The Size and the Power of the Bootstrap Tests for Linear Restrictions in Misspecified Cointegrating Relationships

Alessandra Canepa* and Raymond O'Brien
University of Southampton
Department of Economics
Highfield
Southampton SO17 1BJ
E-mail: acanepa@soton.ac.uk, rjo@soton.ac.uk

July 2000

Abstract

This paper considers computer intensive methods for inference on cointegrating vectors in maximum likelihood analysis. It investigates the robustness of LR , $Wald$ tests and an $F-$type test for linear restrictions on cointegrating space to misspecification of the number of cointegrating relations. In addition, since all the distributional results within the maximum likelihood cointegration model rely on asymptotic considerations, it is important to consider the sensitivity of inference procedures to the sample size. In this paper we use bootstrap hypothesis testing as a way to improve inference for linear restriction on the cointegrating space. We find that the resampling procedure is a very useful device for tests that lack the invariance property such as the $Wald$ test, where the size distortion of the bootstrap test is small even for a sample size $T = 50$. Moreover, it turns out that when the number of cointegrating vectors are correctly specified the bootstrap succeeds where the asymptotic approximation is not satisfactory, that is, for a sample size $T < 200$. The only valid alternative to the resampling procedure is the F-type test proposed by Podivinsky (1992). However, when the number of cointegrating vectors is over-fitted relying on the asymptotic approximation is misleading, since the tests considered exhibit sizes very far from the nominal size. In this situation the bootstrap test is much more robust to misspecifications. The analysis of the power reveals that the procedures have power. However, it is difficult to evaluate the power properties without investigating the asymptotic power, so further work is needed.

Keywords: Bootstrap tests, linear restrictions, cointegration, misspecification, small sample corrections.

JEL classification: C15, C32.

*Partial grant under E.S.R.C. grant R00429924070 is gratefully acknowledged

1 Introduction

The first procedure for testing cointegrating relationships was proposed by Engle and Granger (1987). After their seminal paper cointegration has became an extremely intensive field of research, and in the literature many alternatives to their procedure have been developed.

Among them the Johansen (1988, 1991,1992, 1995) and Johansen and Juselius (1990) procedure for estimation and testing of cointegrating relationships is widely used in applied econometric research. This method applies the maximum likelihood procedure to a multivariate vector autoregressive model written in the error correction form. Maximizing the Gaussian likelihood function leads via reduced rank regression to the analysis of eigenvalues and eigenvectors. To test for linear restrictions on the cointegrating vectors and their weights Johansen (1988) and Johansen and Juselius (1990) proposed likelihood ratio and Wald tests. However, it has been established that the asymptotic χ^2 distributed tests are quite heavily affected by the sample size. Accordingly, Podivinsky (1992) proposed an alternative approximate F-type test. Monte Carlo evidence in Psaradakis (1993) indicates that the application of Podivinsky's (1992) test is worthwhile, since improvements are shown with respect to the size properties of Johansen's (1988) LR and $Wald$ tests. In addition, Psaradakis (1993) proposed a small sample adjustment for LR criterion and the $Wald$ test.

This paper proposes the use of bootstrap hypothesis testing to improve Johansen's (1988) inference for linear restrictions on the cointegrating space in an experimental design where at most two cointegrating vectors are possible. The analysis conducted on the tests considered above allows for potential over-fitting and under-fitting of the number of cointegrating vectors included in the restricted model and particular attention will be given to the study of small samples properties of the tests.

The general idea on which bootstrapping is based is to use the single data set to design a sort of Monte Carlo experiment in which the data themselves are used to generate an approximation to the distribution of the statistics in which we are interested. However, as Veall (1998) suggests there are two main stages in the development of bootstrap theory. The first stage is related to its introduction by Efron (1979) as a computer-based method for evaluating the accuracy of a statistic by using the bootstrap algorithm for estimating standard errors or confidence intervals. This procedure can be useful when the finite-sample distribution of the statistics we are analysing is not known or a good asymptotic approximation is not available. The second stage of the bootstrap literature concerns the case where asymptotic analytic tools are available but in which bootstrap refinements are used to improve finite-sample performance. Good references in this sense are Horowitz (1994) where the bootstrap method is applied to the information matrix tests. For an excellent discussion based on the Edgeworth expansion see Hall (1992). These and other studies have found that bootstrap provides a higher-order asymptotic approximation to critical values based on "smooth" statistics. This means that for bootstrap-based critical values the size distortion (that is the difference between the nominal level and its actual rejection probability) decreases more rapidly with increasing the sample size than if the critical values obtained from first-order asymptotic theory are used.

Davidson and MacKinnon (1999) investigate this issue and they claim that the size distortion of a bootstrap test is of order $T^{-1/2}$ smaller than that of the corresponding asymptotic test. A further refinement of the order $T^{-1/2}$ can be obtained in the case of an asymptotically pivotal statistic, (i.e. a statistic whose limiting distribution is asymptotically independent of unknown nuisance parameters). As far as the consistency of the bootstrap in unit root contest is concerned the debate is still open and more theoretical results are needed since second order improvements have been shown only in the case of stationary time series regressions (see Bose (1988)).

For the non-stationary processes in the literature the asymptotic validity of the bootstrap for least squares estimate of the parameter of an explosive $AR(1)$ process has been established by Basawa *et al.* (1989). However, Basawa *et al* (1991a,b) show that in unit root models the asymptotic bootstrap distribution of the estimator does not converge to the distribution that would be obtained from repeated sampling from the population. Therefore, in this case the standard bootstrap least square estimator is asymptotically invalid. To get around this problem they presented a modified sequential bootstrap which work in this situation. More recently, the paper by Li and Xiao (1999) demonstrate the asymptotic validity for cointegrating regressions and the regression t-ratio statistic (see also Harris (1992)). To our the best of our knowledge, there are not yet theoretical results on the second order refinements in the case of non-stationary time series. One reason for this is the lack of a higher-order asymptotic theory. In fact, non-stationary processes depend on the underlying functional central limit theorem, and developing high order extensions of this theorem is not an easy task.

In this paper we follow the proposal of Gredenhoff and Jacobson (1998) and we use the bootstrap hypothesis testing as a way to reduce the size distortion of the tests for linear restrictions on the cointegrating space. As far as the results are concerned, we find that the resampling procedure is a very useful device for tests that lack the invariance property such as the $Wald$ test, where the size distortion of the bootstrap test is small even for a sample size $T = 50$. Moreover, it turns out that when the number of cointegrating vectors are correctly specified the bootstrap succeeds where the asymptotic approximation is not satisfactory, that is, for a sample size $T < 200$. The correction factors introduced by Psaradakis (1993) help to reduce the size distortion for the $Wald$ and also for the LR test. However, the only valid alternative to the resampling procedure is the F-type test proposed by Podivinsky (1992).

If the number of cointegrating vectors is under-fitted the size distortion is not very different from the one we observe when the model is correctly specified (that is the difference between the empirical and the nominal size asymptotically vanishes). By contrast, if the number of cointegrating vectors is over-fitted relying on the asymptotic approximation is misleading since all the tests considered above exhibit sizes very far away from the nominal size. In this situation the bootstrap test is much more robust to misspecifications.

The outline of this paper is the following. Section 2 briefly introduces the Johansen maximum likelihood estimation and, in particular the likelihood ratio and Wald tests for linear restrictions as well as Podovinsky's (1992) F-type test and Psaradakis (1993) corrected LR and $Wald$ tests. Section 3 describe the bootstrap test. Section 4 describes the data generating process and the Monte Carlo experimental design. In section 5 we compare the powers under certain misspecifications. A brief concluding section offers some recommendation for applied work.

2 Johansen's Maximum Likelihood Procedure

Johansen considers a general vector autoregression in error correction form,

$$\Delta Y_t = \mu + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Pi Y_{t-k} + \epsilon_t, \tag{2.1}$$

where Y_t , and ϵ_t are $(p \times 1)$ vectors, and Γ_1 through Γ_k are $(p \times p)$ matrices of coefficients. $\Delta Y_t = Y_t - Y_{t-1}$. $\epsilon_t \backsim NID(0, \Sigma)$.We specialise to the case $k = 1$, so

$$\Delta Y_t = \mu + \Pi Y_{t-1} + \epsilon_t, \tag{2.2}$$

The matrix Π determines whether or not, and to what extent, the system (2.2) is cointegrated.

