JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **20**: 797–810 (2005)
Published online 9 June 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.811

DISTRIBUTION APPROXIMATIONS FOR COINTEGRATION TESTS WITH STATIONARY EXOGENOUS REGRESSORS

H. PETER BOSWIJK[a] AND JURGEN A. DOORNIK[b]*
[a] *Department of Quantitative Economics, Universiteit van Amsterdam, The Netherlands*
[b] *Nuffield College, Oxford, UK*

SUMMARY

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 with both tabulated critical values and simulated p-values. The proposed procedure is applied to a UK model investigating purchasing power parity.

1. INTRODUCTION

Testing the rank of the cointegrating space is now a common starting point for applied multivariate economic analyses. The likelihood ratio test for cointegrating rank, denoted T_T, has been derived by Johansen (1988, 1995). As the sample size $T \to \infty$, its limiting distribution is characterized by a functional of Brownian motions:

$$\mathsf{T}_T \xrightarrow{\mathcal{D}} \mathsf{T} = \mathrm{tr}\left\{\int_0^1 \mathrm{d}BF' \left(\int_0^1 FF'\mathrm{d}u\right)^{-1} \int_0^1 F\mathrm{d}B'\right\} \quad (1)$$

where $B(u)$ is a standard vector-Brownian motion, and in the most simple case $F(u) = B(u)$. More generally, F depends on the treatment of deterministic terms, see Section 2. No analytical expression is known for (1), so tabulation based on Monte Carlo simulation for a range of values of dimensions and specifications of F has been used instead. See, for example, the accurate tables of MacKinnon *et al.* (1999). A different approach was taken by Doornik (1998), who approximates the distribution of T by a Gamma distribution with the same mean and variance as T. This also provides an accurate approximation, with the benefit that p-values and quantiles are readily available.

Recent extensions to the cointegration framework have given rise to a variant of (1), with the need to allow for correlation between the Brownian motions. This adds another dimension to the tables, making the search for an effective approximation all the more important. Providing that approximation is the objective of the current paper.

The plan of the remainder is as follows. In the next section, we present the distribution that is approximated, and briefly discuss the context in which it has arisen. In Section 3, we show that the

* Correspondence to: Jurgen A. Doornik, Nuffield College, Oxford OX1 1NF, UK.
E-mail: jurgen.doornik@nuffield.ox.ac.uk
Contract/grant sponsor: UK Economic and Social Research Council; Contract/grant number: R000237500.

Received 1 May 2001
Revised 13 April 2004

mean and variance of this distribution 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. In Section 4, this approximation is shown to be very accurate, especially for quantiles and p-values where accuracy is required (in the right-hand tail of the distribution). Section 5 applies the result to the purchasing power parity model of Johansen and Juselius (1992). Section 6 concludes.

2. COINTEGRATION TESTS WITH STATIONARY EXOGENOUS REGRESSORS

First consider the vector autoregression, written as a vector equilibrium correction model (VECM):

$$\Delta X_t = \Pi^* X_{t-1}^* + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \Phi q_t + \varepsilon_t, \qquad t = 1, \ldots, T \tag{2}$$

where $\{X_t\}$ is an n-vector time series, the starting values $(X_{1-k}, \ldots, X_0)$ are fixed, $\{\varepsilon_t\}$ is i.i.d. $N(0, \Omega)$, all parameters vary freely, and $X_{t-1}^* = (X_{t-1}', d_t')'$, where d_t and q_t are deterministic regressors. When rank $\Pi^* = r < n$ such that $\Pi^* = \alpha\beta'$ with α and β of full column rank r, and the characteristic equation of (2) has $n - r$ roots equal to one and all other roots outside the unit circle (such that explosive behaviour and I(2)-ness are ruled out, see Johansen, 1995, theorem 4.2 and corollary 4.3), then the model implies that X_t is cointegrated. The hypothesis of interest is $H(r)$: rank $\Pi^* \leq r$, and the limiting null distribution of the LR statistic T_T is given in (1) above. Now $B(u)$ in (1) is a standard p-vector Brownian motion, $p = n - r$, and F depends on the choice of d_t and q_t. The three deterministic specifications most commonly used[1] 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, $F(u) = B(u)$;
- H_c : $d_t = 1$, q_t is void, restricted constant, $F(u) = \{B(u)', 1\}'$;
- H_l : $d_t = t$, $q_t = 1$, restricted linear trend, $F(u) = \{[B(u) - \int_0^1 B\,\mathrm{d}u]', [u - \frac{1}{2}]\}'$.

Seo (1998) considers an extension of (2), where a weakly exogenous stationary vector process Z_t is added to the regressors (together with some of its lags):

$$\Delta X_t = \Pi^* X_{t-1}^* + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \Phi q_t + \sum_{j=0}^{m} D_j Z_{t-j} + \varepsilon_t, \qquad t = 1, \ldots, T \tag{3}$$

In this case (when Z_t is I(0)), the limiting distribution of the LR statistic for $H(r)$ is characterized by

$$\mathsf{Q}_T \xrightarrow{\mathcal{D}} \mathsf{Q} = \mathrm{tr}\left\{ \int_0^1 \mathrm{d}W F' \left(\int_0^1 F F' \,\mathrm{d}u \right)^{-1} \int_0^1 F\,\mathrm{d}W' \right\} \tag{4}$$

