Testing linear restrictions on cointegrating vectors: Sizes and powers of Wald tests in finite samples

Alfred A. Haug*
University of Canterbury and York University

The Wald test for linear restrictions on cointegrating vectors is compared in finite samples using the Monte Carlo method. The Wald test within the vector error-correction based methods of Bewley et al. (1994) and of Johansen (1991), the canonical cointegration method of Park (1992), the dynamic ordinary least squares method of Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993), the fully modified ordinary least squares method of Phillips and Hansen (1990), and the band spectral techniques of Phillips (1991) are considered. In terms of test size, Johansen's method seems to be preferred, and in terms of test power it is Park's and Phillips'. However, the relatively poor results in the context of cointegrating regressions suggest that improvements on the performance of the Wald tests considered here are needed.

Running title: Performance of Wald tests in finite samples.

---

[0]* Address: Department of Economics, University of Canterbury, Private Bag 4800, Christchurch 8004, New Zealand. E-mail: a.haug@econ.canterbury.ac.nz. Phone: +64-3-364-2018. Financial support from the Social Sciences and Humanities Research Council of Canada under grant # 410-95-1354 is gratefully acknowledged. The author thanks, without implicating, participants at the 1997 Canadian Economics Association Meetings, the 1998 Econometric Society Australasian Meetings, the 1998 New Zealand Econometrics Study Group Meetings, and the 1998 Workshop on Fractional Intergration and Cointegration at the University of Dortmund for helpful comments on an earlier draft.

1. Introduction

Cointegration techniques have been applied widely in empirical economics in recent years. Numerous tests for cointegration and estimation methods for cointegrating vectors have been suggested in the literature. Almost all results are based on asymptotic theory and the performance in finite samples can differ substantially across tests and estimation methods, even though methods might be asymptotically equivalent and efficient. Cheung and Lai (1993), Gregory (1994), Toda (1995), and Haug (1996), among others, provided Monte Carlo comparisons of size distortions and of powers for various tests for cointegration. Stock and Watson (1993), Gonzalo (1994), Kitamura and Phillips (1995), and Ho and Sorensen (1996), among others, compared with the Monte Carlo method the performance of estimators in terms of, e.g., bias in median and dispersion as measured by the interquartile range.

The purpose of this paper is to study the performance in finite samples of tests for parameter restrictions on cointegrating vectors. The Monte Carlo method is employed for these purposes. Testing hypotheses suggested by economic theory is a central concern of econometrics and testing hypotheses about restrictions on parameters in cointegrating vectors is no exception. The goal is to apply tests that have close to correct size and high power.

Wald tests have been proposed for testing linear restrictions on cointegrating vectors for different, though asymptotically equivalent, estimation methods. This Monte Carlo analysis studies the effects of varying the estimation technique on calculating the Wald test. The Wald test statistics are distributed as χ^2 under the null hypothesis and reduce to a t statistic when only one cointegrating vector is present and only a single parameter is involved. The t statistic is then distributed asymptotically as normal.

The asymptotically efficient estimation methods considered for the Wald or t tests in this paper are (in alphabetical order of the chosen abbreviations): Bewley et al.'s (1994) Box–Tiao canonical variates based method (BWLY); Park's (1992) canonical cointegration regression method (CCR); Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson's (1993) dynamic ordinary least squares method (DOLS); Phillips and Hansen's (1990) fully modified ordinary least squares method (FM); Johansen's (1988, 1991) maximum likelihood method (JOH); and Phillips' (1991) band spectral regression methods (PH). The most popular method in empirical applications seems to be JOH. Less often used are CCR, DOLS, FM and PH. Other methods have been suggested in the literature. BWLY has been proposed more recently and is included in this study because it may outperform JOH point estimates in certain cases, as demonstrated by Bewley et al. The above methods are applied to several data generating processes (DGPs) of practical relevance. The Wald or t statistic for a linear restriction on the cointegrating vector is computed from the parameter and variance estimates of each method. Then, empirical sizes and powers of these tests are calculated and compared. The Monte Carlo method is used in connection with a DGP that allows for endogenous, weakly exogenous, and strongly exogenous regressors in the sense of Engle et al. (1983).

In previous research, Stock and Watson compared finite sample critical values of the t statistic for parameter restrictions on cointegrating vectors of five of the six methods considered above. Their DGP revealed relatively modest size distortions. Further, Li and Maddala (1994) suggested to use the moving block bootstrap to correct size distortions for the t statistic for three of the above six methods. However, these studies did not report results on test powers of the t tests. On the other hand, Inder (1993) reported results for powers of t tests for one of the above methods (FM) and other methods not considered in my paper. His preferred choice was a two-stage method combining an error-correction regression with the FM method.

Section 2 briefly outlines the various estimation methods used in the Monte Carlo study. In Section 3, the Monte Carlo design is explained and results are discussed. Section 4 concludes.