We assume first that the eigenvalues of $I + \Pi$ lie on or inside the unit circle. Suppose that Π has rank r. If $r = 0$, and thus Π is a null matrix, Y_t is a vector of random walks related only through the covariances of their innovations ϵ_t . If $r = p$, Y_t is stationary. If $0 < r < p$ (2.2) can be interpreted as an error correction model. The hypothesis of r cointegrating vectors β can be written as:

$$H_0 : \Pi = \alpha\beta',$$

where α and β are $(p \times r)$ matrices. The rows of β' can be interpreted as the distinct cointegrating vectors of Y_t (i.e. such that the linear combinations $\beta' Y_t$ are $I(0)$) and the elements of α represent the weights of each of these r cointegrating relations in the p component equations (2.2).

Johansen (1988) shows that maximising the likelihood function involves solving the eigenvalue problem

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

to give p ordered eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$ and corresponding eigenvectors $\hat{V} = [\hat{v}_1 \ldots \hat{v}_p]$ normalised such that $\hat{V}' S_{kk} \hat{V} = I$. The matrices $S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R'_{jt},\ i, j = 0, k$, where R_{0t} and R_{kt} , are the residuals obtained by regressing ΔY_t and Y_{t-k} on, in general, $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}, D_t$ and 1. In our case the S_{ij} are just mean adjusted moment matrices. A basis for the space spanned by the cointegrating vectors is estimated by $\hat{\beta} = [\hat{v}_1 \ldots \hat{v}_r]$. The corresponding estimate of α is given by $\hat{\alpha}\left(\hat{\beta}\right) = S_{0k}\hat{\beta}$.

A test for the number r of cointegrating vectors can be based on the p eigenvalues $\hat{\lambda}_1 > \ldots > \hat{\lambda}_p > 0$. Johansen (1988) derives a likelihood ratio(LR) test of the hypothesis that there are at most r cointegrating vectors by testing that the $(p - r)$ smallest eigenvalues $\lambda_{r+1}, \ldots, \lambda_p$ are zero against the assumption that $\lambda_i \geq 0$ for $i = 1, \ldots, p$. The LR test statistic for this is known as the trace test, defined as

$$LR(trace)_r = -T \sum_{i=r+1}^{p} \ln\left(1 - \hat{\lambda}_i\right),$$

where $\hat{\lambda}_i$ are the estimates of the λ_i calculated from the maximum likelihood estimator of Π. In addition, the maximum eigenvalue test statistic is given by

$$LR(\max)_r = -T \ln\left(1 - \hat{\lambda}_{r+1}\right)$$

and can be used to test the null $H_0(p) : rank(\Pi) = r$ against the alternative $H_1(p+1) : rank(\Pi) = r + 1$.

The values of r chosen using the LR tests determine the matrices α and β : both are $(p \times r)$. It is then possible to test linear restrictions upon the elements of α and β .

Now we can briefly outline the proposed tests for linear restrictions on the cointegrating vectors. Under the hypothesis $H_0 : \Pi = \alpha\beta'$, the maximised value of the concentrated likelihood function satisfies

$$\hat{L}^{-2/T} = |S_{00}| \prod_{i=1}^{r} \left(1 - \hat{\lambda}_i\right),$$

where S_{00} and $\hat{\lambda}_i$ were defined earlier. Johansen and Juselius (1990) use this to develop LR tests of linear restrictions on the matrices α and β . Here we will consider only the case $\beta = H\varphi$.

To understand how this test is derived, recall that only the ranges of the columns of α and β are identified. If we set $\alpha^* = \alpha B'$ and $\beta^* = \beta B^{-1}$ than $\alpha^* \beta^{*\prime} = \alpha\beta' = \Pi$. Therefore, α and β are identified only up to a non-singular transformation $B(r \times r)$. Now, what enters the model is $\beta' y_{t-k}$, r linear combinations of the p elements in Y_{t-k} . Restricting

$$\beta_{(p \times r)} = \underset{(p \times s)}{H} \underset{(s \times r)}{\varphi}$$

implies that $\beta' y_{t-k} = \varphi' H' y_{t-k}$ and if Y_k is a matrix whose t^{th} row is y'_{t-k} , the column space of $Y_k H$ is now s dimensional.

The maximised value of the concentrated likelihood function subject to the restriction is

$$\tilde{L}^{-2/T} = |S_{00}| \, \Pi_{i=1}^{r} \left(1 - \tilde{\lambda}_i\right)$$

where $\tilde{\lambda}_1 > \ldots > \tilde{\lambda}_s$ are the $s > r$ eigenvalues obtained from solving

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

The LR test of $\beta = H\varphi$ can be obtained from the concentrated likelihood functions above, and is

$$LR\,(\beta) = -2\ln\left(\tilde{L}/\hat{L}\right) = T\sum_{i=1}^{r} \ln\left[\left(1-\tilde{\lambda}_i\right)/\left(1-\widehat{\lambda}_i\right)\right].$$

Johansen (1988) shows that the asymptotic distribution of the LR_r trace test is

$$tr\left(\int_0^1 dBB' \left[\int_0^1 BB'du\right]^{-1} \int_0^1 BdB'\right),$$

where $B\,(u)$ is an $(p-r)$-dimensional Brownian motion with covariance matrix $\mathbf{I}$. He tabulates simulated values of selected percentiles of this asymptotic distribution for a range of values of $(p-1) = 1, 2, 3, 4, 5$. These tabulated values serve for testing $r = 0, r \leq 1, \ldots, r \leq (p-1)$ when p ranges from 2 to 5.

Alternatively, Johansen and Juselius (1990) propose a Wald test. Consider the following null hypothesis $H_0 : K'\beta = 0$ where K is an $(p \times (p-s))$ matrix of full rank, then the W statistic for testing H_0 is:

$$W\,(\beta) = Ttr\left(\left[K'\hat{\beta}\left(\hat{\Lambda}^{-1} - I_r\right)^{-1} \hat{\beta}' K\right]\left[K'\hat{V}_*\hat{V}_*'K\right]^{-1}\right) \tag{2.3}$$

where $\hat{\Lambda} = diag\left(\hat{\lambda}_1, \ldots, \hat{\lambda}_r\right)$ and $\hat{V}_* = [\hat{v}_{r+1}, \ldots, \hat{v}_p]$. Since the limiting distribution of $\hat{\beta}$ is a Gaussian mixture, $LR(\beta)$ and $W(\beta)$ are asymptotically distributed as $\chi^2\,(r\,(p-s))$ under the hypotheses $\beta = H\varphi$ and $K'\beta = 0$ respectively.

It may help to relate the two forms of the restrictions. Given $\underset{(p\times r)}{\beta} = \underset{(p\times s)}{H}\underset{(s\times r)}{\varphi}$, we order the rows of β so that

$$H = \begin{matrix} s \\ (p-s) \end{matrix} \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$$

has H_1 of full rank. So partitioning conformably

$$\begin{matrix} s \\ (p-s) \end{matrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = H\varphi = \begin{bmatrix} H_1\varphi \\ H_2\varphi \end{bmatrix}.$$

Then

$$\beta_1 = H_1\varphi \Rightarrow \varphi = H_1^{-1}\beta_1.$$

Substituting in $\beta_2 = H_2\varphi$

$$\beta_2 = H_2H_1^{-1}\beta_1.$$

Hence $\beta = H\varphi$ implies

$$\begin{bmatrix} -\underset{((p-s)\times s)}{H_2}\underset{(s\times s)}{H_1^{-1}} & I_{(p-s)} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underset{((p-s)\times r)}{0}.$$

This is one way of obtaining

$$K'_{((p-s)\times r)}\beta_{(p\times r)} = 0,$$

i.e. $(p-s)$ common linear restriction on the columns of β.

