

Article

## A Jackknife Correction to a Test for Cointegration Rank

Marcus J. Chambers

Department of Economics, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK;  
E-Mail: mchamb@essex.ac.uk; Tel.: +44-1206-872756; Fax: +44-1206-872724

Academic Editor: Kerry Patterson

Received: 24 July 2014 / Accepted: 15 May 2015 / Published: 20 May 2015

---

**Abstract:** This paper investigates the performance of a jackknife correction to a test for cointegration rank in a vector autoregressive system. The limiting distributions of the jackknife-corrected statistics are derived and the critical values of these distributions are tabulated. Based on these critical values the finite sample size and power properties of the jackknife-corrected tests are compared with the usual rank test statistic as well as statistics involving a small sample correction and a Bartlett correction, in addition to a bootstrap method. The simulations reveal that all of the corrected tests can provide finite sample size improvements, while maintaining power, although the bootstrap procedure is the most robust across the simulation designs considered.

**Keywords:** jackknife correction; bias reduction; cointegration rank test

**JEL classifications:** C12; C32

---

### 1. Introduction

The concept of cointegration has assumed a prominent role in the analysis of economic and financial time series since the pioneering work of Engle and Granger [1], and tests for the cointegration rank of a vector of time series have become an essential part of the applied econometrician's toolkit. The most popular test for cointegration rank is the trace statistic proposed by Johansen [2,3] which exploits the reduced rank regression techniques of Anderson [4] in the context of a vector autoregressive (VAR) model. The limiting distribution of the test statistic can be expressed as a functional of a vector Brownian motion process, the dimension of which depends upon the difference between the number of variables under consideration and the cointegration rank under the null hypothesis. Percentage points of the

limiting distribution have been tabulated by simulation and can be found, for example, in Johansen [5], Doornik [6] and MacKinnon, Haug and Michelis [7].

The accuracy of the limiting distribution as a description of the finite sample distribution has been examined in a number of studies. Toda [8,9] found that the performance of the tests is dependent on the value of the stationary roots of the process, and that a sample of 100 observations is insufficient to detect the true cointegrating rank when the stationary root is close to one (0.8 or above). Doornik [6] proposed Gamma distribution approximations to the limiting distributions, finding that they are more accurate than previously published tables of critical values, while Nielsen [10] used local asymptotic theory to improve the ability of the limiting distribution to act as an approximation in finite samples.

In view of the experimental evidence reported above, there have been a number of further attempts to improve inference in finite samples when using the asymptotic critical values for the trace test of cointegration rank. Johansen [11] demonstrated how Bartlett corrections can be made to the statistic; these rely on various asymptotic expansions of the statistic's expectation and result in complicated functions of the model's parameters that can be estimated from the sample data. A simpler small sample correction factor was suggested by Reinsel and Ahn [12] and involves a degrees-of-freedom type of adjustment to the sample size when calculating the value of the test statistic. This small sample correction was shown to work well by Reimers [13] although, as Johansen [5] (p. 99) notes, the "theoretical justification for this result presents a very difficult mathematical problem." More computationally intensive bootstrap procedures have recently been advocated by, *inter alia*, Swensen [14] and Cavaliere, Rahbek and Taylor [15].

The aim of this paper is to analyse the properties of a simple jackknife-corrected test statistic for cointegration rank. The approach is far less demanding, computationally, than bootstrap methods and does not require an explicit analytical derivation of an asymptotic expansion for the sample moment as in the Bartlett approach, merely relying on its existence. The idea behind the jackknife is to combine the statistic based on the full sample of observations with statistics based on a set of sub-samples, the weights used in the linear combination being chosen so that the leading term in the bias expansion is eliminated. Although intended primarily as a method of bias reduction in parameter estimation following the work of Quenouille [16] and Tukey [17], the jackknife can equally be applied to test statistics in order to achieve the same type of outcome as the Bartlett correction. In the case of stationary autoregressive time series Chambers [18] has shown that jackknife methods are capable of producing substantial bias reductions as well as reductions in root mean square errors compared to other methods in the estimation of model parameters; the jackknife results were also shown to be robust to departures from normality and conditional heteroskedasticity as well as other types of misspecification. However, some care has to be taken when applying these techniques with non-stationary data, as pointed out by Chambers and Kyriacou [19,20], who propose methods that can be used to ensure that the jackknife procedure achieves the bias reduction as intended in the case of non-stationarity.

The paper is organised as follows. Section 2 begins by defining the model as well as the test statistic of interest. Three variants of the model are considered, corresponding to different specifications of the deterministic linear trend, although most attention is given to the two variants of greatest empirical interest. The jackknife-corrected version of the statistic is defined and the appropriate limiting distributions are derived and presented in Theorem 1. Section 3 is devoted to simulation results and

is divided into two subsections. The first sub-section computes (asymptotic) critical values of the limiting distributions of the jackknife-corrected statistics. The second sub-section is concerned with the finite sample size and power properties of the jackknife-corrected statistics, which are compared with the unadjusted statistic as well as the small-sample adjustment of Reinsel and Ahn [12], the Bartlett correction of Johansen [11] and the bootstrap approach of Cavaliere, Rahbek and Taylor [15]. Section 4 concludes and discusses some directions for future research, while an Appendix provides a proof of Theorem 1 and some details on the simulation method used to derive the critical values in Section 3.1.

The following notation will be used. For a  $p \times r$  matrix  $C$  of rank  $r < p$  there exists a full rank  $p \times (p - r)$  matrix  $C_{\perp}$  satisfying  $C'_{\perp}C = 0$ ;  $P_C = C(C'C)^{-1}C'$  denotes the projection matrix of  $C$ ; and  $\|C\| = (\sum_{i=1}^p \sum_{j=1}^r c_{ij}^2)^{1/2}$  denotes the Euclidean norm of  $C$ . In addition, for a square matrix  $A$ ,  $\det[A]$  denotes the determinant, and  $I_p$  denotes the  $p \times p$  identity matrix. The symbol  $\stackrel{d}{=}$  denotes equality in distribution;  $\stackrel{d}{\rightarrow}$  denotes convergence in distribution;  $\stackrel{P}{\rightarrow}$  denotes convergence in probability;  $\Rightarrow$  denotes weak convergence of the relevant probability measures;  $L$  denotes the lag operator such that  $L^j y_t = y_{t-j}$  for some integer  $j$  and variable  $y_t$ ; and  $B(s)$  denotes a standard vector Brownian motion process. Functionals of  $B(s)$ , such as  $\int_0^1 B(s)B(s)'ds$ , are denoted  $\int_0^1 BB'$  for notational convenience.

## 2. The Model and Tests for Cointegration Rank

Following Johansen [5] the model under consideration is the following VAR( $k$ ) system in the  $p \times 1$  vector  $y_t$ :

$$y_t = \sum_{i=1}^k \Pi_i y_{t-i} + \Phi D_t + \epsilon_t, \quad t = 1, \dots, T \tag{1}$$

where  $\epsilon_1, \dots, \epsilon_T$  are independent and identically distributed  $p \times 1$  random vectors with mean vector zero and positive definite covariance matrix  $\Omega$ ,  $D_t$  is a  $q \times 1$  vector of deterministic terms,  $\Pi_1, \dots, \Pi_k$  are  $p \times p$  matrices of autoregressive coefficients,  $\Phi$  is a  $p \times q$  matrix of coefficients on the deterministic terms, and the initial values,  $y_{-k+1}, \dots, y_0$ , are assumed to be fixed. It is convenient, in the analysis of cointegration, to write Equation (1) in the vector error correction model (VECM) form

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \Phi D_t + \epsilon_t, \quad t = 1, \dots, T \tag{2}$$

where  $\Pi = \sum_{i=1}^k \Pi_i - I_p$  and  $\Gamma_i = -\sum_{j=i+1}^k \Pi_j$  ( $i = 1, \dots, k - 1$ ). The following assumption will be made.

**Assumption 1.** (i) Let  $A(z) = (1 - z)I_p - \Pi z - \sum_{i=1}^{k-1} \Gamma_i (1 - z)z^i$ . Then  $\det[A(z)] = 0$  implies that  $|z| > 1$  or  $z = 1$ ; (ii) The matrix  $\Pi$  has rank  $r < p$  and has the representation  $\Pi = \alpha\beta'$  where  $\alpha$  and  $\beta$  are  $p \times r$  with rank  $r$ ; (iii) The matrix  $\alpha'_{\perp} \Gamma \beta_{\perp}$  is nonsingular, where  $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$ .