2. The Wald test in cointegrated systems

2.1 The Box–Tiao Method of Bewley et al.: BWLY

Bossaerts (1988) and Bewley, in several papers, suggested a method for cointegrated systems of equations based on the levels canonical correlation analysis suggested by Box and Tiao (1977).[1] This is in contrast to Johansen's well known method which relates levels to first differences and does therefore incorporate information on the presence of unit roots into the estimation. Bewley et al. used the Monte Carlo method to compare Johansen's estimators to theirs and found for a bivariate first–order model that their estimator is in several relevant cases less dispersed and leptocurtic in small samples than Johansen's.[2] Gonzalo derived for the bivariate first order model the asymptotic distribution of Bewley's Box–Tiao estimator. The distribution is non–symmetric and non–standard. Also, it includes terms that lead to finite–sample bias in the median. Despite these asymptotic problems, hypothesis tests on cointegrating vectors in small samples with this method may outperform those with Johansen's method, parallel to the findings of Bewley et al. for properties of the two estimators.

Following Yang and Bewley, consider a p–dimensional vector autoregressive representation of order k for the cointegrating relationship:

$$\Delta Y_t = -\varrho\beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{k-1}\Delta Y_{t-k+1} + \vartheta + \mathrm{v}_t, \quad t = 1, \ldots, T, \qquad (1)$$

with v_t distributed IIN(0, Υ). Y_t is a $p \times 1$ vector of variables integrated of order one, denoted by I(1), and ϑ is a vector of constants. Δ is the first difference operator and Γ_i is a $p \times p$ matrix. It is assumed that $0 < r < p$. Then, ϱ is a full rank $p \times r$ matrix of error-correction vectors and β is a full rank $p \times r$ matrix of r cointegrating vectors such that $\beta' Y_t$ is integrated of order zero, denoted I(0).

[1]See Bewley and Orden (1994), Bewley et al. (1994), Bewley and Yang (1995), and Yang and Bewley (1996). The last two papers describe cointegration tests within this system.

[2]See Phillips (1994) and Stock and Watson (1993) for a theoretical and an empirical study, respectively, for Johansen's method. Phillips' results also apply to Bewley's Box-Tiao estimator.

The modified Box–Tiao procedure described in Bewley and Orden uses the least squares residuals g_t and h_t from regressing Y_t on ΔY_{t-1}, ..., and ΔY_{t-k+1}, and from regressing $[Y'_{t-1} \quad 1]'$ on the same set of regressors, respectively. The specification considered in this paper allows for a constant in the cointegrating vector only. In other words, the constant ϑ is restricted.[3] Next,

$$G' = [g_1, \ldots, g_T]$$

and

$$H' = [h_1, \ldots, h_T]$$

are formed and the eigen–problem

$$\left[\widehat{\lambda^+}_i G'G - G'H(H'H)^{-1}H'G\right]\widehat{e}_i = 0$$

is solved for p pairs of eigenvalues $\widehat{\lambda^+}_i$ and eigenvectors $\widehat{e}_i$, ordered so that $\widehat{\lambda^+}_1 \leq \ldots \leq \widehat{\lambda^+}_p$. In a model with r cointegrating vectors, the estimator of β is associated with the r smallest eigenvectors:

$$\widehat{\beta} = [\widehat{e}_1, \ldots, \widehat{e}_r]\,.$$

Parallel to Johansen and Juselius (1990) and Johansen (1991), a Wald test for linear restrictions is applicable.[4] The null hypothesis for linear restrictions on the cointegrating vectors is

$$K'\beta = 0,$$

where K is a $(p+1)\times(p+1-s)$ matrix. The Wald test–statistic involves normalizing $K'\widehat{\beta}$ by its 'standard deviation':

$$\mathrm{BWLY}_W = T \;\otimes \mathrm{trace}\left[\{K'\widehat{\beta}(\widehat{D}^{-1} - I)^{-1}\widehat{\beta}'K\}(K'\widehat{\mathbf{e}}\widehat{\mathbf{e}}'K)^{-1}\right],$$

[3]See Johansen (1991) on the role of the constant in equation (1).

[4]Yang (1998) has recently suggested modifications that can be applied to any systems estimator of a cointegrated process with variables integrated of order one. Wald like tests, not considered here, are suggested based on estimators modified to achieve asymptotic efficiency and asymptotic mixed-normality so that the test is asymptotically χ^2–distributed .

with $\hat{\mathbf{e}}$ the eigenvectors corresponding to $\widehat{\lambda^+}_{r+1} > \ldots > \widehat{\lambda^+}_{p+1}$ and $\widehat{D} = \mathrm{diag}(\widehat{\lambda^+}_1, \ldots, \widehat{\lambda^+}_r)$. The BWLY$_W$ statistic is asymptotically distributed as χ^2 with $r(p-s)$ degrees of freedom. In the case of $r = 1$, the Wald statistic reduces to a statistic that is asymptotically distributed as normal:

$$\mathrm{BWLY} = T^{.5} K' \widehat{\beta}_1 \left[(\widehat{\lambda^+}_1^{-1} - 1) \left(\sum_{i=2}^{p+1} (K' \widehat{e}_i)^2 \right) \right]^{-.5}.$$

The Bewley estimator involves choosing the unknown autoregressive order k. Reimers (1993) compared various data based lag selection criteria in cointegrated vector autoregressive systems using the Monte Carlo method and recommended the Schwarz or Hannan–Quinn criterion. These are consistent estimators of the lag order, whereas Akaike's criterion is not. Therefore, I will employ the Schwarz criterion.

