standard \((p-r)\)-vector Brownian motion, with \( E[W(1)B(1)'] = P = \text{diag}(\rho_1, \ldots, \rho_{p-r}) \), where \( \rho_i \in [0, 1] \) are correlation coefficients. Note that \( Q = T \) when \( P = I_{p-r} \).

In order to save notation, but without loss of generality, we shall only consider the case \( r = 0 \) henceforth, so that \( W \) and \( B \) are of dimension \( p \).

When \( p = 1 \) (and \( P = \rho \)), the random variable \( Q \) in model \( H_z \) is the square of

\[
R = \left( \int_0^1 B^2 du \right)^{-1/2} \int_0^1 B dW.
\]

Its distribution was obtained by Kremers et al. (1992) as the limiting distribution of a \( t \)-statistic for cointegration with known cointegrating vector. Because we may decompose \( W \) as \( \rho B + (1 - \rho^2)^{1/2} U \), with \( U \) a standard Brownian motion, independent of \( B \), it follows that

\[
R = \rho \frac{\int_0^1 B dB}{\left( \int_0^1 B^2 du \right)^{1/2}} + (1 - \rho^2)^{1/2} \frac{\int_0^1 B dU}{\left( \int_0^1 B^2 du \right)^{1/2}} = \rho X + (1 - \rho^2)^{1/2} Z, \tag{5}
\]

where \( X \) corresponds to the limiting distribution of the Dickey-Fuller \( t \)-statistic, and \( Z \) is a standard normal random variable, independent of \( B \) and hence \( X \). Kremers et al. (1992) suggested to use critical values from the standard normal distribution by a small-\( \sigma \) (in this case small-\( \rho \)) asymptotic argument.

The same distribution of \( R \) was also obtained by Hansen (1995), in the context of testing for a unit root with stationary exogenous regressors. Hansen tabulated the distribution of \( R \) for \( \rho^2 \in \{0.1, 0.2, \ldots, 1\} \), and suggested to interpolate these critical values for other values of \( \rho^2 \). This approach was extended
by Seo (1998) to the multivariate case, leading to \( Q \). With \( p > 1 \), however, one has to construct tables for different values of \((\rho_1, \ldots, \rho_p)\) which is rather impractical. Seo provided tables for \( p \leq 5 \) and \((\rho_1, \ldots, \rho_p) \in \mathcal{P} \times \cdots \times \mathcal{P} \), where \( \mathcal{P} = \{0, 0.2, 0.4, 0.6, 0.8, 1\} \), which already resulted in 20 pages of tables. Requiring from practitioners to obtain appropriate critical values by interpolation in several dimensions may be too much to ask, and definitely does not add to the ease-of-use of the proposed test. Therefore, we suggest in the next section an alternative approach to obtaining critical values or (preferably) \( p \)-values, based on Doornik’s (1998a) approach to approximate the distribution of \( T \) by a Gamma distribution with the same mean and variance as \( T \), also see Johansen (1988) and Nielsen (1997). This is closely related to Abadir and Lucas’ (1996) approximation of the distribution of \( R \) by a normal distribution with non-zero mean and non-unit variance.

The distribution of \( R \) and \( Q \) also arises when \( H(r) \) is tested in model (1) with non-Gaussian \( \{\varepsilon_t\} \) and corresponding non-Gaussian likelihood function. See, e.g., Lucas (1997), who considers cointegration testing based on a Student-\( t \) likelihood function, and Boswijk and Lucas (1997), who use a semi-nonparametric likelihood function. Furthermore, the distribution of \( R \) also emerges when \( \{\varepsilon_t\} \) is assumed to follow a GARCH process, and an LR test for a unit root is based on the corresponding likelihood function, see Ling and Li (1997, 1998).

The plan of the rest of this paper is as follows. In Section 2, we show that the mean and variance of \( Q \) can be expressed as a function of the mean and variance of \( T \), a covariance parameter, and the correlations \((\rho_1, \ldots, \rho_p)\). It is then suggested to use a Gamma distribution with the same mean and variance as an approximation to the true distribution of \( Q \). In Section 3, this approximation is shown to be very accurate, at least for quantiles and \( p \)-values where accuracy is required (in the right-hand tail of the distribution). Section 4 applies the result to the purchasing-power parity model of Johansen and Juselius (1992). Section 5 concludes.

