Testing Structural Hypotheses on Cointegration Vectors: A Monte Carlo Study

Eriksson, Åsa

2004

Document Version:
Other version

Link to publication

Citation for published version (APA):
Eriksson, Å. (2004). *Testing Structural Hypotheses on Cointegration Vectors: A Monte Carlo Study*. (Working Papers, Department of Economics, Lund University; No. 29). Department of Economics, Lund University.

Total number of authors:
1

Testing Structural Hypotheses on Cointegration Vectors: A Monte Carlo Study

Åsa Eriksson*

December, 2004

Abstract

In this paper, two tests for structural hypotheses on cointegration vectors are evaluated in a Monte Carlo study. The tests are the likelihood ratio test proposed by Johansen (1991) and the test for stationarity proposed by Kwiatkowski et al. (1992). The analysis of the likelihood ratio test is extended with the inclusion of a Bartlett correction factor. Under circumstances common in empirical applications, all tests suffer from large size distortions and have low power to detect a false cointegration vector, but the Johansen (1991) test fares slightly better than the Kwiatkowski et al. (1992) test. Applying a Bartlett correction factor in small samples improves to a large extent the likelihood ratio test.

JEL Classification: C12, C15, C22

Keywords: Cointegration, Structural hypothesis, Monte Carlo simulation

*Department of Economics, Lund University, P.O. Box 7082, SE-220 07 Lund, Sweden. Email: Asa.Eriksson@nek.lu.se. I would like to thank Lisbeth la Cour, Michael Bergman, Pontus Hansson and seminar participants at Lund University for helpful discussions, comments and suggestions.

1 Introduction

Cointegration analysis is widely used in empirical economics. When analysing cointegration relationships, economic theory often predicts specific cointegration vectors that are to be evaluated. Thus, in many applications, the crucial part of the analysis is to test such structural hypotheses on the cointegration vectors. This is, for example, the case when testing the Purchasing Power Parity (PPP) hypothesis. For the PPP hypothesis to hold, a country's real exchange rate has to be stationary. In a two-variable model, this can be stated as follows: the natural logarithm of the foreign price level in domestic currency and the natural logarithm of the domestic price level should be cointegrated with cointegration vector $\begin{bmatrix} 1 & -1 \end{bmatrix}$. Since the theoretical cointegration vector implies a one-to-one relation between the two variables, any other cointegration vector leads to the conclusion that the PPP theory is not valid. Therefore, in order to be able to draw appropriate conclusions regarding economic theory, it is vital to analyse and evaluate different tests used for testing hypothesis on cointegration vectors, on their abilities to distinguish between true and false cointegration vectors.

In this paper, we analyse and compare two econometric methods that are common in cointegration analysis. We focus on situations where we want to test a specific cointegration vector given by economic theory, as in the PPP example above. The first method is the likelihood ratio test for structural hypotheses proposed by Johansen (1991) and Johansen and Juselius (1990, 1992), henceforth called the LR test. Among various likelihood ratio tests, this is the most commonly used test and it has also been shown to posses the most desirable statistical properies (see for example Haug (2002)). The second method is based on the Engle–Granger methodology (Engle and Granger, 1987). Here, we apply the commonly used test proposed by Kwiatkowski et al. (1992) (the KPSS test) to the residuals from the long-run regression and test the residuals for stationarity. The Johansen method and the Engle–Granger method are the two most frequently used methods in cointegration analysis, which makes the comparison between the tests highly important for empirical applications.

The LR test is a multivariate test that tests a null hypothesis of a specified cointegration vector against the alternative hypothesis that the specified vector is not valid. The test is asymptotically χ^2 distributed, but the distribution is not always a good approximation in small samples, as pointed out by, for example, Jacobson (1995) and Zhou (2000). To help reducing the problems in finite samples, a so-called Bartlett correction factor can be applied to the test.[1] In the present study, the LR test is performed both with and without the Bartlett correction factor, in order to establish its consequences for the test.

The KPSS test is univariate and tests the residuals from a long-run relationship for stationarity in a version of the Engle–Granger method. Instead of estimating the long-run relationship, which is the common strategy when using the Engle–Granger method, we specify the long-run coefficients, extract the residuals and apply the KPSS test. We then have a test for a specific cointegration vector in this framework as well. Since the KPSS test has stationarity of the residuals as its null hypothesis, failing to reject the null hypothesis implies that the chosen long-run relation is a cointegration vector. Thus, the null hypotheses for the LR test and the KPSS test are analogous and a straightforward comparison between the two procedures is legitimate.[2]

The analysis in this paper is conducted in a Monte Carlo study. We examine both the size, i.e. the probability of rejecting a null hypothesis if the null hypothesis is true, and the power, i.e. the probability of rejecting a null hypothesis if the null hypothesis is false, of the two tests. Throughout the analysis, we work with a bivariate model. Although

[1] The Bartlett correction originate from Bartlett (1937). A presentation of the ideas behind the Bartlett correction can be found in Jacobson and Larsson (1999).

[2] The most common strategy when applying the Engle–Granger methodology in cointegration analysis, is to test the estimated residuals from the long-run regression with a unit root test, for example an augmented Dickey–Fuller test or a Phillips–Perron test. The null hypothesis of these tests is the presence of a unit root in the data, i.e. no cointegration relation between the variables. Therefore, these tests are not directly comparable with the LR test for structural hypotheses in the Johansen approach, where the null hypothesis implies the presence of a cointegration relation in the model. In this study, however, we compare two tests with analogous null hypotheses. A Monte Carlo study of different cointegration tests with null hypothesis of no cointegration, including the augmented Dickey–Fuller test and the Phillips–Perron, is made by Östermark and Höglund (1999). In the study, they analyse cointegration test but not tests for structural hypotheses on cointegration vectors.

this is a very simple model, it can be appropriate for some economic questions where cointegration analysis is commonly used, for example when testing the PPP hypothesis or the relation between income and consumption. It is important to evaluate statistical tests under circumstances common in empirical studies. The aim is to design the simulation experiments accordingly. We therefore choose values of the parameters in the model often appearing in such studies. One important process in a cointegration model is how fast the variables return back to equilibrium after a deviation from the long-run relation. In many empirical macroeconomic applications, this adjustment process is very slow. We extend earlier research regarding the properties of cointegration tests by analysing slower adjustment processes than before. We also focus on the properties of the KPSS test in applications where we test hypotheses on cointegration vectors.

The paper is organized as follows. Previous studies are discussed in section 2. The LR test and the KPSS test are discussed in section 3 and in section 4, the set-up of the simulation experiment is presented. The empirical results regarding size and power of the tests are presented in section 5. Conclusions are made in section 6.

2 Previous studies

The LR test for testing structural hypotheses on cointegration vectors has been examined by Jacobson (1995), Zhou (2000), Gredenhoff and Jacobson (2001) and Haug (2002). These studies suggest that the LR test is biased towards rejecting a true null hypothesis more often than asymptotic theory suggests. The bias is greater when the sample size is smaller, when the lag length is longer and when the cointegration relationship is highly serially correlated. Jacobson (1995) also studies the power of the LR test and finds that the power is low for detecting a true the cointegration vector close to the one in the null hypothesis. However, the power is fairly high when the sample size is large. Fachin (2000) compares the Johansen (1991) test with a bootstrap procedure for the test in a Monte Carlo study. He finds that the empirical size of test improves with the bootstrap procedure compared to the asymptotical test, but the power of the test may be affected. He therefore propose a combined test based on the outcome of the asymptotic and bootstrap tests that gives small size distortions and high power. Haug (2002) compares several likelihood ratio tests and Wald tests with respect to size and power properties. He finds that Johansen's LR test is preferable regarding size and that it performs well regarding power in comparison to the other tests, although some tests have better power.