This assumption is common in the cointegration literature and implies (see Theorem 4.2 of Johansen [5], for example) that  $y_t$  has the representation

$$y_t = C \sum_{i=1}^t (\epsilon_i + \Phi D_i) + C(L) (\epsilon_t + \Phi D_t) + P_{\beta_{\perp}} y_0, \quad t = 1, \dots, T \tag{3}$$

where  $C(z) = (1 - z)A(z)^{-1} = \sum_{i=1}^{\infty} C_i z^i$  and  $C = C(1) = \beta_{\perp} (\alpha'_{\perp} \Gamma \beta_{\perp})^{-1} \alpha'_{\perp}$ . This representation is convenient because it decomposes  $y_t$  into a stochastic trend component, a stationary component, and a

term that depends on the initial condition  $y_0$ . Assumption 1 also ensures that, although the vector  $y_t$  is  $I(1)$ , both  $\Delta y_t - E(\Delta y_t)$  and  $\beta' y_t - E(\beta' y_t)$  can be given initial distributions such that they are  $I(0)$ . A proof of Equation (3) and a discussion of its implications can be found in Johansen [5] (pp. 49–52).

The specification of the deterministic component  $\Phi D_t$  in Equation (2) has implications not only for the interpretation of the error correction model but also for the level process  $y_t$  itself. The following three leading cases have received most attention in the literature:

Case 1: no deterministic components. In this case  $\Phi D_t = 0$  and so, from Equations (2) and (3),

$$\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \epsilon_t$$

$$y_t = C \sum_{i=1}^t \epsilon_i + C(L)\epsilon_t + P_{\beta \perp} y_0$$

Case 2: restricted intercept. Here  $\Phi D_t = \alpha \rho_0$ , where  $\rho_0$  is  $r \times 1$ , and hence

$$\Delta y_t = \alpha (\beta' y_{t-1} + \rho_0) + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \epsilon_t$$

$$y_t = C \sum_{i=1}^t \epsilon_i + C(L)\epsilon_t + \tau_0 + P_{\beta \perp} y_0$$

where  $\tau_0$  is a vector of intercepts.

Case 3: restricted linear trend. In this specification  $\Phi D_t = \mu_0 + \alpha \rho_1 t$ , where  $\mu_0$  and  $\rho_1$  are  $p \times 1$  and  $r \times 1$  vectors respectively. It follows that

$$\Delta y_t = \alpha (\beta' y_{t-1} + \rho_1 t) + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \mu_0 + \epsilon_t$$

$$y_t = C \sum_{i=1}^t \epsilon_i + C(L)\epsilon_t + \tau_0 + \tau_1 t + P_{\beta \perp} y_0$$

where  $\tau_0$  and  $\tau_1$  are vectors of constants.

Hence in case 1 there are no deterministic terms at all, in case 2 the intercept is restricted to the cointegrating relationships, while in case 3 there is an unrestricted intercept and the time trend only enters through the cointegrating relationships although a linear trend is present in the levels representation.

The trace statistic has become a popular method of testing the null hypothesis of  $r < p$  cointegrating vectors against the maintained hypothesis that  $\Pi$  has full column rank  $p$ , in which case the process  $y_t$  is stationary. In what follows it is convenient to further express the VECM Equation (2) in the form

$$Z_{0t} = \alpha \beta^{*'} Z_{1t} + \Psi Z_{2t} + \epsilon_t, \quad t = 1, \dots, T \tag{4}$$

where  $Z_{0t} = \Delta y_t$  and the remaining terms are defined with respect to the three cases concerning the specification of the deterministic component  $\Phi D_t$  defined above:

Case 1:  $\beta^* = \beta$ ,  $Z_{1t} = y_{t-1}$ ,  $Z_{2t} = (\Delta y'_{t-1}, \dots, \Delta y'_{t-k})'$ ,  $\Psi = [\Gamma_1, \dots, \Gamma_{k-1}]$ .

Case 2:  $\beta^{*'} = (\beta', \rho_0)$ ,  $Z_{1t} = (y'_{t-1}, 1)'$ , and  $\Psi$  and  $Z_{2t}$  are defined as in case 1.

Case 3:  $\beta^{*'} = (\beta', \rho_1)$ ,  $Z_{1t} = (y'_{t-1}, t)'$ ,  $Z_{2t} = (\Delta y'_{t-1}, \dots, \Delta y'_{t-k}, 1)'$ ,  $\Psi = [\Gamma_1, \dots, \Gamma_{k-1}, \mu_0]$ .

Based on Equation (4) it is possible to define the matrices

$$M_{ij} = T^{-1} \sum_{t=1}^T Z_{it}Z'_{jt}, \quad (i, j = 0, 1, 2), \quad S_{ij} = M_{ij} - M_{i2}M_{22}^{-1}M_{2j}, \quad (i, j = 0, 1)$$

The trace statistic is then obtained as

$$S_r = -T \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i) \tag{5}$$

where the (ordered) eigenvalues  $1 > \hat{\lambda}_1 > \dots > \hat{\lambda}_p > 0$  solve the determinantal equation  $\det[\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}] = 0$ . Let  $B(s)$  denote a  $(p - r)$ -dimensional standard Brownian motion process. Then, as  $T \rightarrow \infty$ ,

$$S_r \Rightarrow \text{trace} \left\{ \int_0^1 dB F' \left( \int_0^1 F F' \right)^{-1} \int_0^1 F dB' \right\} \tag{6}$$

where the stochastic process  $F(s)$  is defined for each case as follows:

$$\begin{aligned} \text{Case 1: } & F(s) = B(s); \\ \text{Case 2: } & F(s) = (B(s)', 1)'; \\ \text{Case 3: } & F(s) = \left[ \left( B(s) - \int_0^1 B \right)', s - \frac{1}{2} \right]' \end{aligned} \tag{7}$$

A proof of Equation (6) can be found in Theorem 11.1 of Johansen [5].

The asymptotic distribution given in Equation (6) has been found to provide a poor approximation to the finite sample distribution of  $S_r$  in a number of cases leading to alternative approaches being developed. Recent work has suggested bootstrap techniques as a way of improving inference concerning the cointegration rank in finite samples, for example Swensen [14] and Cavaliere, Rahbek and Taylor [15], while Bartlett corrections have been proposed by Johansen [11]. The idea behind Bartlett correction is to adjust the statistic  $S_r$  so that its finite sample distribution is closer to the limiting distribution. To see how this works, suppose it is possible to expand the expectation of  $S_r$  as

$$E(S_r) = a_0 + \frac{a_1}{T} + O(T^{-2}) \tag{8}$$

where  $a_0$  denotes the limit of the expectation as  $T \rightarrow \infty$  and  $a_1$  is either a known constant (typically a function of the model parameters) or can be estimated consistently from the sample data. Then the adjusted statistic

$$S_r^B = \left( 1 + \frac{a_1}{a_0 T} \right)^{-1} S_r \tag{9}$$

can be shown to satisfy  $E(S_r^B) = a_0 + O(T^{-2})$ , thereby improving the accuracy of the limiting distribution as an approximation to the finite sample distribution (at least in terms of the mean of the

distribution). Johansen [11] derives expressions for the Bartlett correction factor which depend on the parameters of the model in a complicated way and, hence, must be estimated from the sample data. Simulations reveal that the adjusted statistic improves the size of the cointegration rank test in a large part of the parameter space.

A procedure related to the Bartlett approach is the jackknife. It achieves the same order of reduction in the bias of the statistic but does not require a formal derivation of the precise terms in the expansion in Equation (8), merely relying on the existence of such a representation. The idea is to combine the statistic  $S_r$  in a linear combination with the mean of a set of  $m$  statistics obtained from sub-samples, the weights being chosen to eliminate the first order bias term  $a_1/T$ . Suppose, then, that the full sample of observations is divided into  $m$  non-overlapping sub-samples, each sub-sample containing  $\ell = T/m$  observations. If  $S_{rj}$  denotes the statistic computed from sub-sample  $j$  then the jackknife statistic is defined by

$$S_{r,m}^J = \frac{m}{m-1} S_r - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^m S_{rj} \tag{10}$$

Provided that  $E(S_{rj}) = a_0 + a_1\ell^{-1} + O(\ell^{-2})$  for  $j = 1, \dots, m$  and  $\ell = O(T)$ , it is straightforward to show that  $E(S_{r,m}^J) = a_0 + O(T^{-2})$  by substitution of the relevant expressions. Hence both statistics,  $S_r^B$  and  $S_{r,m}^J$ , achieve the same order of bias reduction but by different means.

