Dipartimento Economia e Statistica Cognetti de Martiis

Working Paper Series

07/20

IMPROVEMENT ON THE LR TEST STATISTIC ON THE COINTEGRATING RELATIONS IN VAR MODELS: BOOTSTRAP METHODS AND APPLICATIONS

ALESSANDRA CANEPA

Department of Economics and Statistics "Cognetti de Martiis"
Campus Luigi Einaudi, Lungo Dora Siena 100/A, 10153 Torino (Italy)
www.est.unito.it

UNIVERSITÀ DEGLI STUDI DI TORINO

The Department of Economics and Statistics "Cognetti de Martiis" publishes research papers authored by members and guests of the Department and of its research centers. ISSN: 2039-4004

Improvement on the LR Test Statistic on the Cointegrating Relations in VAR Models: Bootstrap Methods and Applications

Alessandra Canepa

University of Turin*
and Brunel University London

March 26, 2020

Abstract

A Bartlett corrected likelihood ratio test for linear restrictions on the cointegrating relations is examined in Johansen (2000). Simulation results show that the performance of the corrected LR test statistic is highly dependent on the values of the parameters of the model. In order to reduce this dependency, it is proposed that the finite sample expectation of the LR test be estimated using the bootstrap. It is found that the bootstrap Bartlett correction often succeeds in this task.

Keywords: Cointegrating vectors, LR test, Bootstrap methods.

JEL Code: C12, C15, C22.

*Corresponding author: Department of Economic and Statistics Cognetti De Martiis, University of Turin, Lungo Dora Siena 100A, 10153 Turin, Italy, E-mail address: Alessandra.Canepa@unito.it.

1 Introduction

The procedure for estimating and testing cointegrating relationships described in Johansen (1988) is available in virtually all econometric software packages and is widely used in applied research. Briefly this method involves maximizing the Gaussian likelihood function and analysing the eigenvalues and eigenvectors found using the reduced rank regression method. Once that the number of cointegrating vectors has been determined, hypotheses on the structural economic relationships underlying the long-run model can be tested using the likelihood ratio (LR) test.

Although the LR test of linear restriction of cointegrating vectors has the correct size asymptotically, many studies contain reports that the approximation of the χ^2 distribution to the finite sample distribution of the LR test can be seriously inaccurate see, for example, Haug (2002), or Fachin (2000)). In order to address this problem Johansen (2000) proposes a Bartlett adjustment for LR statistic and analytically derives the asymptotic expansions needed to calculate the expectation of the test statistic. Multiplying the unadjusted statistic by a factor derived from an asymptotic expansion of the expectation test provides a closer approximation of the resulting adjusted statistic to the χ^2 distribution, thus reducing the size distortion problem. The draw back of the coin is that the Johansen (2000) Bartlett correction factor is quite difficult to apply. Moreover, simulation results indicate that the correction factor is useful for some parameter values but does not work well for others. As Johansen (2000) points out "the influence of the parameters is crucial [.....] There are parameters points close to the boundary where the order of integration or the number of cointegrating relations change, and where the correction does not work well" (cf. Johansen (2000) p.741).

We believe that the dependency on the parameter values may be reduced by computing the Bartlett adjustment using the non-parametric bootstrap. This method involves calculating a number of bootstrap values of the LR test statistic and estimating the expected value of the test statistic by the average value of the bootstrapped LR statistics. The bootstrap Bartlett method was first proposed in Rocke (1989) where hypothesis testing in seemingly unrelated regression models was considered. Rocke's simulation results showed that the Bartlett adjustment for the LR test determined using the non-parametric bootstrap was considerably more accurate than the Bartlett adjustment from the second-order asymptotic method of Rothenberg (1984).

The purpose of this work is to see if the Bartlett adjustment approximated using the bootstrap method is able to reduce the finite sample dependency of the null rejection probability of the LR test statistic on parameter values. If such an application were to be successful, it would deliver improvements upon the analytic Bartlett correction proposed in Johansen (2000) both in applicability and accuracy.

It is also of interest to compare the bootstrap Bartlett method with the straightforward bootstrap method, in which the significance level assigned to $G_T = -2(\log(LR))$ is the fraction of the $G_{i,T}$ greater than G_T. In principle we may expect the two methods to obtain the same accuracy (i.e. similar empirical sizes), but the former to be less computationally intensive than the latter. Generally speaking, estimating a moment of a distribution requires fewer trials than estimating the tail of the same distribution. This result is formally proven in Rocke (1989).

Thus, in this paper the bootstrap is used in two ways: first, to approximate a Bartlett-type correction; and second to estimate the p-value of the observed test statistic. In addition, we compare the performance of the proposed procedures with the F-type test of Podivinsky (1992). Finally, it is well known that the Bartlett correction factor is designed to bring the actual size of asymptotic tests close to their respective nominal size, but it may lead to a loss in power. Accordingly, the power properties of the proposed procedures will be considered. Throughout the paper, the bootstrap accuracy in small samples is mainly investigated through Monte Carlo studies, although an empirical application is provided to illustrate the performance of the bootstrap with some real data.

We will close this section by a brief presentation of the Bartlett correction. In the next Section we introduce the LR test for linear restrictions on cointegrated space, the Bartlett correction of Johansen (2000), the F-type test of Podivinsky (1992) and the bootstrap inference procedures. In Section 3, the design of the Monte Carlo experiment is explained and the simulation results are reported. In Section 4 an empirical application is considered and Section 5 contains conclusions.

The Bartlett Correction

The Bartlett correction is based on a simple idea, but it can be very effective in reducing the finite sample size distortion problem of the LR tests. Briefly, this method consists of scaling the test statistic by the ratio of its asymptotic and estimated finite sample expectations. In other words, instead of looking directly at $G_T = -2(\log(LR))$, which as $T \to \infty$ tends to G_∞, we focus on the distribution of $\frac{G_T}{E(G_T)}$. Given that $\frac{G_T}{E(G_T)} \to \frac{G_\infty}{E(G_\infty)}$ as $T \to \infty$, we can write

$$G_T \approx \frac{E\left(G_T\right)G_\infty}{E\left(G_\infty\right)}.$$

Typically it is difficult to find an exact expression for $E\left(G_T\right)$, one can instead find an approximation of the form

$$E\left(G_T\right) = q\left(1 + \frac{B\left(\theta\right)}{T}\right) + O\left(T^{-2}\right),$$

where q is equal the degree of freedom parameter for the test. Thus the statistic

$$\frac{G_T}{1 + \frac{B(\hat{\theta})}{T}}$$

has expectation closer to that of χ^2 than G_T. In Lawley (1956) it is proved that, under the assumption of i.i.d. variables, the Bartlett correction not only improves the mean but also implies that all cumulants of the corrected statistic match that of a χ^2 density to the order $O(T^{-3/2})$. The latter result explains why the correction works so well in practice; see, for example, Barndorff-Nielsen and Cox (1984).

The application of the Bartlett's correction in the context of time series models is relatively recent. For example, a correction factor for the LR tests for $AR(1)$ and $MA(1)$ models is given in Taniguchi (1990) and an expression for the Bartlett correction factor for models belonging to the stationary, invertible and Gaussian $ARMA$ family is obtained in Lagos and Morettin (2000). However, the debate on the usefulness (or the legitimate use) of the Bartlett correction for $I(1)$ processes is still unsettled. Although it is established in Jensen and Wood (1997) that, in the $AR(1)$ case, the Bartlett correction only corrects the first moment, results from Bravo (1999) and Nielsen (1997) showed that a Bartlett-type correction to the LR test for unit root in practice can improve the asymptotic approximation considerably.