Johansen (2000) has derived a Bartlett correction factor for the LR test. In simulation experiments, he shows that the correction helps reducing the empirical size of the test. This conclusion is also reached by Haug (2002). Omtzigt and Fachin (2002) analyse the use of the Bartlett correction factor when testing hypotheses on cointegration vectors. They compare a Bartlett correction with two types of bootstrap methods, concluding that the Bartlett correction gives less power losses compared to the bootstrap methods.

3 Statistical tests

3.1 The Johansen method and the LR test

The Johansen method for cointegration analysis is based on the works of Johansen (1988, 1991) and Johansen and Juselius (1990, 1992). The basis for the analysis is a vector autoregressive model (VAR) with dimension p, equal to the number of variables in the model. The VAR model can be written as

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + ... + A_k X_{t-k} + \epsilon_t \tag{1}$$

where k is the number of lags in the model and X_t is a variable vector defined as $X_t = \begin{bmatrix} x_{1t} & ... & x_{pt} \end{bmatrix}'$. The error term ϵ_t is assumed to be independently and identically normally distributed with mean zero and covariance matrix Σ. The VAR model can be rewritten in its vector error correction form as

$$\Delta X_t = A_0 + \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Gamma_2 \Delta X_{t-2} + ... + \Gamma_{k-1} \Delta X_{t-k+1} + \epsilon_t \tag{2}$$

where $\Gamma_i = -\sum_{j=i+1}^{k} A_j$ and $\Pi = \sum_{i=1}^{k} A_i - I_n$. Since the simulation study in this paper is conducted in a two-variable model with one lag, the variable vector $X_t = \begin{bmatrix} x_t & y_t \end{bmatrix}'$ and the vector error correction model simplifies to

$$\Delta X_t = A_0 + \Pi X_{t-1} + \epsilon_t. \tag{3}$$

The Johansen analysis is concerned with the matrix Π. In a cointegrated model, the Π-matrix has reduced rank and can be decomposed into two matrices, α and β, as $\Pi = \alpha\beta'$. Both α and β have the dimension $(p \times r)$, where r is the number of cointegrated vectors in the model. The matrix β is a matrix of long-run cointegration relations and the elements in the matrix α are adjustment coefficients, determining the speed of adjustment back to equilibrium after a deviation from the long-run relation. The number of cointegrated vectors, r, is also equal to the rank of Π. If $r(\Pi) = 2$ in equation (3), Π has full rank and the variables in X_t are stationary. If $r(\Pi) = 0$, X_t is non-stationary, but there is no cointegration relation between the variables, and finally, if $r(\Pi) = 1$, X_t is non-stationary and there exists one cointegration relation among the variables.

Testing for the number of cointegration relations and obtaining estimates of α and β are done with the Johansen maximum likelihood procedure and Johansen's trace statistic. Details about the estimation procedure and the trace statistic are found in Johansen (1988).

After establishing the presence of cointegration in the model and the number of cointegration vectors, the interest turns to testing hypotheses regarding the parameters in the matrices α and β. This kind of structural hypotheses can be tested with the LR test proposed by Johansen (1991) and Johansen and Juselius (1990, 1992).[3] Suppose we want to test the hypothesis that the cointegration vector $\beta = \begin{bmatrix} 1 & -\beta_1 \end{bmatrix}'$. If this is the cointegration vector, $\beta' X_t$ is stationary; otherwise the variables are not cointegrated. The presence of cointegration relations in the model is determined by the rank of the Π-matrix, which also equals the number of eigenvalues of Π that is different from zero. The LR test compares the estimated eigenvalues from the model without any restrictions on α and β, with the eigenvalues estimated with the restrictions defined in the structural hypothesis imposed. If the restriction $\beta = \begin{bmatrix} 1 & -\beta_1 \end{bmatrix}'$ is valid, the eigenvalues of the two models should be the

[3] Johansen (1991) and Johansen and Juselius (1990, 1992) propose various tests for hypotheses on the cointegration vectors. In this paper, we focus on situations where we test if a specific vector belongs to the cointegration space. In addition to this, we can test the same set of restrictions on all cointegration vectors, restrictions on one cointegration vector leaving the other vectors unrestricted and different restrictions on each cointegration vector. In a bivariate model with one cointegration vector, some of these hypotheses result in the same model.

same; if the restriction is not valid, the eigenvalues should differ to a large extent. The LR test statistic is calculated as

$$Q = -T \sum_{i=1}^{r} \ln\left(\frac{1-\hat{\lambda}_{A,i}}{1-\hat{\lambda}_i}\right) \tag{4}$$

where $\hat{\lambda}_i$ are the eigenvalues of the Π-matrix estimated under the null hypothesis and $\hat{\lambda}_{A,i}$ are the eigenvalues estimated under the alternative hypothesis. Asymptotically, this test statistic is χ^2 distributed with degrees of freedom equal to the number of restrictions placed on the vector β. The null hypothesis is rejected if the test statistic exceeds the critical value from the χ^2 distribution.

3.2 The Bartlett correction factor

The LR test for structural hypotheses on cointegration vectors is asymptotically χ^2 distributed. In finite samples, the χ^2 distribution is not always a good approximation of the distribution and to improve the test under such circumstances, a Bartlett correction factor can be used.

The idea behind the Bartlett correction factor is that the test statistic Q_T, converges to Q_∞ when the sample size, T, goes to infinity. In finite samples, Q_T has an error term of order $1/T$. With a Bartlett correction factor, the test statistic Q_T is transformed, to Q_T^*, which converges faster towards Q_∞, with an error term of order $1/T^2$. The correction factor is based on the expectation of the test statistic Q_T, denoted $E[Q_T]$. Following the derivations in Jacobson and Larsson (1999), we take advantage of knowing that $\frac{Q_T}{E[Q_T]}$ approaches $\frac{Q_\infty}{E[Q_\infty]}$ as $T \to \infty$. We then know that

$$Q_T \approx E[Q_T]\frac{Q_\infty}{E[Q_\infty]}. \tag{5}$$

Generally, $E[Q_T]$ is not known and a Taylor expansion of $E[Q_T]$ under the null hypothesis has to be made. The expansion has the form

$$E[Q_T] = E[Q_\infty] + \frac{B}{T} + O\left(\frac{1}{T^2}\right) \tag{6}$$

and by substituting equation (6) into equation (5), and defining $B_0 = \frac{B}{E[Q_\infty]}$, we obtain the Bartlett corrected test statistic as

$$Q^* = Q\left(1 + \frac{B_0}{T}\right)^{-1}. \tag{7}$$

The factor B has to be derived specifically for each model and test statistic. In this paper, the derivation of B is taken from Johansen (2000). The derivation is made for a model with one cointegration vector and one lag, which is identical to the model used in this paper.

3.3 The Engle–Granger method and the KPSS test

Cointegration analysis using the Johansen method takes place in a multi-equation framework. The widely used method proposed by Engle and Granger (1987) is a single-equation method and consists of two steps. For the method to be suitable, all variables in the model have to be integrated of the same order. Assuming that both variables in our bivariate model are integrated of order one, the first step in the analysis is to estimate the long-run relationship between the variables. The long-run relationship between variables y_t and x_t is

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t \tag{8}$$

which can be estimated with ordinary least squares. After the estimation, the residual series $\{\hat{\epsilon}_t\}$ is extracted and saved for further analysis. If two variables are cointegrated, there exists at least one linear combination among them that yields a stationary relation. Therefore, if the residual series $\{\hat{\epsilon}_t\}$ is stationary, the variables y_t and x_t are cointegrated and if $\{\hat{\epsilon}_t\}$ is non-stationary, the variables are not cointegrated. The second step in the Engle–Granger analysis is to test the residuals for stationarity or for the presence of a unit root.