## 2 Mean and Variance of \( Q \)

The Gamma distribution function \( \Gamma(x; b, a) \) is defined here as:

\[
\Gamma(x; b, a) = \int_0^x \frac{t^{b-1} e^{-at}}{\Gamma(b)} dt, \quad x > 0, b > 0, a > 0,
\]

with \( \Gamma(b) = \int_0^\infty t^{b-1} e^{-t} dt \), the Gamma function. A random variable \( X \) with this distribution has mean \( E(X) = b/a \) and variance \( \text{var}(X) = b/a^2 \).

Doornik (1998a) shows that the Gamma distribution with \( b = E(T)^2 / \text{var}(T) \) and \( a = E(T) / \text{var}(T) \) provides an accurate approximation of the distribution of \( T \). The mean and variance of \( T \) could in principle be simulated for the three different deterministic models and many values of \( p \). However, Doornik shows, using estimated response surfaces based on Monte Carlo simulation, that the following approximations are sufficiently accurate:

\[
E(T) \approx \begin{cases} 
2p^2 - p + 0.07 + 0.07 \cdot 1_{\{p=1\}} & \text{for } H_z, \\
2p^2 + 2.01p + 0.06 \cdot 1_{\{p=1\}} + 0.05 \cdot 1_{\{p=2\}} & \text{for } H_c, \\
2p^2 + 4.05p + 0.5 - 0.23 \cdot 1_{\{p=1\}} - 0.07 \cdot 1_{\{p=2\}} & \text{for } H_I,
\end{cases}
\]
and
\[ \text{var}(T) \approx \begin{cases} 
3p^2 - 0.33p - 0.55 & \text{for } H_z, \\
3p^2 + 3.60p + 0.75 - 0.40 \cdot 1_{\{p-1\}} - 0.30 \cdot 1_{\{p-2\}} & \text{for } H_c, \\
3p^2 + 5.70p + 3.20 - 1.30 \cdot 1_{\{p-1\}} - 0.50 \cdot 1_{\{p-2\}} & \text{for } H_l.
\end{cases} \] (8)

Doornik also analyzes
\[ T_i = \int_0^1 dB_i F (\int_0^1 FF' du)^{-1} \int_0^1 F dB_i, \]
where \( B_i \) is the \( i \)th component of \( B \), so that \( T = \sum_{i=1}^p T_i \). Since \( T_i \) and \( T_j \) have the same distribution, we have \( E(T_i) = E(T)/p \). Furthermore, he finds (by simulation) that for \( i \neq j \),
\[ \text{cov}(T_i, T_j) \approx \begin{cases} 
-1.270 & \text{for } H_z, \\
-1.066 & \text{for } H_c, \\
-1.35 & \text{for } H_l.
\end{cases} \] (9)

This can be used to evaluate \( \text{var}(T_i) = \text{var}(T)/p - (p-1)\text{cov}(T_i, T_j) \).

Here we adopt a similar approach for \( Q \). Theorem 1 provides an expression for the mean and variance of \( Q \) in terms of \( E(T), \text{var}(T), \text{cov}(T_i, T_j) \) and \((\rho_1, \ldots, \rho_p)\). Subsequent substitution of the approximations (7)–(9) provides the first two moments of \( Q \). This, in turn, may be used to obtain a Gamma approximation of its distribution.