[1] In these cases the process X_t and the cointegrating relations exhibit the same deterministic pattern. We do not consider H_{lc} (unrestricted constant, $F(u) = \{B_1(u), \ldots, B_{p-1}(u), u\}'$ corrected for a constant) nor H_{ql} (unrestricted trend and constant, $F(u) = \{B_1(u), \ldots, B_{p-1}(u), u^2\}'$ corrected for constant and trend). Rank determination is more complex in these cases (see Doornik *et al.*, 1998, for an example), and the covariance term in (9) is not a constant. Also see Nielsen and Rahbek (2000).

where $F(u)$ is the same as before,[2] and $W(u)$ is a standard p-vector Brownian motion ($p = n - r$), with $E[W(1)B(1)'] = P = \text{diag}(\rho_1, \ldots, \rho_p)$. The nuisance parameters $\rho_i \in [0, 1]$ are defined as the long-run canonical correlation coefficients between $\alpha'_\perp \varepsilon_t$ and $\alpha'_\perp \left(\varepsilon_t + \sum_{j=0}^m D_j Z_{t-j}\right)$, where $\alpha_\perp$ is any $n \times p$ matrix of full column rank such that $\alpha'_\perp \Pi^* = \alpha'_\perp \alpha\beta' = 0$ (see also footnote 8). The motivation for this extension is that taking the stationary regressors into account leads to increased asymptotic efficiency. Seo (1998, pp. 342–343) gives the assumptions under which the limiting distribution (4) holds. In Section 5 we apply this procedure, and compare it with an alternative proposed by Rahbek and Mosconi (1999), which avoids the nuisance parameters. Note that $\mathsf{Q} = \mathsf{T}$ when $P = I_p$.

When $n = 1$ (and $r = 0$, $P = \rho$), the random variable Q in model H_z is the square of

$$\mathsf{V} = \left(\int_0^1 B^2\,\mathrm{d}u\right)^{-1/2} \int_0^1 B\,\mathrm{d}W$$

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

$$\begin{aligned}\mathsf{V} &= \rho \frac{\int_0^1 B\,\mathrm{d}B}{\left(\int_0^1 B^2\,\mathrm{d}u\right)^{1/2}} + (1-\rho^2)^{1/2} \frac{\int_0^1 B\,\mathrm{d}U}{\left(\int_0^1 B^2\,\mathrm{d}u\right)^{1/2}} \\ &= \rho\mathsf{Y} + (1-\rho^2)^{1/2}\mathsf{Z} \end{aligned} \quad (5)$$

where Y corresponds to the limiting distribution of the Dickey–Fuller t-statistic, and Z is a standard normal random variable, independent of B and hence Y. Kremers *et al.* (1992) suggest to use critical values from the standard normal distribution by a small-σ (in this case small-ρ) asymptotic argument.

The same distribution of V was also obtained by Hansen (1995), in the context of testing for a unit root with stationary exogenous regressors. Hansen tabulated the distribution of V for $\rho^2 \in \{0.1, 0.2, \ldots, 1\}$, and suggested to interpolate these critical values for other values of ρ^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 provides tables for $p \le 5$ 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 is 20 pages of tables. Requiring from practitioners to obtain appropriate critical values by interpolation in several dimensions may be prohibitively inconvenient, 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 (1998) 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' (2000) approximation of the distribution of V by a normal distribution with non-zero mean and non-unit variance.

[2] Seo (1998) tabulates H_z, but not H_c or H_l. Instead he considers 'No trend in DGP' (restricted constant in DGP, but unrestricted in model; $F(u) = B(u)$ corrected for a constant) and 'Trend in DGP' (restricted trend/unrestricted constant in DGP, unrestricted trend and constant in model; $F(u) = B(u)$ corrected for constant and trend).

The distribution of V and Q also arises when $H(r)$ is tested in model (2) 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 (2002), who use a semi-non-parametric likelihood function. Furthermore, the distribution of V 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 (1998, 2003) and Seo (1999). In the univariate case ρ now represents the correlation between the errors ε_t and the 'scores' $-\partial\ell_t/\partial\varepsilon_t$, where ℓ_t are the log-likelihood contributions. The definition of ρ_i in the multivariate case is discussed in, e.g., Boswijk and Lucas (2002).

3. APPROXIMATING THE DISTRIBUTION OF Q

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

$$\Gamma(x; b, a) = \int_0^x \frac{a^b}{\Gamma(b)} t^{b-1} \,\mathrm{e}^{-at}\,\mathrm{d}t, \qquad x > 0, b > 0, a > 0 \tag{6}$$