If we want to test a hypothesis about a specific cointegration vector, a version of the Engle–Granger method can be used. We then construct a variable, $u_t = y_t - \beta_0 - \beta_1 x_t$, where β_1 is chosen and not estimated, and test this variable for stationarity. If the variable u_t is stationary, the variables are cointegrated and the cointegration vector is $\begin{bmatrix} 1 & -\beta_0 & -\beta_1 \end{bmatrix}$.

In this paper, we use the test developed by Kwiatkowski et al. (1992), the KPSS test, and test if the variable u_t is stationary.

The KPSS test starts with regressing the variable of interest, in this case u_t, on a set of deterministic components. These can be either an intercept or an intercept and a trend. Here, we specify the model with a constant only, and run the regression

$$u_t = \zeta_t + \rho_t. \tag{9}$$

ζ_t is a random walk given by $\zeta_t = \zeta_{t-1} + \nu_t$, where $\nu_t \sim (0, \sigma_\nu^2)$. The null hypothesis of the KPSS test is

$$H_0 : \sigma_\nu^2 = 0. \tag{10}$$

This implies that ζ_t is a constant, which in turn implies that under the null hypothesis, the variable u_t is level stationary. The test is based on the partial sum series

$$S_t = \sum_{i=1}^{t} \hat{\rho}_i \tag{11}$$

where the series $\{\hat{\rho}_t\}$ is the residuals from regressing equation (9). Kwiatkowski et al. (1992) define the long-run variance

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{T} E(S_T^2) \tag{12}$$

which is estimated by the consistent estimator given by

$$s^2(\ell) = \frac{1}{T}\sum_{t=1}^{T} \hat{\rho}_t^2 + 2\frac{1}{T}\left[\sum_{j=1}^{\ell} w(j,\ell) \sum_{t=j+1}^{T} \hat{\rho}_t \hat{\rho}_{t-j}\right] \tag{13}$$

where $w(j, \ell)$ is a weighting function. Like Kwiatkowski et al. (1992), we use the weighting function

$$w(\ell, j) = 1 - \frac{j}{\ell + 1} \tag{14}$$

and we chose the lag truncation parameter ℓ as $\ell = \sqrt{T}$. The test statistic of the KPSS test, testing the null hypothesis of stationarity in u_t versus a stochastic trend, is then

$$\eta = \frac{1}{T^2}\sum_{t=1}^{T} \frac{S_t^2}{s^2(\ell)} \tag{15}$$

Critical values for the test are available in Kwiatkowski et al. (1992).

4 The Monte Carlo study

4.1 The model

The evaluation of the tests presented in the previous section is made using generated data. The data originate from a two-variable system with variables x_t and y_t, which are defined as

$$x_t = x_{t-1} + \epsilon_{xt} \tag{16}$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t-1} + \epsilon_{yt} \tag{17}$$

The disturbances ϵ_{xt} and ϵ_{yt} are assumed to be normally distributed with mean zero and variance σ_x^2 and σ_y^2, respectively. The covariance between the two disturbances is denoted θ.[4] The two variables are rewritten in error-correction form as

$$\Delta x_t = \epsilon_{xt} \tag{18}$$

$$\Delta y_t = \phi_2(y_{t-1} - \gamma - \beta_1 x_{t-1}) + \epsilon_{yt} \tag{19}$$

where, $\phi_2 = \alpha_1 - 1$, $\gamma = \alpha_0/(1-\alpha_1)$ and $\beta_1 = \alpha_2/(1-\alpha_1)$. With the variance of ϵ_{yt} normalized to one and the relative variance between ϵ_{xt} and ϵ_{yt} denoted σ^2, the joint distribution of the two disturbances can be formulated as

$$\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \theta \\ \theta & 1 \end{pmatrix}\right]. \tag{20}$$

The two-variable system is a rather simple system. It includes x_t, a random walk variable, and y_t, that follows a more complex process, and under certain circumstances, the two variables are cointegrated. If $-1 < \phi_2 < 0$, cointegration is present in the model and the cointegration relationship is given by the expression in parenthesis in equation (19), $y - \gamma - \beta_1 x$. The cointegration vector is thus $\begin{bmatrix} 1 & -\gamma & -\beta_1 \end{bmatrix}$, where γ is a constant. The parameter ϕ_2 is the adjustment coefficient and gives us information about the persistence of the deviations from the long-run relationship. If $|\phi_2|$ is high, y_t rapidly returns to

[4] This structure of the generated data is used in simulation studies on similar topics by, among others, Östermark and Höglund (1999), Zhou (2000) and Haug (2002).

equilibrium after a deviation from the cointegration relationship and if $|\phi_2|$ is low, the adjustment process is slow. As seen from the error correction form in equation (18) and (19), all adjustment takes place in the variable y_t; the variable x_t follows a random walk and does not adjust towards equilibrium.

4.2 Choice of parameter values

The characteristics of the tests depend on a variety of circumstances, such as sample size, parameter values and variances and covariances of the error terms. The aim of this study is to analyse the tests under circumstances common in empirical studies, and we choose values of the different parameters in the data generating process, corresponding to these situations.

While many statistical results concerning a certain test are based on asymptotic theory, in practice the test is carried out in a finite sample. On many occasions, we have a very limited number of observations and therefore, our simulations include situations with small sample sizes. The sample sizes, T, are chosen as $T = \left\{50 \quad 100 \quad 300\right\}$.

As noted in the previous section, a parameter value of ϕ_2 in the range of $-1 < \phi_2 < 0$ implies that the two variables are cointegrated. A high absolute value of the adjustment parameter further implies that the adjustment process towards the cointegration relation is fast, and vice versa. However, in empirical applications concerning the PPP hypothesis, the speed of adjustment is often very low. For example, Rogoff (1996) reports half lives for PPP deviations of about 3 to 5 years when summarizing a number of PPP studies. This corresponds to adjustment parameters of –0.03 to –0.06 on a quarterly basis.[5] The chosen values for the adjustment parameter in this study are $\phi_2 = \left\{-0.01 \quad -0.04 \quad -0.2 \quad -0.5\right\}$. The values are chosen to span over a range of possible parameter values, also including very low parameter values in absolute terms, reflecting the aforementioned empirical results.

The two parameters in the joint distribution of ϵ_{xt} and ϵ_{yt} may also affect the results.

[5] For other empirical studies with low adjustment parameters, see for example Banerjee et al. (2001) for a study of the relation between inflation and the mark-up and Johansen and Juselius (1990), for a study of money demand.

In the simulations, the parameters σ^2 and θ can take the values $\sigma^2 = \begin{Bmatrix} 0.5 & 1 & 2 \end{Bmatrix}$ and $\theta = \begin{Bmatrix} -0.25 & 0 & 0.25 \end{Bmatrix}$. If $\sigma^2 < 1$, the variance of the disturbance in the random walk variable is relatively lower than the variance of the disturbance to the other variable. This implies that y_t, the variable adjusting to the cointegration relationship, has relatively high variance.

The tests presented in section 3 are applied to the bivariate model in equations (18) and (19). Throughout the all analysis we set the value of the constant $\gamma = 0$ and test the cointegration vector $\beta = \begin{bmatrix} 1 & -1 \end{bmatrix}$. This is the vector in the null hypothesis of the LR test and when analysing the KPSS test, the variable u_t is constructed as $y_t - x_t$.

We are interesting in analysing the power of the tests to detect a false cointegration vector. We therefore vary the parameter in the cointegration vector and let β_1 take the values $\beta_1 = \begin{Bmatrix} 0.3 & 0.5 & 0.8 & 0.95 & 1 \end{Bmatrix}$.

