Abstract

The question of the appropriate critical values to be used in conjunction with the trace test for cointegrating rank, when the model contains an unrestricted constant, is clarified, by putting this question into the context of the choice of model. Monte Carlo simulations are used to defend the proposition that the critical values associated with Johansen’s “Case 1” give a correctly sized test if and only if the model is correctly specified. In the process, differences between alternative formulations of the “unrestricted constant” case are elucidated.

1 Introduction

Paul Turner (2009) asks, in relation to testing for cointegrating rank, “Are we using the correct critical values?” This is an important question for practitioners. The Johansen vector error-correction approach is highly popular, promising as it does to tease out long-run equilibrium relationships among time series that appear individually to be random walks. But getting the analysis right is not particularly easy. Johansen (1995) describes five cases regarding the deterministic terms (constant and/or trend) entering potentially cointegrated systems; Pesaran et al. (2000) also describe five cases, two of which are seemingly in competition with the corresponding Johansen cases (as Turner notes). The applied econometrician must therefore choose between, in effect, seven “cases”, without any simple algorithm for deciding which is right. Moreover, the asymptotic distributions of the test statistics for cointegrating rank (never mind the finite-sample distributions) are non-standard and must be...
obtained via simulation. And various different sets of critical values are used by different software packages, some fairly close to each other and some not so close.

So any clarification of the questions facing the practitioner is most welcome. But while Turner asks a pertinent question, he does not answer it satisfactorily. His focus on choice of critical values skirts the prior question of the model in which cointegrating rank is to be tested. We are able to replicate Turner’s simulation results—and have no quarrel with his methodology—but we dispute the interpretation he places on those results. By means of additional simulations of a similar kind, we try to clarify further what is at stake in using one or other set of critical values in testing for cointegration, in relation to the choice of model.

2 Johansen’s five cases

We first give a brief reminder of the Johansen approach to establish notation and situate the problem addressed by Turner. In the notation of Johansen (1995), a VAR of lag order $k$ may be written in error-correction form as

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \mu_0 + \mu_1 t + \epsilon_t$$

(1)

where $X_t$ is a $p$-vector of potentially cointegrated variables, $\mu_0$ and $\mu_1$ are $p$-vectors of coefficients, and $\epsilon_t$ is $p$-dimensional Gaussian noise. If the elements of $X_t$ are in fact cointegrated then $\Pi$ can be written as $\alpha \beta'$ for $p \times r$ matrices $\alpha$ and $\beta$ where $r$, the cointegrating rank, satisfies $1 \leq r < p$. (If $r = p$ then all the elements of $X$ are stationary and there is no error correction; a VAR in levels is the appropriate model.)

If $\mu_0 = \mu_1 = 0$ the model contains no deterministic elements; this is Johansen’s “Case 0”. If $\mu_1 = 0$ but $\mu_0$ is non-zero and unrestricted in its value this generates a linear trend in $X_t$, since the dependent variables in (1) are in differences; this is Johansen’s “Case 1”. Similarly, if $\mu_1$ is non-zero and unrestricted this means that $X_t$ exhibits a quadratic trend (Johansen’s “Case 2”).

Alternatively, a constant and/or trend may be present but confined to the cointegration space. Let $\mu_1 = 0$ but $\mu_0 \neq 0$. If $\mu_0$ is a linear combination of the columns of $\alpha$, then it can be absorbed into
the first term in (1). This is most easily seen for the case of \( p = 2 \) and \( r = 1 \), where

\[
\Pi X_{t-1} = \alpha \beta' X_{t-1} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}
\]

(2)

We can extend this to

\[
\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ 1 \end{pmatrix}
\]

(3)

The constant no longer spills over into a linear trend in \( X_t \). Rather, its role is to allow for a non-zero intercept in the cointegrating relations, the columns of \( \beta \). This is Johansen’s “Case 1*”; his “Case 2*” arises when the linear trend is restricted in a similar manner, adding a trend term to the cointegrating relations without spilling over into a quadratic trend in \( X_t \).