2 Model and Tests

Consider the p-dimensional VAR model

$$\Delta Y_t = \alpha\left(\beta' Y_{t-1} + \rho' D_t\right) + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \phi d_t + \varepsilon_t, \quad t = 1, ..., T \qquad (1)$$

where Y_t and ε_t are $(p \times 1)$ vectors, with $\varepsilon_t \backsim NID(0, \Omega)$, and $\Delta Y_t = Y_t - Y_{t-1}$. Matrices of coefficients have the following dimensions: α and β are $(p \times r)$; ϕ is $(p \times p_d)$; ρ is $(p_d \times r)$; and $\Gamma_1 ...,\ \Gamma_{k-1}$ are $(p \times p)$.Also $d_t\,(p_d \times 1)$ and $D_t\,(p_D \times 1)$ are deterministic terms in (1). Once the cointegrating rank has been established we can test for linear restrictions on cointegrating space. For consistency with the framework of Podivinsky (1992), we focus on the hypothesis $H_0 : \beta = H\varphi$, where $H\,(p \times s)$ for $r \leq s \leq p$ is a known matrix that specifies that the restrictions (s) are imposed on all cointegrating vectors (r); see Johansen (1996) for discussion of tests for other hypotheses. The test statistic for H_0 can be obtained from the concentrated likelihood function and is given by

$$G_T = -T \sum_{i=1}^{r} \ln\left[\left(1 - \tilde{\lambda}_i\right) / \left(1 - \widehat{\lambda}_i\right)\right], \qquad (2)$$

where $\hat{\lambda}_i$ and $\tilde{\lambda}_i$ are the usual eigenvalues implied by the maximum likelihood estimation of the restricted and unrestricted models, respectively. The limiting distribution of $\hat{\beta}$ is a Gaussian mixture and G_T is asymptotically distributed as $\chi^2(r(p-s))$ under the null hypothesis.

A Bartlett correction factor for the LR test is derived in Johansen (2000). This correction depends upon the null hypothesis on the cointegrating coefficients of the VAR. In our case an approximation to order T^{-1} for the Bartlett-correction factor is given by

$$\begin{aligned}\frac{E\left[-2\ln(LR)\right]}{r(p-s)} &= 1+\frac{1}{T}\left[\frac{1}{2}(p+s-r+1+p_d+kp)\right] \\ &\quad +\frac{1}{Tr}\left[2p+s-3r-1\right]v(\alpha)+2c(\alpha)+c_d(\alpha))]\end{aligned} \tag{3}$$

where $v(\alpha) = tr\left\{\alpha'\Omega^{-1}\alpha\sum_{\beta\beta}\right\}$, with $\sum_{\beta\beta} = Var(\beta'Y|\Delta Y_t,...,\Delta Y_{t-k+2})$, $c_d = p_d v(\alpha)$, and the constant $c(\alpha)$ is given in Johansen (2000). Note that, in Johansen (2002), a Bartlett correction factor for the LR test is derived assuming that the adjustment parameter α is known. Although theoretically interesting, this case is less relevant in applied work so we restrict our attention to Johansen (2000) where this assumption is dropped.

Alternative small sample corrections of the LR test for linear restrictions on cointegrating space have been proposed in Podivinsky (1992) and Psaradakis (1994). Podivinsky suggested an F-type test while Psaradakis proposed a simple correction factor for the LR test. However, results given in Canepa (2005) showed that F-type test of Podivinsky (1992) largely outperformed the small sample corrected LR test of Psaradakis (1994). Consequently we restrict our attention to the former procedure.

After considering the analogy with the classical linear regression theory, it is proposed in Podivinsky (1992) to base an approximation on an F-type test. Let

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

and l be the number of parameters estimated subject to the maintained hypothesis $\Pi = \alpha\beta'$, then

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

is taken to have an F distribution with $\left(r\left(p-s\right), T-l\right)$ degrees of freedom under H_0.

Rather than relying upon either asymptotic theory or an analytical Bartlett correction it is proposed below that the non-parametric bootstrap be used to reduce the size distortion of the LR test. The bootstrap is used to obtain two tests. The first of these procedures is in the spirit of Rocke (1989) and the second is a straightforward application of the bootstrap p-value approach. We will examine them in turn (note: the subscript $(*)$ will be used to indicate the bootstrap analog throughout the paper).

Calculating the bootstrap Bartlett corrected LR test (BSB) involves undertaking a simulation study using the constrained estimates of θ, denoted by $\hat{\theta} = \left(\hat{\alpha}, \hat{\beta}, \hat{\Gamma}_i, \hat{\phi}, \hat{\rho}, \hat{\Omega}\right)$, conditional on the initial values Y_0 and ΔY_0, as the true values. Given these estimates and any required starting values, bootstrap data can be generated recursively after resampling residuals. From each generated sample, one obtains a bootstrap value of the LR statistic whose average, denoted by $\overline{G}^*_T$, estimates the mean of the LR statistic under the null hypothesis. The corrected statistic $G^*_T = \frac{G_T(r(p-s))}{\overline{G}^*_T}$ is then referred to a $\chi^2\left(r\left(p-s\right)\right)$ distribution. An heuristic explanation of why this procedure is asymptotically valid is the following.

Let $\phi d_t = \mu$ (i.e. a constant term) and $\rho = 0$ in (1). Also define $\alpha_\perp$ and $\beta_\perp$ as the $p \times (p-r)$ matrices, such that $\alpha'\alpha_\perp = 0$ and $\beta'\beta_\perp = 0$. By the Granger representation theorem the process $\{Y_t\}$ has the following Wold vector moving average (MA) representation

$$Y_t = C\sum_{i=1}^{t}\varepsilon_i + C\mu t + C_1\left(L\right)\left(\varepsilon_t + \phi d_t\right) + Y_0 \tag{4}$$

where $C = \beta_\perp\left(\alpha'_\perp\Gamma\beta_\perp\right)^{-1}\alpha'_\perp$, and $C_1\left(L\right)$ is a stable $p \times p$ lag polynomial. From the representation (4) it follows that $\{Y_t\}$ can be rewritten as the sum of an $I(1)$ component given by

$$Y_t^P = C\left(\sum_{i=1}^{t} \varepsilon_i + \mu t\right) = \beta_\perp \left(\alpha'_\perp \Gamma \beta_\perp\right)^{-1} \left(\alpha'_\perp \mu + \sum_{i=1}^{t} \alpha'_\perp \epsilon_t\right),$$

where $\left(\alpha'_\perp \mu + \sum_{i=1}^{t} \alpha'_\perp \epsilon_t\right)$ represents the $(p-r)$ common trends along with their coefficients $\beta_\perp \left(\alpha'_\perp \Gamma \beta_\perp\right)^{-1}$, an $I(0)$ component given by

$$Y_t^s = C_1(L)\left(\varepsilon_t + \phi d_t\right),$$

and an initial values denoted by Y_0. Thus, the asymptotic properties of $\{Y_t\}$ depend on which linear combination of the process we consider. Cointegration implies multiplying Y_t^P by $\beta' C = 0$ so that linear combinations of $\beta' Y_t$ are stationary (note that the initial values cancel). In other words, the cointegrating vectors act as a detrending model and $\beta' C_1(L)\left(\varepsilon_t + \phi d_t\right)$ is a representation of the disequilibrium error $\beta' Y_t$ (see for example Johansen (1996) for more details).