Although valid in cases 2 and 3, the above argument concerning the jackknife statistic  $S_{r,m}^J$  falters in case 1 on the assumption that the sub-sample statistics,  $S_{rj}$  ( $j = 1, \dots, m$ ), all share the same expansion in Equation (8) as the full-sample statistic  $S_r$ . It has been shown by Chambers and Kyriacou [19,20] that, in a univariate setting with a unit root, the sub-sample statistics (for  $j \neq 1$ ) have different properties to the full sample statistic in the limit due to the stochastic order of magnitude of the pre-sub-sample value, and the same phenomenon also arises here in case 1. The implication is that the limiting distributions of the sub-sample statistics  $S_{rj}$  (at least for  $j \neq 1$ ) will differ from that of  $S_r$ ; the expansions of  $E(S_{rj})$  will differ from that of  $E(S_r)$ ; and, hence, the jackknife statistic (as defined above) will not fully eliminate the  $O(T^{-1})$  term in the bias. These problems do not arise in cases 2 and 3 because the presence of the intercept and/or time trend ensures that the distributions are invariant to the initial (pre-sub-sample) conditions. Although it is possible to overcome these problems in case 1 by simply subtracting  $y_{(j-1)\ell}$  from the observations in sub-sample  $j$ —an idea proposed in the univariate unit root setting by Chambers and Kyriacou [19]—we do not pursue this avenue any further in view of the limited applicability of case 1 in practice. Instead, we focus on the application of the jackknife correction in the more empirically relevant cases 2 and 3.

In order to economise on notation it is convenient to define the functional

$$Q(U, V, \delta) = \text{trace} \left\{ \int_{\delta} dUV' \left( \int_{\delta} VV' \right)^{-1} \int_{\delta} VdU' \right\}$$

where  $U(s)$  and  $V(s)$  are vector stochastic processes defined on  $s \in \delta$ , and to define the intervals  $\delta_0 = [0, 1]$  and  $\delta_{j,m} = [(j-1)/m, j/m]$  ( $j = 1, \dots, m$ ). With this notation the limiting distribution in Equation (6), for example, can be represented as

$$S_r \Rightarrow Q(B, F, \delta_0)$$

The formal statement of the main result is as follows.

**Theorem 1.** Let  $y_1, \dots, y_T$  be generated according to Equation (2) and let Assumption 1 hold. Then, as  $T \rightarrow \infty$ :

(a) If  $m$  is fixed:

Case 2.  $S_{rj} \Rightarrow Q(B, F_2, \delta_{j,m})$  ( $j = 1, \dots, m$ ), and

$$S_{r,m}^J \Rightarrow \frac{m}{m-1} Q(B, F_2, \delta_0) - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^m Q(B, F_2, \delta_{j,m})$$

where  $F_2(s) = [B(s)', 1]'$ . Furthermore,  $Q(B, F_2, \delta_{j,m})$  and  $Q(B, F_2, \delta_{k,m})$  are independent for  $j \neq k$ .

Case 3.  $S_{rj} \Rightarrow Q(B, F_{j,m}, \delta_{j,m})$  ( $j = 1, \dots, m$ ), and

$$S_{r,m}^J \Rightarrow \frac{m}{m-1} Q(B, F_3, \delta_0) - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^m Q(B, F_{j,m}, \delta_{j,m})$$

where

$$F_3(s) = \left[ \left( B(s) - \int_0^1 B \right)', s - \frac{1}{2} \right]'$$

and

$$F_{j,m}(s) = \left[ \left( B(s) - m \int_{(j-1)/m}^{j/m} B(s) ds \right)', s - \frac{j - \frac{1}{2}}{m} \right]', \quad j = 1, \dots, m$$

Furthermore,  $Q(B, F_{j,m}, \delta_{j,m})$  and  $Q(B, F_{k,m}, \delta_{k,m})$  are independent for  $j \neq k$ .

(b) If  $m^{-1} + mT^{-1} \rightarrow 0$ :

Case 2.  $S_{r,m}^J \Rightarrow Q(B, F_2, \delta_0)$ .

Case 3.  $S_{r,m}^J \Rightarrow Q(B, F_3, \delta_0)$ .

Theorem 1(a) shows that, when the number of sub-samples  $m$  is fixed, the limiting distribution of the jackknife-corrected statistic is the same linear combination of the limiting distributions of the full- and sub-sample statistics. Note that the length of each sub-sample,  $\ell = T/m$ , increases with  $T$  for fixed  $m$ . However, when  $m$  is allowed to increase with  $T$  but at a slower rate, Theorem 1(b) shows that the limiting distribution is equivalent to that of the full-sample statistic alone. In this case note that, if  $mT^{-1} \rightarrow 0$  as  $T \rightarrow \infty$ , then  $\ell = m^{-1}T \rightarrow \infty$ . In order to use these distributions for inference it is necessary to obtain the appropriate critical values; these are provided in Section 3 for a range of values of  $m$  and  $p - r$ . Further analysis of the finite sample properties of the jackknife-corrected statistics is provided in the next section by means of Monte Carlo simulations.

### 3. Simulation Results

#### 3.1. Critical Values of Limiting Distributions

The limiting distributions of the jackknife statistic  $S_{r,m}^J$  presented in Theorem 1 for fixed values of the jackknife parameter  $m$  are nonstandard and, in order to be useful in practice, it is necessary to obtain the appropriate critical values. Tables 1 and 2 provide the 90%, 95% and 99% points of the distributions for values of  $p-r$  ranging from 1 to 12 for the empirically relevant cases 2 and 3, respectively, and for values of  $m \in \{2, 3, 4, 5, 6, 8, 10, 12, 16, 20\}$ . A total of 100,000 replications were carried out with a sample size of  $T = \max\{1200, 100m\}$  spanning the interval  $[0, 1]$  ensuring that when  $m \geq 12$  each sub-sample (of length  $\ell = T/m$ ) contains 100 points. The method described in Chapter 15 of Johansen [5] was employed, with suitable modifications to allow for the sub-sample statistics used in constructing the jackknife corrections; details are provided in the Appendix. It can be seen that, in all cases, the critical values for  $S_{r,m}^J$  are larger than those for  $S_r$  that are reported in, for example, Johansen [5], Doornik [6] and MacKinnon, Haug and Michelis [7]. The critical values are also seen to decrease as  $m$  increases for a given value of  $p-r$ .

**Table 1.** Percentage points of limiting distributions of  $S_{r,m}^J$ : case 2.

$m :$	2	3	4	5	6	8	10	12	16	20
$p-r$	90%									
1	10.05	9.08	8.66	8.42	8.26	8.06	7.96	7.88	7.80	7.75
2	22.25	20.50	19.76	19.38	19.11	18.78	18.62	18.50	18.36	18.28
3	38.21	35.81	34.79	34.22	33.88	33.43	33.20	33.04	32.85	32.74
4	58.09	54.96	53.67	52.98	52.54	51.98	51.66	51.46	51.20	51.06
5	82.03	78.19	76.62	75.75	75.19	74.55	74.18	73.92	73.61	73.43
6	109.89	105.40	103.53	102.51	101.85	101.06	100.61	100.32	99.97	99.75
7	141.58	136.48	134.26	133.11	132.40	131.53	130.98	130.65	130.24	129.98
8	177.55	171.72	169.24	167.89	167.05	165.99	165.43	165.06	164.59	164.30
9	217.22	210.62	207.94	206.44	205.49	204.39	203.73	203.29	202.75	202.41
10	260.95	253.74	250.84	249.23	248.23	246.90	246.18	245.69	245.07	244.68
11	308.49	300.83	297.62	295.80	294.65	293.28	292.50	291.92	291.25	290.82
12	360.20	351.68	348.22	346.20	345.04	343.50	342.64	342.07	341.29	340.80
$p-r$	95%									
1	12.56	11.26	10.68	10.35	10.14	9.87	9.71	9.62	9.50	9.43
2	25.89	23.65	22.74	22.18	21.82	21.38	21.16	20.98	20.81	20.69
3	42.93	39.85	38.50	37.74	37.29	36.71	36.38	36.17	35.91	35.75
4	63.91	59.93	58.27	57.41	56.83	56.08	55.69	55.40	55.05	54.87
5	89.01	84.13	82.07	80.92	80.19	79.32	78.78	78.46	78.05	77.82
6	117.86	112.19	109.73	108.37	107.55	106.49	105.92	105.53	105.06	104.80
7	150.83	144.14	141.37	139.81	138.84	137.70	137.02	136.57	136.03	135.69
8	187.76	180.10	177.07	175.32	174.21	172.93	172.21	171.70	171.08	170.71



Table 1. Cont.