The five Johansen cases are summarized in Table 1.

3 The trace test

Johansen describes two statistics that can be used to assess the value of the cointegrating rank, namely the trace test and the maximum eigenvalue test. Following Turner (and much of the recent literature), we will consider the trace test only. This is a likelihood ratio test; for the null hypothesis \( r = 0 \) it is computed as

\[
-T \sum_{i=1}^{P} \log(1 - \hat{\lambda}_i)
\]

(4)

where \( T \) is the sample size and the \( \hat{\lambda}_i \)'s are the solutions of the eigen-problem

\[
|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0
\]

(5)

The \( S_{ij} \) matrices are defined as

\[
S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt}
\]

The “spillover” into a trend is the degree in which \( \mu \) is not expressible as a linear combination of the columns of \( \alpha \), that is, \( \alpha' \mu_0 \) where \( \alpha_{\perp} \) is the left null space of \( \alpha \), such that \( \alpha' \alpha_{\perp} = 0 \).
and for the cases where the deterministic terms (if any) are unrestricted, $R_0$ is the matrix of residuals from least squares regression of $\Delta X_t$ on $\Delta X_{t-i}$ ($i = 1, \ldots, k - 1$) and the deterministic terms, while $R_1$ is the matrix of residuals from regression of $X_{t-1}$ on the same set of regressors.

As mentioned above, no analytical expression is available for the asymptotic distribution of the trace test. Johansen shows that it can be written as a function of integrals of Brownian motions, where the function differs depending on the deterministic case. The various distributions can be simulated on a computer via discretization of the Brownians.\(^3\)

Since the distribution of the trace test depends on the case selected, it is obviously important that the selection be appropriate for the data. In the Johansen approach there is really no such thing as a null hypothesis of $r = 0$ (no cointegration) \textit{in general}; the null must be framed in the context of a model, in which the deterministic terms are specified.

4 Choosing the wrong model?

Turner is mostly concerned with the case of the unrestricted constant (Johansen’s Case 1). His worry is related to the fact that Pesaran et al. (2000) (PSS) discuss a variant of this case, which they label “Case III” (more on this below), and produce critical values for the trace test which differ substantially from those of Johansen Case 1. He wonders: when one performs the trace test using commonly available econometric software packages, for the “unrestricted constant” case, can we be sure that the right critical values are being used?

Turner’s initial approach is to generate an artificial bivariate dataset in which the two variables are not cointegrated, and to test the data for cointegration using various popular programs, selecting the “unrestricted constant” case. He finds that all the programs he tests (Microfit, EViews, PcGive and Stata) agree on the value of the trace test, but they do not agree on its marginal significance level. Specifically, Microfit is the odd man out, using the PSS Case III critical values; the other programs are in rough agreement, and more or less in conformity with best-practice critical values pertaining to Johansen’s Case 1.

Who is right, Microfit or the rest? Turner proposes that a “possible method for tackling this question

\(^3\)Osterwald-Lenum (1992) and Johansen both produced early tables of critical values, but in modern software there is no reason to use these when superior alternatives are now available, namely the highly accurate results in MacKinnon et al. (1999) and the parsimonious gamma-approximation devised by Doornik (1998). These alternatives are based on much more extensive simulation and the estimation of response surfaces.
is to make use of Monte Carlo methods to establish whether the critical values used generate rejection frequencies which are consistent with the size of the test assumed” (p. 828). He then proceeds to simulate non-cointegrated data (10,000 replications) and run it through EViews; he selects EViews’ “Case 3” specification—i.e. the unrestricted constant case—and examines the rejection rates using two sets of critical values: (a) those for Johansen Case 1 and (b) those for PSS Case III.\(^4\) He finds that the rejection rate using set (a) is 12.01 percent, substantially exceeding the nominal size of the test (5 percent), while the rate using set (b) is about right, at 5.59 percent. So it appears that Microfit was right, despite being odd man out.