Turning to the non-parametric bootstrap, let $\hat{F}_\varepsilon$ denotes the empirical density function of the residuals. The resampling scheme imposes that the characteristic polynomial

$$\Psi^*(z) = (1-\hat{z})I - \hat{\alpha}\hat{\beta}' \ \hat{z} - \sum_{i=1}^{k-1} \hat{\Gamma}_i (1-\hat{z})\hat{z}^i,$$

has $p-r$ roots equal to 1 and all the other roots outside the unit circle $(0 < r < p)$. Thus the process generated by the resampling scheme

$$\Delta Y_t^* = \hat{\alpha}\hat{\beta}' Y_{t-1}^* + \sum_{i=1}^{k-1} \hat{\Gamma}_i \Delta Y_{t-i}^* + \hat{\phi} d_t + \varepsilon_t^*,$$

where $\varepsilon_t^* \backsim \hat{F}_\varepsilon$, has the following MA representation

$$Y_t^* = \hat{C} \sum_{i=1}^{t} \varepsilon_i^* + \hat{C}\hat{\mu} t + \hat{C}_1(L)\left(\varepsilon_t^* + \hat{\phi}\ d_t\right) + Y_0^*,$$

and $\hat{\beta}' Y_t^*$ is stationary. Moreover, under weak regularity conditions, the partial sum of the ε_i^* satisfies the functional limit theorem

$$T^{-1/2}\sum_{t=1}^{[Tr]}\varepsilon_i^* \xrightarrow{d} W(r),$$

where $W(r)$ denotes an $(p-r)$-dimensional Wiener process with covariance Ω and $\xrightarrow{d}$ stands for weak convergence conditional on the sample Y_t.

Turning now to the Bartlett correction, the bootstrap Bartlett estimator is consistent for all $\xi > 0$, and for all $\varpi > 0$ if

$$\lim_{T\to\infty} P\left[\sup\left|\left[\sqrt{T}\left(\overline{G}_T^* - E(G_T)\right) \le \varpi\right] - \left[\sqrt{T}\left(E(G_T) - E_\infty\left(G_T\right)\right) \le \varpi\right]\right| > \xi\right] = 0.$$

Using the fundamental properties of the Mallows distance d_2,see Bickel and Freedman (1981), it can be proved that

$$l_2\left(\sqrt{T}\left(\overline{G}_T^* - E(G_T)\right) \le \varpi, \sqrt{T}\left(E(G_T) - E_\infty\left(G_T\right)\right) \le \varpi\right)^2 \le l_2(M, M_\infty)^2,$$

which shows that the distance between the bootstrap distribution and the finite-sample distribution can be bounded by the distance between the empirical density function (M) and the underlying distribution function (M_∞). However, the following result holds true

$$\lim_{T\to\infty} P\left[l_2(M, M_\infty)^2 > \xi\right] = 0.$$

by the Glivenko-Cantelli theorem and the strong law of large numbers. Thus, the first order asymptotic validity of the procedure results from the fact that the parameters of model (1) are consistently estimated, the asymptotic distribution is a smooth function of the disturbances, and $l_2(M, M_\infty)^2 \to 0$.

The second bootstrap procedure is a straightforward application of the bootstrap p-value approach. In this case the bootstrap values of LR are employed to approximate the p-value of the observed value of the test statistic. The bootstrap p-value is then compared with the desired null rejection probability: this second test is denoted by BSP. By using the empirical distribution function in place of some specific parametric distribution, the non-parametric bootstrap does not require a choice of error distribution and this feature may be appealing to the applied worker. In the literature it has been shown that in many cases the bootstrap delivers an automatic approximation to the Edgeworth expansion; see Hall (1992). Thus it can be considered as a numerical approximation to analytical calculations of one-term Edgeworth expansion. For cointegrated VAR models, due to the intricate analysis, the ability of the bootstrap test to provide second order refinement is still an open question. An important breakthrough in this literature is given in Park (2000) where asymptotic expansions for the unit root models are developed and it is shown that the bootstrap test provides asymptotic refinements for the Dickey-Fuller tests. Considering that Johansen's rank tests are the multivariate extensions of the (augmented) Dickey-Fuller tests, Park's results are quite promising. In practice, however, whether a particular asymptotically valid technique is useful in small samples can only be evaluated by simulation. Thanks to the increases in the power of computers, the number of studies evaluating the usefulness and the limitations of bootstrap inference in cointegrating systems is growing rapidly. For example, computer intensive methods for inference on cointegrating vectors are considered in Gredenhoff and Jacobson (2001) and it is shown that the parametric bootstrap is very effective in reducing the finite sample error in rejection probability. The non-parametric bootstrap is examined in Fachin (2000) and the resulting bootstrap test is found to be overly conservative. The results of Fachin (2000) are, however, not confirmed in this study.

The steps used to implement the non-parametric bootstrap can be summarized as follows:

Step (1) : Estimate (1) and compute G_T.

Step (2) : Resample the residuals from $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)$ independently 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 given in (1).

Step (3) : Compute the bootstrap replication of $\{G_T^*\}$ using $(y_1^*, ..., y_t^*)$.

If B is the number of bootstrap samples, the two bootstrap alternatives to the use of asymptotic critical values can be implemented as follows.

Step (4) : Calculate the mean value of the bootstrap values of the test, $\overline{G}_T^* = T^{-1} \sum_{i=1}^{T} G_{i,T}^*$. A Bartlett-type corrected statistic is therefore $BSB = \frac{r(p-s)G_T}{\overline{G}_T^*}$. The BSB statistic is then compared with critical values from the asymptotic $\chi^2 \left(r \left(p - s \right) \right)$ distribution.

Step (5) : Compute the p-value function

$$P(G_T) = B^{-1} \sum_{i=1}^{B} I \left(G_{i,T}^* \geq G_T \right),$$

where $I(\cdot)$ is the indicator function that equals one if the inequality is satisfied and zero otherwise. The bootstrap test (BSP) is carried out by comparing $P^*(G_T)$ with the desired critical level and rejecting the null hypothesis if $G_{i,T}^*$ in not greater than G_T .

3 The Monte Carlo experiment

Can the bootstrap be successfully applied to approximate the finite sample expectation of the LR test? Also how does it compare to the available analytical correction? Moreover, what can we say about alternative inference procedures? These are the questions addressed in this section. With this target in mind, two groups of experiments have been designed. In the first group of experiments the effects of varying the model complexity (i.e. sample size, number of lags, number of cointegrating vectors) in the VAR model will be investigated. The second group of experiments is in the spirit of Haug (2002); also see Gonzalo (1994). The performance of the test statistics will be evaluated on the basis of their behavior in term of empirical sizes in response to variations of the key parameters influencing the DGP.