<i>m</i> :	2	3	4	5	6	8	10	12	16	20
9	228.63	220.14	216.60	214.86	213.53	212.13	211.23	210.73	210.02	209.60
10	273.43	264.11	260.24	258.12	256.76	255.20	254.19	253.58	252.77	252.32
11	321.89	311.80	307.72	305.46	303.90	302.13	301.13	300.43	299.55	299.01
12	374.52	363.51	359.00	356.62	355.07	353.12	351.97	351.20	350.28	349.67
<i>p</i> - <i>r</i>	99%									
1	17.99	16.02	15.21	14.64	14.30	13.90	13.65	13.48	13.28	13.15
2	33.52	30.41	28.92	28.17	27.67	26.97	26.62	26.36	26.08	25.90
3	52.54	48.08	46.04	45.00	44.30	43.44	43.02	42.65	42.23	42.02
4	75.56	69.88	67.45	66.17	65.22	64.23	63.61	63.20	62.70	62.43
5	102.65	95.74	92.60	91.02	89.99	88.76	88.05	87.54	86.95	86.58
6	133.52	125.28	121.70	119.75	118.60	117.10	116.25	115.73	115.05	114.65
7	168.40	159.21	155.18	152.98	151.64	149.90	148.84	148.21	147.48	146.98
8	207.38	196.97	192.68	190.16	188.48	186.65	185.54	184.77	183.87	183.36
9	249.98	238.76	233.74	231.05	229.56	227.31	225.92	225.21	224.21	223.54
10	296.32	283.65	278.30	275.06	273.38	271.14	269.66	268.83	267.75	267.06
11	347.61	334.18	327.61	324.53	322.64	320.06	318.46	317.51	316.24	315.52
12	402.54	387.33	380.59	377.15	374.75	372.10	370.54	369.34	368.02	367.27

Table 2. Percentage points of limiting distributions of  $S_{r,m}^J$ : case 3.

<i>m</i> :	2	3	4	5	6	8	10	12	16	20
<i>p</i> - <i>r</i>	90%									
1	18.76	14.91	13.51	12.79	12.35	11.85	11.57	11.39	11.18	11.05
2	35.82	29.69	27.56	26.49	25.85	25.10	24.68	24.41	24.10	23.92
3	56.12	47.98	45.28	43.88	43.04	42.07	41.55	41.20	40.79	40.55
4	80.08	70.16	66.83	65.15	64.14	62.95	62.28	61.85	61.32	61.02
5	107.99	96.36	92.42	90.42	89.22	87.82	87.03	86.52	85.92	85.56
6	139.70	126.38	121.86	119.54	118.11	116.50	115.59	114.99	114.27	113.85
7	175.42	160.26	155.23	152.63	151.08	149.19	148.17	147.49	146.67	146.18
8	215.34	198.49	192.84	189.80	188.06	185.95	184.79	184.05	183.11	182.56
9	258.79	240.45	234.13	230.83	228.88	226.55	225.27	224.43	223.37	222.73
10	306.55	286.65	279.68	276.15	273.91	271.35	269.92	268.96	267.81	267.10
11	358.30	336.52	329.07	325.17	322.77	319.95	318.35	317.36	316.08	315.31
12	413.57	390.29	382.24	377.98	375.45	372.47	370.74	369.57	368.15	367.31
<i>p</i> - <i>r</i>	95%									
1	22.34	17.62	15.94	15.09	14.56	13.95	13.61	13.39	13.13	12.99
2	40.58	33.37	30.91	29.64	28.88	28.00	27.50	27.18	26.81	26.59
3	61.90	52.53	49.38	47.76	46.80	45.67	45.05	44.63	44.13	43.85

Table 2. Cont.

<i>m</i> :	2	3	4	5	6	8	10	12	16	20
4	86.92	75.52	71.70	69.72	68.56	67.21	66.43	65.93	65.33	64.98
5	115.89	102.49	98.05	95.75	94.35	92.70	91.79	91.17	90.45	90.03
6	148.70	133.59	128.43	125.74	124.12	122.26	121.16	120.47	119.65	119.17
7	185.48	168.61	162.76	159.67	157.88	155.67	154.43	153.64	152.69	152.11
8	226.37	207.59	201.08	197.68	195.59	193.16	191.79	190.91	189.85	189.20
9	270.96	250.37	243.27	239.51	237.30	234.62	233.07	232.11	230.93	230.18
10	320.06	297.55	289.65	285.71	283.12	280.20	278.50	277.40	276.05	275.23
11	372.32	347.92	339.37	334.93	332.33	329.16	327.29	326.12	324.66	323.74
12	428.72	402.54	393.18	388.61	385.63	382.25	380.20	378.93	377.35	376.35
<i>p</i> - <i>r</i>	99%									
1	30.24	23.55	21.28	20.10	19.36	18.50	18.02	17.73	17.38	17.16
2	50.52	40.95	37.71	36.13	35.12	33.99	33.31	32.89	32.40	32.10
3	73.77	61.94	57.89	55.85	54.50	53.10	52.28	51.75	51.14	50.75
4	100.20	86.44	81.56	79.09	77.66	75.85	74.89	74.23	73.47	72.99
5	131.58	115.32	109.74	106.87	105.06	102.93	101.78	100.97	100.02	99.55
6	166.69	147.92	141.48	138.09	136.03	133.67	132.33	131.43	130.33	129.73
7	205.79	184.90	177.73	173.88	171.62	168.78	167.29	166.32	165.11	164.36
8	248.25	225.64	217.36	213.17	210.70	207.55	205.74	204.69	203.27	202.46
9	294.75	269.95	261.04	256.53	253.66	250.36	248.36	247.22	245.70	244.75
10	344.98	317.65	308.29	303.22	300.04	296.43	294.31	293.04	291.35	290.38
11	399.79	370.35	359.97	354.38	351.15	347.01	344.72	343.19	341.39	340.35
12	458.22	426.77	415.51	409.54	405.96	401.61	399.07	397.46	395.48	394.31

### 3.2. Finite Sample Properties

The finite sample size and power properties of the jackknife-corrected test statistics were investigated using the simulation model adopted by Cavaliere, Rahbek and Taylor [15] (denoted CRT12) for the purpose of evaluating their bootstrap procedure. The model takes  $p = 4$  and is given by

$$\Delta y_t = \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \epsilon_t, \quad t = 1, \dots, T$$

where  $\epsilon_t$  is a vector of normally distributed independent random variables with covariance matrix  $I_4$ , the sample size  $T \in \{50, 100, 200\}$ , and the initial condition is  $y_0 = \Delta y_0 = 0$ . The short-run adjustment matrices are defined as  $\alpha = (a, 0, 0, 0)'$  and

$$\Gamma_1 = \begin{pmatrix} \gamma & \delta & 0 & 0 \\ \delta & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma \end{pmatrix}$$

where  $a$ ,  $\gamma$  and  $\delta$  are scalar parameters defined below for each of the three data generation processes (DGPs) considered:

DGP1:  $a = -0.4$ ,  $\beta = (1, 0, 0, 0)'$ ,  $\gamma = 0.8$ ,  $\delta \in \{0, 0.2\}$ .