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

Doornik (1998) demonstrated by simulation that the Gamma distribution with $b = E(\mathsf{T})^2/\mathrm{var}(\mathsf{T})$ and $a = E(\mathsf{T})/\mathrm{var}(\mathsf{T})$ provides an accurate approximation of the distribution of T. The mean and variance of T could in principle be tabulated for the three different deterministic models and many values of p. However, Doornik estimates response surfaces for the Monte Carlo results, which were found (by simulation) to be sufficiently accurate:[3]

$$E(\mathsf{T}) \approx \begin{cases} 2p^2 - p + 0.07 + 0.07 \cdot 1_{\{p=1\}} & \text{for } H_z \\ 2p^2 + 2.01p + 0.06 \cdot 1_{\{p=1\}} + 0.05 \cdot 1_{\{p=2\}} & \text{for } H_c \\ 2p^2 + 4.05p + 0.5 - 0.23 \cdot 1_{\{p=1\}} - 0.07 \cdot 1_{\{p=2\}} & \text{for } H_l \end{cases} \tag{7}$$

$$\mathrm{var}(\mathsf{T}) \approx \begin{cases} 3p^2 - 0.33p - 0.55 & \text{for } H_z \\ 3p^2 + 3.60p + 0.75 - 0.40 \cdot 1_{\{p=1\}} - 0.30 \cdot 1_{\{p=2\}} & \text{for } H_c \\ 3p^2 + 5.70p + 3.20 - 1.30 \cdot 1_{\{p=1\}} - 0.50 \cdot 1_{\{p=2\}} & \text{for } H_l \end{cases} \tag{8}$$

Doornik also analyses

$$\mathsf{T}_i = \int_0^1 \mathrm{d}B_i F' \left(\int_0^1 FF'\,\mathrm{d}u \right)^{-1} \int_0^1 F\,\mathrm{d}B_i$$

where B_i is the ith component of B, so that $\mathsf{T} = \sum_{i=1}^p \mathsf{T}_i$. Since T_i and T_j have the same distribution, we have $E(\mathsf{T}_i) = E(\mathsf{T})/p$. Furthermore, he finds (by simulation) that for $i \neq j$,

$$\mathrm{cov}(\mathsf{T}_i, \mathsf{T}_j) \approx \begin{cases} -1.270 & \text{for } H_z \\ -1.066 & \text{for } H_c \\ -1.35 & \text{for } H_l \end{cases} \tag{9}$$

This can be used to evaluate $\mathrm{var}(\mathsf{T}_i) = \mathrm{var}(\mathsf{T})/p - (p-1)\mathrm{cov}(\mathsf{T}_i, \mathsf{T}_j)$.

[3] The variance entries in table VII of Doornik (1998) should be labelled $n - p, 1, n - p = 1, n - p = 2$.

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

Theorem 1 *Let q denote*[4] *the dimension of the vector F. Then*

$$E(\mathsf{Q}) = \frac{E(\mathsf{T})}{p}\sum_{i=1}^{p}\rho_i^2 + \left(1 - \frac{1}{p}\sum_{i=1}^{p}\rho_i^2\right) pq \tag{10}$$

and

$$\begin{aligned}\text{var}\,(\mathsf{Q}) = \text{var}\,(\mathsf{T}_i)\sum_{i=1}^{p}\rho_i^4 + 2\text{cov}\,(\mathsf{T}_i, \mathsf{T}_j)\sum_{i=2}^{p}\sum_{j=1}^{i-1}\rho_i^2\rho_j^2 \\ + \frac{4E(\mathsf{T})}{p}\sum_{i=1}^{p}\rho_i^2(1-\rho_i^2) + 2q\sum_{i=1}^{p}(1-\rho_i^2)^2\end{aligned} \tag{11}$$

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

$$\mathsf{Z} = \left(\int_0^1 FF'\,\mathrm{d}u\right)^{-1/2}\int_0^1 F\,\mathrm{d}U' \sim N(0, I_{q\times p})$$

independently of B. Defining

$$\mathsf{Y} = \left(\int_0^1 FF'\,\mathrm{d}u\right)^{-1/2}\int_0^1 F\,\mathrm{d}B'$$

it follows that (remembering that both P and R are diagonal)

$$\begin{aligned}\mathsf{Q} &= \text{tr}([\mathsf{Y}P + \mathsf{Z}R]'[\mathsf{Y}P + \mathsf{Z}R]) \\ &= \text{tr}(P\mathsf{Y}'\mathsf{Y}P + P\mathsf{Y}'\mathsf{Z}R + R\mathsf{Z}'\mathsf{Y}P + R\mathsf{Z}'\mathsf{Z}R) \\ &= \text{tr}(P\mathsf{Y}'\mathsf{Y}P) + 2\ \text{tr}(P\mathsf{Y}'\mathsf{Z}R) + \text{tr}(R\mathsf{Z}'\mathsf{Z}R) \\ &= \sum_{i=1}^{p}\rho_i^2\mathsf{T}_i + 2\ \text{tr}(P\mathsf{Y}'\mathsf{Z}R) + \sum_{i=1}^{p}(1-\rho_i^2)\xi_i\end{aligned} \tag{12}$$

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

$$\begin{aligned}E(\mathsf{Q}) &= \sum_{i=1}^{p}\rho_i^2 E(\mathsf{T}_i) + \sum_{i=1}^{p}(1-\rho_i^2)q \\ &= \frac{E(\mathsf{T})}{p}\sum_{i=1}^{p}\rho_i^2 + \left(1 - \frac{1}{p}\sum_{i=1}^{p}\rho_i^2\right) pq\end{aligned}$$

[4] So $q = p + 1$ for models H_c, H_l and $q = p$ for H_z.