Two models were simulated; the first imposes only one cointegrating vector (i.e. $\boldsymbol{\beta} = \left[\begin{array}{cccc} \beta_{11} - 1 & \beta_{21} & \beta_{31} & 0 \end{array} \right]'$), while model 2 allows for two cointegrating vectors (i.e. $r = 2$) among four integrated variables. Both models allow for one to three lags (k). Thus, the DGPs are given by:

$DGP = DGP1:$

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \\ \Delta y_{4t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} & 0 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \\ y_{4t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix} \tag{5}$$

$DGP = DGP2$

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \\ \Delta y_{3t} \\ \Delta y_{4t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} & 0 \\ 0 & \beta_{22} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \\ y_{4t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}$$

where $\begin{bmatrix} \varepsilon_{1t}, & \varepsilon_{2t} & \varepsilon_{3t} & \varepsilon_{4t} \end{bmatrix}'$ are i.i.d. with $\boldsymbol{\varepsilon}_{it} \sim N(0, \Omega)$, $Var(\varepsilon_{it}) = \sigma^2_{\varepsilon i}$ and $Cov(\varepsilon_{it}, \varepsilon_{jt}) = 0$.

All simulations were carried out using the matrix programming language GAUSS. The Monte Carlo experiment is based $n = 10,000$ replications for the LR, BRT, F-type tests and (unless otherwise specified) on $n = 1,000$ replications for BSB and BSP. All the bootstrap distributions are generated from resampling and calculating the test statistic 400 times, (i.e. $B = 400$). The random number generator was restarted for each T values and the initial value set equal to zero. Note that in Johansen's procedure the maximum likelihood estimator of β in equation (1) is calculated by as the set of eigenvectors corresponding to the s largest eigenvalues of $S'_{0k}S^{-1}_{00}S_{0k}$ with respect to S_{kk} , where S_{00}, S_{kk} and S_{0k} are the moment matrices formed from the residuals Δy_t and y_{t-k}, respectively onto the Δy_{t-j}. In this paper in place of the conventional algorithm for cointegration analysis (i.e. the algorithm for maximum likelihood estimation that use the second moment matrices) all the simulation results reported are obtained using an algorithm based on QR decomposition In particular, we follow "Algorithm 3" in Doornik and O'Brien (2002), (p. 189). The rationale of doing so is to obtain simulation results that are more numerically stable.

3.1 The probability of type I error

As far as the first group of experiments is concerned the Monte Carlo results can be summarized as follows. The first thing to note in Table 1 is that inference based on first order asymptotic critical values is markedly inaccurate with excessively high rejection rates. In general, the proportionate error in null rejection probability is higher the lower the nominal size of the test. For example, for $r = 1$, $T = 50$, $k = 1$ and nominal size of 1%, the empirical size can be nearly four times as large as the reference nominal size, whereas when the nominal size is 10% the empirical size is approximately twice as much the nominal size. Increasing the number lags, k, dramatically increases the deviation from the nominal levels. By contrast, allowing more cointegrating vectors, r, in the system slightly reduces the size distortion. Turning to empirical sizes for BRT, we can see that they are much closer to the nominal sizes than the first order asymptotic critical values. The F-type test shows empirical sizes similar to those of BRT, whereas BSB and BSP seem to be more sensitive to the number of lags in the DGP, with the latter slightly outperforming the former.

Overall, the results in Table 1 show that the small sample adjusted LR tests and their bootstrap counterparts are quite effective in reducing the small size distortion problem. However, introducing many nuisance parameters in the model affects the size accuracy of all test statistics under consideration. The impression is that BRT, BSB, BTP and F-type act in a similar fashion to the asymptotic tests, only to a much less extent. Thus the performance of these tests deteriorate when the performance of the asymptotic LR test worsens.

PLEASE INSERT TABLE 1 ABOUT HERE.

Coming to the second group of experiments, we now evaluate the sensitivity of the test statistics to variations of key parameters in the DGP. Our target is to detect regions of the parameter space where a large size distortion is more likely to occur. With this in mind we restrict our attention to the DGP with $r = 1$ and $k = 1$ (i.e. $DGP = DGP1$). This may be seen as a limitation of our study. However, one has to keep in mind that a test or a class of tests that does not perform well in simple situations cannot be expected to do well for more complicated models. Moreover, including more nuisance parameters would involve losing control of the experimental design.

What are the target parameters? Calculating the Bartlett correction factor using equation given in (3) with $p = 4$, $k = 1$, $s = 1$, $p_d = 0$, $p_D = 0$, $c_d(\alpha) = 0$, $v(\alpha) = -\frac{\beta'\alpha\left(2+\beta'\alpha\right)}{\alpha'\Omega^{-1}\alpha\beta'\Omega\beta}$, and $c(\alpha) = -2\frac{\beta'\alpha\left(1+\beta'\alpha\right)}{\alpha'\Omega^{-1}\alpha\beta'\Omega\beta}$ gives

$$\begin{aligned}\frac{E\left[-2\ln\left(LR\right)\right]}{3} &= 1 + \frac{14}{2T} - \frac{1}{T}\frac{\beta'\alpha}{\alpha'\Omega^{-1}\alpha\beta'\Omega\beta} \times \\ &\quad \left[5\left(2+\beta'\alpha\right) + 4(\left(1+\beta'\alpha\right)\right].\end{aligned}$$

Thus, the Bartlett correction factor depends on $\alpha'\beta$, $\alpha'\Omega^{-1}\alpha\beta'\Omega\beta$ and T. However, it can be shown that not only

$$E\left[-2\ln\left(LR\right)\right]/3 = f\left(\alpha'\beta,\ \alpha'\Omega^{-1}\alpha\beta'\Omega\beta,\ T\right),$$

but also the that the parameters enter into distribution of the LR test through the same function; see Johansen (2000). Therefore, by varying the arguments in $f(\cdot)$, one is able to analyse the effect of variations of the parameters on the empirical sizes of the test statistics. One way of doing it is to change the coordinate of the system and find a canonical form of the DGP as in Johansen (2000, 2002). This implies fixing the cointegrating vector and change the coordinate of the system so that only two parameters, both function of all parameters in the VAR are left. However, in our case we are interested in controlling the loading ($\boldsymbol{\alpha}$) and the cointegrating coefficients ($\boldsymbol{\beta}$) and the covariates (Ω) separately so that regions where the dependency of the size distortion of the LR test from the parameter space is stronger can be detected.

Experiments are carried out with $T = (50, ..., 150)$; $\alpha_{11} = (0.2,\ 0.5,\ 0.7)$ in $\boldsymbol{\alpha} = \left[\begin{array}{ccccc}\alpha_{11} & 0 & 0 & 0 & 0\end{array}\right]$; and $\sigma = (0.5,\ 1,\ 2)$ in Ω. Coming to $\boldsymbol{\beta}$, one problem we face is that the long-run cointegration relationship depends on a relatively large number of parameters. A possible way to address this issue is to change the coordinates of the system and rotate the $VAR(1)$. This transformation would leave the statistical analysis of the model unchanged, but it may lead to a reduction in the number of parameters in $\boldsymbol{\beta}$ without loss of generality. Under $H_0 : \beta = H\varphi$, model (5) can be rotated into

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \\ \Delta X_{4t} \end{bmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} & 0 & 0 \end{pmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \\ X_{3t-1} \\ X_{4t-1} \end{bmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix}$$

where $\boldsymbol{\varepsilon}_t \sim N(0, \mathbf{I})$ and X_t is used (instead of Y_t) to indicate the rotated $VAR(1)$. Thus a simple rotation of the process highlights the fact that not all the parameters in $\boldsymbol{\beta}$ are equally important. Accordingly, the parameter β_{31} can safely be set equal to zero.