5 Empirical results

In this section, the results from the Monte Carlo study are presented. The simulation experiment is carried out as follows. Normally distributed random numbers are drawn in the program GAUSS and two variables are generated according to equations (18) and (19). Test statistics for the LR test for structural hypotheses, with and without the Bartlett correction factor, and the KPSS test, are calculated and compared to critical values at the 5 percent significance level. The procedure is repeated 10000 times. Finally, rejection frequencies are calculated for the different parameter combinations, presented in section 4.2. The tests are applied to the generated data for the possible parameter combinations.[6]

The empirical size of the tests is calculated by counting the number of times the null hypothesis is rejected if the null hypothesis is true. Since we place one restriction on the cointegration vector in the LR test, i.e. that the parameter $\beta_1 = 1$, critical values from the χ^2 distribution with one degree of freedom is used. Kwiatkowski et al. (1992) provide critical values for the KPSS test.

[6]The total number of parameter combinations is 540. For space-saving reasons, the result regarding all parameter combinations will not be presented the paper.

If the null hypothesis is false, we want to reject it. The rejection frequencies of the tests when the null hypothesis is false is the power of the tests. When investigating the power of the tests, we use empirical critical values from our simulations, i.e. we use as critical values test statistics from our 10000 trials that would make us reject the null hypothesis in 5 percent of the trials when the null hypothesis is true. If the empirical size of a test depends on the parameter combination used, we have used parameter-specific critical values. By doing so, we correct the sizes to be the same across all tests, which is necessary in order to be able to make a valid comparison of the power and the rejection frequencies both within and between the tests.

5.1 Empirical size

5.1.1 The LR test for structural hypotheses

We start by examining the empirical size of the LR test. Since the empirical size is the probability of rejecting a true null hypothesis, we evaluate the parameter combinations from the simulations where the parameter β_1 in equation (19) is $\beta_1 = 1$. In this case, the null hypothesis of the LR test is true.

In table 1, the empirical size of the test is presented for some of the parameter combinations. The empirical size of the test turns out to depend on all parameters in the model. The two parameters with most impact on the size are the adjustment parameter, ϕ_2, and the number of observations, T. In the first panel of table 1, the result when $\sigma^2 = 1$ and $\theta = 0$ is shown. We see that the empirical size of the test is often much higher than the nominal 5 percent, especially when the number of observations is small and the absolute value of the adjustment parameter is low. For $\phi_2 = -0.01$ and $\phi_2 = -0.04$, the empirical size is between 28 and 43 percent for a sample size of 50 or 100, while it is 31 percent and 12 percent, respectively, for 300 observations. The empirical size decreases when the number of observations increases, and is close to 5 percent for 100 and 300 observations and values of the adjustment parameter that imply a fast adjustment process.

The size effects of a change in the value of the parameter σ^2 are shown in the two lower panels in table 1. Compared to $\sigma^2 = 1$, the empirical size is higher when $\sigma^2 < 1$ and lower

Table 1: Empirical size of the LR test for structural hypotheses

β_1	σ^2	θ	ϕ_2	$T=50$	$T=100$	$T=300$
1	1	0	-0.01	0.429	0.401	0.312
			-0.04	0.361	0.275	0.120
			-0.2	0.141	0.085	0.060
			-0.5	0.076	0.062	0.053
1	0.5	0	-0.01	0.433	0.410	0.330
			-0.04	0.376	0.300	0.139
			-0.2	0.162	0.094	0.063
			-0.5	0.084	0.064	0.054
1	2	0	-0.01	0.424	0.389	0.278
			-0.04	0.335	0.232	0.099
			-0.2	0.115	0.076	0.058
			-0.5	0.071	0.060	0.054

Note: The table shows the empirical size of the test, i.e. the rejection frequencies for the test when the null hypothesis is true.

when $\sigma^2 > 1$. This means that if the relative variance in the variable adjusting to the cointegration relation is high, the case when $\sigma^2 < 1$, we more often erroneously reject a true null hypotheses. The value of θ influences the empirical size to a minor degree and the result is not shown in the table.

The results above point in the same direction as the results in Zhou (2000), Jacobson (1995) and Haug (2002); the LR test is biased towards rejecting a true null hypothesis too many times. For all parameter combinations, the size is higher than the nominal 5 percent, but the problem is particularly severe if the adjustment parameter ϕ_2 is low in absolute value, so that the adjustment process to the cointegration relationship is slow. In these cases, we reject up to 40 percent of the true null hypotheses. This is a relevant concern in empirical economics, since few observations and a slow speed of adjustment are common in empirical applications.

In figure 1, the empirical size of the LR test is presented in a different way. The level curves in the figure show combinations of the adjustment parameter, ϕ_2, and the number of

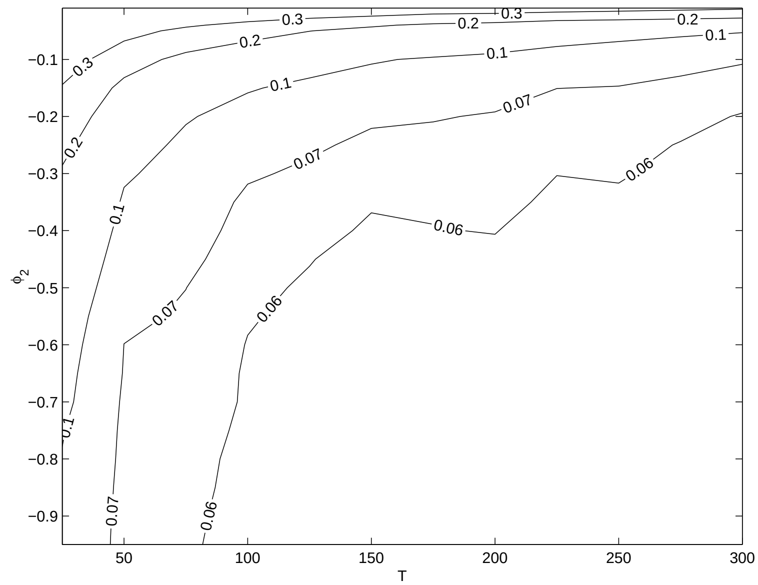

Figure 1: Empirical size of the LR test for structural hypotheses

observations, T, that yield a certain empirical size. The figure is based on simulations of the variables x_t and y_t in section 4.1, but with a different set of parameters than before. Since the result presented earlier showed that the two parameters affecting the empirical size the most was ϕ_2 and T, the relative variance and the covariance between the disturbances in equation (18) and (19) are kept fixed at $\sigma^2 = 1$ and $\theta = 0$, in the figure.[7] Two different aspects of the empirical size can be seen in the figure. First, for a given level curve, all parameter combinations above the curve yield a higher empirical size and all combinations below the curve yield a lower size. Worth noticing is that no combination of ϕ_2 and T yields an empirical size of 5 percent. We also see the trade-off between values of the adjustment parameter and the sample size, since we see which empirical size different combinations of the adjustment parameter and the number of observations yield. For example, with 50 observations, we need an adjustment parameter of –0.6 to get an empirical size of 7 percent, but with 300 observations, the adjustment parameter only has to be –0.1. With $|\phi_2| < 0.1$, we seldom get an empirical size less than 10 percent, regardless of the number of observations.

[7] In the simulations behind figure 1, the parameter values of ϕ_2 and T are $\phi_2 = \{-0.01 \quad -0.04 \quad -0.05 \quad -0.1 \quad -0.15 \quad ... \quad -0.9 \quad -0.95\}$ and $T = \{25 \quad 50 \quad 75 \quad ... \quad 275 \quad 300\}$.