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 Y. Next, the covariance between the second and third term in (12) is zero (conditionally on Y), because elements of Z are uncorrelated with squared elements of Z (the normal distribution has third moment equal to zero). Hence all covariances are zero, and the variance of Q can be reduced to

$$\text{var}(\mathsf{Q}) = \text{var}\left(\sum_{i=1}^{p} \rho_i^2 \mathsf{T}_i\right) + 4\ \text{var}(\text{tr}[P\mathsf{Y}'\mathsf{Z}R]) + \text{var}\left(\sum_{i=1}^{p}(1-\rho_i^2)\xi_i\right) \tag{13}$$

For the first term of (13) we find

$$\text{var}\left(\sum_{i=1}^{p} \rho_i^2 \mathsf{T}_i\right) = \sum_{i=1}^{p} \rho_i^4 \text{var}(\mathsf{T}_i) + 2\sum_{i=2}^{p}\sum_{j=1}^{i-1} \rho_i^2\rho_j^2 \text{cov}(\mathsf{T}_i, \mathsf{T}_j)$$

To evaluate the second variance term, we use

$$\text{tr}(P\mathsf{Y}'\mathsf{Z}R) = \text{tr}(\mathsf{Y}'\mathsf{Z}RP) = \text{vec}(P)'(\mathsf{Y}' \otimes R)\text{vec}(\mathsf{Z}')$$

Hence, because $E(\text{tr}[P\mathsf{Y}'\mathsf{Z}R]|Y) = 0$, and because P and R commute, being diagonal matrices:

$$\begin{aligned}
\text{var}(\text{tr}[P\mathsf{Y}'\mathsf{Z}R]) &= E\{\text{var}[\text{vec}(P)'(\mathsf{Y}' \otimes R)\text{vec}(\mathsf{Z}')|\mathsf{Y}]\} \\
&= E[\text{vec}(P)'(\mathsf{Y}'\mathsf{Y} \otimes R^2)\text{vec}(P)] \\
&= E[\text{tr}(\mathsf{Y}'\mathsf{Y}P^2R^2)] \\
&= \sum_{i=1}^{p} \rho_i^2(1-\rho_i^2)E(\mathsf{T}_i) = \frac{\sum_{i=1}^{p} \rho_i^2(1-\rho_i^2)}{p}E(\mathsf{T})
\end{aligned}$$

The final term in (13) follows immediately from the fact that $\xi_i \sim$ 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. □

The approximation to the distribution of Q is $\Gamma(\cdot; b, a)$, with $b = E(\mathsf{Q})^2/\text{var}(\mathsf{Q})$ and $a = E(\mathsf{Q})/\text{var}(\mathsf{Q})$. The arguments are computed using Theorem 1 in combination with the response surfaces (7)–(9) obtained from prior simulations.

When using the proposed approximation in practice, several errors will be made.

1. Approximation error: the Gamma distribution is used instead of the exact limiting distribution.
2. Estimation error: the correlation coefficients are not known, and replaced by estimates.
3. Small sample error: the exact small sample distribution is not known. The actual distribution is not only different because the sample size is small, but also because there is only asymptotic

similarity with respect to all the nuisance parameters (other than the deterministic terms and correlations, which are already handled).

Only (1) is relevant for the approximation provided in this paper, because it concerns the limiting distribution. In particular, if one wanted to avoid the first approximation error and simulate the exact asymptotic distribution (or interpolate Seo's (1998) tables), then errors (2) and (3) would remain. Although we emphasize the importance of these other errors in practice, they do not affect the main contribution of the present paper. Therefore, the next section studies the approximation error (1), which will turn out to be small.

4. EVALUATION OF THE APPROXIMATION

Our proposal is to approximate the limiting distribution of Q_T by a Gamma distribution, with arguments provided by Theorem 1 and response surfaces. The first source of approximation error is that the actual distribution is not Gamma. A possible second source is that the arguments of the approximating distribution are obtained from a response surface.

Evaluation of the approximation is hampered by the fact that the actual distribution is a functional of Brownian motions. Such a distribution is usually simulated by a discretized representation. However, using a fixed sample size introduces a distortion. MacKinnon *et al.* (1999) solve this by capturing the curvature with respect to the discretization sample size of the quantiles, while Doornik (1998) captures the curvature in the mean and variance of the statistic. Then, letting $T \to \infty$ gives the asymptotic quantities. This indicates that simulation is not trivial, providing an additional argument for the use of our procedure.

To illustrate our approach to assess the accuracy, we first consider one particular model: $H_l(0)$, the test for rank zero in the presence of an unrestricted constant and a restricted trend, with $p = 3$ and $\rho = (0.8, 0.2, 0.0)$. Details of the simulation procedure are given below.[5] Write $\hat{\Gamma}_{M,T}$ for the Gamma distribution that is fitted by matching the mean and variance to that found in this particular Monte Carlo experiment, and $\hat{\Gamma}_\infty$ for the approximation proposed in this paper.

Figure 1a gives the simulated distribution for discretization sample size $T = 1000$ and $M = 100\,000$ replications. On the horizontal axis are the probability levels, and on the vertical axis the corresponding simulated critical values. Also shown are its approximate 95% confidence interval, computed as $[p(1-p)/(f^2 M)]^{0.5}$, where f is the density corresponding to $\hat{\Gamma}_{M,T}$, together with $\hat{\Gamma}_\infty$ and $\hat{\Gamma}_{M,T}$. These are all indistinguishable on the current scale. Figure 1b draws the corresponding QQ plot, with the approximation as reference on the horizontal axis. Finally, Figure 1c draws the 'tilted' QQ plot, i.e. in deviation from the 45° line. Then the reference distribution becomes a horizontal line at zero. The tilted QQ plot allows us to distinguish between the lines that are drawn. It shows that $\hat{\Gamma}_{M,T}$ (the dashed line) is mostly inside the confidence band, except in the left tail, while $\hat{\Gamma}_\infty$ (the line at zero) is mostly outside, except for the right tail of the distribution. Note that the confidence bands of the simulated critical values can be made arbitrarily small as M grows, putting $\hat{\Gamma}_\infty$ outside. However, the true limiting distribution will then also be outside, because it does not coincide with $T = 1000$.

