UNIVERSITY OF SOUTHAMPTON

COINTEGRATION IN MISSPECIFIED MODELS

Nicoletta Pashourtidou

Doctor of Philosophy

Department of Economics

August 2002

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

ECONOMICS

Doctor of Philosophy

COINTEGRATION IN MISSPECIFIED MODELS

by

Nicoletta Pashourtidou

This thesis examines analytically (using asymptotic theory) and via Monte Carlo simulations the effects of two types of misspecifications on the LR tests for cointegration proposed by Johansen (1988, 1996).

The first type of misspecification is intercept shifts, represented by step dummy variables. It is assumed that the DGP consists of $I(1)$ processes which are cointegrated and some of them contain intercept shifts. The presence of intercept shifts is ignored in the construction of the statistical model (SM) used for cointegration testing. It is shown that under the above misspecification the tests overestimate the cointegrating rank with probability one as the sample size tends to infinity. An upper bound is found for the number of spurious cointegrating vectors that arise in the limit, and it is given by the number of distinct intercept shifts in the DGP. The attainment of the bound depends on the weak exogeneity status of the variables. Monte Carlo experiments designed in a way that allows control over the local power show that as the sample size and the magnitude of the shift become larger the frequency of accepting a bigger cointegrating rank than that in the DGP, increases. The impacts of intercept shifts are quite noticeable for sample sizes and model specifications used in empirical works.

The second type of misspecification is the presence of irrelevant variables in the SM or omission of relevant variables from the SM used for cointegration testing. We show that the inclusion of irrelevant variables does not affect the inference about the cointegrating rank or the consistency of the estimators of the cointegrating vectors, adjustment coefficients and variance of the errors, but simulations show a reduction in the power of the tests. We also show that the omission of relevant variables from the SM leads to either failure in detecting cointegration or underestimation of the cointegrating rank. Although in the latter case the estimator of the detected cointegrating vectors is shown to be consistent, this is not the case for the estimators of the adjustment coefficients and the variance of the errors.

Contents

Acknowledgements

I am greatly indebted to Raymond O'Brien who inspired and patiently supervised this thesis. Thanks to John Aldrich, Grant Hillier, Grayham Mizon and Jan Podivinsky for useful comments. A Postgraduate Studentship from the Department of Economics, University of Southampton, is gratefully acknowledged. Finally, my greatest thanks go to my family for their support.

List of figures

List of tables

List of abbreviations

(A)DF	(Augmented) Dickey-Fuller
AR	Autoregressive
CMT	Continuous Mapping Theorem
DGP	Data Generating Process
ECM	Error Correction Model
LM	Lagrange Multiplier
LR	Likelihood Ratio
SM	Statistical Model
VAR	Vector Autoregressive
WLLN	Weak Law of Large Numbers

Chapter 1
Introduction

Many economic variables appear to be integrated of order one (with a drift), denoted by $I(1)$, (see e.g. Nelson and Plosser (1982)) that is they have non-constant unconditional means and divergent unconditional variances as the date of the observation becomes large. In other words they show a type of stochastic non-stationarity where random shocks have a permanent effect. Even though the variables individually might exhibit the non-stationary behaviour mentioned above, it is possible that certain linear combinations of them have lower order of integration and in this instance these linear combinations are integrated of order zero, denoted by $I(0)$, or loosely, stationary. If this is the case, the variables are said to be cointegrated. The existence of linear combinations of variables with lower order of integration than the component variables i.e. cointegration, is implied by the notion of long-run equilibrium in economic theory, see Engle and Granger (1987).

Ascertaining the existence of cointegration among non-stationary ($I(1)$) variables, has particular consequences for their econometric analysis as well. Various tests for detecting cointegration have been proposed in the literature. These tests can be divided into two categories: tests that are based on single equation regression, e.g. Engle and Granger (1987), Phillips (1987) and tests based on systems of equations (vector autoregression (VAR) models), e.g. Johansen (1988, 1991a, 1994, 1996) (henceforth Johansen (1988, 1996)), Perron and Campbell (1993), Reinsel and Ahn (1992), Stock and Watson (1988). The testing procedures in the first category are essentially unit root-type tests, for example (augmented) Dickey - Fuller ((A)DF) or Phillips' Z_α or Z_t tests, on the residuals of the static (cointegrating) regression. In the case which more than two variables are jointly under examination, more than one cointegrating relation might be present. So, the tests in the second category can be used to determine the number of cointegrating relations or alternatively the number of common stochastic trends.