Turner’s proposal of the Monte Carlo approach is reasonable, but for this to work the artificial data must be generated in conformity with the null hypothesis under test. If the null is \(r = 0\) in Johansen’s model 1, the series should be independent random walks with non-zero drift. Thanks to the *Journal of Applied Econometrics* policy of making pertinent details available for download, we’re able to see exactly how Turner generated his data: he constructed time series of length 500, and the (non-cointegrated) data were simulated as random walks *without* any drift.

The basic problem with Turner’s procedure is therefore that he simulates data in conformity with the null of \(r = 0\) in the model with *no* constant, but then calculates the test for the model with an unrestricted constant. It’s then not surprising that the test comes out wrongly sized.

We are able to replicate Turner’s results fairly closely as shown in Table 2. The data (with \(p = 2\), as in Turner) were generated according to

\[
y_{i,t} = y_{i,t-1} + \epsilon_{i,t} \quad i = 1, 2
\]

with \(\epsilon_1\) and \(\epsilon_2\) both NID(0, 1). For 50,000 replications we then calculate the trace test using the procedure for Johansen’s Case 1 (as Turner does using EViews), and also using the procedure for no constant (Case 0). In the first case we apply two rejection criteria: (a) the p-value calculated via Doornik’s (1998) gamma approximation,\(^5\) with adjustment for sample size, falls below 0.05; and (b) the trace test exceeds the 5 percent critical value given for PSS Case III in MacKinnon et al. (1999), namely 18.1058. In the second case we just record the rejection rate using the Doornik p-value. All calculations were performed using gretl.\(^6\)

---

\(^4\)In each case Turner employs the critical values given by MacKinnon et al. (1999).
\(^5\)Throughout, the Doornik p-values are calculated in conformity with Johansen Case.
The results are clear: if you calculate as if the unrestricted constant model were appropriate when it is not, the test using Johansen Case 1 p-values over-rejects by a factor of two at all sample sizes, but if you use the right model for the data (no constant) the test is correctly sized. As it happens, if you apply the PSS case III critical values to trace statistics generated using the wrong model, you get something close to a 5 percent rejection rate. Is this just luck, or what’s happening? To answer this we need to examine PSS Case III in more detail.

5 PSS Case III

The VECM given in (1) can be obtained by re-writing a VAR in levels. The VAR formulation from which Johansen starts

\[ X_t = \sum_{i=1}^{k} A_i X_{t-i} + \mu_0 + \mu_1 t + \epsilon_t \]  

(8)

or

\[ (I - A(L)) X_t = \mu_0 + \mu_1 t + \epsilon_t \]  

(9)

Pesaran et al., however, start from a different formulation of the VAR in levels, namely

\[ (I - A(L)) (X_t - \mu_0 - \mu_1 t) = \epsilon_t \]  

(10)

As they explain (2000, pp. 296–7), the formulation (10), in conjunction with auxiliary assumptions that rule out explosive roots and \( I(2) \) behavior of the \( X_s \), ensures that the “spillover” from \( \mu_0 \) to a linear trend in \( X_t \) or from \( \mu_1 \) to a quadratic trend in \( X_t \)—as discussed in section 2—cannot occur. PSS are happy with this maintained restriction; they argue that the alternative implies an “unsatisfactory conclusion”, namely that “quite different deterministic trending behavior should be observed in the levels process . . . for differing values of the cointegrating rank, \( r \), the number of independent quadratic deterministic trends, \( m - r \), decreasing as \( r \) increases” (p. 298). (In PSS’s notation \( m \) is the number of rows in the levels matrix that we are calling \( X_t \).)