**Theorem 1** Let \( q = \dim(F) \). Then
\[ E(Q) = \frac{\sum_{i=1}^p \rho_i^2}{p} E(T) + \left( 1 - \frac{\sum_{i=1}^p \rho_i^2}{p} \right) pq, \] (10)

and
\[ \text{var}(Q) = \sum_{i=1}^p \rho_i^4 \text{var}(T_i) + 2 \sum_{i=2}^p \sum_{j=1}^{i-1} \rho_i^2 \rho_j^2 \text{cov}(T_i, T_j) + 4 \sum_{i=1}^p \rho_i^2 (1 - \rho_i^2) E(T) + 2q \sum_{i=1}^p (1 - \rho_i^2)^2. \] (11)

**Proof.** Decompose \( W \) as \( PB + RU \), where \( R = \text{diag} \{(1 - \rho_1^2)^{1/2}, \ldots, (1 - \rho_p^2)^{1/2}\} = (I - P^2)^{1/2} \), and where \( U \) is a standard \( p \)-vector Brownian motion, independent of \( B \). This implies that
\[ Z = \left( \int_0^1 FF' du \right)^{-1/2} \int_0^1 F dU' \sim N(0, I_{qp}), \]

independently of \( B \). Defining
\[ X = \left( \int_0^1 FF' du \right)^{-1/2} \int_0^1 F dB', \]

\(^1\)The variance entries in Table 7 of Doornik (1998a) should be labelled \((n - p), 1, n - p = 1, n - p = 2\).
it follows that (remembering that both $P$ and $Q$ are diagonal)

$$Q = \text{tr} \left( [XP + ZR]'[XP + ZR] \right)$$

$$= \text{tr} \left( PX'XP + PX'ZR + RZ'XP + RZ'ZR \right),$$

$$= \text{tr} \left( PX'XP \right) + 2\text{tr} \left( PX'ZR \right) + \text{tr} \left( RZ'ZR \right)$$

$$= \sum_{i=1}^{p} \rho_i^2 T_i + 2\text{tr} \left( PX'ZR \right) + \sum_{i=1}^{p} (1 - \rho_i^2) \xi_i,$$

(12)

where $\xi_i = \sum_{j=1}^{q} Z_{ji}^2$ are independent $\chi^2(q)$ random variables. Because $Z$ is independent of $X$, $E(2\text{tr}[PX'ZR]) = 0$. Thus we find

$$E(Q) = \sum_{i=1}^{p} \rho_i^2 E(T_i) + \sum_{i=1}^{p} (1 - \rho_i^2) q$$

$$= \frac{\sum_{i=1}^{p} \rho_i^2}{p} E(T) + \left( 1 - \frac{\sum_{i=1}^{p} \rho_i^2}{p} \right) pq.$$

To obtain the variance of $Q$, we first note that the first and third term in (12) are independent, and hence uncorrelated. Furthermore, the second term is uncorrelated with the first term, because it has mean zero conditionally on $X$. Next, the covariance between the second and third term in (12) is zero (conditionally on $X$), because elements of $Z$ are uncorrelated with squared elements of $Z$. Hence all covariances are zero, and the variance of $Q$ can be reduced to

$$\text{var}(Q) = \text{var} \left( \sum_{i=1}^{p} \rho_i^2 T_i \right) + 4\text{var} \left( \text{tr}[PX'ZR] \right) + \text{var} \left( \sum_{i=1}^{p} (1 - \rho_i^2) \xi_i \right).$$

(13)

For the first term of (13) we find

$$\text{var} \left( \sum_{i=1}^{p} \rho_i^2 T_i \right) = \sum_{i=1}^{p} \rho_i^4 \text{var}(T_i) + 2 \sum_{i=2}^{p} \sum_{j=1}^{i-1} \rho_i^2 \rho_j^2 \text{cov}(T_i, T_j).$$

To evaluate the second variance term, we use

$$\text{tr}(PX'ZR) = \text{vec}(P)'(X' \otimes R)\text{vec}(Z').$$

Hence

$$\text{var} \left( \text{tr}[PX'ZR] \right) = E \left\{ \text{var} \left[ \text{vec}(P)'(X' \otimes R)\text{vec}(Z') \right] \mid X \right\}$$

$$= E \left[ \text{vec}(P)'(X'X \otimes R^2)\text{vec}(P) \right]$$

$$= E \left[ \text{tr} \left( X'XP^2 R^2 \right) \right]$$

$$= \sum_{i=1}^{p} \rho_i^2(1 - \rho_i^2) E(T_i)$$

$$= \frac{\sum_{i=1}^{p} \rho_i^2(1 - \rho_i^2)}{p} E(T).$$

The final term in (13) follows immediately from the fact that $\xi_i \sim \text{i.i.d.} \chi^2(q)$:

$$\text{var} \left( \sum_{i=1}^{p} (1 - \rho_i^2) \xi_i \right) = 2q \sum_{i=1}^{p} (1 - \rho_i^2)^2.$$