The most commonly used tests in applied works, for the number of cointegrating relations/vectors or cointegrating rank are the likelihood ratio (LR) tests proposed by Johansen (1988, 1996), which can be implemented using either the trace or the maximal eigenvalue statistic. The asymptotic distributions of these statistics are non-standard, involving integrals of Brownian motion, and they depend on the number of unit roots in the model. These distributions also depend on the nature of the deterministic terms (such as the intercept or linear time trend) included in the data generating process (DGP) and in the statistical model (SM) used for cointegration testing. Therefore, these asymptotic tests are not similar but they can be asymptotically similar with respect to certain parameters, see Nielsen (1997), Nielsen and Rahbek (2000).

In the literature there are works that investigate approximations to the asymptotic distribution of the LR tests for the cointegrating rank. Johansen (1988) suggests approximating the asymptotic distribution by $c\chi^2(f)$ where f equals twice the square of common trends (unit roots) in the model and $c = 0.85 - 0.58/f$. Doornik (1998) gives an approximation using the Gamma distribution and Larsson (1999) derives tail approximation of the asymptotic distribution using multivariate saddlepoint approximation.

The asymptotic nature of the LR tests for cointegration has triggered various Monte Carlo studies in the literature. For example Toda (1995) investigates through Monte Carlo simulation the finite sample performance of LR tests for cointegration proposed by Johansen (1991a). His findings suggest that a sample size of at least 300 observations is required in order for the tests to achieve good performance, over the values of the nuisance parameters and the cointegrating rank. In addition, Toda (1994) examines the finite sample properties of LR tests for 'stochastic cointegration' (i.e. linear combinations of variables are trend stationary) proposed by Johansen (1994) and Perron and Campbell (1993). He finds (as in Toda (1995)) that these asymptotic test procedures exhibit sensitivity to the value of the stationary root(s) of the process and to the correlation between the errors of the stationary and non-stationary components. In both papers (Toda (1994, 1995)) the analysis takes place in the framework of a bivariate VAR transformed into 'canonical form'.

Moreover, there are works that examine small sample corrections of the LR tests for the cointegrating rank. For the bivariate first order model Nielsen (1997) provides a new asymptotic distribution whose moments approximate well the finite sample moments. Johansen (1999) derives the Bartlett correction for the LR test for the cointegrating rank in a VAR model.

Since the LR tests for the cointegrating rank have been so widely applied in empirical research it is of interest to study their behaviour under various types of misspecification of the SM, used for cointegration testing. The robustness of the LR tests for cointegration was investigated using Monte Carlo simulations under omitted or irrelevant (redundant) step or impulse dummy variables (Andrade et al. (1994)), dynamic misspecification using a DGP with autoregressive and moving average dynamics (Boswijk and Franses (1992), Cheung and Lai (1993)), and non-normality assuming non-symmetric and leptokurtic innovations (Cheung and Lai (1993)).

The purpose of this thesis is to investigate the behaviour of LR tests for cointegration proposed by Johansen (1988, 1996), under misspecifications, analytically (asymptotic analysis) and via Monte Carlo simulations. By misspecification we mean that the model fitted to the data (the SM) and used for cointegration testing differs from the DGP in terms of specification of either the deterministic components or the endogenous variables included in the VAR model. Thus, we examine the effects of two types of misspecification: intercept shifts and irrelevant or omitted variables.