DGP2:  $a = -0.4$ ,  $\beta = (1, 0, 0, 0)'$ ,  $\gamma = 0.5$ ,  $\delta \in \{0, 0.2\}$ .

DGP3:  $a = 0$ ,  $\delta = 0$ ,  $\gamma \in \{0, 0.5, 0.8, 0.9\}$ .

In DGPs 1 and 2 there is a single cointegrating vector while in DGP3 there is no cointegration and  $y_t$  is an I(1) VAR(2) process (or, equivalently,  $\Delta y_t$  is an I(0) VAR(1) process). The form of cointegration in DGPs 1 and 2 was considered in CRT12 and implies that  $y_{1t}$  is I(0). The value of  $\delta$  that appears in the matrix  $\Gamma_1$  was used by CRT12 because it is related to some auxiliary conditions relevant for the bootstrap procedure of Swensen [21]. These conditions are satisfied for  $\delta = 0$  but not for  $\delta = 0.2$ . CRT12 also included two additional values of  $\delta$  (equal to 0.1 and 0.3) but, as will be seen, the value of  $\delta$  does not have a major impact on the test procedures under consideration and so we restrict attention to just two of the four values used in CRT12. Note that, in DGP3,  $\delta = 0$  and the matrix  $\Gamma_1$  is diagonal with the scalar  $\gamma$  forming the diagonal elements.

It is also necessary for the DGPs to satisfy the three parts of Assumption 1. The first requires the roots of the equation  $\det[A(z)] = 0$  to have modulus greater than or equal to one, where in this case  $A(z) = (1 - z)I_4 - \alpha\beta'z - \Gamma_1(1 - z)z$ . In DGPs 1 and 2 there are three unit roots and in DGP3 there are four; the moduli of the non-unit roots are reported in Table 3, where it can be seen that Assumption 1(i) is satisfied in all cases. Comparing DGP1 with DGP2, the effect of reducing  $\gamma$  from 0.8 to 0.5 is to increase the modulus of each of the non-unit roots. In DGP3, increasing the parameter  $\gamma$  reduces the non-unit roots towards unity, and in the extreme case of  $\gamma = 1$ ,  $y_t$  becomes an I(2) process. It is well known that the rank test performs poorly as this extreme case is approached; see, for example, the simulation evidence in Johansen (2002). Note that, when  $\gamma = 0$ , there are no roots in addition to the four unit roots because, in this case,  $A(z) = (1 - z)I_4$  and hence  $\det[A(z)] = (1 - z)^4$ . Assumption 1(ii) is obviously satisfied, while it can be shown that  $\det[\alpha'_\perp \Gamma \beta_\perp] = (1 - \gamma)^3$  and hence Assumption 1(iii) is satisfied provided  $\gamma \neq 1$ .

A total of seven test statistics for cointegration rank were considered. The first is the standard (unadjusted) trace statistic  $S_r$  defined in Equation (5). The second uses the small sample correction proposed by Reinsel and Ahn [12]; the resulting statistic, denoted  $S_r^{RA}$ , is defined by

$$S_r^{RA} = -(T - pk) \sum_{i=r+1}^p \log(1 - \hat{\lambda}_i) = \frac{(T - pk)}{T} S_r \quad (11)$$

The third statistic is the Bartlett-corrected statistic defined in Equation (9); full details concerning computation of the correction factors can be found in Johansen [11]. The fourth method is based on the bootstrap procedures of CRT12. The bootstrap samples are obtained by estimating the VECM under the null hypothesis, checking that the roots of the estimated matrix polynomial equation  $\det[A(z)] = 0$  satisfy Assumption 1(i), and then generating a total of  $N_{BS}$  samples recursively using an appropriate

method. In the simulations reported here a wild bootstrap was employed in which, if  $\hat{\epsilon}_{it}$  denotes element  $i$  of the residual vector  $\hat{\epsilon}_t$ , then the residuals used for the bootstrap samples were of the form

$$\hat{\epsilon}_{it}^{BS} = u_{it} \left( \hat{\epsilon}_{it} - T^{-1} \sum_{t=1}^T \hat{\epsilon}_{it} \right), \quad i = 1, \dots, 4, \quad t = 1, \dots, T$$

where the  $u_{it}$  are independent normal variates. The statistic  $S_r$  is computed in each bootstrap sample and the critical value is obtained from the distribution of  $S_r$  over the  $N_{BS}$  bootstrap samples. We denote this test by  $S_r^{BS}$  but emphasise that the test statistic is actually  $S_r$  which is compared with the critical value from the finite sample bootstrap distribution rather than the critical value from the asymptotic distribution. Full details of the procedure can be found in CRT12.

**Table 3.** Moduli of non-unit roots in simulations.

$\delta/\gamma$	Moduli				
DGP1					
0.0	1.1180	1.1180	1.2500	1.2500	1.2500
0.2	1.1335	1.1335	1.2500	1.2500	1.2972
DGP2					
0.0	1.4142	1.4142	2.0000	2.0000	2.0000
0.2	1.3639	1.3639	2.0000	2.0000	2.5599
DGP3					
0.5	2.0000	2.0000	2.0000	2.0000	
0.8	1.2500	1.2500	1.2500	1.2500	
0.9	1.1111	1.1111	1.1111	1.1111	

In addition to the above statistics, three versions of the jackknife statistic are considered. The first is  $S_{r,m}^J$  defined in Equation (10). This was computed for a range of values of  $m$  where practicable, although we report mainly the results for  $m = 2$ ; details of how the tests perform for other values of  $m$  are also provided. The remaining two jackknife statistics are based on small sample adjustments to either the full sample statistic  $S_r$  and/or the sub-sample statistics  $S_{rj}$  upon which the jackknife is based. In particular the two additional jackknife statistics are defined by

$$S_{r,m}^{J1} = \frac{m}{m-1} S_r^{RA} - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^m S_{rj}$$

$$S_{r,m}^{J2} = \frac{m}{m-1} S_r^{RA} - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^m S_{rj}^{RA}$$

in which the small sample adjusted sub-sample statistics are defined analogously to Equation (11) by

$$S_{rj}^{RA} = \frac{(\ell - pk)}{\ell} S_{rj}, \quad j = 1, \dots, m.$$

The first of these statistics uses the small sample adjustment purely on  $S_r$  while the second also uses it on the sub-sample statistics.

A total of  $R = 10,000$  replications were performed for each combination of parameter values for each DGP and, as in CRT12, the VAR model was fitted with a restricted intercept (case 2). The bootstrap procedure is the most computationally intensive component in the simulations, requiring a sufficiently large number ( $N_{BS}$ ) of bootstrap samples in each replication in order to compute the critical value from the bootstrap distribution; CRT12, for example, set  $N_{BS} = 399$ . The bootstrap computations are therefore  $O(RN_{BS})$  compared to  $O(R)$  for the other statistics. While using a large number of bootstrap samples poses no problem in a single empirical application, it is more of a computational (and time) burden when computing a large number of bootstrap samples for each one of a large number of Monte Carlo replications. We therefore employed the approach of Davidson and MacKinnon [22] and Giacomini, Politis and White [23] and used only one bootstrap sample per replication, *i.e.*,  $N_{BS} = 1$ . Instead of using a large number of bootstrap samples to determine the critical value from the bootstrap distribution in each replication, the critical value is obtained from the bootstrap distribution across the  $R$  Monte Carlo replications. This ‘warp-speed’ method reduces substantially the number of bootstrap computations from  $O(RN_{BS})$  to  $O(R)$  in accordance with the other non-bootstrap statistics.