This completes the proof.
3 Evaluation of the Approximation

The accuracy of the Gamma approximation based on the results from Theorem 1 is assessed using simulation. The experimental design is as follows. Following Doornik (1998a, Section 4), the distribution of $Q$ is simulated from $-T \sum \log(1 - \lambda_i)$. Here, $\lambda_i$ are the eigenvalues of $T^{-1}E'R(R'R)^{-1}R'E$, which is the discrete approximation to the expression inside the trace of (4). The dimensions and correlations are respectively: $p = 1, \ldots, 5$, and $(\rho_1, \ldots, \rho_p) \in \mathcal{P} \times \cdots \times \mathcal{P}$, where $\mathcal{P} = \{0, 0.2, 0.4, 0.6, 0.8, 1\}$ (the ordering of the correlations is irrelevant). The design consists of 461 specifications for each of the three treatments of deterministic terms, so 1383 experiments in total. This corresponds to the tables in Seo (1998). The number of Monte Carlo replications was chosen as $M = 10\,000$, and the sample size in the discretization as $T = 2000$.\footnote{All experiments are done using Ox 2.0, see Doornik (1998b).}

Table 1 compares the absolute relative error in the mean and standard deviation of $Q$ when using Theorem 1 together with (7)–(9) to that found in the simulations. The table reports the mean of the absolute relative errors for each dimension separately as a percentage. The proposed procedure is very accurate indeed. The discrepancy is somewhat higher at dimension one. Here, the distribution is very skewed (as discussed in Doornik, 1998a), requiring more precise estimates of the mean and variance, corresponding to the dummies for low dimensions needed in (7) and (8).

<table>
<thead>
<tr>
<th></th>
<th>$n = 1$</th>
<th>$n = 2$</th>
<th>$n = 3$</th>
<th>$n = 4$</th>
<th>$n = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.83%</td>
<td>0.35%</td>
<td>0.37%</td>
<td>0.23%</td>
<td>0.21%</td>
</tr>
<tr>
<td>std.deviation</td>
<td>1.27%</td>
<td>0.75%</td>
<td>0.75%</td>
<td>0.70%</td>
<td>0.71%</td>
</tr>
<tr>
<td>count</td>
<td>18</td>
<td>63</td>
<td>168</td>
<td>378</td>
<td>756</td>
</tr>
</tbody>
</table>

To evaluate the whole procedure, we contrast the $p$-values obtained from the Gamma approximation based on (7)–(9) and Theorem 1 with the empirical rejection frequencies. We record the absolute difference between the $p$-value from the Gamma approximation and the simulated values. Table 2 reports the percentage of experiments where this absolute difference exceeds 0.0025, 0.005, 0.01 respectively. In no experiment did the difference exceed 0.02. For example, at 0.95 (the most commonly used $p$-value), 27% of the experiments (379 out of 1383) have a difference exceeding 0.0025, and only 2% a difference greater than 0.005 (but never exceeding 0.01). If the simulated distribution were exact, this implies that only 2% from the Gamma approximation are outside the range 0.945 – 0.955. Actually, we cannot rule out that the Gamma approximation is more accurate than the simulated values, because the latter uses $T = 2000$, rather than the range of sample sizes which were used to obtain (7) and (8).
4 Application