Another issue we face is how to select the set of parameters for the remaining parameters in $\boldsymbol{\beta}$. These have to chosen in such a way that the stability of the system is preserved. In our case we have

$$\boldsymbol{\beta} = \begin{Bmatrix} \beta_{11} \in -1.9, -1.5, -0.9, -0.5, -0.3, -0.2, \\ \beta_{21} \in 0.2, 0.3, 0.4, 0.9, 1.5 \end{Bmatrix}.$$

In Table 2 we report the empirical sizes of the tests for a nominal 5% level. Other parameters given, we can see that the size distortion of the LR test highly depends on the magnitude of α (i.e. the parameter that controls the speed of the adjustment to the equilibrium cointegrating relationship). When the speed of the adjustment is low (i.e. $\alpha_{11} = 0.2$) the size distortion of the test can be large and the convergence to the nominal size slow. Generally speaking, the results show that the higher the magnitude of α (i.e. the faster the adjustment) the lower the size distortion. This result is consistent with the literature; see, for example, Gonzalo (1994), Haug (2002) and Fachin (2000). Coming to BRT, the size distortion of the Bartlett corrected LR test is again highly dependent on the magnitude of α, whereas its bootstrap counterpart is much less so: for $\alpha_{11} = 0.5$

(which is quite close to value found in empirical studies), no matter the sample size the empirical size is within the 95% confidence interval (i.e. 3.6-6.4). The ordinary bootstrap test also performs well. By contrast, the F-type test while able to capture the dependency on the number of the estimated parameters does not take into account the magnitude of the parameters in itself, thus generally speaking it performs only slightly better than the asymptotic LR test.

Turning to σ, it is clear from Table 2 that the distribution of the LR test statistic is invariant to changes of the standard error of the shocks ε_{it}, and, of course, so are the other test statistics. It would be of interest to consider an experimental design where $Cov(\varepsilon_{it}, \varepsilon_{jt}) \neq 0$ since it is well know that correlation between the noise terms can adversely affect the finite sample performance of the LR test; see for example Toda (1995). However, in this case we would not expect the ordinary bootstrap to work because the correlation between the innovations would be lost in the resampling and a different kind of bootstrap procedure, e.g., the stationary bootstrap, should be applied. For this reason we shall leave this issue for future study.

PLEASE INSERT TABLE 2 ABOUT HERE

Figure 1 shows the distribution of the test statistics considered as a function of β_{11} and β_{21} with the sample size fixed at $T = 50$. From Figure 1 it is clear that empirical size of the LR test (reported on the z-axis) rapidly increases when β_{11} and β_{21} get closer to 0. Note that if β_{11} is in the interval $(-2, 0]$, then the process Y_t is $I(1)$. When $\beta_{11} = 0$, the process is a pure $I(1)$ process which does not cointegrate, whereas if $\beta_{11} < -2$ or $\beta_{11} > 0$ the process Y_t is explosive. Thus, as one may expect, the LR test does not perform well for parameter points close to the boundary where the order of integration changes or the number of cointegrating relations changes. The F-type test mimics the behavior of the LR test but with empirical sizes closer to the nominal ones. By contrast, BSB works remarkably well, largely outperforming BRT and partially outperforming BSP.

3.2 The probability of type II error

The evaluation of the power of the test statistics has been carried out by generating the data under the following alternatives: $H_1 : \beta_{41} = 0.15$ and $H_1 : \beta_{41} = 0.3$ with $r = 1$, $k = 1, 2, 3$, and $T = 50, 100, 150$. From Table 3, we can see that, in general, as expected, the rejection frequencies increase with the sample sizes and the distance between the null and the alternative. The power estimates for the larger sample size $T = 100$ are reasonable for all the alternatives. Turning to the comparison of the power among the different procedures, overall we found that correcting the test statistics for the size shifts the estimated power function down. There is evidence that the tests BRT, BSP, BSP and F-type share similar power properties, with no test uniformly outperforming its competitors. The results for the sensitivity of the parameter space are not reported in detail here but show that a slow adjustment to the equilibrium worsens the rejection frequencies for BRT, BSP and BSP, whereas the changes of the standard deviation of the errors do not have an important impact on the power estimates.

PLEASE INSERT TABLE 3 ABOUT HERE

4 An empirical example

In this section we analyse the finite sample properties of the inference procedure considered using data from the Danish economy on the demand for money as used in Johansen and Juselius (1990). These are quarterly data from 1974:1 to

1987:3 on log M2 monetary aggregate (m_t), the real log income (y_t), the bond rate (i_t^b) and the deposit rate (i_t^d).

Once again, the performance of the test statistics has been investigated by Monte Carlo. The starting point of the simulation experiment is to replicate the results of Johansen and Juselius (1990) in order to get the *DGP* parameter values; also see Johansen (1996). Accordingly, using the Danish data, the model in (1) has been estimated with a restricted constant term and four seasonal seasonal dummies. Using the rank test procedure, we have found no evidence in the Danish data for more than one cointegration relation. Thus, solving the eigenvalue problem, we have found that the estimated long-run β coefficients associated to the first eigenvalue (i.e. $\hat{\lambda}_1 = 0.4332$) are

$$\hat{\boldsymbol{\beta}}' = \left[\begin{array}{ccccc} 1.0000 & -1.03295 & 5.20692 & -4.21588 & -6.05993 \end{array} \right]$$

while the normalised adjustment coefficients of $\hat{\boldsymbol{\alpha}}$ are

$$\hat{\boldsymbol{\alpha}}' = \left[\begin{array}{cccc} -0.21295 & 0.11502 & 0.02318 & 0.02941 \end{array} \right].$$

Finally,

$$\hat{\boldsymbol{\Omega}} = \begin{pmatrix} 0.01965 & 0 & 0 & 0 \\ 0.01150 & 0.01706 & 0 & 0 \\ -0.00875 & 0.00402 & 0.01818 & 0 \\ -0.00148 & -0.00061 & 0.00095 & 0.00490 \end{pmatrix}.$$

We then have investigated the rejection probability of the test for linear restrictions on $\hat{\boldsymbol{\beta}}$ using $\boldsymbol{\theta} = \left(\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\alpha}}', \hat{\boldsymbol{\Omega}}\right)$ in the *GDP* with the initial values taken from the estimated initial values, and to simplify, the parameters for $\hat{\boldsymbol{\Gamma}}$ have been set equal to zero. As far as the Monte Carlo design is concerned, the simulation experiment has been carried using the bootstrap algorithm described in Section 2, whereas the Bartlett correction factor for the *LR* test has been calculated using (3) with $p = 4$, $r = 1$, $s = 1$, $k = 2$, $p_D = 1$, $p_d = 0$ and the estimated parameters in $\boldsymbol{\theta}$. The restriction under consideration is $H_0 : \beta_1 = 1$ in $\hat{\boldsymbol{\beta}}' = \left[\begin{array}{ccccc} \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \end{array} \right]$ (i.e. the estimated coefficient is the same as the value in the *DGP*).

In Table 4 the simulation results are reported for the rejection probability for the size of the test statistics at the 5% nominal level, while for the experiment evaluating the power the data has been generated under the alternative H_1 : $\beta_1 = 1.3$.