2.2 Park's canonical cointegration regression method: CCR

Park (1992) derived a canonical cointegration regression estimator, β^+, for the cointegrating vector β (normalized) in the following single equation cointegration model:

$$y_t = \beta x_t + u_t, \tag{2}$$

where u_t is I(0) with mean zero. Park's canonical regression procedure is based on the idea that cointegrating vectors are not unique and transformations using stationary components of the model do not alter the cointegrating relation. Nonparametric data transformations are used to remove asymptotically the cross serial correlations between the regression errors and the innovations of the regressors.

It is assumed that x_{0t} is one and $\Delta x_{it} = v_{it}, \;\; i = 1, \ldots, m$ (so that $p = m + 1$) with the v_{it} representing mean-zero random errors. Define $z_t' = (u_t \;\; v_t')$ and

$$\begin{aligned} \Omega &= \lim_{T \to \infty} (1/T) \sum_{t=1}^{T} \sum_{j=1}^{T} E(z_j z_t') \\ \Lambda &= \lim_{T \to \infty} (1/T) \sum_{t=1}^{T} \sum_{j=1}^{t} E(z_j z_t'). \end{aligned}$$

The matrices Ω and Λ are partitioned in conformity with z_t:

$$\Omega = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} \Lambda_{uu} & \Lambda_{uv} \\ \Lambda_{vu} & \Lambda_{vv} \end{bmatrix}.$$

Next, y_t and x_t are modified in order to eliminate nuisance parameters:

$$y_t^+ = y_t - \widehat{z}_t' \left(\widehat{\Phi}^{-1} \widehat{\Lambda}_2 \widehat{\beta} + \begin{bmatrix} 0 \\ \widehat{\Omega}_{vv}^{-1} \widehat{\Omega}_{vu} \end{bmatrix} \right)$$

and

$$x_t^+ = x_t - \widehat{z}_t' \left[\widehat{\Phi}^{-1} \widehat{\Lambda}_2 \right],$$

where

$$\widehat{\Lambda}_2 = \begin{pmatrix} \widehat{\Lambda}_{vu} & \widehat{\Lambda}_{vv} \end{pmatrix}',$$

$$\widehat{\Phi} = T^{-1} \sum_1^T \widehat{z}_t \widehat{z}_t',$$

and $\widehat{\beta}$ is the least squares estimate from equation (2). The next step is to apply least squares estimation to equation (2) with y_t^+ and x_t^+ instead of y_t and x_t in order to get the asymptotically efficient estimators β^+ and the associated variance-covariance matrix

$$\left[\widehat{\Omega}_{uu} - \widehat{\Omega}_{vu}' \widehat{\Omega}_{vv}^{-1} \widehat{\Omega}_{vu} \right] \left(x_t'^+ x_t^+ \right)^{-1}.$$

The Wald statistic for H_0: $h(\beta) = 0$ with $H(\beta) = \partial h / \partial \beta'$ of full rank q, the number of restrictions, is:

$$\mathrm{CCR} = \left\{ h'(\widehat{\beta}^+) \left[H(\widehat{\beta}^+)(C'C)^{-1} H(\widehat{\beta}^+)' \right]^{-1} h(\widehat{\beta}^+) \right\} / \widehat{\Omega}_{u\bullet v},$$

where

$$\Omega_{u\bullet v} = \Omega_{uu} - \Omega_{uv} \Omega_{vv}^{-1} \Omega_{vu}$$

and is the long–run variance and $\widehat{\Omega}_{u\bullet v}$ its estimate . C is the design matrix of the CCR. The statistic has a limiting χ^2 distribution with q degrees of freedom. When only one parameter is involved and $r = 1$, this test reduces again to a t statistic with an asymptotic normal distribution.

The estimations of the long–run variance–covariance matrix Ω and Λ are carried out using non–parametric methods. The method of Andrews (1991) is used to calculate the test–statistic denoted by CCR-A. A quadratic spectral kernel with the associated automatic, data–dependent, plug–in bandwidth estimator is employed. Also, this kernel estimator is prewhitened with a first order vector autoregression, as suggested by Andrews and Monahan (1992). Furthermore, to provide a comparison for the performance of Andrews' estimators, the Bartlett window with four lags is used instead to calculate the variances and covariances, denoted by CCR-B.

2.3 Phillips and Loretan, Saikkonen, and Stock and Watson's dynamic ordinary least squares method: DOLS

Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1993) suggested the dynamic ordinary least squares (DOLS) method for estimating cointegrating vectors. Stock and Watson compared different asymptotically efficient estimators and recommended, based on a limited Monte Carlo study for U.S. money demand, the DOLS estimator. If the variables are I(1) and there are r cointegrating vectors among the p variables, then there are r least squares regressions. Each regression has $(p - r)$ regressors in levels, a constant, contemporaneous values, leads, and lags of the first difference of each regressor. The DOLS estimator has a mixture normal distribution and the Wald statistic for restrictions on the parameters in the cointegrating vectors is distributed as χ^2. Again, the test reduces to a t statistic with a limiting normal distribution when $r = 1$ and only one parameter is involved. I use Schwarz's criterion in order to determine the appropriate lead and lag lengths for the DOLS regressions. I calculate for the Wald (or t) statistics the variances again with the quadratic kernel estimator of Andrews, denoted by DOLS-A.[5] The Bartlett method is used too, denoted by DOLS-B.

[5] See Stock and Watson.