To illustrate the proposed procedure, we use a variant of the model estimated by Johansen and Juselius (1992). They estimate a VAR(2) with unrestricted constant (model $H_{lc}$) and seasonal dummies using the UK wholesale price index ($p_1$), the trade-weighted foreign wholesale price index ($p_2$), the UK effective exchange rate ($e_{12}$), the three-month treasury bill rate in the UK ($i_1$), and the three-month Eurodollar interest rate in the UK ($i_2$). In addition, the change in oil price ($\Delta p_{oil}$) and its lag were used as conditioning variables. Johansen and Juselius (1992) found two cointegrating vectors. Hansen and Juselius (1995) use a transformed version in terms of $p_1$, $p_2$; $\Delta p_{oil}$ to avoid I(2)-ness. Following Rahbek and Mosconi (1998) we adopt model $H_l$ by allowing a trend to enter the cointegrating space. In terms of (3) our specification is a VAR(2):

$$X_t = (p_{1t} - p_{2t}, \Delta p_{1t}, e_{12t}, i_{1t}, i_{2t})', \quad X_t^* = (X_t', t)', \quad q_t = (1, S_{1t}, S_{2t}, S_{3t})', \quad Z_t = \Delta p_{oil,t},$$

with $k - 1 = m = 1$, and where $S_{it}$ are seasonal dummy variables. The effective sample size, after taking all lags into account is 1972 (4) – 1987 (2).

Table 3 lists the test values for each rank, together with $p$-values. The fourth column, labelled $1 - P(T)$, gives the asymptotic $p$-value under the assumption that the presence of $Z_t$ does not affect the distribution. The second cointegrating vector is only present if we are willing to adopt a 10% significance level. However, the sample is very small: 59 observations with 17 regressors in each equation. There

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$0.01$</th>
<th>$0.05$</th>
<th>$0.1$</th>
<th>$0.2$</th>
<th>$0.3$</th>
<th>$0.4$</th>
<th>$0.5$</th>
<th>$0.6$</th>
<th>$0.8$</th>
<th>$0.9$</th>
<th>$0.95$</th>
<th>$0.99$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0025 difference</td>
<td>$6%$</td>
<td>$30%$</td>
<td>$43%$</td>
<td>$55%$</td>
<td>$70%$</td>
<td>$73%$</td>
<td>$70%$</td>
<td>$63%$</td>
<td>$45%$</td>
<td>$27%$</td>
<td>$2%$</td>
<td></td>
</tr>
<tr>
<td>0.005 difference</td>
<td>$1%$</td>
<td>$5%$</td>
<td>$12%$</td>
<td>$27%$</td>
<td>$43%$</td>
<td>$44%$</td>
<td>$43%$</td>
<td>$29%$</td>
<td>$11%$</td>
<td>$2%$</td>
<td>$0%$</td>
<td></td>
</tr>
<tr>
<td>0.01 difference</td>
<td>$0%$</td>
<td>$1%$</td>
<td>$1%$</td>
<td>$3%$</td>
<td>$9%$</td>
<td>$11%$</td>
<td>$11%$</td>
<td>$3%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td></td>
</tr>
<tr>
<td>0.02 difference</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td>$0%$</td>
<td></td>
</tr>
</tbody>
</table>
may be a tendency to overreject the true rank in small samples. Indeed, when adopting the sample-size adjusted \( p \)-value for \( H_l(1) \) (using the formulae in Doornik, 1998a), we find that it changes from 7.7\% to 14\%. Correcting for the stationary exogenous regressors shrinks the distribution towards zero, so that the \( p \)-values will always decrease. Now rank one is firmly rejected with a \( p \)-value of 1.6\%. If we use the small sample mean and variance of \( T \) in the formulae for \( Q \) (which is somewhat ad hoc), the \( p \)-values for ranks 0–2 change to 1\%, 3\%, 11\% respectively.

The canonical correlations \( \hat{p}_i \) are computed using the kernel method of Andrews (1991), as suggested in Seo (1998). We use the quadratic spectral kernel with automatic bandwidth and an AR(1) for each component. However, we found that the standard long-run covariance matrix gave nearly identical results.

To illustrate the procedure to obtain the distribution, consider \( H_l(r = 4) \), which has only one canonical correlation. The following steps are involved:

- Use \( p - r = 1 \) for \( p \) in (7) and (8) to compute \( E(T) \) and \( \text{var}(T) \). This yields 6.32 and 10.6. The next step requires \( \text{var}(T_1) \) which is also 10.6 in this case.