Similarly, given an arbitrary K, and $K'\beta = 0$ we can write

$$\begin{bmatrix} \underset{((p-s)\times s)}{K_1} & \underset{((p-s)\times(p-s))}{K_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = 0,$$

hence, if we order the rows of β so that K_2 is invertible then

$$K_1\beta_1 + K_2\beta_2 = 0 \Rightarrow -K_1\beta_1 = K_2\beta_2 \Rightarrow \beta_2 = -K_2^{-1}K_1\beta_1.$$

Thus

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underset{((p-s)\times s)}{\begin{bmatrix} I \\ -K_2^{-1}K_1 \end{bmatrix}} \underset{(s\times r)}{\beta_1} = \begin{bmatrix} I \\ -K_2^{-1}K_1 \end{bmatrix} H_1 H_1^{-1} \underset{(s\times s)}{\beta_1}$$

$$= \begin{bmatrix} H_1 \\ -K_2 K_1^{-1} H_1 \end{bmatrix} \varphi = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \varphi = H\varphi.$$

Thus to move from $H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ to K', we have to set

$$K' = \left[-K_2 H_2 H_1^{-1} : K_2 \right],$$

where K_2 is an arbitrary non singular $(p-s)\times(p-s)$. In the same way, to move from $K' = \begin{bmatrix} K_1 & K_2 \end{bmatrix}$ to H_1 we have to set $H = \begin{bmatrix} H_1 \\ -K_2^{-1}K_1H_1 \end{bmatrix}$, where H_1 is an arbitrary non singular $(s\times s)$matrix.

2.1 Podivinsky and Psaradakis corrections to the tests for linear restrictions: "A variation on a theme "

The Johansen (1988) simulated critical values are based on the percentiles of the appropriate asymptotic distribution, and may not be appropriate when used with relatively small sample sizes. In the literature a lot of work has been done on the procedure for inference in cointegrated systems. Podivinsky (1992) and Psaradakis (1993) investigated the adequacy of these asymptotic critical values in moderately sized samples.

They consider a simple DGP with limited number of lags, and just one cointegrating vector. Their simulation analysis indicate that the asymptotic χ^2 distributed LR tests are quite heavily affected by the size of the sample. Accordingly, they proposed small sample adjustments respectively for an F-type test and for the LR criterion and the $Wald$ test.

First, consider the Podivinsky (1992) approximate F-type test. If again we denote estimation under the null by tilde, and unrestricted estimation by a circumflex, and

$$\widehat{S} = \Pi_{i=1}^{r}\left(1-\widehat{\lambda}_i\right),$$

$$\widetilde{S} = \Pi_{i=1}^{r}\left(1-\widetilde{\lambda}_i\right),$$

then the F-type statistics for testing the linear restriction hypothesis $\beta = H\varphi$ is

$$F(\beta) = \frac{\left(\tilde{S}-\widehat{S}\right)/\left(r\left(p-s\right)\right)}{\widehat{S}/\left(T-l\right)}$$

where l is the number of parameters estimated subject to the maintained hypothesis $\Pi = \alpha\beta'$. In our case $l = 2pr - r^2 + p$, when estimating α, β, and μ. Then $F(\beta)$ is approximately distributed as $F\left(r\left(p-s\right), T-l\right)$.

Psaradakis (1993) proposes the application of LR and $Wald$ tests adjusted by certain correction factors. Letting

$$C(\beta) = (l/p) + (1/2)\left[p - r\left(p-s\right)/p + 1\right],$$

the modified statistics are defined as

$$\begin{aligned} LR_c(\beta) &= LR(\beta)\left[T-(l/p)\right]/T, \\ LR_a(\beta) &= LR(\beta)\left[T-C(\beta)\right]/T, \\ W_c(\beta) &= W(\beta)\left[T-(l/p)\right]/T, \end{aligned}$$

where the $LR_c(\beta)$ and $W_c(\beta)$ are obtained by replacing T by $T-(l/p)$ in standard likelihood ratio test.

Monte Carlo evidence in Podivinsky (1992) indicates that the application of the modified F-type test is worthwhile, since improvement are shown with respect to the size properties of LR tests proposed by Johansen (1988). These results are mainly confirmed by Psaradakis (1993), but in addition he shows that the small-sample behavior of LR statistics may be improved by the use of simple scale corrections as indicated above.

More recent work in the literature points out that the problem of size distortion can be substantial when more complex DGP are considered (i.e. when $r > 1$, and more lags and seasonal dummy are inserted), see for instance Fachin (1997), Jacobson and Gredenhoff (1998). As Nielsen (1998) suggests the problem arises because the finite sample distribution of the LR test depends continuously on the nuisance parameters so the test is not asymptotically similar.

To improve the approximation to the asymptotic χ^2 distribution Bartlett adjustment of the likelihood ratio test statistics has recently received interest in this contest.(See for instance Johansen (1999) and Nielsen (1998) Jacobson and Larsson (1996)). However, although Bartlett correction is quite effective in correcting the size distortion of the test, it is not addressed to increase the power of the tests, and it may lead to a loss in power.

Instead of modifying the test statistic an alternative method is considering a corrected distribution which is closer to the true null distribution of our test statistic than the first order limiting distribution. This is usually done by replacing the critical values of the limit distribution with transformations of critical values obtained from the Edgeworth expansions of the distribution function. Larsson (1999) applies multivariate saddlepoint techniques to approximate small sample corrections of the lower tails of the distributions for the LR statistic.

However, this approach is analytically rather demanding. In this sense estimating critical values using simulated-based method is a plausible numerical alternative.

3 The bootstrap test

A key objective in the classical testing of statistical hypotheses is achieving good power while controlling the size of the tests. As seen above the first-order asymptotic approximation can be very inaccurate when we are dealing with small samples. One reason is that for the asymptotic theory to be valid it is necessary that p-value function does not depend on the DGP, which is not usually the case in small samples. As a result, the true and the nominal probabilities that a test rejects a correct H_0 can be very different when the p-value is obtained from the asymptotic distribution of the test statistic. Since, the bootstrap distribution is able to mimic possible skewness of the finite sample distribution it may account for deviations of the actual distribution from the χ^2 distribution. Therefore, it can be used to approximate the finite-sample distribution of the tests considered above.

As seen above, the LR and W test proposed by Johansen (1988) and Johansen and Juselius (1990) enable a researcher to test for linear restrictions on β after having accepted cointegration among variables and Podivinsky (1992) and Psaradakis (1995) propose small sample adjustment for these tests and for an $F-$type test. In this paper, we investigate on the size distortion of these tests in finite sample. Moreover, we are interested in analysing the robustness of the bootstrap test to misspecification in the number of cointegrating relationships. In particular we evaluate the bootstrap tests via Monte Carlo simulation experiments in situations where there is a possible mismatch between the number of cointegrating vectors entering the restricted model and the number of cointegrating vectors entering the DGP (i.e. the number of cointegrating relationship is underfitted or overfitted).

The model estimated is a $VAR(1)$ defined by

$$\Delta y_t \;= \Pi y_{t-1} + \mu + \epsilon_t, \tag{3.1}$$

where y_t, and y_{t-1} are (4×1) vectors, μ is a vector of intercepts and $\epsilon_t \backsim i.i.d.N\,(0, I)$

When testing for linear restrictions on cointegrating vectors, the true DGP is not known. Since the null model, and consequently the DGP is unknown, the estimated DGP is used. In our case the estimated error correction model is

$$\Delta y_t = \widehat{\alpha}\widehat{\beta}' y_{t-1} + \hat{\mu} + \hat{\varepsilon}_t \tag{3.2}$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are the estimates for a given cointegrating rank r.

The idea behind the parametric bootstrap is to approximate the finite sample distribution of the $\widehat{LR}, \widehat{W}, \widehat{F}-$type tests by drawing several B bootstrap realizations $\left\{\widehat{LR}_i^{\,*}\right\}, \left\{\widehat{W}_i^{\,*}\right\}$,or $\left\{\widehat{F}_i^{\,*}\right\}$ for $i = 1, 2, ..., B$ bootstrap samples $\left\{\left(\Delta y^*, y^*_{t-1}\right)_i\right\}$. In order to do this we re-sample the residuals $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_t)$ from (3.2). Denote the bootstrap sample $(\varepsilon^*_1, ..., \varepsilon^*_t)$. The bootstrap algorithm can be summarised as follows:

1) Estimate the error correction model given by (3.2) and compute $\widehat{LR}$, $\widehat{W}$, $\widehat{F}$ as described in Section 2.

2) Re-sample the residuals from $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)$ with replacement to obtain a bootstrap sample $(\varepsilon^*_1, ..., \varepsilon^*_T)$. Generate the bootstrap sample $(y^*_1, ..., y^*_T)$ recursively from $y_0 = 0$ and $(\varepsilon^*_1, ..., \varepsilon^*_t)$ using the estimated restricted model

$$\Delta y_t = \tilde{\alpha}\tilde{\beta}' y^*_{t-1} + \hat{\mu} + \varepsilon^*_t$$

where $\tilde{\alpha}$ and $\tilde{\beta}$ denote the restricted estimates under the null hypothesis $\beta = H\varphi$.