PLEASE INSERT TABLE 4 ABOUT HERE

Simulation results from Table 4 show that the estimated size for the LR test based on the first order asymptotic approximation is more than 4 times the reference nominal size, thus confirming the previous results. Among the corrected versions of the test, BRT outperforms the F-type test as well as the bootstrapped counterpart (BSB), although only marginally. Turning to estimated rejection rates under the alternative, the F-type test appears to be most powerful among modified LR tests, but this is possibly a spurious result due to larger size distortion under the null. By contrast, BRT, BSP, and BSP share similar power properties, with BRT slightly outperforming the competitors.

5 Concluding remarks

This paper investigates through Monte Carlo simulation the finite sample properties of modified versions of the LR test for linear restrictions on cointegrated relationships. The Bartlett corrected LR test of Johansen's (2000) and the F-type test of Podivinsky (1992) are both studied. The performance of these test statistics is also compared with the finite sample behavior of the tests obtained using the bootstrap procedure. The bootstrap is used in two ways: first, to approximate a Bartlett-type correction; and second, to estimate the p-value of the observed test statistic. The need for modified versions is indicated by the poor performance of the standard asymptotic check which is also included. The assessment of the small sample rejection probabilities of the five test statistics is then carried out: i) by varying the model complexity (i.e. sample size, number of lags, number of cointegrating vectors) in the VAR model; and ii) by evaluating their behavior in response to variations of the key parameters influencing the DGP.

The main simulation results can be summarized as follows. As far as the asymptotic LR test is concerned, our study mainly confirms previous research findings; see, for example, Haug (2002), Podivinsky (1992), or Gredenhoff and Jacobson (2001)). More precisely, it is found that, for small to moderate sample sizes, inference based on first order asymptotic approximation is largely inaccurate and the size distortion of the asymptotic test dramatically worsens when more nuisance parameters are included in the DGP. In addition, the size distortion of the LR test is sensitive to the magnitude of the cointegrating coefficients and also to the magnitude of the parameter that controls the speed of the adjustment to the equilibrium cointegrating relationships (the faster the adjustment to the equilibrium the lower is the size distortion of the test statistic). The Bartlett correction of Johansen (2000) dramatically improves the behavior of the LR procedure in several instances, but still leaves something to be desired particularly for parameter points close to the boundary where the order of integration changes or the number of cointegrating relations changes. More generally, as expected, the distribution of the levels of the corrected test mimics the distribution of the asymptotic LR test. Thus the performance of the former deteriorates in regions of the parameter space where the the performance of the latter worsens. From our simulation results, it appears that calculating an approximation to the finite sample expectation of the LR test by using the bootstrap can be worthwhile. The bootstrap Bartlett corrected test appears to be less sensitive to the values of the parameters of the DGP than its analytical counterpart, although it is slightly more sensitive to the inclusion of nuisance parameters. The bootstrap p-value test also works fine, however (as Rocke (1989) points out) the bootstrap Bartlett adjustment requires fewer trials to obtain the same order of accuracy than bootstrap p-value test and is therefore less computationally demanding. Finally, for small sample sizes ($T \leq 50$), the results for the F-type test are mixed and it performs only slightly better than the asymptotic LR test in some cases.

As concluding remarks we give some practical considerations. Based on our Monte Carlo experiment, when testing for linear restrictions of the cointegrating vectors, the Bartlett corrected LR test, the bootstrap Bartlett LR test, and the F-type all deliver reasonable accurate inference for sample sizes of approximately 100 observations (or about 25 years quarterly data). Thus, if this or a larger sample size is available, all of these tests are reliable. However, when the sample size is smaller, a possible strategy for improving the performance of the LR test is the following: first, calculate the roots of the characteristic polynomial and determine under what conditions on the parameters the process fulfil the root assumption (see Johansen,1996, p. 14); second, estimate the model applying the Johansen reduced rank regression procedure; and third, if the estimated parameters are close to the boundary of the parameter space, use the bootstrap to approximate the finite sample expectation of the LR test, otherwise the analytical Bartlett correction would probably be satisfactory.

References

[1] Bartlett, M.S. (1937). "Properties of Sufficiency and Statistical tests". *Proceedings of the Royal Society.* 160, 263-268.

[2] Barndorff-Nielsen, O.E. and Cox, D.R. (1984). "Bartlett adjustments to Likelihood Ratio Statistics and the Distribution of the Maximum Likelihood Estimator". *Journal of the Royal Statistical Society.* 46,483-495.

[3] Bickel, P. J. and Freedman D. A. (1981). "Some Asymptotic Theory for the Bootstrap". *The Annals of Statistics*, 6, 1196-1217.

[4] Bravo, F. (1999). "A Correction Factor for Unit Root Test Statistics". *Econometric Theory*, 15, 218-227.

[5] Canepa, A. (2006). "Small Sample Corrections for Linear Restrictions on Cointegrating Vectors: a Monte Carlo Comparison", *Economics Letters, 91,330-336*.

[6] Doornik, J. A. and O'Brien, R. J. (2002). "Numerically Stable Cointegration Analysis", *Computational Statistics & Data Analysis*, 41, 185-193.

[7] Fachin, S. (2000) . "Bootstrap and Asymptotic Tests of Long-run Relationships in Cointegrated Systems". *Oxford Bulletin of Economics and Statistics*, 62, 543-551.

[8] Gonzalo, J. (1994). "Comparison of Five Alternative Methods of Estimating Long-run Equilibrium Relationships". *Journal of Econometrics*, 60, 203-233.

[9] Gredenhoff, M., Jacobson, T. (2001). "Bootstrap Testing Linear Restrictions on Cointegrating Vectors", *Journal of Business & Economics Statistics*, 19, 63-72.

[10] Haug, A. (2002). "Testing Linear Restrictions on Cointegrating vectors: Sizes and Powers of Wald and Likelihood Ratio Tests in Finite Samples". *Econometric Theory*, 18, 505-524.

[11] Jensen, J. L. and Wood, T.A. (1997). "On the Non-existence of a Bartlett Correction for Unit Root Tests", *Statistics and Probability Letters*, 35, 181-225.

[12] Johansen, S. (1988), "Statistical Analysis of Cointegrated Vectors", *Journal of Economic Dynamic and Control,* 12, 231-254.

[13] 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*, 52, 169-210.

[14] Johansen, S. (1996). *Likelihood Inference in Cointegrated Vector Auto-Regressive Models*. Oxford University Press, Oxford.

[15] Johansen, S. (2000). "A Bartlett Correction Factor for Tests of on the Cointegrating Relations". *Econometric Theory*, 16, 740-778.

[16] Johansen, S. (2002). "A Small Sample Correction for Tests of Hypotheses on the Cointegrating Vectors", *Journal of Econometrics*, 111, 195-221.

[17] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

[18] Lagos, B. M. and Morettin, P. (2000). "Improvement of the Likelihood Ratio Test Statistic in ARMA Models". *Journal of Time Series Analysis*, 25, 83-101.

[19] Lawley, A.N. (1956). "A General Method for Approximating to the Distribution of the Likelihood Ratio Criteria". *Biometrika*, 43, 555-563.

[20] Nielsen, B. (1997). "Bartlett Correction of the Unit Root Test in Autoregressive Models". *Biometrika*, 84, 500-504.