Now, PSS discuss a “Case I” and “Case II” which are equivalent to Johansen’s cases 0 and 1*, respectively. But when it comes to their “Case III”, how does this relate to Johansen’s Case 1

\footnote{Unlike Johansen (1995), but like Harbo et al. (1998), PSS develop an analysis that can handle \( I(1) \) exogenous variables in the model to be tested for cointegration. But that is not our focus here. The difference between Johansen’s and PSS’s representation of the VAR in levels is independent of the issue of exogenous regressors.}
(unrestricted constant)?

The VECM that PSS estimate under Case III (their equation 3.11 on p. 301) is identical apart from notation to that estimated by Johansen in Case 1. This is confirmed on p. 303, where PSS describe the construction of the residuals matrices that go into the calculation of the Case III trace test: these are exactly as Johansen describes for his Case 1 (consistent with our account of the $R_{ij}$ matrices in section 3 above). PSS gloss this by saying that the “the intercept restriction [implied by their VAR formulation] is ignored” in constructing the Case III test statistic (p. 300).

So the difference between Johansen 1 and PSS III does not lie in the calculation of the trace test, rather it lies in the distribution to which this test is referred. The Johansen critical values are determined on the assumption that there really is an unrestricted constant, spilling over into a trend in $X_t$, while PSS critical values are determined on the basis that there is no such trend.

If the data contain no trend—and Turner’s simulated data do not—then testing the null in the wrong model will be much less consequential when PSS III critical values are used than when Johansen 1 values are used; and so we find.

6 Simulations with non-zero drift

A further question of interest is, given a dataset that conforms to the null of $r = 0$ in Johansen’s Case 1, how do the rejection rates differ depending on the set of critical values used, when the test is correctly chosen?

To answer this, simulated data were generated according to

$$y_{i,t} = y_{i,t-1} + \mu_i + \epsilon_{i,t} \quad i = 1, 2$$  \hspace{1cm} (11)

with $\epsilon_1$ and $\epsilon_2$ both NID(0, 1) and the $\mu_i$s non-zero.

Table 3 shows rejection rates for the null of no cointegration on the trace test, for three sets of $\mu$ values and a nominal significance level of 5 percent. The columns headed “Doornik” show rejection rates when the criterion is that the Doornik p-value for Johansen Case 1, adjusted for sample size, is less than 0.05; for those headed “PSS III” the criterion is that the trace test exceeds the 5 percent critical value for Pesaran et al. Case III. In addition, Figure 1 plots rejection rates just using Doornik p-values, with the maximum sample size extended to 10,000.
Focusing first on the results using Doornik p-values, when the $\mu$ (drift) values are fairly substantial we get a test of approximately correct size. But when the drift is very small substantial over-rejection occurs. This is not surprising. As noted above, the asymptotic distribution of the trace test is not invariant with respect to the presence or absence of a trend. With $\mu = (0.05, 0.03)$ a trend is in fact present, but it is so small relative to the variance of the disturbance (1.0) as to be barely statistically detectable. It therefore takes a massive sample to get close to the correct test size.

For reference, taking the log of real quarterly GDP in the USA (series \texttt{gdpc96} from the FRED system at the St Louis Federal Reserve Bank), over the period 1947:2–2010:4, regression of its first difference on a constant produces a drift coefficient of 0.008 as against a standard error of regression of 0.01. That is, if we treat real GDP as a random walk with drift, the drift is of the same order of magnitude as the standard deviation. This corresponds to the maximal $\mu$ used in the simulations.

As for the PSS III critical values, the fact that they give a roughly correct rejection rate for very small $\mu$ and moderate sample size is a matter of luck. For larger $\mu$ and/or large samples, it’s clear that (inappropriate) use of these critical values would lead to consistent under-rejection of the null. We doubt, however, that this would trouble PSS, because the data-generating process in question falls outside of the class of stochastic processes that they regard as proper candidates for cointegration, as represented by the VAR (10) above.