3) Compute the bootstrap replication of $\left\{\widehat{LR}^*\right\}$, $\left\{\widehat{W}^*\right\}$,or $\left\{\widehat{F}^*\right\}$, using $(y_1^*, ..., y_t^*)$

4) Repeat steps 2-4 B times. Defining the bootstrap $p-$values function by the quantity

$$p^*\left(\hat{\theta}\right) = B^{-1}\sum_{i=1}^{B} I\left(\theta^* \geq \hat{\theta}\right) \tag{3.3}$$

where $i = 1, ...B$, $\hat{\theta}$ is the test statistic considered, and $I(\cdot)$ is the indicator function that equals one if the inequality is satisfied and zero otherwise.

6) Reject the null hypothesis if the selected significance level exceeds $p^*\left(\hat{\theta}\right)$.

Therefore, in this way we approximate the distribution of $T^{1/2}\left(\hat{\theta} - \theta\right)$ by the bootstrap distribution of $T^{1/2}\left(\hat{\theta}^* - \hat{\theta}\right)$. Asymptotic validity of the bootstrap requires that with probability one the asymptotic distribution of $T^{1/2}\left(\hat{\theta}^* - \hat{\theta}\right)$ conditional on $\{F_t : t \geq 1\}$ equals the distribution of $T^{1/2}\left(\hat{\theta} - \theta\right)$.

In the literature it has been established that the bootstrap provides a higher-order asymptotic approximation to critical values based on "smooth" statistics. A further refinement of the order $T^{-1/2}$ can be obtained in the case of an asymptotically pivotal statistic. As seen before $LR(\beta)$ and $W(\beta)$ are asymptotically pivotal since they are asymptotically distributed as χ^2. Therefore, the we may expect refinements of order T^{-1}.

4 Design of the Monte Carlo experiments

This section deals with the design of simulation experiments. In order to keep an high degree of experimental control the DGP used are simple VAR(1) processes with small dimension. We consider three different DGP , the first is given by: $DGP1$:

$$\begin{aligned} \Delta\mathbf{y}_{1t} &= \boldsymbol{\epsilon}_{1t}, \\ \Delta\mathbf{y}_{2t} &= \boldsymbol{\epsilon}_{2t}, \end{aligned}$$

where $\epsilon_t = \begin{bmatrix} \epsilon'_{1t} & \epsilon'_{2t} \end{bmatrix}' \approx i.i.d.N(0, \sum)$, $\mathbf{y}_{2t}, \mathbf{y}_{1t}$ are (2×1) vectors and $\sum$ is a (4×4) matrix. The variance-covariance matrix of the disturbances is set to a unit matrix throughout. So, we have four unrelated random walks and $r = 0$.

The second GDP is given by $DGP2$:

$$\begin{aligned}
\Delta y_{1t} &= \epsilon_{1t}, \\
\Delta y_{2t} &= \epsilon_{2t}, \\
\Delta y_{3t} &= \epsilon_{3t}, \\
y_{4t} &= \beta_{23} y_{2,t-1} + \beta_{33} y_{3,t-1} + \beta_{43} y_{4,t-1} + \epsilon_{4t},
\end{aligned}$$

with $\beta_{23}, \beta_{33,} \beta_{43} < 1$, and $\epsilon_t = \begin{bmatrix} \epsilon_{1t} & \epsilon_{2t} & \epsilon_{3t} & \epsilon_{4t} \end{bmatrix}' \approx i.i.d.N(0, \mathbf{I})$. So that we have one cointegrating vector $\begin{bmatrix} 0 & \beta_{23} & \beta_{33} & \beta_{43} - 1 \end{bmatrix}'$.

The third is given by $DGP3$:

$$\begin{aligned}
\Delta y_{1t} &= \epsilon_{1t}, \\
\Delta y_{2t} &= \epsilon_{2t}, \\
y_{3t} &= \beta_{22} y_{2t-1} + \beta_{32} y_{3t-1} + \beta_{42} y_{4t-1} + \epsilon_{3t}, \\
y_{4t} &= \beta_{23} y_{2t-1} + \beta_{33} y_{3t-1} + \beta_{43} y_{4t-1} + \epsilon_{4t},
\end{aligned}$$

with $\epsilon_t = \begin{bmatrix} \epsilon_{1t}, & \epsilon_{2t} & \epsilon_{2t}, & \epsilon_{3t} \end{bmatrix}' \approx i.i.d.N(0, \mathbf{I})$. So that we have two cointegrating vectors.

Two possible situations are investigated:

a) The model is correctly specified:

-DGP is $DGP2$ and in model estimated $r = 1$

-DGP is $DGP3$ and in model estimated $r = 2$

b) The number of cointegrating vectors is over-fitted or under-fitted:

-DGP is $DGP1$ but we are assuming $r = 1$

-DGP is $DGP2$ but we are assuming $r = 2$

-DGP is $DGP3$ but we are assuming $= 1$

All simulations were carried out on 400MHz Pentium PC using the matrix programming language GAUSS Version 3.2.32. The random numbers were generated by the function *rndns*. For each sample we calculated the six tests considered above in a VAR(1) model with intercept and we generated $B = 400$ bootstrap samples according to the algorithm given in the previous section. Then the bootstrap is evaluated by Monte Carlo, and each

Monte Carlo experiment is based on 1,000 replications. Obviously, the level of accuracy of the experiment could be improved using a larger number of bootstrap replications and a larger number Monte Carlo replicates, (a 95% confidence interval around a 5% nominal size is $[3.6 - 6.4]$ for 1,000 replicates). However, 1,000 replications with B=400, T =800, uses 3.2×10^8 random deviates of the 4×10^9 distinct deviates available from *rndns*. For the non-bootstrapped tests, 100,000 Monte Carlo replications were used. The random number generator was restarted for each T value.

According to Davidson and MacKinnon (1996b), in some situations $B = 400$ is the smallest number of replications that guarantees a reasonable trade off between the gains in power and computational costs. However, increasing the number of bootstrap replications involves increasing computational costs, consequently it is necessary to reduce them to a number that minimizes the loss of power. To explore the sensitivity of the estimated size to the number of bootstrap replications we made a pilot experiment for $B \in \{100, 200, 400, 600, 800, 1200\}$ (the results are reported in Appendix 2) and this simulation confirms that $B = 400$ is adequate for our purposes.

5 Preliminary Monte Carlo Results

In Table 1-5 we report the results of the Monte Carlo experiment with respect to the sizes of the tests. The notation is the following: T is the sample size, LR is the uncorrected likelihood ratio test; LC_c and LR_a are the likelihood ratio tests adjusted by Psaradakis (1993) correction factors; W and W_c are respectively the uncorrected and corrected $Wald$ tests; F is the F-type test proposed by Podivinsky (1992). Therefore, from column 2 to column 7 we report the Monte Carlo estimated sizes, and column 8 and 9 report the bootstrap corrected likelihood ratio and the bootstrap Wald tests. These results are preliminary, and the conclusions thus tentative; so far we have only investigated a few points in parameter space.

The first thing it is important to note is that the empirical sizes of $BootLR$ are equal to those for $BootF$, the bootstrap corrected F statistic, as the F statistic is a one to one function of the LR statistic. Hence, the columns of $BootF$ have been omitted.

We find the poorest performance for both the W and W_c versions of the $Wald$ statistic. A reason for this may be the non-invariance property of the $Wald$ test.

The invariance properties states that the decision reached by the hypothesis testing procedure should remain unchanged under transformation of the parameters. So, the Wald statistic varies with the parametrisation of the null hypothesis being tested and its numerical value can vary greatly according to the specification of H_0 that is being used. As a result, the finite sample level of the $Wald$ test can be greatly different from the nominal level, and using the asymptotic distribution of the $Wald$ statistic can be misleading. In this sense the bootstrap provides a better approximation to the finite sample distribution than first order asymptotic theory and therefore smaller size distortion. The problem of the non invariance of the Wald test has been discussed by Gregory and Veall (1985), Lafontaine and White (1991) and Horowitz (1997).

A second issue is the following: since for practical purposes any bootstrap procedure involves computational costs, might an investigation avoid resampling methods and rely on the application of LR and $Wald$ tests adjusted by the correction factors proposed by Psaradakis (1993) or on the Podivinsky's (1992) F-type test?

Monte Carlo evidence in Table 1 and 2 in some sense confirms their results in the case where the number of cointegrating vectors is correctly specified and this is particularly true for the F-type. For the W_c test the actual significance level is much higher then the 5% nominal level, and as a consequence the true null hypothesis will be rejected too often.