Table 2: Empirical size of the Bartlett corrected LR test for structural hypotheses

β_1	σ^2	θ	ϕ_2	$T = 50$	$T = 100$	$T = 300$
1	1	0	-0.01	0.239	0.219	0.165
			-0.04	0.202	0.147	0.073
			-0.2	0.089	0.061	0.051
			-0.5	0.058	0.052	0.049
1	0.5	0	-0.01	0.236	0.225	0.181
			-0.04	0.210	0.162	0.082
			-0.2	0.098	0.064	0.051
			-0.5	0.062	0.053	0.050
1	2	0	-0.01	0.236	0.210	0.149
			-0.04	0.179	0.126	0.063
			-0.2	0.076	0.057	0.050
			-0.5	0.056	0.051	0.050

Note: The table shows the empirical size of the test, i.e. the rejection frequencies for the test when the null hypothesis is true.

5.1.2 The Bartlett corrected LR test for structural hypotheses

In this section, the LR test is adjusted with the Bartlett correction factor, with the aim of improving the distribution of the test statistic in finite samples. The empirical size of the Bartlett corrected LR test is shown in table 2. Comparing the tables 1 and 2, we see that the Bartlett correction has a significant impact on the empirical size of the test. The correction makes us reject the null hypothesis more seldom, lowering the empirical size closer to the nominal 5 percent. With the Bartlett correction, the size is close to 5 percent for several parameter combinations, even with few observations. However, for $\phi_2 = -0.01$ and $\phi_2 = -0.04$, the size is still considerably higher than 5 percent. The pattern arising from changing the values of the parameters σ^2 and θ is similar to the pattern without the correction.

In figure 2, level curves for the empirical size of the Bartlett corrected LR test are presented. Compared to figure 1, the curves have moved in an up-left direction. This reflects the fact that the empirical size of the Bartlett corrected LR test is lower for a

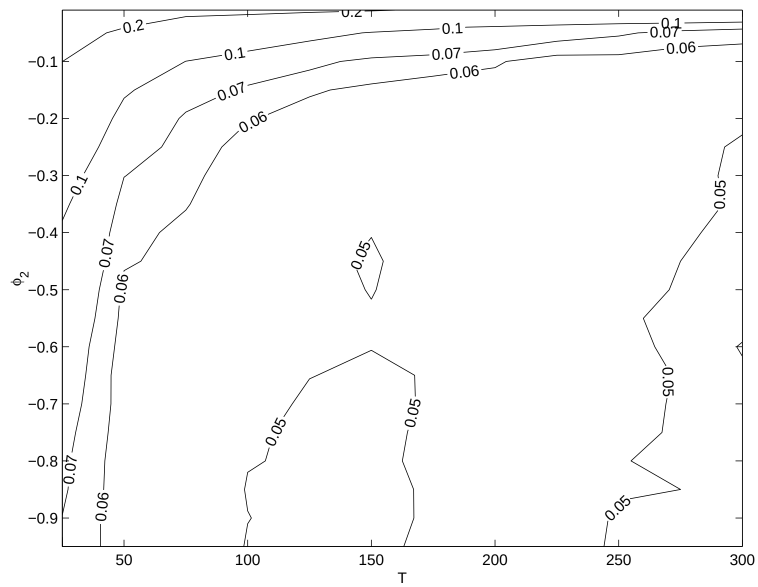

Figure 2: Empirical size of the Bartlett corrected LR test for structural hypotheses

given parameter combination than the test without the correction. Now, some parameter combinations yield an empirical size of 5 percent. Around the 5 percent level, the variation in the empirical size is very low, which explains the form of the level curve for 5 percent.

5.1.3 The KPSS test

Finally, we apply the KPSS test of stationarity on the constructed variable $u_t = y_t - x_t$ and analyse the empirical size of the test. As for the LR test, the null hypothesis is true when $\beta_1 = 1$.

In table 3, the empirical size of the KPSS test is presented, and we see that the test suffers from large size distortions for some parameter combinations. If the adjustment parameter is –0.01, the empirical size of the test lies in the range 44 percent to 63 percent. For these parameter combinations, the size is higher for a large sample size compared to a smaller. This pattern for the empirical size is found by Kwiatkowski et al. (1992) and Amano and van Norden (1992). In simulation studies, they found that the empirical size increased with an increasing sample size for a fixed level of truncation parameter ℓ. In our simulations, there are some exceptions from this pattern, for example when $\phi_2 = -0.2$, the

Table 3: Empirical size of the KPSS test

β_1	σ^2	θ	ϕ_2	$T = 50$	$T = 100$	$T = 300$
1	1	0	−0.01	0.439	0.528	0.627
			−0.04	0.342	0.395	0.359
			−0.2	0.111	0.112	0.084
			−0.5	0.043	0.054	0.053

Note: The table shows the empirical size of the test, i.e. the rejection frequencies for the test when the null hypothesis is true.

empirical size is lower for 300 observations than for 50 and 100 observations. The influence on the empirical size from σ^2 and θ is of minor importance and the results are therefore not presented in the table.

Level curves for the empirical size of the KPSS test are shown in figure 3. The curves have a different appearance compared to the curves for the LR test. Except when the number of observations is very small, the curves are fairly constant across the number of observations, for given values of the adjustment parameter. This implies that for many values of the adjustment parameter, the test will never reach the nominal size even with a large number of observations. When the absolute value of the adjustment parameter increases, the empirical size decreases. In the area below the lowest curve, the empirical size is below 5 percent.

The KPSS test has large size distortions for parameter combinations common in empirical studies, especially for small absolute values of the adjustment parameter. Under these circumstances, the distortions are larger than for the LR test and the Bartlett corrected LR test. With fast adjustment processes, the empirical size of the KPSS test is somewhat closer to 5 percent, compared to the LR test. For many values of ϕ_2 the empirical size depends almost solely on the value of the adjustment parameter and not on the number of observations.

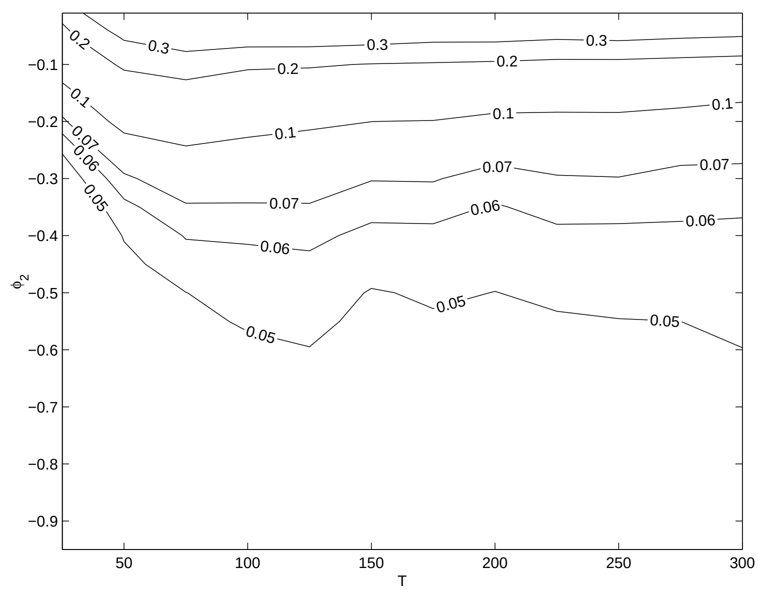

Figure 3: Empirical size of the KPSS test

5.2 Power

5.2.1 The LR test for structural hypotheses

The power of a test is the ability to detect a false null hypothesis. In this section, we evaluate the LR test when the parameter in the cointegration relation $\beta_1 \neq 1$, so that the generated data does not support the restriction in the null hypothesis.