To obtain a better representation of the limiting distribution, we simulate the critical values for sample sizes $T = 200, 400, \ldots, 1800, 2000$. The asymptotic critical value is then the intercept in

[5] All experiments are done using Ox 3, see Doornik (2001).

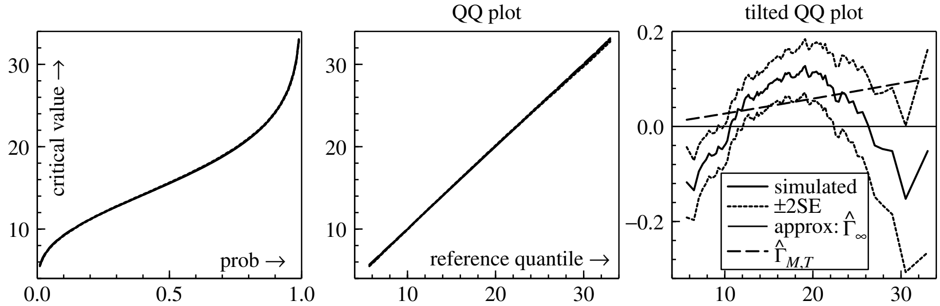

Figure 1. Simulated quantiles of $H_l(0)$, $p = 3$, $M = 10^5$, $T = 1000$ and approximation $\hat{\Gamma}_\infty$

a regression on 1, $1/T$, and the estimated standard error is used for the confidence bands. Because we need to compare many different distributions in the next stage, we change the graphical representation from QQ plots to probability plots (called PP plots in the remainder). Instead of transforming the horizontal axis of Figure 1a with the quantiles of the Gamma approximation, we transform the vertical axis with the distribution of the Gamma approximation. Tilting of the plot is again necessary to show sufficient detail. The outcome for the experiment considered before is in Figure 2. It shows that the approximation $\hat{\Gamma}_\infty$ works remarkably well: the distance between the distributions never exceeds 0.0025.

To evaluate the approximation more generally, we consider dimensions up to five and a range of correlations: $p = 1, \ldots, 5$ 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 fully correlated case, which coincides with the standard trace test, is omitted. This design consists of 456 specifications for each of the three treatments of deterministic terms, so 1368 experiments in total. The test statistic[6] is simulated at sample sizes 200 (200) 2000, and the asymptotic critical value is again the intercept of a

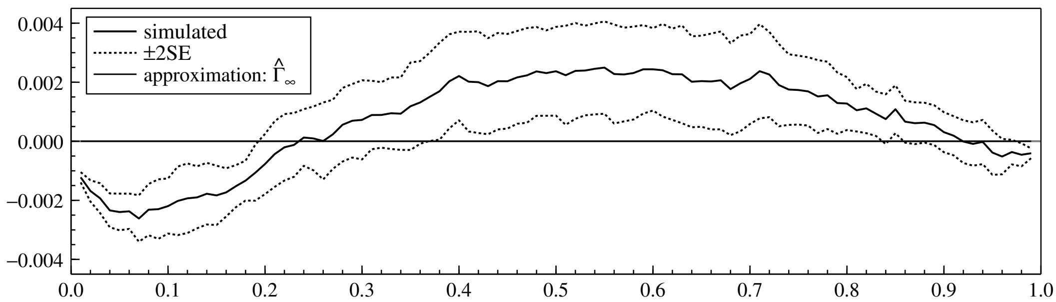

Figure 2. Tilted PP plot of simulated asymptotic quantiles for $H_l(0)$, $p = 3$, $M = 10^5$, with reference to $\hat{\Gamma}_\infty$

[6] Following Doornik (1998, section 4), the distribution of Q is simulated from $-T\sum\log(1-\lambda_i)$. Here, λ_i are the eigenvalues of $T^{-1}W_T'F_T(F_T'F_T)^{-1}F_T'W_T$, which is the discrete approximation to the expression inside the trace of (4). Let E_T be a $T \times p$ matrix of standard normal random numbers, standardized such that $E_T'E_T = TI_p$. B_T is the cumulated sum of E_T, lagged one period, and F_T is B_T for H_z, (B_T, ι) for H_c, and (B_T, τ) in deviation from mean for H_l, where ι and τ represent the constant and trend term, respectively. Finally, dW is approximated by $W_T = E_TP + D_TR$, where D_T is an independent $T \times p$ matrix of standard normal random numbers. Using the eigenvalues instead of the trace mimics the statistical procedure, and has been found to be a more accurate approximation for given T.

regression on 1, $1/T$. The number of replications is 10^5, except for dimensions four and five, where 10^4 is used (10^5 replications would take about a month of computation on a 1.6 MHz Pentium IV).