The overall impression is that when the number of cointegrating relationships is correctly specified the size distortion asymptotically vanishes, and the asymptotic theory is uniformly satisfactory for $T \geq 200$. However, for $T < 200$ the only tests that provide nearly exact α level is Podivinsky's (1993) F-type test, $BootLR$ and $BootW$.

Table 1.
Sizes for tests of $\beta_{1,1} = 0$ assuming correct cointegrating rank of 1.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.1000	0.0907	0.0827	0.1860	0.1740	0.0611	0.046	0.050
75	0.0804	0.0747	0.0699	0.1290	0.1230	0.0567	0.055	0.05
100	0.0711	0.0672	0.0642	0.1040	0.0994	0.0549	0.050	0.049
150	0.0639	0.0614	0.0594	0.0835	0.0810	0.0539	0.049	0.048
200	0.0607	0.0590	0.0576	0.0746	0.0725	0.0532	0.050	0.048
400	0.0544	0.0537	0.0529	0.0605	0.0598	0.0510	0.044	0.043
800	0.0511	0.0507	0.0504	0.0543	0.0538	0.0496	0.052	0.05

$DGP2 : \beta_{23} = 0.5, \beta_{33} = 0.4, \beta_{43} = 0.1$

Table 2.
Sizes for tests of $[\beta_{11}, \beta_{21}] = [0, 0]$ assuming correct cointegrating rank of
2

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.1000	0.0906	0.0824	0.171	0.160	0.0597	0.045	0.047
75	0.0804	0.0749	0.0705	0.125	0.118	0.0574	0.062	0.061
100	0.0738	0.0699	0.0664	0.105	0.101	0.0570	0.055	0.059
150	0.0666	0.0640	0.0620	0.0868	0.0841	0.0560	0.048	0.051
200	0.0621	0.603	0.0589	0.0772	0.0752	0.0545	0.054	0.059
400	0.0567	0.0558	0.0552	0.0640	0.0632	0.0532	0.049	0.052
800	0.0529	0.0523	0.0519	0.0564	0.0560	0.0510	0.058	0.058

$DGP3 : \beta_{23}, \beta_{33}, \beta_{43}$ as Table 1, $\beta_{22} = 0, \beta_{32} = 0.9, \beta_{42} = 0.1$

Table3.
Probability of rejecting $\beta_{1,1} = 0$ when true but assuming $r = 1$ when $r = 0$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.412	0.394	0.379	0.681	0.673	0.327	0.132	0.191
75	0.405	0.393	0.383	0.674	0.668	0.351	0.124	0.176
100	0.406	0.398	0.390	0.672	0.668	0.367	0.152	0.215
150	0.401	0.395	0.390	0.670	0.667	0.375	0.138	0.172
200	0.399	0.395	0.391	0.669	0.667	0.380	0.138	0.205
400	0.398	0.396	0.394	0.666	0.665	0.388	0.122	0.204
800	0.397	0.396	0.396	0.664	0.664	0.393	0.134	0.211

$DGP1$

Table 4.
Probability of rejecting $\beta_{1,1} = 0$ when true, but assuming $r = 1$ when $r = 2$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.0998	0.0817	0.0727	0.189	0.165	0.0345	0.045	0.044
75	0.0792	0.0685	0.0636	0.130	0.117	0.0403	0.061	0.062
100	0.0703	0.0630	0.0591	0.105	0.0961	0.0424	0.061	0.057
150	0.0631	0.0583	0.0560	0.0843	0.0787	0.0453	0.051	0.05
200	0.0603	0.0569	0.0554	0.0758	0.0716	0.0476	0.052	0.047
400	0.0535	0.0520	0.0513	0.0606	0.0590	0.0478	0.042	0.043
800	0.0519	0.0510	0.0506	0.0550	0.0543	0.0488	0.056	0.054

$DGP3$, as Table 2

Table 5.
Probability of rejecting $[\beta_{11}, \beta_{21}] = [0, 0]$ when true, but assuming $r = 2$ when $r = 1$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.375	0.337	0.318	0.706	0.691	0.207	0.103	0.179
75	0.357	0.332	0.319	0.689	0.679	0.252	0.108	0.154
100	0.351	0.333	0.323	0.681	0.673	0.274	0.100	0.163
150	0.342	0.330	0.324	0.677	0.671	0.293	0.100	0.158
200	0.338	0.329	0.324	0.673	0.669	0.301	0.090	0.165
400	0.334	0.330	0.328	0.668	0.666	0.317	0.098	0.168
800	0.332	0.330	0.329	0.666	0.665	0.323	0.123	0.162

$DGP2$, as Table 1

Tables 3-5 report the Monte Carlo sizes for the tests considered in situations where the number of cointegration vectors is over-fitted or under-fitted. The overall impression is that the bootstrap tests appear to be robust to misspecifications, whereas the asymptotic tests in most occasions is not satisfactory. For example, in Table 3 the LR test has empirical sizes that vary between 39.7% and 41.2%. In this case the $BootLR$ though still affected by size distortion has $12.2\% \leq \alpha \leq 15.2\%$.

As far as the Podivinsky's test (1992) and Psaradakis (1993) corrections to Johansen's (1988) and Johansen and Juselius (1990) tests are concerned they are heavily affected by misspecifications, so in this case the F-type test cannot be considered as an alternative inference procedure to the bootstrap test.

The same considerations above apply for the Wald test, where the reduction is particularly remarkable, since in the over-fitting case the Monte Carlo estimated sizes are between 66.4% and 68.1% in Table 3 and between 66.6% and 70.6% in Table 5.

The size distortion caused by overfitting is so large that it calls into question the use of the tests. A size greater than.0.5 implies that one is more often wrong than right when using the test.

One explanation of the excessive size of the Wald test when overfitting the cointegrating rank is as follows. In the case of a single constraint, $\beta_{11} = 0$, $K' = \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \end{array} \right]$,with $r = 1$ assumed, one can write the Wald test in the form

$$\widehat{W} = \widehat{\beta}_{11}^2 \widehat{\sigma}_1^2 / \sum_{j=2}^{p} \widehat{v}_{1j}^2$$

where

$$\widehat{V} = \left[\begin{array}{cccc} \widehat{\beta}_1 & \widehat{v}_2 & \cdots & \widehat{v}_p \end{array} \right]$$

$\widehat{\beta}_{11}$ is the first element of $\widehat{\beta}_1$and

$$\sigma_1^2 = \widehat{\lambda}_1 / (1 - \widehat{\lambda}_1).$$

Under $H_0 : K'\beta = 0$, if $r = 0$ $\widehat{\beta}_{11}$ is an $O_p(T^{-1/2})$ estimator of 0; if $r = 1$ it is $O_p(T^{-1})$. Thus $\widehat{\beta}_{11}$ is more variable, and $\widehat{\beta}_{11}^2$ on average larger, if $r = 0$. While the other two terms in $\widehat{W}$ also change, both being on average smaller when $r = 0$, in simulations it seems that the effect on $\widehat{\beta}_{11}^2$ dominates by an order of magnitude.

Considering the general case of $\widehat{W}$ as defined in (2.3), this intuition suggests that overfitting can be regarded as misclassifying the columns of $\widehat{V}$. If one assumes that the rank of Π is $r+1$ when it is r, one erroneously regards $\widehat{v}_{r+1}$ as $\widehat{\beta}_{r+1}$ and includes it in the 'numerator' of $\widehat{W}$ rather than the 'denominator'. As it is $O_p(T^{-1/2})$ rather than $O_p(T^{-1})$, and its 'square' enters $\widehat{W}$, this shifts the distribution of $\widehat{W}$ to the right.

This only explains the behaviour of the likelihood ratio directly insofar as $\widehat{W}$ and $\widehat{LR}$ are correlated. Turning to the bootstrap tests, when overfitting their size is around 10%, and does not converge to the correct value. Why does boostrapping fail?

In the correctly specified model $\beta' y_{t-1}$ and $\alpha\beta' y_{t-1}$are stationary. If we overfit, we include in $\widehat{\beta}' y_{t-1}$ linear combinations of y_{t-1} which are not stationary, and when generating Δy_t^* from the resampled residuals $\Delta y_t - \widehat{\mu} - \widehat{\alpha}\widehat{\beta}' y_{t-1}$ both the residuals and Δy_t^* will be $I(1)$. Thus bootstrapping fails. The size is not as distorted as it is for the non-bootstrapped tests, but there is no reason to think the power properties will be desirable.