[21] Podivinsky, J. M. (1992). "Small Sample Properties of Tests of Linear Restrictions on Cointegrating Vectors and their Weights", *Economics Letters*, 39, 13-18.

[22] Psaradakis, Z. (1994). "A Comparison of Tests of Linear Hypotheses in Cointegrated Vector Autoregressive Model", *Economics Letters* 45, 137-144.

[23] Rocke, D.M. (1989). "Bootstrap Bartlett Adjustment in Seemingly Unrelated Regression". *Journal of the American Statistical Association*, 84, 598-601.

[24] Rothemberg, T.J. (1984). "Hypothesis Testing in Linear Models When the Error Covariance Matrix is Non-scalar". *Econometrica*, 52, 827-842.

[25] Taniguchi, M. (1988). "Third-order Asymptotic Properties of a Class of Test Statistics Under Local Alternative". *Journal of Multivariate Analysis*, 37, 223-238.

[26] Toda, H. (1995). "Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Ranks in Vector Autoregressions". *Econometric Theory*, 11, 1015-1932.

.

Table 1. Size: Rejection frequencies (in percent) under the null of $\beta_{41} = 0$ and $\begin{bmatrix} \beta_{41} & \beta_{42} \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$ ($r = 1$ and $r = 2$, respectively). Parameters in the *DGP*: $\alpha_{11} = \alpha_{22} = 1$, $\beta_{11} = 0.4$, $\beta_{21} = 1.7$, $\beta_{31} = 0.1$, $\beta_{22} = 0.4$, $\beta_{32} = 0.5$, $\Omega = (0, \mathbf{I})$.

$Nominal\ Size:$			10%		
	LR	BRT	BSB	BSP	F
$r=1, k=1, T=50$	19.8	11.4	11.8	9.1	12.6
$r=1, k=2, T=50$	33.8	21.9	25.6	21.5	24.3
$r=1, k=3, T=50$	50.2	37.1	37.4	37.0	40.2
$r=2, k=1, T=50$	19.1	11.6	12.3	12.2	8.8
$r=2, k=2, T=50$	30.3	19.8	24.5	24.7	17.1
$r=2, k=3, T=50$	44.2	31.4	33.2	33.5	28.9
$r=1, k=1, T=100$	13.3	9.7	11.6	9.8	10.3
$r=1, k=2, T=100$	18.4	13.7	16.2	15.2	14.7
$r=1, k=3, T=100$	24.1	18.4	20.3	19.6	19.8
$r=2, k=1, T=100$	13.5	10.5	12.4	11.7	9.2
$r=2, k=2, T=100$	17.6	13.5	15.0	15.0	12.6
$r=2, k=3, T=100$	22.3	17.2	18.0	18.1	16.7
$r=1, k=1, T=150$	12.8	10.1	10.2	9.0	10.6
$r=1, k=2, T=150$	15.7	12.6	12.9	12.0	13.4
$r=1, k=3, T=150$	18.8	15.1	15.2	14.9	16.3
$r=2, k=1, T=150$	12.4	10.2	10.8	11.2	9.6
$r=2, k=2, T=150$	14.7	11.8	14.7	15.1	11.3
$r=2, k=3, T=150$	17.5	14.2	15.6	16.0	14.1

Table 1 Continue. Size: Rejection frequencies (in percent) under the null of $\beta_{41} = 0$ and $\begin{bmatrix} \beta_{41} & \beta_{42} \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}$ ($r = 1$ and $r = 2$, respectively). Parameters in the DGP: $\alpha_{11} = \alpha_{22} = 1$, $\beta_{11} = 0.4$, $\beta_{21} = 1.7$, $\beta_{31} = 0.1$, $\beta_{22} = 0.4$, $\beta_{32} = 0.5$, $\Omega = (0, \mathbf{I})$

Nominal Size :			5%					1%		
	LR	BRT	BSB	BSP	F	LR	BRT	BSB	BSP	F
$r = 1, k = 1, T = 50$	12.2	5.6	5.7	5.7	6.9	3.9	1.4	1.4	0.8	1.8
$r = 1, k = 2, T = 50$	23.8	11.9	13.2	13.0	15.2	10.1	3.9	3.7	3.7	5.0
$r = 1, k = 3, T = 50$	40.3	25.1	26.3	26.2	29.7	22.6	11.4	11.2	11.1	13.6
$r = 2, k = 1, T = 50$	11.2	6.0	6.2	6.8	4.2	3.4	1.1	1.7	1.6	0.5
$r = 2, k = 2, T = 50$	20.6	11.9	15.7	15.1	9.61	8.7	3.5	5.3	5.7	2.1
$r = 2, k = 3, T = 50$	33.1	23.4	24.5	24.0	18.9	16.3	7.6	8.5	8.7	5.6
$r = 1, k = 1, T = 100$	7.5	4.5	5.1	5.1	5.2	2.0	1.0	1.1	0.8	1.1
$r = 1, k = 2, T = 100$	11.2	7.8	7.6	7.5	8.1	3.3	1.8	2.1	1.9	1.9
$r = 1, k = 3, T = 100$	15.7	8.1	8.3	8.2	12.0	5.8	3.4	3.6	3.5	3.7
$r = 2, k = 1, T = 100$	7.5	5.1	5.8	6.2	4.4	2.1	1.0	1.0	1.2	0.8
$r = 2, k = 2, T = 100$	10.7	7.3	8.0	7.6	6.7	3.3	1.9	2.1	2.1	1.5
$r = 2, k = 3, T = 100$	14.6	10.7	11.2	10.9	9.6	4.8	3.0	3.2	3.3	2.6
$r = 1, k = 1, T = 150$	6.5	4.8	4.6	4.6	4.8	1.3	0.8	1.1	0.8	1.0
$r = 1, k = 2, T = 150$	9.0	6.7	7.0	6.8	7.0	2.3	1.4	1.5	1.3	1.6
$r = 1, k = 3, T = 150$	11.3	8.3	8.6	8.4	9.2	3.2	2.2	2.4	2.2	2.4
$r = 2, k = 1, T = 150$	6.7	5.0	6.1	6.5	4.4	1.3	0.9	1.3	1.4	0.7
$r = 2, k = 2, T = 150$	8.1	6.3	7.8	8.0	5.9	2.1	1.3	1.4	1.2	1.0
$r = 2, k = 3, T = 150$	10.1	8.0	9.1	9.5	7.1	2.7	1.7	1.8	1.7	1.5

Table 2. Size: Sensitivity analysis; rejection frequencies (in percent) under the null of $\beta_{41} = 0$. Results for α and σ (DGP with $r = 1$, $k = 1$, $T = 50, ..., 150$, nominal significance level 5%).