This experiment yields 1368 analogues to Figure 2, which are displayed in the second column of Figure 3, separate for each dimension. The first figure in column two is for $p = 1$, which has 15 PP plots, the second for $p = 2$ has 60 PP plots, and the third has 165. Dimensions 4 and 5 are similar, and combined; however, the 1128 PP plots form just a black cloud, and are not shown.

The first column of Figure 3 shows the (smoothed) envelope of the cloud of PP plots in the dashed line. The reference approximation $\widehat{\Gamma}_\infty$ is again the line at zero. Next, the average of the PP plots is drawn, together with the standard error.

These plots enable us to assess the accuracy of the approximation. On average, there is very little systematic bias. There is some systematic error in the left tail at dimension 1, but this disappears as the dimension grows. The envelope gives an indication of the maximum error that is made when using the proposed distribution. Except for the left tail of $p = 1$, this is always smaller than 0.01. For dimension 3 and higher the maximum error is less than 0.005 ($p = 4, 5$ uses a lower number of replications, and we believe this would be more concentrated with $M = 10^5$). In the right tail of the distribution, which is most relevant for testing, the maximum error is smaller still: less than 0.002 for critical values with probabilities in excess of 0.9. So if, for example, the Gamma approximation is used to attach a p-value of 0.05 to a test outcome, the true value will be between 0.048 and 0.052 (and potentially even closer to 0.05).

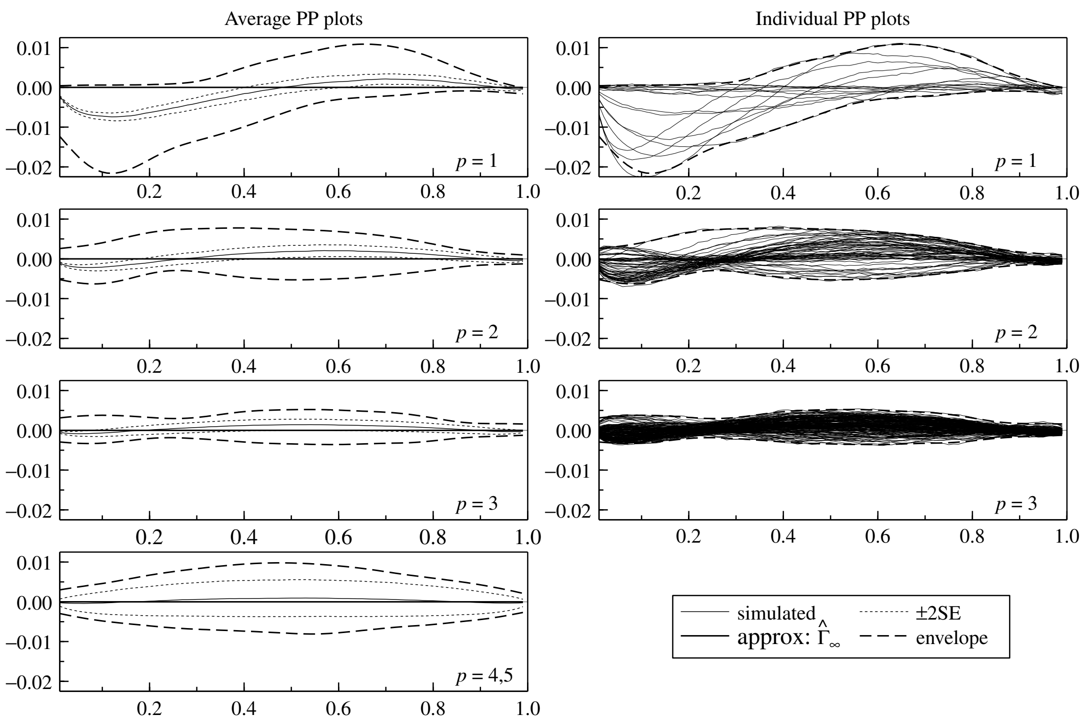

Figure 3. Average and individual tilted PP plots, with reference to $\widehat{\Gamma}_\infty$

5. APPLICATION

To illustrate the proposed procedure, we use a variant of the model estimated by Johansen and Juselius (1992) to investigate purchasing power parity in the UK. 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 (Δp_{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 $p_1 - p_2$, Δp_1 to avoid I(2)-ness. Following Rahbek and Mosconi (1999) we adopt model H_l by allowing a trend to enter the cointegrating space. In terms of (3) our specification is a VAR(2):

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

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). This model falls in the class considered by Seo (1998), which allows us to take the presence of $\Delta p_{\text{oil},t}$ into account. The test statistic is the standard cointegration trace test (noting that $\Delta p_{\text{oil},t}$ and its lag are also partialled out from the levels and differences), but the limiting distribution is different.