One might try to recover the situation by using the parametric bootstrap. If one does so, the residuals are replaced by independent and identically distibuted Normal vectors, and Δy_t^* has the properties implied by the cointegrating rank r assumed and $\beta_1 = 0$. However, the test statistic, $\widehat{W}$ or $\widehat{LR}$, being compared with this bootstrap distribution is calculated from data with a smaller r, and the equivalent of Table 5 for the parametric bootstrap shows sizes from 23% to 63%.

When underfitting, as in Table 4, the performance of all the tests is much better, although the F test is a little undersized.

5.1 Response surface regressions

As was seen in the previous section Johansen's test for linear restrictions are heavily affected by misspecification in the number of cointegrating relationships. However, the size distortion when underfitting is not very different, either for magnitude or direction, from the size distortion when the model is correctly specified. This suggests that the difference between the nominal and the empirical size is more likely to be due to finite sample effects than to misspecification. In fact, in both cases the size of the tests depends on the sample size and on the many parameters of the model, and in both cases this dependence asymptotically vanishes (even though the adjustment is quite slow).

By contrast, when overfitting the asymptotic theory does not help. However, Johansen (1999) derives a Bartlett correction factor[1] for the likelihood ratio test for linear hypotheses on the cointegrating space which depends on the sample size and the parameters of the model. Since we are interested in analysing situations where the rank test is giving incorrect answers we calculate Johansen's correction factor for the model with $DGP = DGP2$ to identify regions of the parameter space where this error is likely.

The correction factor for the hypothesis $\beta = H\varphi$ is given by

$$\frac{E\left[-2\log LR|\alpha_{\perp}'\varepsilon\right]}{r\left(p-s\right)} =$$

$$1+\frac{1}{T}\left[\left(p_d+kp\right)+\frac{1}{2}\left(p+1+s-r\right)\right]+\frac{1}{Tr}\left[\left(p-2r+s+2p_d-1\right)v+2c\right],$$

where p is the dimension, k is the lag length, r the cointegrating rank, p_d the number of deterministic trends,

$$v=\frac{-\alpha'\beta\left(2+\alpha'\beta\right)}{\beta'\Omega\beta\alpha'\Omega^{-1}\alpha},$$

$$c=-2\frac{\alpha'\beta\left(1+\alpha'\beta\right)}{\beta'\Omega\beta\alpha'\Omega^{-1}\alpha},$$

and for $DGP = DGP2$ in the previous section the correction factor is given by

$$\frac{E\left[-2\log LR|\alpha_{\perp}'\varepsilon\right]}{r\left(p-s\right)}=1+\frac{1}{T}\left[8+\frac{14+9\beta_{43}}{\beta_{23}^2+\beta_{33}^2+\beta_{43}^2}\right]$$

[1] 'Bartlett'or 'Bartlett-type' correction provides a better approximation to the limiting distribution of a statistic by adjusting the statistic so that its finite sample distribution has the same mean as the limiting distribution. Johansen's (1999) correction factor is based on two ideas:

1) Inference on β is asymptotically independent of inference on α. So calculation can be done fixing the parameter α.

2) Since $\hat{\beta}$ is asymptotically mixed Gaussian and the asymptotic inference involves conditioning on the asymptotic common trends, he conditions on the common trends when making inference. See Johansen (1999) for more details.

for $\alpha' = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$ and $\beta' = \begin{bmatrix} 0 & \beta_{23} & \beta_{33} & \beta_{43} \end{bmatrix}$, the matrix $\alpha'\beta$ reduces to the scalar $\alpha'\beta = \beta_{43}$. So that the parameter β_{43} is the most influential.

In $DGP2$, in the original simulation $\beta_{23}, \beta_{33}, \beta_{43}$, were set as $\beta_{23} = 0.5, \beta_{33} = 0.4, \beta_{43} = 0.1$. In order to evaluate the sensitivity of the empirical sizes to variations of the parameter values and the sample size of the DGP we estimate the response surface regression of $size = f(\beta_{33}, \beta_{43}, T)$. However, before doing it we need to analyse the constraint on the parameters in order to preserve the stability of the system. Hence, we calculate the characteristic polynomial for $DGP = DGP2$:

$$\begin{aligned} A(z) &= I(1-z) - \alpha\beta' z = \\ &= \begin{pmatrix} 1-z & 0 & 0 & 0 \\ 0 & 1-z & 0 & 0 \\ 0 & 0 & 1-z & 0 \\ 0 & -\beta_{23}z & -\beta_{33}z & 1-(1+\beta_{43})z \end{pmatrix}, \end{aligned}$$

such that $|A(z)| = (1-z)^3 (1-(1+\beta_{43})z) = 0$ if and only if

$$z = \begin{cases} 1 \quad \text{or} \quad 1/(1+\beta_{43}), & \text{if } \beta_{42} \neq -1 \\ 1, & \text{if } \beta_{43} = -1 \end{cases}$$

Therefore, if β_{43} is in the interval $(-2, 0]$, then the process Y_t is $I(1)$. (In the case $\beta_{43} = 0$, the process is a pure $I(1)$ process which does not cointegrate. For $\beta_{43} < -2$ or $\beta_{43} > 0$ the process Y_t is explosive).

SECTION TO BE FINISHED

5.2 Power

In this section we consider the power of the bootstrap tests. Since, in 1,000 Monte Carlo replications, the bootstrapped procedure has power 1 if $T \geq 75$, we report in Table 6 the rejection frequencies at 5% level only for $T = 50$ and $DGP = DGP3$.

Table 6. Rejection frequencies for $\beta_1 = \mathbf{0}$,conditional on acceptance of one or two cointegrating vector, Max and Trace tests

$T = 50$	$H_0 : r = 0$	$H_0 : r = 1$	$H_0 : r = 2$
$BootLR$	1	0.943	0.996
$BootW$	0.999	0.995	0.720.

An inspection of the table above seems to reveal that $BootLR$ and $BootW$ have power. However, an informative investigation of parameter space requires the use of asymptotic theory.

SECTION TO BE COMPLETED

6 Conclusion

This paper propose the use of bootstrap hypothesis testing as a way of improving inference for linear restriction on cointegrating space. We analyse the sensitivity of the LR , $Wald$, and $F-$type to misspecification on the number of the cointegrating vectors, and both the cases of over-fitting and under-fitting have been considered. Particular attention has been given to the analysis of small sample properties of these tests.

The Monte Carlo evaluation of the bootstrap tests show that the resampling procedure provides empirical sizes which are much closer to the nominal size. This is particular true when $T < 200$. One reason for this might be the poor correspondence in the small and moderate size between the exact distribution of the test statistic and its reference distribution. In addition, we find that the size distortion of the bootstrap $Wald$ test converges to zero even for a sample size $T = 50$. Therefore, for practical purposes the bootstrap procedure for this test is strongly recommended.

Such a partial Monte Carlo investigation makes any conclusions provisional. What is suggested is that if there is any uncertainty about the cointegrating rank r, tests on β should be conducted under different assumptions about r. If the conclusions change when r is increased, especially if the bootstrap test results start to diverge from the those of the asymptotic tests, then only the results for smaller r should be relied upon. This is in contrast to the suggestion in Podivinsky (1998), that "possible overspecification of the number of variables in a model has less serious consequences" (then underspecification): we (provisionally) argue that overestimating cointegrating rank seriously biases tests on β. Be generous with the variables, by all means, but overfitting r leads to failure of the testing procedure.

References

[1] Andrade, I.C., O'Brien, R.J. and Podivinsky, J.M. (1994) Cointegration Tests and Mean Shifts, University of Southampton Discussion Paper 9405.

[2] Basawa, I. V., Mallik, A. K. , McCormick, W.P. and Taylor, R.L. (1989), Bootstrapping explosive autoregressive processes. Annals of Statistics 17,1479-1486.

[3] Basawa, I. V., Mallik, A. K. , McCormick, W.P. and Taylor, R.L. (1991a), Bootstrapping unstable first order autoregressive processes. Annals of Statistics 19,1098-1101.

[4] Basawa, I. V., Mallik, A. K. , McCormick, W.P. and Taylor, R.L. (1991b), Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive processes, Communication in Statistics: theory and methods, Vol.20 pp. 1015-1026.

[5] Bose A. (1988), Edgeworth correction by bootstrap in autoregressions. Annals of statistics, 16, 1709-1722.