The simulation results are summarised in Tables 4–7; in all cases the tests are based on a nominal size of 5%. Table 4 contains the empirical size of each of the seven test statistics in the case of the VAR with a single cointegrating vector. The value of  $\delta$  has a relatively small impact on the performance of the tests but the reduction in  $\gamma$  from 0.8 to 0.5 has a much larger impact. It is apparent that, in all DGPs, the unadjusted Johansen statistic  $S_1$  has large size distortions, with the empirical size being as large as 45% in DGP1 with  $\delta = 0.2$  and  $T = 50$ . The small sample adjustment that results in  $S_1^{RA}$  reduces the size closer to its nominal level in all cases, particularly in DGP 2 (where  $\gamma = 0.5$ ) but less so in DGP 1 ( $\gamma = 0.8$ ) where size distortions remain. The Bartlett correction has a tendency to over-compensate, leading to empirical sizes below 5% (and around 2% for  $T = 50$ ) in most cases. The bootstrap produces sizes around 5% in DGP 1 but in DGP 2 the empirical size tends to be slightly lower than the nominal size. The jackknife statistic  $S_{1,2}^J$  manages to reduce the size towards the 5% level compared to the unadjusted statistic  $S_1$  with empirical sizes around 6%–7% in DGP 2 and a bit higher in DGP 1. The small sample adjustment in  $S_{1,2}^{J1}$  reduces the empirical size in all cases, compared to  $S_{1,2}^J$ , while  $S_{1,2}^{J2}$  produces sizes close to the nominal level in DGP 2 but shows little improvement (if any) over  $S_{1,2}^J$  in DGP 1.

The power performance of the tests in the cointegrated VAR is summarised in Table 5 in which the probability of rejecting the null hypothesis that  $r = 0$  is reported. Beginning with the unadjusted statistic  $S_0$ , the high power at the smaller sample sizes in DGP 1 is a reflection of the large size distortions reported in Table 4. All of the adjusted statistics are less powerful than  $S_0$  but it should be remembered that they do have better size properties. The statistic  $S_{0,2}^{J1}$  has particularly low power for  $T = 50$  in DGP 2.

The size properties of the tests in a non-cointegrated VAR are reported in Table 6. As the value of  $\gamma$  increases from 0 to 0.9 the unadjusted statistic  $S_0$  suffers from huge size distortions, rising to 92% for  $T = 50$  when  $\gamma = 0.9$ . The size properties of the adjusted statistics are all better than  $S_0$  with the bootstrap test controlling size best over this range of parameters. The Bartlett adjustment again tends to reduce empirical size to below its nominal level as  $\gamma$  increases while, for the jackknife statistics,  $S_{0,2}^{J2}$

performs best for smaller values of  $\gamma$  while  $S_{0,2}^{J1}$  produces the best performance of the three for larger values of  $\gamma$ .

**Table 4.** Empirical size: cointegrated VAR.

$\delta$	$T$	$S_1$	$S_1^{RA}$	$S_1^B$	$S_1^{BS}$	$S_{1,2}^J$	$S_{1,2}^{J1}$	$S_{1,2}^{J2}$
DGP1								
0.0	50	44.68	18.80	2.10	5.49	14.26	2.53	14.37
	100	23.02	13.36	3.91	4.64	10.04	4.83	9.37
	200	13.03	9.87	4.73	5.22	7.85	5.03	7.28
0.2	50	45.26	19.07	2.42	4.98	14.53	2.30	14.61
	100	22.39	13.28	4.38	5.24	9.99	4.67	9.38
	200	12.61	9.73	5.02	5.35	8.01	5.46	7.49
DGP2								
0.0	50	14.35	3.15	2.11	2.64	6.00	0.59	5.01
	100	10.44	5.38	4.68	4.62	7.62	3.31	6.58
	200	7.14	5.21	4.75	5.11	6.03	3.95	5.50
0.2	50	15.40	3.42	2.21	2.69	6.27	0.60	5.16
	100	10.50	5.37	4.90	4.58	7.30	3.18	6.42
	200	7.50	5.29	4.94	4.86	5.87	3.88	5.29

**Table 5.** Empirical power: cointegrated VAR.

$\delta$	$T$	$S_0$	$S_0^{RA}$	$S_0^B$	$S_0^{BS}$	$S_{0,2}^J$	$S_{0,2}^{J1}$	$S_{0,2}^{J2}$
DGP1								
0.0	50	97.57	85.20	30.93	51.76	70.05	27.76	74.00
	100	99.99	99.92	98.78	99.11	99.59	98.26	99.62
	200	100.00	100.00	100.00	100.00	100.00	100.00	100.00
0.2	50	97.03	83.09	30.90	46.41	65.77	23.63	69.81
	100	99.99	99.93	98.21	99.02	99.40	97.45	99.45
	200	100.00	100.00	100.00	100.00	100.00	100.00	100.00
DGP2								
0.0	50	62.52	27.08	18.39	17.79	26.14	3.63	26.36
	100	92.80	83.40	77.70	78.20	81.03	61.70	79.93
	200	100.00	100.00	100.00	100.00	99.98	99.95	99.98
0.2	50	66.70	29.74	19.45	18.88	27.87	4.19	28.26
	100	94.25	77.06	82.00	82.14	84.40	66.73	83.67
	200	100.00	100.00	99.99	99.99	99.99	99.98	99.99

**Table 6.** Empirical size: non-cointegrated VAR (DGP3).

$\gamma$	$T$	$S_0$	$S_0^{RA}$	$S_0^B$	$S_0^{BS}$	$S_{0,2}^J$	$S_{0,2}^{J1}$	$S_{0,2}^{J2}$
0.0	50	17.30	2.75	5.40	3.76	6.33	0.27	5.19
	100	9.37	4.03	5.22	4.36	6.18	2.00	5.29
	200	7.12	4.71	5.30	4.98	5.89	3.36	5.41
0.5	50	37.19	9.93	4.44	4.10	8.92	0.77	8.94
	100	16.94	8.07	4.90	4.73	7.73	2.62	6.82
	200	9.45	6.26	4.75	4.58	6.32	3.48	5.74
0.8	50	78.48	41.96	1.67	6.33	22.52	2.88	25.87
	100	44.30	27.61	3.64	5.90	13.84	5.33	13.77
	200	21.33	15.45	4.82	5.48	8.97	5.77	8.61
0.9	50	92.73	66.06	0.76	8.56	39.16	7.08	44.76
	100	75.26	58.00	1.10	7.96	27.61	12.09	28.55
	200	44.69	35.42	3.09	5.89	14.69	9.78	14.49

**Table 7.** Empirical size of  $S_{1,m}^J$  for varying  $m$ .

$\delta/\gamma$	$T$	$m$				
		2	4	5	8	10
DGP1						
0.0	100	10.04	9.95	10.12		
	200	7.85	7.64	8.00	7.83	7.92
0.2	100	9.99	9.87	9.94		
	200	8.01	7.70	8.01	7.95	8.01
DGP2						
0.0	100	7.62	6.60	6.01		
	200	6.03	5.90	5.85	5.50	5.32
0.2	100	7.30	6.32	5.87		
	200	5.87	5.94	5.95	5.68	5.44
DGP3						
0.0	100	6.18	5.18	4.56		
	200	5.89	5.64	5.35	5.17	4.81
0.5	100	7.73	6.56	5.90		
	200	6.32	5.76	5.52	5.37	5.05
0.8	100	13.84	16.46	16.74		
	200	8.97	9.55	9.78	10.83	10.97
0.9	100	27.61	38.04	40.20		
	200	14.69	19.45	21.55	25.45	26.62

The results for the jackknife tests in Tables 4–6 are based on  $m = 2$  sub-samples, but it is of interest to ascertain how the performance of the tests is affected using different values of  $m$ . For  $T = 50$  there is little scope to increase  $m$  much further; with  $m = 2$  each sub-sample has only  $\ell = 25$  observations,

so increasing  $m$  soon makes sub-sample estimation infeasible. However, for larger sample sizes some experimentation is possible, and so Table 7 reports the empirical size of  $S_{1,m}^J$  for  $m \in \{2, 4, 5\}$  when  $T = 100$  and  $m \in \{2, 4, 5, 8, 10\}$  when  $T = 200$ ; in each case, for the largest value of  $m$ , the sub-samples contain just  $\ell = 20$  observations. Table 7 shows that the empirical size of  $S_{1,m}^J$  is remarkably robust to the value of  $m$ , with the exception of DGP3 when  $\gamma = 0.9$ .

To summarise the simulation results, it appears that rank test statistics based on some form of correction factor can provide size improvements over the unadjusted Johansen statistic while still maintaining good power properties, although a bootstrap approach offers the most consistent performance over the range of DGPs considered. It should be stressed, however, that the corrected statistics and the bootstrap operate in rather different ways. All of the corrected statistics—whether the correction is a simple small sample adjustment, a (parametric) Bartlett correction or a (nonparametric) jackknife correction—aim to adjust the raw statistic in such a way that the distribution of the corrected statistic matches better the asymptotic distribution, the critical values of which the corrected statistic is compared with. The bootstrap, on the other hand, uses as the test statistic the unadjusted statistic itself, but by generating bootstrap samples whose size is equal to the given finite number of observations, uses critical values from the finite sample bootstrap distribution against which to compare the statistic. The evidence obtained here suggests that the latter approach is the most robust in practice.