The power of the LR test is presented in table 4. The rejection frequencies shown are the number of times we reject the null hypothesis $\beta = \begin{bmatrix} 1 & -1 \end{bmatrix}$ for each parameter combination, divided by the total number of trials the test is done, i.e. 10000.[8] The power of the LR test depends to a large extent on the value of the adjustment parameter. For very small absolute values of ϕ_2, the rejection frequencies are around 5 percent, regardless of the number of observations. In the case when $\sigma^2 = 1$, $\theta = 0$ and $\phi_2 = -0.2$, the power is never larger than 20 percent, even if the number of observations are 300, i.e. we reject only 20 percent of the false null hypotheses with 300 observations and a fairly fast adjustment process. The power is also largely affected by the sample size. With few observations, the

[8]When calculating the rejection frequencies we use critical values from our simulations, forcing the empirical size of the test to be 5 percent.

Table 4: Power of the LR test for structural hypotheses

β_1	σ^2	θ	ϕ_2	$T = 50$	$T = 100$	$T = 300$
0.95	1	0	−0.01	0.050	0.051	0.051
			−0.04	0.050	0.050	0.055
			−0.2	0.053	0.063	0.203
			−0.5	0.074	0.151	0.671
0.95	0.5	0	−0.01	0.050	0.050	0.051
			−0.04	0.050	0.051	0.054
			−0.2	0.051	0.058	0.124
			−0.5	0.062	0.102	0.469
0.95	2	0	−0.01	0.053	0.051	0.052
			−0.04	0.052	0.053	0.060
			−0.2	0.056	0.079	0.326
			−0.5	0.093	0.246	0.850
0.8	1	0	−0.01	0.050	0.052	0.056
			−0.04	0.052	0.058	0.105
			−0.2	0.090	0.226	0.851
			−0.5	0.319	0.758	0.999
0.8	0.5	0	−0.01	0.050	0.051	0.053
			−0.04	0.052	0.054	0.081
			−0.2	0.071	0.146	0.680
			−0.5	0.200	0.575	0.993
0.8	2	0	−0.01	0.050	0.053	0.058
			−0.04	0.056	0.065	0.158
			−0.2	0.120	0.362	0.948
			−0.5	0.492	0.895	1.0

Note: The table shows the power of the test, i.e. the rejection frequencies for the test when the null hypothesis is false. The critical values used are simulated, giving an empirical size of the test of 5 percent for every parameter combination.

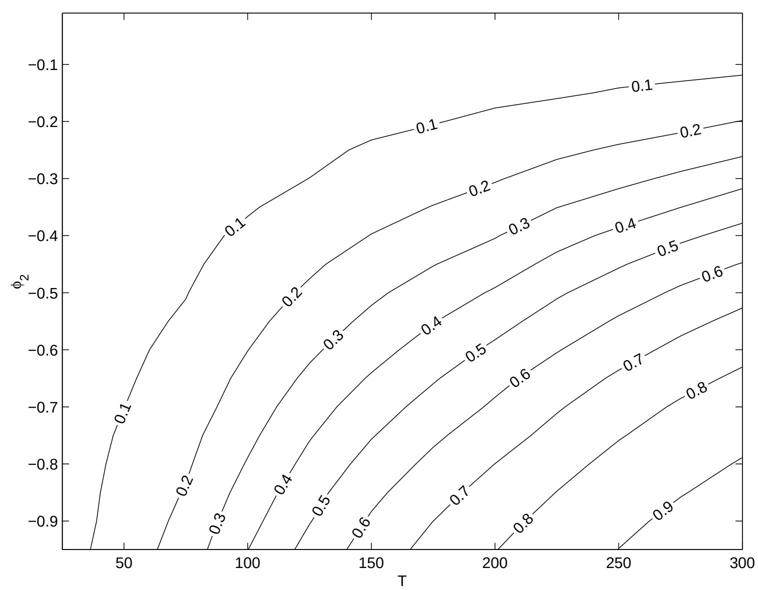

Figure 4: Power of the LR test for structural hypotheses when $\beta_1 = 0.95$

test has very low power, but as the the number of observations grow, the power of the test increases.

For the power of the LR test, the relative variance of the two disturbances, σ^2, matters to a large extent. The power is lower if $\sigma^2 < 1$, compared to if $\sigma^2 = 1$, and higher if $\sigma^2 > 1$. This indicates that if the relative variance of the variable adjusting to the cointegration relationship is high, we more often get an incorrect outcome from the test. The incorrectness in the test arises from the high variation in ϵ_{yt}, which leads to a relatively high variation in y_t, so that $y_t - \beta_1 x_t$ is more often far from the cointegration relationship.[9]

In figure 4, we present level curves for the power of the LR test when $\beta_1 = 0.95$. The level curves are calculated based on the same simulations as the level curves for the empirical size in section 5.1. For many combinations of adjustment parameters and sample sizes, the power is lower than 10 percent, represented by the area above the curve marked '0.1'. A power over 40 percent is never reached if the absolute value of the adjustment parameter is larger than –0.3, regardless of the number of observations.

[9]The value of the covariance parameter, θ, also matters for the power of the test. The result for different values of θ is, for space-saving reasons, not presented in the table. The differences in power are very small, but the pattern is that the power of the test increases if the absolute value of the θ-parameter increases.

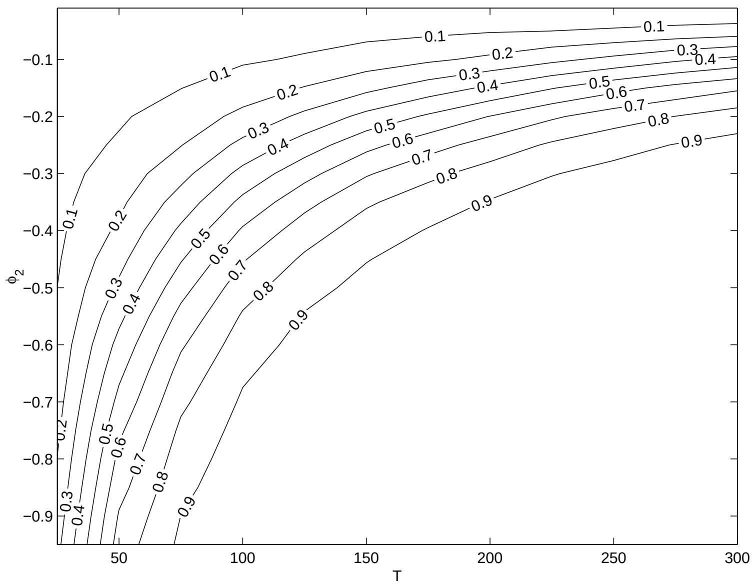

Figure 5: Power of the LR test for structural hypotheses when $\beta_1 = 0.8$

We also examine the power when the true value of the parameter β_1 in equation (19) is further away from 1. The power of the test for $\beta_1 = 0.8$ is shown in the lower panel of table 4. Not surprisingly, the power is generally higher compared to $\beta_1 = 0.95$. For large numbers of observations and high absolute values of the adjustment parameter, the power is close to 100 percent when $\sigma^2 = 1$. Troublesome is the low power still arising when the number of observations is small and the adjustment process is slow. For $\sigma^2 = 1$, the power is still about 5 to 10 percent if $\phi_2 = -0.01$ or $\phi_2 = -0.04$, even for 100 and 300 observations. Changing the value of σ^2 yields the same pattern as for $\beta_1 = 0.95$; when $\sigma^2 < 1$, the power is lower compared to when $\sigma^2 > 1$.

The power level curves for $\beta_1 = 0.8$ are shown in figure 5. Compared to figure 4, the curves have moved in an up-left direction, reflecting a higher power to detect a value of the β_1-parameter of 0.8 than 0.95. Still, in the upper left corner, we see that many combinations of ϕ_2 and T yield a power less than 10 percent.