2.4 Phillips and Hansen's fully modified regression method: FM

The procedure of Phillips and Hansen is similar to Park's. It is also a two–step procedure and the asymptotic distributions of the two estimators are identical. Park's procedure is to correct both, y_t and x_t, before applying least squares. In contrast, Phillips and Hansen first modified y_t to get y_t^{++} and then corrected the least squares estimates from the regression of y_t^{++} on x_t in order to eliminate nuisance parameters, leading to $\widehat{\beta}^{++}$. Phillips and Hansen's method employs semi-parametric corrections that also lead to asymptotically median-unbiased estimates.

Phillips and Hansen's procedure applies least squares to equation (2) to get the residuals $\widehat{z}_t' = (\widehat{u}_t \;\; \Delta x_t')$. Define

$$\Lambda_{vu}^{++} = \Lambda_{vu} - \Lambda_{vv}\Omega_{vv}^{-1}\Omega_{vu}.$$

Next, the variance–covariance matrices are estimated again with Andrews' procedure. The term Λ_{vu}^{++} represents the bias (due to endogeneity) of the regressors x_t. The fully modified estimator of β is given by

$$\widehat{\beta}^{++} = \left[\sum_{t=1}^{T}\left(y_t^{++}x_t' - (0 \quad \widehat{\Lambda}_{vu}^{++\prime})\right)\right]\left[\sum_{t=1}^{T} x_t x_t'\right]^{-1},$$

where

$$y_t^{++} = y_t - \widehat{\Omega}_{uv}\widehat{\Omega}_{vv}^{-1}\Delta x_t.$$

The Wald test is

$$\text{FM-A} = h'(\widehat{\beta}^{++})\left[H(\widehat{\beta}^{++})(x_t'x_t)^{-1}\widehat{\Omega}_{u\bullet v}H'(\widehat{\beta}^{++})\right]h(\widehat{\beta}^{++}).$$

The test statistic with the Bartlett window instead is denoted FM-B. The asymptotic corrections of the least squares estimator β^{++} and of β^{+} are equivalent. Both estimators eliminate nuisance parameters asymptotically.

2.5 Johansen's maximum likelihood method: JOH

Johansen (1991) derived the Wald test within the following vector autoregressive representation of order k:

$$Y_t = \Pi_1 Y_{t-1} + \cdots + \Pi_k Y_{t-k} + \vartheta + \mathrm{v}_t.$$

The system is rewritten as an error-correction model, as in equation (1):

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Pi Y_{t-k} + \vartheta + \mathrm{v}_t,$$

where

$$\Gamma_i = -\left(I - \Pi_1 - \cdots - \Pi_i\right), \quad i = 1, 2, \cdots, k-1,$$

and

$$\Pi = -\left(I - \Pi_1 - \cdots - \Pi_k\right).$$

The rank of the matrix Π determines the number r of cointegrating vectors among the variables in Y_t, $0 \leq$ rank $(\Pi) = r < p$. If $r = 0$, then $\Pi = 0$ and all variables appear only in first differences in the model and there are no cointegrating vectors. If $0 < r < p$, then the matrix $\Pi = \varsigma\beta'$ and $\beta' Y_t$ is I(0).

For this procedure, the ordinary least squares residuals R_{kt} and R_{0t} are calculated from regressions of $[Y_{t-k}' \quad 1]'$ on ΔY_{t-1}, . . ., and ΔY_{t-k+1} and of ΔY_t on the same set of regressors, respectively, to purge the system of short–run dynamics. Reduced rank regressions are then employed to estimate the cointegrating vectors. The major difference to Bewley's method is that it relates in the canonical correlation analysis the levels of lagged Y_t, instead of first differences, to Y_t. Bewley's method extracts first the most nonstationary components and then the stationary canonical variates, whereas Johansen's method extracts first the stationary canonical variates. Bewley et al. argued that their method ensures small sample (in addition to asymptotic) orthogonality between the estimated stationary and most nonstationary variates.

The cross correlation matrix of the residuals is given by

$$S_{ij} = T^{-1} \sum_{i=1}^{T} R_{it} R_{jt}',$$

where $i,j = 0,k$. The eigenvalues $\widehat{\lambda^*}_1 > \ldots > \widehat{\lambda^*}_{p+1}$ are the solutions of

$$\left|\lambda^* S_{kk} - S_{k0} S_{00}^{-1} S_{0k}\right| = 0$$

and represent the squared canonical correlations.

For a given r, the cointegrating vectors in β are given by the eigenvectors associated with the r largest eigenvalues, $\widehat{\lambda^*}_1 > \ldots > \widehat{\lambda^*}_r$ and these are the reduced rank estimators of β. It can be shown that this estimator is equivalent to the maximum likelihood estimator when errors are Gaussian. Johansen and Juselius (1990) and Johansen (1991) suggested to use a Wald test for the linear restrictions as described in Section 2.1. For the JOH statistic, $\hat{\mathbf{e}}$ is replaced by the eigenvectors corresponding to $\widehat{\lambda^*}_{r+1} > \ldots > \widehat{\lambda^*}_{p+1}$ and $\widehat{D} = \mathrm{diag}(\widehat{\lambda^*}_1, \ldots, \widehat{\lambda^*}_r)$. The JOH statistic is asymptotically distributed as χ^2 with $r(p-s)$ degrees of freedom. In the case of $r = 1$, the Wald statistic reduces again to a t statistic that is asymptotically distributed as normal.