7 Tests using cointegrated data

We have so far concentrated on testing the null of no cointegration. Turner also presents the results of simulations in which he constructs bivariate data that embody a single cointegrating vector, and tests the null of $r = 1$ against the alternative $r = 2$ (stationarity). Again, he finds that if he conducts this test in Johansen model 1 there is substantial over-rejection of the null (worse than in the test of $H_0 : r = 0$ for non-cointegrated data) but if he applies the PSS Case III critical values to the trace test the size is about right.

Here we can be brief: the problem is exactly the same as before. Turner’s cointegrated data contain no trend and so Johansen’s model 0 is appropriate; if the test is conducted in that model it is correctly sized. (This is confirmed in simulations not reported here but available on request.)

Again, however, it is of interest to see what happens if the data do contain drift. To investigate this
we modified Turner’s cointegrated DGP. Given NID series $\epsilon_1$ and $\epsilon_2$, he constructs a random walk, $\Delta u_{1,t} = \epsilon_{1,t}$, and a stationary AR(1) process, $(1 - 0.85L)u_{2,t} = \epsilon_{2,t}$. His observable series (with $r = 1$) are then

\[
y_{1,t} = 2u_{1,t} - u_{2,t}
\]
\[
y_{2,t} = -u_{1,t} + u_{2,t}
\]

We maintain this set-up with the single modification $\Delta u_1 = \epsilon_{1,t} + \mu$.

The results with $\mu = 0.5$ are given in Table 4. We show rejection rates on the trace test for both $r = 1$ (true) and $r = 0$ (false), in Johansen’s model 1 using Doornik p-values (“Joh 1”) and also using PSS Case III critical values. The test for $r = 1$ is correctly sized using p-values for Johansen 1 while the PSS III critical value is surpassed much less than 5 percent of the time. The test for $r = 0$ is interesting: if the sample is large enough the false null is rejected in all trials, but in smaller samples Johansen 1 has greater power than PSS III, rejecting about twice as often.

8 Conclusion

We have shown the following. First, if we take two random walks without drift and apply the trace test for cointegration in Johansen’s model 1 we will over-reject the null roughly two-fold, potentially producing spurious findings of cointegration. But if we test in Johansen’s model 0 our tests will be correctly sized.

Second, if we take two random walks with statistically detectable drift (which depends on both the magnitude of the drift and the sample size) and apply a cointegration test in Johansen’s model 1, we will get a test of correct size. About five percent of the time the realizations of this DGP will mimic cointegration sufficiently closely to reject the null of no cointegration at the five percent level, as seems right and proper. Or at least, it’s right and proper if we accept Johansen’s VAR (equation 9) as representing the class of stochastic processes potentially exhibiting cointegration.

Third, if we follow Pesaran et al. (2000) in delimiting the class of processes that are viable candidates for cointegration more narrowly than Johansen, then we should (as Paul Turner recommends) refer the trace test calculated as per Johansen 1 to its distribution under PSS Case III—provided, of course, that a “lower” (Johansen 0 or 1∗) or “higher” (Johansen 2∗) case is not more appropriate for the data.
In closing, one further point might be noted. We have deliberately held the issue of exogenous regressors out of the foregoing discussion, yet in many practical applications we want to investigate cointegration while controlling for the influence of exogenous factors. In this context the distribution of the trace test differs from the basic Johansen cases discussed above, and moreover differs depending on whether the exogenous terms are $I(0)$ or $I(1)$.\(^{8}\) Now here’s the point most relevant to the topic of this paper: we have mostly focused on Johansen Case 1, but Harbo et al. (1998) show that this case plus $I(1)$ exogenous variables is problematic, as the asymptotic distribution of the trace test cannot be purged of nuisance parameters.\(^{9}\) For that reason they recommend that if the data seem to exhibit trends, testing for cointegration is better done in the larger model $2^\ast$ (constant plus restricted trend). If we were to follow that recommendation generally, puzzles over Case 1 would recede in importance.

References


\(^{8}\)The $I(1)$ case is discussed by Pesaran et al. (2000) and Harbo et al. (1998); for the $I(0)$ case see for example Boswijk and Doornik (2005).