Table I lists the test values for each rank, together with p-values. The fourth column, labelled $1 - \Pr(\mathsf{T}_T)$, 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 may be a tendency to over-reject the true rank in small samples: when adopting the sample size-adjusted[7] p-value for $H_l(1)$, 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 1 is firmly rejected with a p-value of 1.7%. If we use the small sample mean and variance of T_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 final column lists the estimated

Table I. p-Values for the tests T_T and Q_T

r	$p = n - r$	Trace test	$1 - \Pr(\mathsf{T}_T)$	$1 - \Pr(\mathsf{Q}_T)$	Canonical correlations, $\hat{\rho}$
0	5	95.3	0.014	0.002	1, 1, 1, 0.849, 0.385
1	4	61.4	0.077	0.017	1, 1, 0.893, 0.412
2	3	37.8	0.150	0.084	1, 0.932, 0.802
3	2	16.7	0.445	0.300	0.970, 0.818
4	1	5.27	0.567	0.516	0.960

[7] Using the formulae in Doornik (1998). This is not the exact distribution, but the limiting distribution evaluated using a discretization at the same sample size as the estimations.

canonical correlations, following the suggestion of Seo (1998, p. 348).[8] So, when ignoring the role of the stationary regressor (i.e. using $1 - \Pr(\mathsf{T}_T)$), we find one cointegrating relation. However, basing inference on Q_T results in two relations.

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 $n - r = 1$ for p in (7) and (8) to compute $E(\mathsf{T})$ and $\text{var}(\mathsf{T})$. This yields 6.32 and 10.6. The next step requires $\text{var}(\mathsf{T}_i)$, which is also 10.6 in this case.
- To compute $E(\mathsf{Q})$ and $\text{var}(\mathsf{Q})$, again use $n - r = 1$ for p; q is one more, corresponding to the trend that has been added to the cointegrating vector. With $\hat{\rho} = 0.96$, the result is 5.98 and 10.85, respectively.
- The approximating distribution is $\Gamma(\cdot; 5.98^2/10.85 = 3.30, 5.98/10.85 = 0.55)$. Or roughly: 1.1×5.27 comes from a $\chi^2(6.6)$; this uses $X \sim \Gamma(\cdot, b, a)$ then $2aX \sim \chi^2(2b)$.

Rahbek and Mosconi (1999) suggest an alternative approach to handle the presence of the stationary regressor, which avoids the nuisance parameters in the limiting distribution. Their solution, when Z_t denotes the stationary regressor (possibly present with additional lags), is to add $C_{t-1} = \sum_{i=1}^{t-1} Z_i$ to the cointegrating vectors (i.e. adding C_t to X_t^*). 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_T here, was also considered by Doornik (1998), and the p-values are listed in Table II. As Rahbek and Mosconi (1999) note, these results barely support the hypothesis that the rank is 1.

This discrepancy merits further investigation, because both approaches address the same issue. The small sample size is a possible limitation to the power of the tests, and therefore we shall work with a model that is more parsimonious. However, we first note that, although Δp_{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 variable for 1980(2). With this adjustment, all the vector and univariate misspecification tests of PcFiml (Doornik and

[8] Recall that $\alpha_\perp$ is an $n \times p$ matrix orthogonal to α. Define $\widehat{e}_t = \widehat{\alpha}'_\perp \widehat{\varepsilon}_t$ and $\widehat{u}_t = \widehat{\alpha}'_\perp \left(\sum_{j=0}^{m} \widehat{D}_j Z_{t-j} + \widehat{\varepsilon}_t \right)$. Then the kernel method of Andrews (1991) is applied:

$$\widehat{\Sigma} = \sum_{j=-K}^{K} k(j/\widehat{S}_T)\widehat{\Gamma}(j) \quad \text{where} \quad \widehat{\Gamma}(j) = \begin{cases} T^{-1} \sum_{t=j+1}^{T} \widehat{v}_t \widehat{v}'_{t-j} & \text{for } j \geq 0 \\ T^{-1} \sum_{t=-j+1}^{T} \widehat{v}_{t+j} \widehat{v}'_t & \text{for } j < 0 \end{cases}$$

and $\widehat{v}_t = (\widehat{e}'_t, \widehat{u}'_t)' = (\widehat{v}_{1,t}, \ldots, \widehat{v}_{2p,t})'$. K is chosen as $\min(T-1, \lfloor 50\widehat{S}_T \rfloor)$. We use the quadratic spectral kernel with automatic bandwidth and an AR(1) for each component:

$$k(u) = \frac{25}{12\pi^2 u^2} \left[\frac{\sin(6\pi u/5)}{6\pi u/5} - \cos(6\pi u/5) \right], \qquad \widehat{S}_T = 1.3321[\widehat{\eta} T]^{1/5}, \qquad \widehat{\eta} = \sum_{i=1}^{2p} \frac{4\widehat{\phi}_i^2 \widehat{\sigma}_i^4}{(1-\widehat{\phi}_i)^8} \Big/ \sum_{i=1}^{2p} \frac{\widehat{\sigma}_i^4}{(1-\widehat{\phi}_i)^4}$$

where $\widehat{\phi}_i$ is the OLS estimate from regressing $\widehat{v}_{i,t}$ on $\widehat{v}_{i,t-1}$ ($T-1$ observations) and $\widehat{\sigma}_i^2$ the corresponding residual variance. Partitioning $\widehat{\Sigma}$ according to e and u, the canonical correlations are found from the generalized eigenproblem $|\rho^2 \widehat{\Sigma}_{ee} - \widehat{\Sigma}_{eu} \widehat{\Sigma}_{uu}^{-1} \widehat{\Sigma}_{ue}| = 0$.

The results in Table I are not much changed when using $\widehat{\Gamma}(0)$ instead of $\widehat{\Sigma}$. The smallest roots are then for $r = 0, \ldots, 4$, respectively: 0.528, 0.545, 0.871, 0.887, 0.958, with small impact on the p-values: 0.003, 0.021, 0.091, 0.331, 0.513.