If the number of observations is large and the absolute value of the adjustment parameter is high, the power of the LR test to detect a false null hypothesis is high, but in empirical applications, these circumstances are not fairly common. In the PPP literature, values of the adjustment parameter of –0.04 on a quarterly basis are common and under this circumstance, the power of the test is only around 5 percent. In table 4, the result for $\beta_1 = 0.95$ and $\beta = 0.8$ is presented. If the true value of β_1 is 0.5 or 0.3, the test has higher power. However, when the absolute value of the adjustment parameter is very low, the power is around 5 to 10 percent for 50 or 100 observations even for those values of the β_1-parameter.

5.2.2 The Bartlett corrected LR test for structural hypotheses

The power of the LR test corrected with the Bartlett factor is shown in table 5. The power differences compared to the test without the correction are marginal, which can be seen after a comparison between table 4 and 5.[10] In section 5.1.2, we saw that the Bartlett correction had a significant impact on the empirical size. Since the correction does not have a large negative impact on power, it is advisable to apply the Bartlett correction to the LR test, especially if the sample size is small.

5.2.3 The KPSS test

Finally, in table 6, the power of the KPSS test to detect a false cointegration vector is presented. Starting with $\beta_1 = 0.95$, shown in the top panel of the table, we see that the power of the test is often very low. For 50 observations, the power is never higher than 6 percent and for 100 observations never higher than 8 percent. For $\phi_2 = -0.01$ and $\phi_2 = -0.04$, the power is around 5 percent across all sample sizes. Only if the number of observations is 300 and the adjustment process is fast, the power reaches 30 percent. The picture emerging from changing the value of the relative variance σ^2 is the same as for the LR test. Compared to when $\sigma^2 = 1$, the power is lower if $\sigma^2 < 1$ and the power i higher if $\sigma^2 > 1$. For the KPSS test, changing the value of σ^2 results in rather small power differences. Also changing the values of the covariance parameter θ results in very small power differences and this is not shown in the table.

Level curves for the power of the test when $\beta_1 = 0.95$ are shown in figure 6. The power

[10] Since the power is almost the same as for the test without the correction, the power level curves for the Bartlett corrected test are not presented from the paper.

Table 5: Power of the Bartlett corrected LR test for structural hypotheses

β_1	σ^2	θ	ϕ_2	$T = 50$	$T = 100$	$T = 300$
0.95	1	0	-0.01	0.050	0.051	0.051
			-0.04	0.049	0.051	0.056
			-0.2	0.053	0.063	0.202
			-0.5	0.073	0.151	0.670
0.95	0.5	0	-0.01	0.050	0.050	0.050
			-0.04	0.051	0.052	0.054
			-0.2	0.052	0.058	0.122
			-0.5	0.063	0.101	0.469
0.95	2	0	-0.01	0.050	0.050	0.051
			-0.04	0.051	0.052	0.058
			-0.2	0.056	0.078	0.326
			-0.5	0.093	0.245	0.850
0.8	1	0	-0.01	0.050	0.053	0.057
			-0.04	0.051	0.059	0.102
			-0.2	0.086	0.215	0.848
			-0.5	0.307	0.753	0.999
0.8	0.5	0	-0.01	0.050	0.052	0.052
			-0.04	0.051	0.056	0.078
			-0.2	0.070	0.140	0.673
			-0.5	0.195	0.568	0.993
0.8	2	0	-0.01	0.050	0.051	0.058
			-0.04	0.056	0.063	0.146
			-0.2	0.113	0.345	0.947
			-0.5	0.484	0.893	1

Note: The table shows the power of the test, i.e. the rejection frequencies for the test when the null hypothesis is false. The critical values used are simulated, giving an empirical size of the test of 5 percent for every parameter combination.

Table 6: Power of the KPSS test

β_1	σ^2	θ	ϕ_2	$T = 50$	$T = 100$	$T = 300$
0.95	1	0	-0.01	0.051	0.050	0.052
			-0.04	0.051	0.051	0.053
			-0.2	0.053	0.055	0.098
			-0.5	0.060	0.079	0.294
0.95	0.5	0	-0.01	0.050	0.051	0.052
			-0.04	0.050	0.050	0.051
			-0.2	0.052	0.053	0.078
			-0.5	0.055	0.068	0.227
0.95	2	0	-0.01	0.051	0.051	0.052
			-0.04	0.052	0.051	0.054
			-0.2	0.053	0.056	0.115
			-0.5	0.060	0.091	0.342
0.8	1	0	-0.01	0.052	0.054	0.059
			-0.04	0.053	0.055	0.085
			-0.2	0.074	0.123	0.446
			-0.5	0.160	0.340	0.689
0.8	0.5	0	-0.01	0.051	0.054	0.057
			-0.04	0.052	0.053	0.074
			-0.2	0.066	0.100	0.378
			-0.5	0.125	0.279	0.648
0.8	2	0	-0.01	0.054	0.055	0.063
			-0.04	0.057	0.057	0.100
			-0.2	0.082	0.145	0.496
			-0.5	0.187	0.387	0.715

Note: The table shows the power of the test, i.e. the rejection frequencies for the test when the null hypothesis is false. The critical values used are simulated, giving an empirical size of the test of 5 percent for every parameter combination.

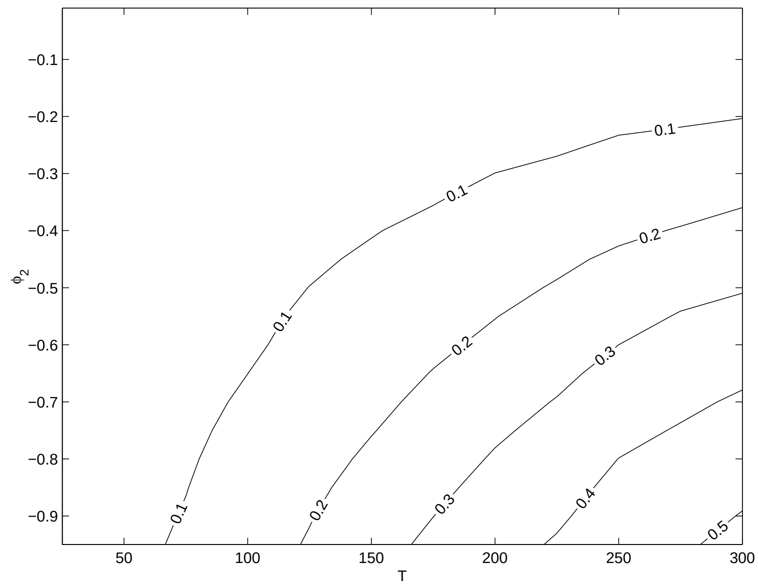

Figure 6: Power of the KPSS test when $\beta_1 = 0.95$

increases with an increasing number of observations and increasing absolute values of the adjustment parameter. For many parameter combinations, represented by the area above the highest power curve, the power is less than 10 percent and the power almost never reaches 50 percent.

Turning to $\beta_1 = 0.8$, in the lower panel of table 6, the power is higher compared to when $\beta_1 = 0.95$. If the adjustment process is very slow, the power is still very low and the power is lower than 10 percent for $\phi_2 = -0.01$ and $\phi_2 = -0.04$.[11] Comparing the figures 6 and 7, we see the higher power of the test to detect the false cointegration vector when $\beta_1 = 0.8$. For $\beta_1 = 0.8$, the power reaches 70 percent for some parameter combinations, but is still less than 10 percent for many values of the parameters ϕ_2 and T.