2.6 Phillips' spectral regression method: PH

Phillips (1991) proposed to employ a block triangular representation of the cointegrated system and to apply nonparametric methods to the regression errors from the system. The advantage of this approach is that it is not necessary to be explicit about the generating mechanism of the errors. Phillips suggested to use so–called Hannan–efficient spectral regressions. Because cointegration is concerned with long–run relationships, it is possible to focus on the most relevant frequency by using band spectral regression at zero frequency. In other words, the regressors are I(1) processes whose power is concentrated at the origin. Full frequency band regression is not needed for efficient estimation in large samples. However, it may be useful in small samples. Furthermore, the system spectral method leads to cointegration estimators that are asymptotically median unbiased and symmetrically distributed and an optimal theory of inference applies. Hypothesis tests can be carried out using asymptotic χ^2 tests. Also, full spectral estimation is asymptotically equivalent to maximum likelihood.

The block triangular error correction representation is given by

$$\Delta Y_t = \gamma\alpha' Y_{t-1} + \psi_t, \tag{3}$$

with $\gamma' = (-1,\ \ 0)$, $\alpha' = (1,\ \ -\beta')$. Further,

$$\psi_t = \begin{bmatrix} 1 & \beta' \\ 0 & I \end{bmatrix} \zeta_t$$

and

$$Y_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix}$$

with Y_{1t} an I(1) variable and Y_{2t} a vector of m variables, each I(1), so that $p = m+1$. $\zeta_t' = (\zeta_{1t}\ \ \zeta_{2t})$ and ζ_t is I(0).

The first step is to apply least squares to equation (3) to get the residuals

$$\widehat{\psi}_t = \Delta Y_t - \gamma\widehat{\alpha}' Y_{t-1}.$$

Next, finite Fourier transforms are calculated:[6]

$$\begin{aligned}
\omega_\Delta(\lambda) &= (2\pi T)^{-.5} \sum_{t=1}^{T} \Delta Y_t \exp^{it\lambda} \\
\omega_*(\lambda) &= (2\pi T)^{-.5} \sum_{t=1}^{T} Y_{*t} \exp^{it\lambda} \\
\omega_Y(\lambda) &= (2\pi T)^{-.5} \sum_{t=1}^{T} Y_{t-1} \exp^{it\lambda} \\
\omega_{\widehat{\psi}}(\lambda) &= (2\pi T)^{-.5} \sum_{t=1}^{T} \widehat{\psi}_t \exp^{it\lambda}
\end{aligned}$$

for $\lambda\ \epsilon\ [-\pi, \pi]$, $Y_{*t}' = (Y_{1t}, \Delta Y_{2t}')$, and $\omega_Y'(\lambda) = (\omega_1(\lambda),\ \ \omega_2(\lambda))$', partitioned conformably with Y_t. Also, it holds that $\omega_{\widehat{\psi}}(\lambda) = \omega_\Delta(\lambda) - \gamma\widehat{\alpha}'\omega_Y(\lambda)$. Next, the smoothed periodogram estimates are computed.[7] The efficient weight function

$$\Xi(\lambda_s) = \widehat{f}_{\psi\psi}(\omega_j)^{-1}$$

[6] The Cooley–Tukey Fast Fourier algorithm in GAUSS is used.

[7] Instead of the smoothed periodogram estimates, other conventional spectral estimates could be used.

is used for all

$$\lambda_s \;\epsilon\; \mathcal{B}(j) = \left(\omega_j - \frac{\pi}{2M} < \lambda \leq \omega_j + \frac{\pi}{2M}\right)$$

for a frequency band with width $\frac{\pi}{M}$ so that $M \to \infty$ when $\frac{M}{T} \to 0$:

$$\begin{aligned}
\widehat{f}_{\psi\psi}(\omega_j) &= \frac{2M}{T}\sum_{\mathcal{B}(j)} \left[\omega_\Delta(\lambda_s) - \gamma\widehat{\alpha}'\omega_Y(\lambda_s)\right]\left[\omega_\Delta(\lambda_s) - \gamma\widehat{\alpha}'\omega_Y(\lambda_s)\right]^* \\
\widehat{f}_{22}(\omega_j) &= \frac{2M}{T}\sum_{\mathcal{B}(j)} \omega_2(\lambda_s)\omega_2(\lambda_s)^* \\
\widehat{f}_{2*}(\omega_j) &= \frac{2M}{T}\sum_{\mathcal{B}(j)} \omega_2(\lambda_s)\omega_*(\lambda_s)^*.
\end{aligned}$$

The estimate $\widehat{\alpha}$ is consistent so that

$$f_{\psi\psi}(\omega_j) \xrightarrow{p} f_{\psi\psi}(\omega)$$

as $T \to \infty$. The full spectrum estimator of β is

$$\tilde{\beta} = -\left[\frac{1}{2M}\sum_{j=-M+1}^{M} \gamma'\widehat{f}_{\psi\psi}^{-1}(\omega_j)\gamma\widehat{f}_{\psi\psi}'(\omega_j)\right] \times \left[\frac{1}{2M}\sum_{j=-M+1}^{M} \widehat{f}_{2*}^{-1}(\omega_j)\widehat{f}_{\psi\psi}^{-1}(\omega_j)\gamma\right]$$