[6] Davidson, R. and MacKinnon, J.G. (1996a), The Size and Power of Bootstrap Tests. Discussion paper, Queen's University, Kingston, Ontario.

[7] Davidson, R. and MacKinnon, J.G. (1996b), The Power of Bootstrap Tests. Discussion paper, Queen's University, Kingston, Ontario.

[8] Davidson, R. and MacKinnon, J.G. (1998),The size Distortion of Bootstrap Tests. Econometric theory, 15, 1999, 361-376.

[9] Efron, B. (1979) Bootstrap method: another look at the Jackknife. Annals of Statistics. Vol. 7, pp.1-26

[10] Fachin, S. (1997) Bootstrap and asymptotic tests of long-run relationships in cointegrated systems, discussion paper, University if Rome

[11] Fisher, N.I. and Hall, P. (1990), On bootstrap hypothesis testing. Australian Journal of Statistics, Vol. 84, pp 996-1002.

[12] Gredenhoff, M., Jacobson, T. (1998). Bootstrap approximate finite sample distributions for tests of linear restrictions on cointegrating vectors. Discussion paper, Stockholm School of Economics.

[13] Gregory, A.W. and Veall, M.R. (1985), Formulating Wald tests of non-linear restrictions. Econometrica, 53:1465-8.

[14] Hall, P. (1992), The bootstrap and Edgeworth expansion, New York, Springer-Verlag.

[15] Harris, R.I.D. (1992) Small sample testing for unit roots. Oxford Bulletin of Economics and Statistics. Vol. 54, pp. 615-625.

[16] Horowitz, J.L. (1994), Bootstrap-based critical values for the Information Matrix test, Journal of Econometrics, Vol. 61, pp 395-441.

[17] Horowitz, J.L. (1997), Bootstrap methods in econometrics: theory and numerical performance, in D.M. Kreps and K.F. Wallis, Eds. Advances in Economics and Econometrics: theory and applications, 7th World Congress of the Econometric Society Vol. III, Cambridge University Press.

[18] Jacobson and Larsson (1996) Bartlett Corrections in Cointegration Testing. Discussion Paper University of Stockholm.

[19] Jacobson, T. (1998) Bootstrap Testing and Approximate Finite Sample Distributions for Tests of Linear Restrictions on Cointegrating Vectors, paper delivered at the European Meeting of the Econometric Society, Berlin.

[20] Johansen, S. (1988), Statistical analysis of cointegrated vectors, Journal of Economic Dynamic and Control 12, 231-254; reprinted as Chapter 7 in R.F.

[21] Johansen, S. and Juselius, K. (1990), Maximum likelihood estimation and inference on cointegration with application to the demand for money. Oxford Bulletin of Economics and Statistics , Vol.52, pp 169-210.

[22] Johansen, S. , Likelihood Inference in Cointegrated Vector Auto-Regressive Models. Oxford University Press,Oxford, 1995. University Press,Oxford, 1995.

[23] Johansen, S.(1999). A small sample correction for tests of hypotheses on the cointegrating vectors. EUI Working Paper ECO No. 99/9.

[24] Lafontaine, F. and White, K. J.(1986). Obtaining any Wald statistic you want. Economics Letters, 21:35-48

[25] Li, H. and Xiao, Z. (1999) Bootstrapping Time Series Regressions with Integrated Processes, mimeo, paper delivered at the Far Eastern Meeting of the Econometric Society, Singapore, July 1999.

[26] Li, H. and Maddala, G.S. (1997) Bootstrapping cointegrating regressions, Journal of Econometrics, Vol. 80,No.2, pp. 297-318.

[27] Nielsen, B. (1999). On the distribution of tests for cointegrating rank. Mimeo, Nuffield Colege

[28] O'Brien, R.J. (1996), The curious case of spurious cointegration. University of Southampton Department of Economics Discussion Paper 9405.

[29] O'Brien, R.J. (1997), Further curiosa in spurious cointegration. University of Southampton Department of Economics Discussion Paper 9704.

[30] O'Brien, R.J. (1999), The curious case of spurious cointegration version for Far Eastern Meeting of Econometric Society, Singapore, 1999 accessible via http://www.soton.ac.uk/∼econweb/biograph/ro.html

[31] Psaradakis, Z. (1994), A comparison of tests of linear hypotheses in cointegrated vector autoregressive model. Economics Letters vol. 45, pp 137-144.

[32] Podivinsky, J.M. (1992), Small sample properties of tests of linear restrictions on cointegrating vectors and their weights, Economics Letters, 39,13-18.

[33] Podivinsky, J.M. (1998), Testing misspecified cointegrating relationships, Economics Letters 60, 1-9

[34] Veall, M.R. (1998), "Application of the Bootstrap in Econometrics and Economic Statistics", in Handbook of Applied Economic Statistics, Edited by Ullah A. and Giles, D.E.A.

Appendix 1: Computation details

The implementation of Johansen's cointegrating tests was not conducted using Johansen original algebra, but using QR and singular value decomposition as employed in O'Brien (1996). So, in this appendix we show how Johansen cointegration analysis can be rewritten in term of QR decomposition [2].

For easy of notation we report here the model in Section 3

$$\Delta y_t = \Pi_1 y_{t-1} + \alpha_t + \epsilon_t$$

where y_t and is y_{t-1} are (4×1), α_i is a vector of intercepts, and $\boldsymbol{\epsilon}_t \approx N(0, I)$. The $VAR(1)$ model can be rewritten as

$$w_t' = \left[\alpha_t', y_{t-1}', \Delta y_t'\right], \tag{A.1}$$

which forms the $t-$th row of the matrix W. Then a QR decomposition of the matrix W yields a Cholesky factorisation[3] R such that such that $R'R = W'W$. We partition

$$R = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}$$

where R_{11} has $p(k-1)+q$ rows and columns, while $R22$and $R33$ are each $(p \times p)$.

Using the Cholesky factorisation we can estimate $\hat{\Pi}$ in a reasonably straightforward way. First note that if $R'R = W'W$, $W = [W_A, W_B]$, and R is conformably partitioned into$\begin{bmatrix} R_{AA} & R_{AB} \\ 0 & R_{BB} \end{bmatrix}$ then,

$$R_{AA}' R_{AA} = W_A' W_A, \tag{A.2}$$

$$R_{AA}' R_{AB} = W_A' W_B, \tag{A.3}$$

and

$$R_{AB}' R_{AB} + R_{BB}' R_{BB} = W_A' W_B. \tag{A.4}$$

[2]For further details see O'Brien (1996).

[3]If A is a positive definite $(m \times m)$ matrix there exists a lower triangular matrix P such that or $A = P'P$. The decomposition $A = P'P$ is called a Cholesky decomposition.

Thus from $(A.3)$ solving for R_{AB}

$$R_{AB} = (R_{AA})^{-1} W_A' W_B, \tag{A.5}$$

solving $(A.4)$ for $R_{BB}' R_{BB}$ and substituting $(A.5)$ in $(A.4)$ we get

$$\begin{aligned} R_{BB}' R_{BB} &= W_B' W_B - R_{AB}' R_{AB} \\ &= W_B' W_B - \left[(R_{AA})^{-1} W_A' W_B \right]' \left[(R_{AA})^{-1} W_A' W_B \right] \end{aligned} \tag{A.6}$$

$$= W_B' W_B - W_B' W_A \left(R_{AA}' R_{AA} \right)^{-1} W_A' W_B$$

$$= W_B' W_B - W_B' W_A \left(W_A' W_A \right)^{-1} W_A' W_B.$$

Identifying R_{AA} with R_{11}, and $\begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}$ with R_{BB}, and conformably partitioning $W = \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix}$ so that W_2 and W_3 each have p columns, we can rewrite $(A.4)$ as

$$\begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix}' \begin{bmatrix} R_{22} & R_{23} \\ 0 & R_{33} \end{bmatrix} = [W_2, W_3]' \left[I - W_1 \left(W_1' W_1 \right)^{-1} W_1' \right] [W_2, W_3]$$

which in Johansen's notation is the product moment matrix

$$T \begin{bmatrix} S_{kk} & S_{k0} \\ S_{0k} & S_{00} \end{bmatrix}.$$

Thus,

$$R_{22}' R_{22} = T S_{kk}, R_{22}' R_{23} = T S_{k0}, R_{23}' R_{23} + R_{33}' R_{33} = T S_{00}. \tag{A.7}$$