\(^{9}\)This applies whenever the highest-order deterministic term is unrestricted, so it also applies to Johansen Case 2 (unrestricted trend). Not coincidentally, this is another case that bothers Turner.

<table>
<thead>
<tr>
<th>Label</th>
<th>Constant</th>
<th>Trend</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( \mu_0 = 0 )</td>
<td>( \mu_1 = 0 )</td>
</tr>
<tr>
<td>2</td>
<td>1*</td>
<td>( \alpha \perp' \mu_0 = 0 )</td>
<td>( \mu_1 = 0 )</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>( \mu_0 \neq 0 )</td>
<td>( \mu_1 = 0 )</td>
</tr>
<tr>
<td>4</td>
<td>2*</td>
<td>( \mu_0 \neq 0 )</td>
<td>( \alpha \perp' \mu_1 = 0 )</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>( \mu_0 \neq 0 )</td>
<td>( \mu_1 \neq 0 )</td>
</tr>
</tbody>
</table>

Table 1: Johansen's five cases
<table>
<thead>
<tr>
<th>$T$</th>
<th>Unrestricted const.</th>
<th>No constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>Doornik 0.1103</td>
<td>PSS III 0.0552</td>
</tr>
<tr>
<td>100</td>
<td>Doornik 0.1129</td>
<td>PSS III 0.0529</td>
</tr>
<tr>
<td>500</td>
<td>Doornik 0.1149</td>
<td>PSS III 0.0508</td>
</tr>
<tr>
<td>1000</td>
<td>Doornik 0.1134</td>
<td>PSS III 0.0495</td>
</tr>
<tr>
<td>2000</td>
<td>Doornik 0.1164</td>
<td>PSS III 0.0507</td>
</tr>
<tr>
<td>5000</td>
<td>Doornik 0.1150</td>
<td>PSS III 0.0508</td>
</tr>
</tbody>
</table>

Table 2: Rejection rates, given data with no drift ($N = 50,000$)
\[
\begin{array}{cccccc}
T & \mu = (0.05, 0.03) & \mu = (0.50, 0.30) & \mu = (1.00, 0.60) \\
50 & 0.107 & 0.053 & 0.060 & 0.028 & 0.052 & 0.025 \\
100 & 0.109 & 0.049 & 0.054 & 0.025 & 0.053 & 0.024 \\
500 & 0.094 & 0.041 & 0.052 & 0.021 & 0.051 & 0.022 \\
1000 & 0.082 & 0.035 & 0.049 & 0.021 & 0.051 & 0.022 \\
2000 & 0.069 & 0.028 & 0.053 & 0.022 & 0.050 & 0.021 \\
5000 & 0.057 & 0.022 & 0.052 & 0.022 & 0.050 & 0.021 \\
\end{array}
\]

Table 3: Empirical dependence of rejection rates on $\mu$ in Johansen Case 1 ($N = 50,000$)
Figure 1: Rejection rates against sample size, Johansen Case 1 with Doornik p-values (N = 50,000)
<table>
<thead>
<tr>
<th>$T$</th>
<th>$H_0: r = 1$ Joh 1</th>
<th>$H_0: r = 1$ Joh 1</th>
<th>$H_0: r = 0$ PSS III</th>
<th>$H_0: r = 0$ PSS III</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.050</td>
<td>0.002</td>
<td>0.108</td>
<td>0.052</td>
</tr>
<tr>
<td>100</td>
<td>0.049</td>
<td>0.003</td>
<td>0.280</td>
<td>0.146</td>
</tr>
<tr>
<td>500</td>
<td>0.051</td>
<td>0.004</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>1000</td>
<td>0.051</td>
<td>0.004</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>2000</td>
<td>0.050</td>
<td>0.004</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4: Rejection rates, bivariate data with $r = 1$ and non-zero drift ($N = 50,000$)