Nonlinear estimation is not necessary because γ is known. The spectral estimator at the origin (zero frequency) is

$$\tilde{\beta}_{(0)} = \left[-\widehat{f}_{22}^{-1}(0)\widehat{f}_{2*}(0)\widehat{f}_{\psi\psi}^{-1}(0)\gamma\Big/\left[\gamma'\widehat{f}_{\psi\psi}^{-1}(0)\gamma\right].\right.$$

The usual Wald statistic (denoted PH(zero))is constructed for $\tilde{\beta}_{(0)}$ with the variance defined by[8]

$$V_{T0} = \frac{2M}{T}\left[\gamma'\widehat{f}_{\psi\psi}^{-1}(0)\gamma\widehat{f}_{22}(0)\right]^{-1}.$$

3. The Monte Carlo Design and Results

The DGP used in my Monte Carlo study is similar to the one used by, among many others, Gonzalo (1994). It is given for $p = 2$ by

$$y_t - \beta x_{1t} \;=\; \phi_t$$

[8]Results for the PH statistic calculated with the full spectrum estimators, denoted PH(full), are not reported in all Tables because, in general, PH(zero) performs better.

$$\begin{aligned} a_1 y_t + a_2 x_{1t} &= \epsilon_t \\ \phi_t &= \rho\phi_{t-1} + \mathrm{u}_t \\ \epsilon_t &= \epsilon_{t-1} + \theta(\epsilon_{t-12} - \epsilon_{t-13}) + \mathrm{e}_t \end{aligned}$$

and

$$\begin{bmatrix} \mathrm{u}_t \\ \mathrm{e}_t \end{bmatrix} = \text{iid N}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & \eta\sigma \\ \eta\sigma & \sigma^2 \end{bmatrix}\right].$$

Gonzalo showed that this DGP can be expressed as a DGP with moving average errors or, alternatively, as an error–correction model. The DGP can be extended to $p > 2$.[9] The parameter space experimented with is:

$$\begin{aligned} p &= (2,3) \\ \mathrm{T} &= (50,100,250) \\ \beta &= (1,1.08,1.14) \\ a_1 &= (0,1) \\ a_2 &= 1 \\ \rho &= (.3,.6,.8,.9) \\ \theta &= (0,.8) \\ \eta &= (-.5,0,.5) \\ \sigma &= (.5,1,2). \end{aligned}$$

The pseudo-normal variates u_t and e_t are generated by the RNDN function in GAUSS. A sample of size T+200 is generated and 5000 replications are used for every experiment. I start at $\mathrm{u}_0 = 0$ and $\mathrm{e}_0 = 0$ and discard the first 200 observations to mitigate startup effects. The parameter value choices for the DGP are motivated by choosing realistic values so that the DGP should come close to actual processes found in non–artificial data.[10]

I used GAUSS, COINT 2.0, and own code for all simulations. When $a_1 = 0$,

[9] See Haug, footnote 27, for details on one possibility.

[10] See Haug for more details.

x_{1t} is weakly exogenous with respect to the parameter of interest and with $a_1 = 1$ it is endogenous. It is strongly exogenous when $a_1 = 0$ and $\eta = 0$.

The inclusion of stationary autoregressive errors (AR) at a long lag (twelve lags in the above DGP) is motivated by a study by Rossana and Seater (1995) who demonstrated this to be a feature of many macroeconomic time–series at a disaggregated level.[11] This case will be considered only in Table 4 and θ will be set to zero otherwise.

Table 1 reports the empirical sizes of the tests for a nominal 5% level two–sided t test when $\rho = .8$. For BWLY and JOH, the distributions of the t statistic are invariant to changes in the value of σ, the sign of η, and to whether $a_1 = 1$ or $a_1 = 0$. Various values for T, σ, and η are considered. In general, size distortions increase as the sample size decreases, except for BWLY, where distortions change little. For T=250, the t statistic calculated with Johansen's method has overall the most stable size with the least distortion across the various values of σ and η when $a_1 = 1$. It has the least size distortion of all methods in seven out of the nine cases considered. In the two other cases, it ranked second and third. The size distortion of JOH ranges from .093 to .109 for the nominal .05 level test size. This distortion is not trivial, however, compared to the other tests, which reach empirical sizes in the 90% range, it has rather good size properties. When $a_1 = 0$, the preferred tests in terms of size are CCR-A and JOH, followed by FM-A. Differences are not very large.

Table 1 also gives results for T=100 and T=50. The relative rankings change somewhat for $a_1 = 1$. However, overall JOH is still the preferred test with the least size distortions. On the other hand, the FM-A test is preferred when $a_1 = 0$. It leads to much less size distortions in smaller samples than CCR-A and JOH.