- To compute \( E(Q) \) and \( \text{var}(Q) \), again use \( p - r = 1 \) for \( p \); \( q \) is one more, corresponding to the trend which has been added to the cointegrating vector. With \( \hat{p} = 0.96 \), the result is 5.98 and 10.85 respectively.

- The approximating distribution is \( \Gamma(-5.98^2/10.85, 5.98/10.85) \). Or roughly: \( 1.1 \times 5.27 \) comes from a \( \chi^2(6.6) \).

Rahbek and Mosconi (1998) suggest to add \( p_{oil} \) to the cointegrating space to avoid the need to compute the nuisance parameters. In that case, the analysis is conditional on an I(1) variable, and the analysis of Harbo et al. (1998) pertains. This test statistic, denoted \( S \) here, was considered by Doornik (1998a), and the \( p \)-values are listed in Table 4. As Rahbek and Mosconi (1998) note, these results barely support the hypothesis that the rank is one.

<table>
<thead>
<tr>
<th>( r )</th>
<th>( p - r )</th>
<th>trace test</th>
<th>( 1 - P(S) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>99.1</td>
<td>0.059</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>65.2</td>
<td>0.179</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>41.5</td>
<td>0.246</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>18.8</td>
<td>0.570</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6.10</td>
<td>0.682</td>
</tr>
</tbody>
</table>

This discrepancy merits further investigation. The small sample size is a possible limitation to the power of the tests, and therefore we shall work with a model which is more parsimonious. However, we first note that, although \( \Delta p_{oil} \) was added by Johansen and Juselius (1992) to avoid non-normality, there
is still strong non-normality in the $i_1$ equation, caused by a single outlier. Therefore we add a dummy for 1980(2). With this adjustment, all the vector and univariate misspecification tests of PcFiml (Doornik and Hendry, 1997) are passed. A joint test on the seasonals supports their deletion; the same holds for the second lags, with the exception of $i_{2,t-2}$. The seasonals and second lags are deleted, but $\Delta i_{2,t-1}$ is entered unrestrictedly.

<table>
<thead>
<tr>
<th>$r$</th>
<th>$p - r$</th>
<th>trace test</th>
<th>$1 - P(T)$</th>
<th>$1 - P(Q)$</th>
<th>trace test</th>
<th>$1 - P(S)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>171.6</td>
<td>0.000</td>
<td>0.000</td>
<td>176.3</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>73.5</td>
<td>0.000</td>
<td>0.002</td>
<td>78.3</td>
<td>0.018</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>41.2</td>
<td>0.072</td>
<td>0.039</td>
<td>45.0</td>
<td>0.137</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>16.4</td>
<td>0.468</td>
<td>0.348</td>
<td>18.9</td>
<td>0.566</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3.22</td>
<td>0.841</td>
<td>0.816</td>
<td>3.47</td>
<td>0.932</td>
</tr>
</tbody>
</table>

Table 5: $p$-values for the tests $T$, $Q$, and $S$

The new test results, using the same sample period, are in Table 5. (Doornik et al., 1998, noted that an impulse dummy, when entered unrestrictedly, does not affect the distribution.) The outcomes are no longer contradictory: a rank of two or more (the original conclusion of Johansen and Juselius), is clearly supported. There is some evidence of a third cointegrating vector, but the small sample argument leads us to reject this. Accepting $r = 2$, we can test whether the oil price can be deleted in the model corresponding to $S$. The test supports this: $\chi^2(2) = 0.92 \ [0.63]$. A weak form of purchasing power parity, namely that $p_1 = p_2$ and $e_{12}$ have equal but opposite coefficients in both cointegrating vectors, is not rejected: $\chi^2(2) = 4.38 \ [0.11]$ (without the trend, it would be strongly rejected).

## 5 Conclusion

We have derived a convenient way to tabulate the distribution of cointegration tests in the presence of additional stationary regressors. The proposed method is very accurate, and avoids the need for interpolation required with previous tabulations. In addition, the method is compact and easy-to-use, making it suitable for application in computer programs.

## References