Both the LR test for structural hypotheses and the KPSS test have very low power to detect a false null hypotheses if the adjustment parameter is very low in absolute terms. The low power is present if the true value of the β_1-parameter is 0.95 or 0.8, but also if the true value of β_1 is far from the null hypotheses ($\beta_1 = 0.5$ or $\beta_1 = 0.3$). For faster

[11] If the β_1-parameter is equal to 0.5 or 0.3, the power increases. For the lowest values of the adjustment parameter, the power is still around 5 to 10 percent for 50 and 100 observations, and somewhat higher for 300 observations.

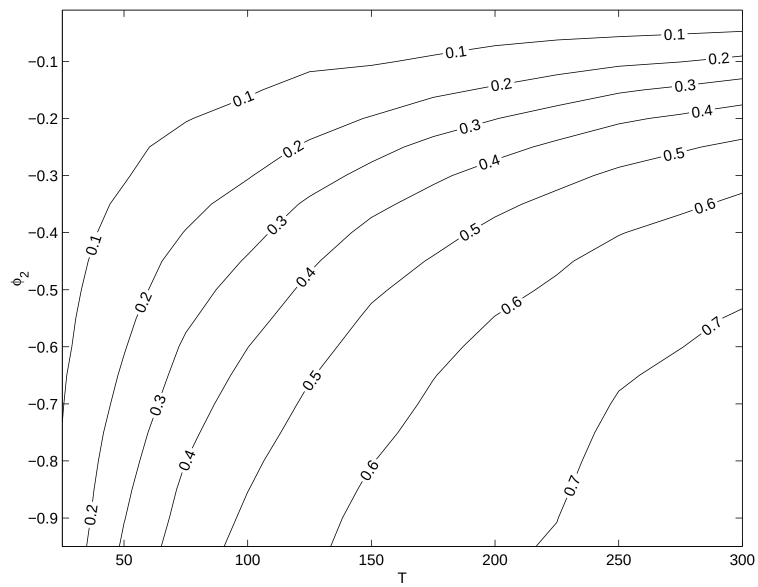

Figure 7: Power of the KPSS test when $\beta_1 = 0.8$

adjustment behaviour, the power of the tests is higher and sometimes reaches high levels. The power of the KPSS test is generally lower than the power of the LR test.

6 Conclusions

In this study, we examine two different tests that can be used when testing structural hypotheses on cointegration vectors. The two tests are the Johansen LR test and the KPSS test for stationarity applied to residuals from a long-run relationship in an Engle–Granger framework. We also analyse the consequences of applying a Bartlett correction factor to the LR test. The tests are evaluated according to their size and power properties and the analysis is performed in a Monte Carlo study. When setting up the simulations, the aim is to choose parameter values for our generated data that corresponds to common results in empirical applications. Therefore, values of the adjustment parameter in the cointegration equation that correspond to a slow adjustment process are included in the analysis, and we also work with small sample sizes.

The results show that both tests are influenced by the value of the adjustment parameter to a large extent. Both the empirical size and the power of the tests are heavily affected by this parameter. The sample size also plays an important role for the properties of the tests.

The empirical size of the LR test is often higher than the nominal one. With a slow adjustment process and few observations, this size distortion is very large. For some parameter combinations, using critical values intended to make us reject 5 percent of the true null hypotheses actually make us reject 30 to 40 percent of the true null hypotheses. The Bartlett correction factor helps to correct the empirical size of the test, but the empirical size is still larger than the nominal size for parameter combinations common in empirical applications, particularly when the speed of adjustment is slow. The KPSS test suffers from large size distortions when the speed of adjustment is slow as well. With faster adjustment, however, the empirical size of the KPSS test is close to the nominal 5 percent. For many parameter combinations, the KPSS test will never reach the nominal size, regardless of the number of observations.

When the speed of adjustment is slow, the size distortions are largest for the KPSS test and smallest for the Bartlett corrected LR test. Otherwise, the empirical size for all tests decreases to around 5 percent. The KPSS test is then closer to the nominal size than the LR test but although the KPSS test has empirical sizes close to 5 percent for few observations, the Bartlett corrected LR test seems to have the best size properties.

The power of the LR test can be rather high even for detecting values of the parameters in the cointegration vector close to the one specified in the null hypothesis, if we have a large number of observations and fast adjustment behaviour. When the adjustment parameter is low, the power of the test is very low, even for detecting values of the cointegration parameters far from the values in the null hypothesis. The Bartlett corrected LR test has almost the same power as the test without the correction. The KPSS test also exhibits very low power to detect a false cointegration vector if the speed of adjustment is low. The low power is still present if the true value of the cointegration parameter is far from the one specified in the null hypothesis. With faster adjustment, the power of the KPSS test increases and reaches fairly high levels, but it never reaches the power of the LR test.

Under favourable circumstances, the power of the two tests to detect a false null hypothesis can be rather high. The LR test seems to have best power properties, although both tests have low power under circumstances frequently arising in empirical applications, i.e. low adjustment parameters in absolute terms and small sample sizes. The low power is important to keep in mind when drawing conclusions from empirical studies.

The size and power properties are better for the LR test compared to the KPSS test, even if the differences sometimes are rather small. The LR test would therefore be preferred in empirical studies. The Bartlett correction factor helps reducing the size distortions, but does not influence the power. Therefore, this correction of the LR test is desirable when applying the test in empirical studies.

References

Amano, R. A. and van Norden, S. (1992). Unit-Root Tests and the Burden of Proof. Working Paper no. 92-7, Bank of Canada.

Banerjee, A., Cockerell, L., and Russell, B. (2001). An I(2) Analysis of Inflation and the Markup. *Journal of Applied Econometrics*, 16(3):221–240.

Bartlett, M. S. (1937). Properties of Sufficiency and Statistical Tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 160(901):268–282.

Engle, R. F. and Granger, C. W. J. (1987). Co-Integration and Error-Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276.

Fachin, S. (2000). Bootstrap and Asymptotic Tests of Long-run Relationships in Cointegrated Systems. *Oxford Bulletin of Economics and Statistics*, 62(4):543–551.

Gredenhoff, M. and Jacobson, T. (2001). Bootstrap Testing Linear Restrictions on Cointegrating Vectors. *Journal of Business and Economic Statistics*, 19(1):63–72.

Haug, A. A. (2002). Testing Linear Restrictions on Cointegrating Vectors: Sizes and Powers of Wald and Likelihood Ratio Tests in Finite Samples. *Econometric Theory*, 18(2):505–524.

Jacobson, T. (1995). Simulating Small-Sample Properites of the Maximum Likelihood Cointegration Method: Estimation and Testing. *Finnish Economic Papers*, 8(2):96–107.

Jacobson, T. and Larsson, R. (1999). Bartlett Corrections in Cointegration Testing. *Computational Statistics and Data Analysis*, 31(2):203–225.

Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12(2/3):231–54.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6):1551–1580.

Johansen, S. (2000). A Bartlett Correction Factor for Tests on the Cointegrating Relations. *Econometric Theory*, 16(?):740–778.

Johansen, S. and Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 52(2):169–210.

Johansen, S. and Juselius, K. (1992). Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK. *Journal of Econometrics*, 53(1–3):211–244.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root? *Journal of Econometrics*, 54(1–3):159–178.

Omtzigt, P. and Fachin, S. (2002). Bootstrapping and Bartlett Correction in the cointegrated VAR model. Working Paper 2002/24, Faculty of Economics, University of Insubria.

Rogoff, K. (1996). The Purchasing Power Parity Puzzle. *Journal of Economic Literature*, 34(2):647–668.

Östermark, R. and Höglund, R. (1999). Simulating Competing Cointegration Tests in an Bivariate System. *Journal of Applied Statistics*, 26(7):831–846.

Zhou, S. (2000). Testing Structural Hypotheses on Cointegration Relations with Small Samples. *Economic Inquiry*, 38(4):629–640.