In general, BWLY does not perform better in Table 1 than JOH. Also, CCR, DOLS, and FM perform in all cases considered better with Andrews method (A) than with the Bartlett window (B). In summary, the results for size distortions suggest to

[11] They further showed that temporal aggregation can distort this underlying process and lead to an integrated moving average process instead.

employ the JOH test when $a_1 = 1$ and the FM-A test when $a_1 = 0$.[12] Alternatively, the results suggest that test sizes should be corrected. The literature has suggested bootstrap techniques for this purpose.[13] Li and Maddala (1994) suggested to use the moving block bootstrap and showed that it works well for t statistics in cointegrated systems.[14]

Tables 2 and 3 report powers of t tests. The null hypothesis is $H_0 : \beta = 1$ and the alternative hypothesis is $H_1 : \beta = 1.14$. Powers depend crucially on the value of β and larger values would produce much higher powers and the reverse holds for lower values.[15] For the power studies, the untrue null hypothesis that $\beta = 1$ is tested when the data are generated under the alternative, and the Tables report the rejection frequency for two–sided t tests at the 5% significance level. Size–adjusted or empirical powers are reported throughout the paper. These powers are based on critical values calculated as quantiles under the null hypothesis, for every sample size and DGP used (instead of using asymptotic critical values).

Table 2 depicts powers for $\rho = .8$ when $a_1 = 1$. For a sample of 250, the BWLY statistic produces in several cases the highest powers. However, it also produces outliers with the lowest powers among all test. The same holds true for samples of T=100 and T=50, however, the performance of the BWLY test deteriorates somewhat relative to the other tests as T falls. Contrary to Bewley et al.'s findings, the earlier mentioned asymptotic problems of Bewley's estimator come to bear when testing hypotheses on cointegrating vectors. When T=250, CCR-B leads overall to the highest and most stable powers across the various values of σ and η. When T=100, FM-B performs relatively well, followed closely by PH(zero), DOLS-B, and CCR-B. When T=50, the performance of DOLS-B, FM-B, CCR-B, and PH(zero) is very similar. Overall, the CCR-B test is preferable in terms of power when $a_1 = 1$. In general, the

[12]See Banerjee et al. (1993, Chapter 8) on testing for $a_1 = 0$.

[13]Algebraic derivations of Edgeworth expansions for size corrections seem to be too cumbersome here.

[14]See also Davidson and MacKinnon (1996) on sizes of bootstrap tests in general.

[15]Table 4 considers $H_1 : \beta = 1.08$.

Bartlett window (B) outperforms Andrews' method (A) in the power studies.

Table 3 studies the powers of the t tests for $\rho = .8$ when $a_1 = 0$. The BWLY test performs well in many cases but again produces outliers with the lowest powers among all tests. Regardless of sample size, the PH(zero) test leads overall to the highest and most stable powers among all tests and it is the preferred test when $a_1 = 0$.

Finally, Table 4 studies the behavior of the tests for $\rho = .6$ instead of $\rho = .8$, three instead of two cointegrated variables, AR errors, and $\beta = 1.08$ instead of 1.14 under the alternative hypothesis. To save space, Table 4 reports results only for one value of σ and η. Before discussing the effect of changing ρ, I will discuss the other cases. Increasing p leads to more size distortion and lower powers in general, but leaves the relative rankings of the tests unchanged. Similarly, the introduction of AR errors into the DGP increases size distortions and lowers powers in general without changing relative rankings. As expected, a lower value of β under the alternative hypothesis leads to substantially lower powers.

For $\rho = .6$ and $a_1 = 1$, the results for powers in Table 4 differ from those in the other Tables where $\rho = .8$. The parameter ρ measures the speed of adjustment to the equilibrium cointegrating relationship. A high value indicates a slower speed of adjustment and vice versa. Bewley et al. (1994) reported experimental results for Johansen's estimator that showed that it performs well when speeds of adjustment are high and that it produces outliers when the speed of adjustment is slow.[16] Table 4 confirms these results for the t tests. JOH produces the highest and most stable powers when ρ takes on low values of .6 or .3 (not reported), however, it performs worse than other tests when ρ takes on larger values of .8 or .9 (not reported). On the other hand, when $a_1 = 0$, changes in ρ do not affect the previous results.

[16]Phillips (1994) provided a theoretical analysis showing that the finite sample distribution is leptocurtic in the general case.

4. Conclusion

This paper used the Monte Carlo method to study the performance of tests of linear restrictions on cointegrating vectors. The t statistics were calculated for several cointegration estimators and size distortions and test powers were compared. In terms of size distortions, Johansen's t test is preferred.[17] However, the size is often double of its nominal value and bootstrap or other techniques should be considered to correct sizes, as suggested by Li and Maddala (1994). Johansen (1998) proposed recently a Bartlett type correction factor for the likelihood ratio instead of the Wald test in cointegrating systems.

In terms of size–adjusted test powers, the JOH test performance depends critically on the speed of adjustment to the cointegration equilibrium and produces relatively low powers when the adjustment speed is slow. Instead, the CCR-B test is in general preferred when regressors are not weakly exogenous, and the PH(zero) when they are weakly or strongly exogenous. These results suggest to explore size corrections for the CCR and PH based tests. Xiao and Phillips (1998a) developed recently asymptotic expansions for Wald tests. They proposed a modified Wald test that uses a bandwidth selection criterion to minimize second order effects and is modified by using consistent estimates of second order terms.[18]

Overall, the Monte Carlo results indicate serious size distortions of Wald tests. Also, powers in samples of size 100 and below are very low for the Wald tests in cointegrated systems. The paper shows that the problems of Wald tests found in stationary cases are compounded in the cointegrating cases.[19]

[17] When $a_1 = 0$, the FM-A test has somewhat better size in samples of 50 and 100 observations than JOH.

[18] See also Xiao and Phillips (1998b) on the issue of using second order expansions and mean squared error approximations for efficient frequency domain regression estimators.