Using the latent root of $\hat{\lambda}_i$ and the latent vector $\boldsymbol{\epsilon}_i$ of $C^{-1} S_{k0} S_{00}^{-1} S_{0k} \left(C' \right)^{-1}$, where $CC' = S_{kk}$, then defining $E = [e_1 e_2 ... e_r]$ and $S = \left[\begin{smallmatrix} I_r \\ O \end{smallmatrix} \right]$,we have

$$\hat{\beta} = \left(C' \right)^{-1} E S_r$$

and

$$\alpha(\beta) = S_{0k} \beta \left(\beta' S_{kk} \beta \right)^{-1}$$

this gives us

$$\left(\hat{\beta}' S_{kk} \hat{\beta}\right) = S_r' E' C^{-1} S_{kk} \left(C'\right)^{-1} E S_r = S_r' E' I_k E S_r = S_r' S_r = I_r$$

and

$$\alpha\left(\hat{\beta}\right) = S_{0k} \hat{\beta}.$$

Identifying $\sqrt{T}C$ with R_{22}' from equation $(A.7)$

$$\begin{aligned} C^{-1} S_{k0} S_{00}^{-1} S_{0k} \left(C'\right)^{-1} &= \left(R_{22}'\right)^{-1} R_{22}' R_{23} \left(R_{23}' R_{23} + R_{33}' R_{33}\right)^{-1} R_{23}' R_{22} \left(R_{22}^{-1}\right) \quad (A.8) \\ &= R_{23} \left(R_{23}' R_{23} + R_{33}' R_{33}\right)^{-1} R_{23}' \\ &= I - \left[I + R_{23} \left(R_{33}' R_{33}\right)^{-1} R_{23}'\right]^{-1} \end{aligned}$$

Using the singular value decomposition, let

$$R_{23} R_{33}^{-1} = U \Sigma_R V$$

where $U'U = I_k = V'V$ and Σ_R is diagonal with the singular values σ_i of $R_{23} R_{33}^{-1}$ as its diagonal elements. Thus

$$R_{23} R_{33}^{-1} \left(R_{23} R_{33}^{-1}\right)' = U \Sigma_R^2 U'$$

and

$$R_{23} \left(R_{33}' R_{33}\right)^{-1} R_{23}' (U')^{-1} = U \Sigma_R^2$$

so if u is a column of U, and σ the corresponding diagonal element of Σ_R,

$$R_{23} \left(R_{33}' R_{33}\right)^{-1} R_{23}' u = \sigma^2 u$$

so that σ^2, u are the eigenvalues and the eigenvectors of $R_{23} \left(R_{33}' R_{33}\right)^{-1} R_{23}'$ respectively. Thus rearranging equation $(A.8)$ we have

$$\left\{I - (I + R_{23} \left(R_{33} R_{33}^{-1}\right)^{-1} R_{23}')^{-1}\right\} u = \left\{1 - \left(1 + \sigma^2\right)^{-1}\right\} u$$

so that u is a latent vector and $\left\{1-\left(1+\sigma^2\right)^{-1}\right\} u = \frac{\sigma^2}{1+\sigma^2}$ a latent root of $R_{23}\left(R_{33}' R_{33}\right)^{-1} R_{23}'$. Thus the Johansen required quantities are $\hat{\lambda}_i = \frac{\sigma_i^2}{1+\sigma_i^2}$, and $\hat{\beta} = \left(C'\right)^{-1} E S_r = \sqrt{T} R_{22}^{-1} U S_r$,with

$$\hat{\alpha} = \alpha\left(\hat{\beta}\right) = S_{0k} = \sqrt{T}\left(R_{22}' R_{23}\right)' \hat{\beta} = 1/\sqrt{T} R_{23}' U S_r.$$

Moreover, for the LR likelihood test of $H_0 : \beta = H\varphi$, where $\beta = H\varphi$ is a set of restrictions, with $H(p \times s)$, we can again use a QR decomposition. First, adapting equation $(A.1)$ we have:

$$\underset{(2p\times(s+p))}{\begin{bmatrix} R_{22}H & R_{23} \\ 0 & R_{33} \end{bmatrix}}$$

so that

$$\begin{bmatrix} R_{22}H & R_{23} \\ 0 & R_{33} \end{bmatrix}' \begin{bmatrix} R_{22}H & R_{23} \\ 0 & R_{33} \end{bmatrix} = T \begin{bmatrix} H' S_{kk} H & H' S_{k0} \\ S_{0k} H & S_{00} \end{bmatrix} \tag{A.9}$$

then we can perform a QR decomposition of this matrix to produces, $R_\beta = \begin{bmatrix} R_{\beta_{22}} & R_{\beta_{23}} \\ 0 & R_{\beta_{33}} \end{bmatrix}$ where $R_{\beta_{22}}$ is $(s \times s)$, and $R_{\beta_{33}}$ is $(p \times p)$. Then if we replaces R_{22}, R_{23} and R_{33} in our initial analysis with $R_{\beta_{22}}$, $R_{\beta_{23}}$, and $R_{\beta_{33}}$ this will yield $\lambda_i, \hat{\varphi}_0$, $\hat{\beta} = H\hat{\varphi}_0$ and $\widehat{\alpha}$. Tests on α are handled in a similar way.

Appendix 2: Supplementary simulations

Alternative values for B in Table 1.

Table A1. Sizes(%) for tests of $\beta_1 = 0$ assuming correct cointegrating rank of $r = 1$ and $N = 1000$.

BootLR

T \ B	100	200	400	600	800	1000	1200
50	5.1	4.8	4.7	5.1	5.0	5.2	5.7
75	4.9	6.5*	5.1	5.1	4.7	4.9	6.0
100	5.6	5.2	5.0	6.3	4.7	4.3	5.1
150	5.2	4.7	5.6	4.3	4.7	3.7	5.2
200	4.3	4.3	5.3	4.4	5.0	5.7	5.5

BootW

T \ B	100	200	400	600	800	1000	1200
50	6.9*	6.0	5.6	5.2	5.6	5.5	6.1
75	5.0	6.3	5.6	5.2	4.9	5.4	6.0
100	5.7	5.2	5.2	6.3	5.1	5.1	4.6
150	4.9	5.3	6.0	4.6	4.8	3.4*	5.2
200	4.6	4.6	5.4	4.6	5.1	5.6	5.8

Monte Carlo precision $\pm 1.35\%$; values marked * are significantly different from the nominal size of 5% when testing at a 5% level of significance. Time required, 18.5 hours (400 MHz Pentium).

For the bootstrapped likelihood ratio ($BootLR$), $B = 400$ is only slightly improved on by $B = 800$. The results for the bootstrapped $Wald$ test, $BootW$, also suggest $B = 400$ is a reasonable compromise.

Table A2

Probability of rejecting $[\beta_{11}, \beta_{21}] = [0, 0]$ when true, but assuming $r = 2$ when $r = 0$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.573	0.532	0.510	0.880	0.872	0.371	0.14	0.203
75	0.564	0.535	0.520	0.878	0.872	0.438	0.128	0.190
100	0.565	0.543	0.532	0.877	0.872	0473	0.152	0.206
150	0.559	0.545	0.538	0.874	0.871	0.500	0.117	0.171
200	0.556	0.546	0.541	0.875	0.873	0.513	0.145	0.201
400	0.553	0.548	0.545	0.875	0.874	0.533	0.145	0.176
800	0.553	0.551	0.549	0.874	0.873	0.543	0.145	0.196

$DGP1$

Table A3. Parametric bootstrap

Probability of rejecting $\beta_1 = [\beta_{11}, \beta_{21}]' = 0 \ \ (2 \times 1) = 0$ when true, but assuming $r = 2$ when $r = 1$.

T	LR	LR_c	LR_a	W	W_c	F	$BootLR$	$BootW$
50	0.392	0.357	0.333	0.728	0.713	0.221	0.226	0.467
75	0.370	0.345	0.335	0.698	0.682	0.265	0.238	0.489
100	0.357	0.335	0.324	0.669	0.664	0.269	0.228	0.507
150	0.313	0.307	0.297	0.671	0.668	0.265	0.232	0.557
200	0.348	0.338	0.332	0.681	0.672	0.300	0.259	0.567
400	0.315	0.310	0.304	0.655	0.653	0.292	0.268	0.586
800	0.338	0.336	0.333	0.675	0.674	0.327	0.304	0.631

$DGP2$, as Table 1