#### 4. Conclusions

This paper has investigated the asymptotic properties and finite sample performance of jackknife-corrected test statistics for cointegration rank in a VAR system. In particular, the limiting distributions of jackknife-corrected test statistics have been derived for the two trend specifications of most empirical relevance; the asymptotic critical values for these cases have been tabulated; and the finite sample size and power properties of the jackknife-corrected tests have been compared with the usual (unadjusted) rank test statistic as well as statistics using various small sample corrections and a bootstrap approach. The simulations reveal that all the corrected statistics, including the jackknife variants, can provide size improvements over the unadjusted statistic while still maintaining good power properties, although a bootstrap approach offers the most consistent performance over the range of DGPs considered.

There are a number of ways in which the analysis of this paper can be built upon. In practice the precise form of the VAR (*i.e.*, the specification of the deterministic trend function and the number of lags) is unknown, and various pre-tests are often conducted, including the use of information criteria to determine the VAR order. This has an impact on the performance of the rank tests, and it would be of interest to ascertain how well the jackknife methods perform relative to other tests in such a scenario. Another potentially fruitful area of investigation concerns the use of jackknife methods in estimating the cointegrating parameters themselves. Additionally, bootstrap methods have been shown to be adaptable to situations where heteroskedasticity (both conditional and unconditional) is present as well as breaks in variance and correlations; see, for example, Cavaliere, Rahbek and Taylor [24,25]. Jackknife methods have also been found to be robust to conditional heteroskedasticity in stationary autoregressions by



Chambers [18], and so a further comparison with bootstrap methods would be of interest in the context of cointegration. Such avenues are left for future work.

**Acknowledgments**

I am grateful to the Editor-in-Chief, Kerry Patterson, and three anonymous referees for helpful comments on this paper. In particular I thank one of the referees for pointing out the independence properties of the functionals  $Q(B, F_2, \delta_{j,m})$  and  $Q(B, F_{j,m}, \delta_{j,m})$  that appear in parts (a) and (b), respectively, of Theorem 1. I also thank Giuseppe Cavaliere for providing Matlab code for implementing the bootstrap methods of Cavaliere, Rahbek and Taylor [15], upon which I was able to base my own Gauss code, and Rob Taylor for suggesting I explore ‘warp-speed’ methods of speeding up the simulations involving the bootstrap. The initial research upon which this paper is based was funded by the Economic and Social Research Council under grant number RES-000-22-3082.

**Appendix**

*A.1. Proof of Theorem 1*

(a) Johansen [5] (Theorem 11.1) shows that  $S_r \Rightarrow Q(B, F, \delta_0)$  as  $T \rightarrow \infty$  under the stated conditions, where  $F(s)$  is defined in Equation (7) for each case. Taking each of the two relevant cases in turn:

*Case 2*

Here,  $S_r \Rightarrow Q(B, F_2, \delta_0)$  and the sub-sample statistics have the same distribution defined on  $\delta_{j,m}$ , i.e.,  $S_{rj} \Rightarrow Q(B, F_2, \delta_{j,m})$ . The result for  $S_{r,m}^J$  follows straightforwardly as  $m$  is fixed.

To demonstrate the independence of  $Q(B, F_2, \delta_{j,m})$  and  $Q(B, F_2, \delta_{k,m})$  for  $j \neq k$ , consider

$$Q(B, F_2, \delta_{j,m}) = \text{trace} \left\{ \int_{(j-1)/m}^{j/m} dB(s)F_2(s)' \left[ \int_{(j-1)/m}^{j/m} F_2(s)F_2(s)' ds \right]^{-1} \times \int_{(j-1)/m}^{j/m} F_2(s)dB(s)' \right\}$$

and recall that  $F_2(s) = [B(s)', 1]'$ . This expression is a function of  $B(s)$  for  $s \in \delta_{j,m}$  which is clearly not independent of the process  $B(r)$  for  $r \in \delta_{k,m}$  that enters  $Q(B, F_2, \delta_{k,m})$ . However, let

$$A_{j,m} = \begin{pmatrix} I_{p-r} & -m \int_{(j-1)/m}^{j/m} B(s)ds \\ 0'_{p-r} & 1 \end{pmatrix}$$

where  $0_{p-r}$  denotes a  $(p - r) \times 1$  vector of zeros. We can then write

$$Q(B, F_2, \delta_{j,m}) = \text{trace} \left\{ \int_{(j-1)/m}^{j/m} dB(s)F_2(s)' A'_{j,m} \times \left[ \int_{(j-1)/m}^{j/m} A_{j,m}F_2(s)F_2(s)' A'_{j,m} ds \right]^{-1} \int_{(j-1)/m}^{j/m} A_{j,m}F_2(s)dB(s)' \right\}$$

which is a function of the process

$$A_{j,m}F_2(s) = \begin{pmatrix} B(s) - m \int_{(j-1)/m}^{j/m} B(s)ds \\ 1 \end{pmatrix}$$

This shows that  $Q(B, F_2, \delta_{j,m})$  can be represented in terms of the quasi-demeaned process  $B_{j,m}(s) = B(s) - m \int_{(j-1)/m}^{j/m} B(s)ds$  for  $s \in \delta_{j,m}$ , and it follows that  $Q(B, F_2, \delta_{k,m})$  can be written in terms of the process  $B_{k,m}(r) = B(r) - m \int_{(k-1)/m}^{k/m} B(r)dr$  for  $r \in \delta_{k,m}$ . The two functionals of interest will be independent if  $B_{j,m}(s)$  and  $B_{k,m}(r)$  are independent which, due to them being Gaussian processes, only requires their covariance to be zero. The covariance of interest is

$$C_{j,k} = E(B_{j,m}(s)B_{k,m}(r)') = E(B(s)B(r)') - mE\left(\int_{(j-1)/m}^{j/m} B(s)dsB(r)'\right) - mE\left(B(s)\int_{(k-1)/m}^{k/m} B(r)'dr\right) + m^2E\left(\int_{(j-1)/m}^{j/m} B(s)ds\int_{(k-1)/m}^{k/m} B(r)'dr\right)$$

Suppose, without loss of generality, that  $k > j$  which implies  $r > s$ . Then we have  $E(B(s)B(r)') = \min(s, r)I_{p-r} = sI_{p-r}$ ,

$$E\left(\int_{(j-1)/m}^{j/m} B(s)dsB(r)'\right) = \int_{(j-1)/m}^{j/m} \min(s, r)dsI_{p-r} = \int_{(j-1)/m}^{j/m} sdsI_{p-r} = \frac{j - \frac{1}{2}}{m^2}I_{p-r},$$

$$E\left(B(s)\int_{(k-1)/m}^{k/m} B(r)'dr\right) = \int_{(k-1)/m}^{k/m} \min(s, r)drI_{p-r} = s\int_{(k-1)/m}^{k/m} drI_{p-r} = \frac{s}{m}I_{p-r},$$

$$E\left(\int_{(j-1)/m}^{j/m} B(s)ds\int_{(k-1)/m}^{k/m} B(r)'dr\right) = \int_{(j-1)/m}^{j/m} \int_{(k-1)/m}^{k/m} \min(s, r)drds = \int_{(j-1)/m}^{j/m} sds\int_{(k-1)/m}^{k/m} dr = \frac{j - \frac{1}{2}}{m^2} \frac{1}{m}I_{p-r}$$

Combining these expressions we find that

$$C_{j,k} = \left(s - m \cdot \frac{j - \frac{1}{2}}{m^2} - m \cdot \frac{s}{m} + m^2 \cdot \frac{j - \frac{1}{2}}{m^2} \cdot \frac{1}{m}\right) I_{p-r} = 0_{(p-r) \times (p-r)}$$

as required, where  $0_k$  denotes a  $k \times k$  matrix of zeros.