[19] See for example Bera et al. (1981).

REFERENCES

Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica 59, 817–858.

Andrews, D. W. K., and Monahan, J. C., 1992, An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator, Econometrica 60, 953–966.

Banerjee, A., Dolado, J. J., Galbraith, J. W., and Hendry, D. F., 1993, Cointegration, error–correction, and the analysis of non–stationary data (Oxford University Press, Oxford, UK).

Bera, A. K., R. P. Byron, and C. M. Jarque, 1981, Further evidence on asymptotic tests for homogeneity and symmetry in large demand systems, Economics Letters 8, 101–105.

Bewley, R., and Yang, M., 1995, Test for cointegration based on canonical correlation analysis, Journal of the American Statistical Association 90, 990–996.

Bewley, R., and Orden, D., 1994, Alternative methods of estimating long–run responses with applications to Australian import demand, Econometric Reviews 13, 179–204.

Bewley, R., Orden, D., Yang, M., and Fisher, L. A., 1994, Comparison of Box–Tiao and Johansen canonical estimators of cointegrating vectors in VEC(1) models, Journal of Econometrics 64, 3–27.

Bossaerts, P., 1988, Common nonstationary components of asset prices, Journal of Economic Dynamics and Control 12, 348–364.

Box, G. P. E., and Tiao, G. C., 1977, A canonical analysis of multiple time series, Biometrika 64, 355–365.

Cheung, Y.–W., and K. S. Lai, 1993, Finite sample sizes of Johansen's likelihood ratio tests for cointegration, Oxford Bulletin of Economics and Statistics 55, 313–328.

Davidson, R., and MacKinnon, J. G., 1996, The size distortion of bootstrap tests, Queen's University, Institute for Economic Research, Discussion Paper No. 936.

Engle, R. F., Hendry, D. F., and Richard, J.–F., 1983, Exogeneity, Econometrica 51, 277–304.

Gonzalo, J., 1994, Comparison of five alternative methods of estimating long–run equilibrium relationships, Journal of Econometrics 60, 203–233.

Gregory, A., 1994, Testing for cointegration in linear quadratic models, Journal of Business and Economic Statistics 12, 347–360.

Haug, A. A., 1996, Tests for cointegration: A Monte Carlo comparison, Journal of Econometrics 71, 89–115.

Ho, M. S., and Sorensen, B. E., 1996, Finding cointegration rank in high dimension systems using the Johansen test. An illustration using data based Monte Carlo simulations, Review of Economics and Statistics 78, 726–732.

Inder, B., 1993, Estimating long-run relationships in economics, Journal of Econometrics, 57, 53 68.

Inder, B., 1995, Finite sample arguments for appropriate estimation of cointegrating vectors, Monash University, Department of Econometrics.

Johansen, S., 1998, A small sample correction for tests of hypotheses on cointegrating vectors, European University Institute.

Johansen, S., 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, Econometrica 59, 1551–1580.

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

Johansen, S., and Juselius K., 1990, Maximum likelihood estimation and inference on cointegration - With applications to the demand for money, Oxford Bulletin of Economics and Statistics 52, 169–210.

Kitamura, Y., and P. C. B. Phillips, 1995, Efficient IV estimation in nonstationary regression, Econometric Theory 11, 1095–1130.

Li, H., and G. S. Maddala, 1994, Small sample inference in cointegrated systems using bootstrap methods, paper presented at the Econometric Society Summer Meetings in Quebec City.

Park, J. Y., 1992, Canonical cointegrating regressions, Econometrica 60, 119–143.

Phillips, P. C. B., 1994, Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error–correction models, Econometrica 62, 73–93.

Phillips, P. C. B., 1991, Spectral regression for cointegrated time series, in: Non-parametric and semi–parametric methods in economics and statistics, ed. by W. Barnett, Cambridge University Press, Cambridge.

Phillips, P. C. B., and Hansen, B. E., 1990, Statistical inference in instrumental variable regression with I(1) processes, Review of Economic Studies 57, 99–125.

Phillips, P. C. B., and M. Loretan, 1991, Estimating long run economic equilibria, Review of Economic Studies 58, 407 436.

Reimers, H., 1993, Comparison of tests for multivariate cointegration, Statistical Papers 33, 335-359.

Rossana, R. J., and Seater, J. J., 1995, Temporal aggregation and economic time series, Journal of Business and Economic Statistics 13, 441–451.

Saikkonen, P., 1991, Asymptotically efficient estimation of cointegrating vectors, Econometric Theory 7, 1–21.

Stock, J. H., and Watson, M. W., 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica 61, 783–820.

Toda, H. Y., 1995, Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions, Econometric Theory 11, 1015 1032.

Xiao, Z., and P. C. B. Phillips, 1998a, Higher order approximations for Wald statistics in cointegrating regressions, Working Paper, available from http://korora.econ.yale.edu/phillips.htm.

Xiao, Z., and P. C. B. Phillips, 1998b, Higher order approximations for frequency domain time series regressions, Journal of Econometrics 86, 297 336.

Yang, M., 1998, System estimators of cointegrating matrix in absence of normalising information, Journal of Econometrics 85, 317 337.

Yang, M., and Bewley, R., 1996, On cointegration tests for VAR models with drift, Economics Letters 51, 45–50.