Table II. p-Values for the test conditional on the I(1) variable p_{oil}

r	$p = n - r$	Trace test	$1 - \Pr(S_T)$
0	5	99.1	0.059
1	4	65.2	0.179
2	3	41.5	0.246
3	2	18.8	0.570
4	1	6.10	0.682

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.

The new test results, using the same sample period, are in Table III. (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 2 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 from the cointegrating vectors in the model corresponding to S_T. The test supports this: $\chi^2(2) = 0.92$ (p-value: 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: $\chi^2(2) = 4.38$ (p-value: 0.11) (without the trend, it would be strongly rejected).

6. CONCLUSION

We have derived a convenient way to tabulate the distribution of cointegration tests in the presence of additional stationary regressors, analysed by Seo (1998). 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. Moreover, we tabulated the most commonly used treatments of deterministic terms, which have not yet been available.

We considered an application to a model investigating UK purchasing power. The test that we tabulated requires estimates of nuisance parameters. We also considered the approach of Rahbek and Mosconi (1999), which has the benefit of avoiding the nuisance parameters. Whether this is sufficient reason to prefer this method over the approach of Seo is an open question, since it is not clear what the loss of power of the Rahbek–Mosconi approach is in typical applications. In our application it was small, but only after more careful modelling. Now, at least, we can do inference

Table III. p-Values for the tests T_T, Q_T and S_T

r	$p = n - r$	Trace test	$1 - \Pr(\mathsf{T}_T)$	$1 - \Pr(\mathsf{Q}_T)$	Trace test	$1 - \Pr(S_T)$
0	5	171.6	0.000	0.000	176.3	0.000
1	4	73.5	0.005	0.003	78.3	0.018
2	3	41.2	0.072	0.043	45.0	0.137
3	2	16.4	0.468	0.352	18.9	0.566
4	1	3.22	0.841	0.820	3.47	0.932

in the Seo framework more easily using our approximation.[9] Finally, our approximation will also be useful in other settings where no alternative is readily available, in particular when testing for a unit root or cointegration in models with non-Gaussian or GARCH errors.

ACKNOWLEDGEMENTS

Doornik gratefully acknowledges financial support from the UK Economic and Social Research Council (grant R000237500). We are grateful to the editor and anonymous referees for their helpful comments.

REFERENCES

Abadir KM, Lucas A. 2000. Quantiles for *t*-statistics based on M-estimators of unit roots. *Economics Letters* **67**: 131–137.

Andrews DWK. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**: 817–858.

Boswijk HP, Lucas A. 2002. Semi-nonparametric cointegration testing. *Journal of Econometrics* **108**: 253–280.

Doornik JA. 1998. Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys* **12**: 573–593.

Doornik JA. 2001. *Ox–An Object-Oriented Matrix Programming Language*. Timberlake Consultants Press: London.

Doornik JA, Hendry DF. 1997. *Modelling Dynamic Systems using PcFiml 9 for Windows*. Timberlake Consultants Press: London.

Doornik JA, Hendry DF, Nielsen B. 1998. Inference in cointegrating models: UK M1 revisited. *Journal of Economic Surveys* **12**: 533–572.

Hansen B. 1995. Rethinking the univariate approach to unit root testing: using covariates to increase power. *Econometric Theory* **11**: 1148–1171.

Hansen H, Juselius K. 1995. *CATS in RATS: Cointegration Analysis of Time Series*. Estima: Evanston, IL.

Harbo I, Johansen S, Nielsen B, Rahbek A. 1998. Asymptotic inference on cointegrating rank in partial systems. *Journal of Business Economics and Statistics* **16**: 388–399.

Johansen S. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**: 231–254.

Johansen S. 1995. *Likelihood Based Inference on Cointegration in the Vector Autoregressive Model*. Oxford University Press: Oxford.

Johansen S, Juselius K. 1992. Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics* **53**: 211–244.

Kremers J, Ericsson NR, Dolado JJ. 1992. The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* **54**: 325–348.

Ling S, Li WK. 1998. Limiting distributions of maximum likelihood estimators for unstable autoregressive moving average time series with general autoregressive heteroskedastic errors. *Annals of Statistics* **26**: 84–125.

Ling S, Li WK. 2003. Asymptotic inference for unit root processes with GARCH(1,1) errors. *Econometric Theory* **19**: 541–564.

Lucas A. 1997. Cointegration testing using pseudo likelihood ratio tests. *Econometric Theory* **13**: 149–169.

MacKinnon JG, Haug AA, Michelis L. 1999. Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics* **14**: 563–577.

Nielsen B. 1997. Bartlett correction of the unit root test in autoregressive models. *Biometrika* **84**: 500–504.

[9] As pointed out by an anonymous referee, if a confidence interval for the correlations is available, this could now easily be used to provide a corresponding range of p-values.

Nielsen B, Rahbek A. 2000. Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics* **62**: 5–22.
Rahbek A, Mosconi R. 1999. Cointegrating rank inference with stationary regressors in VAR models. *Econometrics Journal* **2**: 82–97.
Seo B. 1998. Statistical inference on cointegration rank in error correction models with stationary covariates. *Journal of Econometrics* **85**: 339–385.
Seo B. 1999. Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* **91**: 113–144.