Case 3

For the full-sample statistic,  $S_r \Rightarrow Q(B, F_3, \delta_0)$ , while the appropriate process  $F_{j,m}$  for the sub-samples is obtained as the residual from a continuous time projection of  $B(s)$  and  $s$  on a constant on the interval  $\delta_{j,m}$  (the process  $F_3(s)$  is obtained in the same way but with the projections taking place on  $\delta_0$ ). The first projection is given by

$$B(s) - \frac{\int_{(j-1)/m}^{j/m} B(s)ds}{\int_{(j-1)/m}^{j/m} ds} = B(s) - \frac{\int_{(j-1)/m}^{j/m} B(s)ds}{1/m} = B(s) - m \int_{(j-1)/m}^{j/m} B(s)ds$$

which provides the first  $p - r$  elements of  $F_{j,m}(s)$ . The second projection residual, which is the final element of  $F_{j,m}(s)$ , is obtained as

$$s - \frac{\int_{(j-1)/m}^{j/m} sds}{\int_{(j-1)/m}^{j/m} ds} = s - \frac{(j - \frac{1}{2})/m^2}{1/m} = s - \frac{j - \frac{1}{2}}{m}$$

The sub-sample statistics satisfy  $S_{rj} \Rightarrow Q(B, F_{j,m}, \delta_{j,m})$  and the result for  $S_{r,m}^J$  follows.

The independence of  $Q(B, F_{j,m}, \delta_{j,m})$  and  $Q(B, F_{k,m}, \delta_{k,m})$  follows from the arguments used for case 2 above, noting that  $F_{j,m}$  already contains the process  $B_{j,m}$ .

(b) In both cases, when  $m \rightarrow \infty$  as  $T \rightarrow \infty$  such that  $m^{-1} + mT^{-1} \rightarrow 0$  (so that  $\ell = m^{-1}T \rightarrow \infty$ ), it follows that  $S_{r,m}^J = S_r + o_p(1)$  and the results are straightforward.  $\square$

A.2. Method of Simulation of Limiting Distributions

The objective is to simulate distributions of the form

$$Q(B, F, \delta) = \text{trace} \left\{ \int_{\delta} dB F' \left( \int_{\delta} F F' \right)^{-1} \int_{\delta} F dB' \right\}$$

where  $B(s)$  is a  $(p - r)$ -dimensional standard Brownian motion process and  $F(s)$  is a stochastic processes whose precise form is given in Theorem 1 and depends on the specification of the deterministic component in the model. Consider, first,  $Q(B, F, \delta_0)$ , and define the  $(p - r)$ -dimensional process  $\Delta y_t = \epsilon_t, t = 1, \dots, T$ , where the elements of  $\epsilon_t$  are independent standard normal variates and  $y_0 = 0$ . Then, following Johansen [5] (chapter 15), the distribution of  $Q(B, F, \delta_0)$  is approximated by

$$\hat{Q}_T = \text{trace} \left\{ \sum_{t=1}^T \epsilon_t P_t' \left( \sum_{t=1}^T P_t P_t' \right)^{-1} \sum_{t=1}^T P_t \epsilon_t' \right\}$$

for an appropriate choice of  $P_t (t = 1, \dots, T)$ , as follows:

Case 2:  $P_t = (y'_{t-1}, 1)'$ ;

Case 3:  $P_t = [(y_{t-1} - \bar{y})', t - \frac{1}{2}(T + 1)]'$ , where  $\bar{y} = T^{-1} \sum_{t=1}^T y_{t-1}$ .

In a similar way the distributions of the sub-sample statistics are simulated using

$$\hat{Q}_{T,j,m} = \text{trace} \left\{ \sum_{t=(j-1)\ell+1}^{j\ell} \epsilon_t P_t' \left( \sum_{t=(j-1)\ell+1}^{j\ell} P_t P_t' \right)^{-1} \sum_{t=(j-1)\ell+1}^{j\ell} P_t \epsilon_t' \right\}$$

again subject to an appropriate choice of  $P_t$  ( $t = (j-1)\ell + 1, \dots, j\ell$ ):

Case 2:  $P_t = [y_{t-1}', 1]'$ ;

Case 3:  $P_t = [(y_{t-1} - m\bar{y}_j)', t - (j - \frac{1}{2})\ell - \frac{1}{2}]'$ , where  $\bar{y}_j = \ell^{-1} \sum_{t=(j-1)\ell+1}^{j\ell} y_{t-1}$ .

The simulated values are combined to approximate the distribution in Theorem 1 (a) using

$$\frac{m}{m-1} \hat{Q}_T - \frac{1}{m-1} \frac{1}{m} \sum_{j=1}^m \hat{Q}_{T,j,m}$$

for a range of values of  $m$ .

## Conflicts of Interest

The author declares no conflict of interest.

## References

1. Engle, R.F.; Granger, C.W.J. Cointegration and error correction: Representation, estimation and testing. *Econometrica* **1987**, *55*, 251–276.
2. Johansen, S. Statistical analysis of cointegration vectors. *J. Econ. Dyn. Control* **1988**, *12*, 231–254.
3. Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* **1991**, *59*, 1551–1580.
4. Anderson, T.W. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Stat.* **1951**, *22*, 327–351.
5. Johansen, S. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*; Oxford University Press: Oxford, UK, 1995.
6. Doornik, J.A. Approximations to the asymptotic distribution of cointegration tests. *J. Econ. Surv.* **1998**, *12*, 573–593.
7. MacKinnon, J.G.; Haug, A.A.; Michelis, L. Numerical distribution functions of likelihood ratio tests for cointegration. *J. Appl. Econom.* **1999**, *14*, 563–577.
8. Toda, H.Y. Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present. *Rev. Econ. Stat.* **1994**, *76*, 66–79.
9. Toda, H.Y. Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions. *Econom. Theory* **1995**, *11*, 1015–1032.
10. Nielsen, B. On the distribution of likelihood ratio test statistics for cointegration rank. *Econom. Rev.* **2004**, *23*, 1–23.
11. Johansen, S. A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica* **2002**, *70*, 1929–1961.

12. Reinsel, G.C.; Ahn, S.K. Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting. *J. Time Series Anal.* **1992**, *13*, 353–375.
13. Reimers, H.E. Comparisons of tests for multivariate cointegration. *Stat. Pap.* **1992**, *33*, 335–359.
14. Swensen, A.R. Bootstrap algorithms for testing and determining the cointegration rank in VAR models. *Econometrica* **2006**, *74*, 1699–1714.
15. Cavaliere, G.; Rahbek, A.; Taylor, A.M.R. Bootstrap determination of the co-integration rank in VAR models. *Econometrica* **2012**, *80*, 1721–1740.
16. Quenouille, M.H. Notes on bias in estimation. *Biometrika* **1956**, *43*, 353–360.
17. Tukey, J.W. Bias and confidence in not-quite large samples. In Proceedings of the Institute of Mathematical Statistics, Ames, USA, 3–5 April 1958.
18. Chambers, M.J. Jackknife estimation of stationary autoregressive models. *J. Econom.* **2013**, *172*, 142–157.
19. Chambers, M.J.; Kyriacou, M. *Jackknife Bias Reduction in the Presence of a Unit Root*; Discussion Paper 685, University of Essex Department of Economics: Colchester, UK, 2010.
20. Chambers, M.J.; Kyriacou, M. Jackknife estimation with a unit root. *Stat. Probab. Lett.* **2013**, *83*, 1677–1682.
21. Swensen, A.R. Corrigendum to “Bootstrap algorithms for testing and determining the cointegration rank in VAR models.” *Econometrica* **2009**, *77*, 1703–1704.
22. Davidson, R.; MacKinnon, J.G. Improving the reliability of bootstrap tests with the fast double bootstrap. *Comput. Stat. Data Anal.* **2007**, *51*, 3259–3281.
23. Giacomini, R.; Politis, D.N.; White, H. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econom. Theory* **2013**, *29*, 567–589.
24. Cavaliere, G.; Rahbek, A.; Taylor, A.M.R. Testing for co-integration in vector autoregressions with non-stationary volatility. *J. Econom.* **2010**, *158*, 7–24.
25. Cavaliere, G.; Rahbek, A.; Taylor, A.M.R. Bootstrap determination of the co-integration rank in heteroskedastic VAR models. *Econom. Rev.* **2014**, *33*, 606–650.