		LR	BRT	BSB	BSP	F
$\alpha = 0.2, \sigma = 1$	$T = 50$	30.6	12.0	8.1	9.2	23.0
	$T = 100$	17.8	9.2	6.5	6.4	14.5
	$T = 150$	12.1	6.9	5.6	6.1	10.3
$\alpha = 0.5, \sigma = 1$	$T = 50$	15.1	7.4	5.5	6.0	10.2
	$T = 100$	8.4	5.1	5.3	5.1	6.5
	$T = 150$	7.2	5.1	5.7	5.5	6.2
$\alpha = 0.7, \sigma = 1$	$T = 50$	11.7	5.8	4.5	4.7	7.4
	$T = 100$	7.0	4.7	5.3	5.6	5.4
	$T = 150$	6.5	4.7	5.3	5.3	5.4
$\alpha = 1, \sigma = 0.5$	$T = 50$	9.6	4.9	5.1	5.1	5.5
	$T = 100$	6.7	4.7	5.4	5.4	5.1
	$T = 150$	6.3	4.9	4.4	4.7	5.8
$\alpha = 1, \sigma = 1$	$T = 50$	9.5	5.1	4.4	4.5	5.8
	$T = 100$	6.6	4.6	4.9	4.7	5.0
	$T = 150$	5.9	4.4	5.4	5.4	4.8
$\alpha = 1, \sigma = 2$	$T = 50$	9.3	5.0	5.3	4.7	5.7
	$T = 100$	6.6	4.4	5.8	5.6	4.7
	$T = 150$	6.4	4.8	4.6	4.7	5.2
$\alpha = 0.2, \beta_{11} = -0.2, \beta_{21} = 0.5$	$T = 50$	38.1	29.1	8.1	10.5	29.5
	$T = 100$	30.0	7.2	8.0	9.5	26.3
	$T = 150$	21.8	7.4	5.4	6.4	19.4

Note: The following values are chosen, unless stated otherwise: $\beta_{11} = -0.9$, $\beta_{21} = 0.9$.

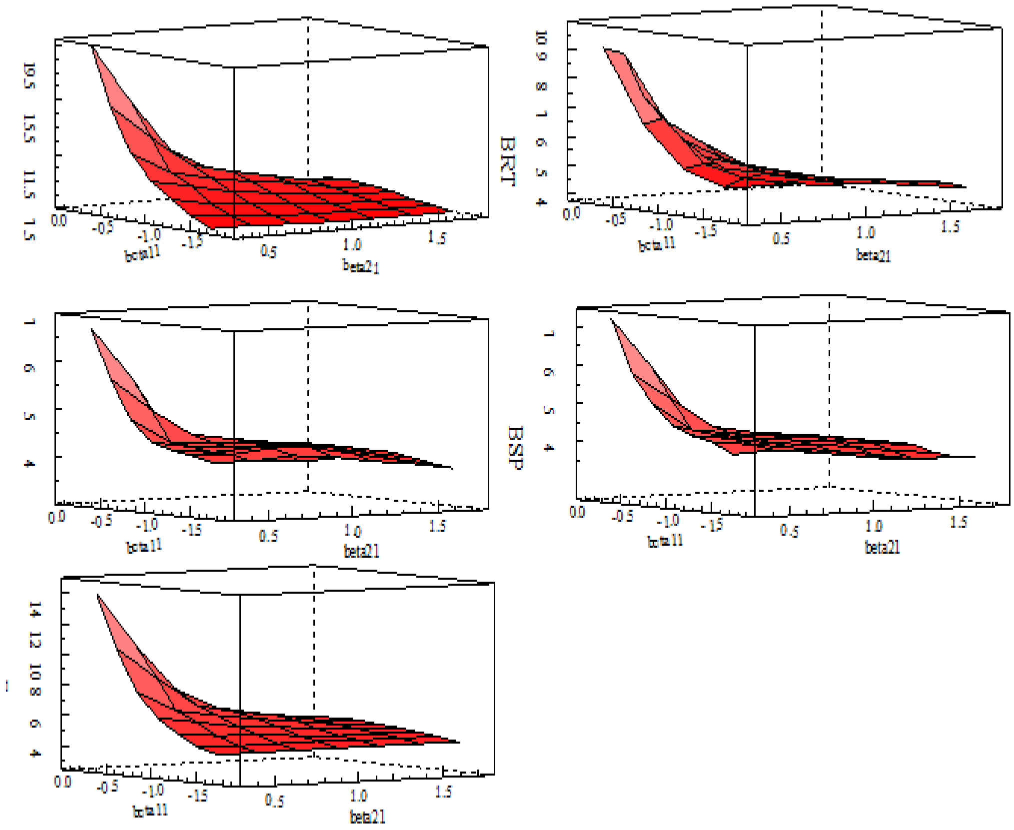

Figure 1: Size: Sensitivity analysis for β (rejection frequencies at nominal level 5% and $T = 50$).

Table 3. Power: Rejection frequencies under the alternative. Parameters in the DGP: $\alpha_{11} = \alpha_{22} = 1$, $\beta_{21} = 1.7$, $\beta_{31} = 0.1$, $\beta_{22} = 0.4$, $\beta_{32} = 0.5$, $\Omega = (0, \mathbf{I})$.

	H_1: β_{41}= 0.15					H_1: β_{41}= 0.3				
	LR	BRT	BSB	BSP	F	LR	BRT	BSB	BSP	F
$r=1, k=1, T=50$	51.5	38.1	39.4	38.7	40.2	85.4	77.5	77.3	76.6	78.9
$r=1, k=2, T=50$	54.1	39.9	43.1	43.7	42.8	79.4	69.0	73.3	73.4	71.6
$r=1, k=3, T=50$	61.4	46.7	48.7	49.1	50.9	78.3	66.0	72.4	72.6	69.8
$r=2, k=1, T=50$	55.1	38.8	41.9	41.1	33.8	85.8	78.5	80.6	80.6	75.0
$r=2, k=2, T=50$	52.5	39.3	44.5	44.5	34.8	79.6	70.2	72.9	72.8	66.1
$r=2, k=3, T=50$	57.3	43.2	50.3	50.2	39.8	76.6	65.4	66.0	66.1	62.1
$r=1, k=1, T=100$	88.1	85.0	87.1	87.2	85.6	99.7	99.6	99.6	99.6	99.6
$r=1, k=2, T=100$	84.8	80.9	82.1	81.7	81.9	98.9	98.4	98.9	98.6	98.5
$r=1, k=3, T=100$	82.2	76.8	80.2	79.9	78.6	97.1	95.7	96.2	96.0	96.1
$r=2, k=1, T=100$	88.4	85.8	86.9	86.4	84.9	99.6	99.5	99.5	99.4	99.4
$r=2, k=2, T=100$	85.6	82.1	83.7	83.4	81.2	99.0	98.5	98.7	98.6	98.4
$r=2, k=3, T=100$	82.8	78.1	80.0	79.9	77.4	97.4	96.4	96.7	96.5	96.3
$r=1, k=1, T=150$	98.3	98.0	96.8	98.7	98.0	100	100	100	100	100
$r=1, k=2, T=150$	97.5	96.8	96.9	96.8	97.0	100	100	100	100	100
$r=1, k=3, T=150$	96.2	95.3	95.8	95.6	95.3	99.9	99.8	100	100	99.9
$r=2, k=1, T=150$	98.1	97.8	98.2	98.0	97.7	100	100	100	100	100
$r=2, k=2, T=150$	97.3	96.6	97.5	97.5	96.5	100	100	100	100	100
$r=2, k=3, T=150$	96.2	95.3	95.9	95.6	95.2	99.9	99.8	99.8	99.8	99.8

Table 4. The Danish data: rejection frequencies for the size and the power, in percent, at nominal level 5%.

$T = 53$	LR	BRT	BSB	BSP	F
Size	20.5	10.4	11.5	11.6	12.6
Power	83.6	73.6	73.1	73.1	77.1

Note: $n = 10,000$, $B = 400$. The asymptotic distribution is a $\chi^2(3)$.