1.1 Intercept shifts

The LR tests for cointegration proposed by Johansen (1988, 1996) were derived under the assumption of constant parameters in the error correction model (ECM) and correctly specified deterministic components. However structural changes in the economy such as oil shocks or policy regime shifts induce the estimated parameters in a SM to change over time. If those changes are left undetected and therefore unmodelled, their presence will invalidate the use of inferential procedures that assume parameter constancy. One of the issues we seek to explore is the robustness of Johansen's (1988, 1996) LR tests for cointegration when intercept shifts, represented by step dummies, are present in the DGP but not in the VAR model used for performing cointegration tests (the SM). Thus, the DGP consists of non-stationary ($I(1)$) variables, which are cointegrated, and some of them possess intercept shifts. Under these assumptions we find that the test statistics tend to infinity as the sample size increases, therefore the true null hypothesis of cointegrating rank r is rejected with probability that tends to unity as the sample size tends to infinity. In addition, we investigate the impacts of intercept shifts via Monte Carlo simulations in a setup that allows us to control, to some extent, the local power of the tests. We conclude that the tests are not robust to this form of misspecification since they lead to the acceptance of spurious cointegrating relations as the sample size and the magnitude of the shift are increased. The results of the simulations are in accordance with the asymptotic analysis.

Andrade et al. (1994) study the effects of impulse and step dummy variables on the Engle and Granger (1987) procedure and Johansen (1988) tests for cointegration, using a bivariate model. In the case of the Engle and Granger procedure they examine dummies included in the SM and therefore in testing, but not in the DGP and vice versa, whereas for the Johansen procedure they investigate only the case of dummies in the SM. They find that the impact of a step dummy on the performance of the tests is greater than that of an impulse dummy. When a step dummy is included in the DGP they find that the Engle and Granger procedure suggests spurious stationarity. When the step dummy is present in testing all tests under consideration are found to over-reject the true null hypothesis of no cointegration and one cointegrating vector for the Engle and Granger, and Johansen procedures respectively.

O'Brien (1996, 1997, 1999) provides an asymptotic analysis and Monte Carlo simulations for the case that step dummy variables (operative at a common or different dates) are present in the DGP but not in the SM. He proves that among a set of uncorrelated random walks (thus he considers only the null case of no cointegration) the false hypothesis of cointegration is accepted with probability one as the sample size tends to infinity.

The presence of a structural break or a shift in the intercept of univariate autoregressive (AR) processes, has considerable consequences on unit root tests, which can be viewed as the predecessors of cointegration tests. Hendry and Neale (1991), conduct a Monte Carlo investigation and conclude that the power of standard ADF or DF tests, to distinguish between a non-stationary series and a stationary one with an intercept shift, is low. Perron (1989) shows analytically and by a small Monte Carlo experiment, that the unit root hypothesis cannot be rejected by the standard tests against trend stationary alternatives, when the DGP is a trend stationary AR process with a break, in either the intercept or the slope of the linear trend, or both. He extends the DF testing procedure to the case that there is one-time exogenous (known) shift (in the intercept or in the slope of the trend or both) under both the null of unit root and the alternative of trend stationarity. Perron and Vogelsang (1992) provide a class of test statistics along with simulated critical values for testing the null of a unit root when there is a change in the mean at an unknown date, under both the null and the alternative. Other relevant works in the literature include Zivot and Andrews (1992) and Banerjee et al. (1992). Zivot and Andrews (1992) consider a variation of Perron's (1989) test in which the structural break is only under the alternative and the choice of the date of the break is estimated rather than being exogenous. For this purpose they suggest the minimum one-sided t statistic (for testing the null of unit root) over all possible dates of the break. Banerjee et al. (1992) give the asymptotic distributions and simulated critical values for recursive rolling, and sequential tests for unit roots and/or shift in the coefficients of the AR process.

There are also studies that investigate the effects of structural breaks on the tests for cointegration and some of them propose tests for parameter stability in cointegrated models. For example, Campos et al. (1996) compare via Monte Carlo experiments, the power of the DF test for cointegration and of t-test on the coefficient of the error correction term. They consider a dynamic model reparametrised as an ECM, where the marginal process of one of the cointegrated variables is stationary with a structural break. Their analysis suggests that the t-test based on the ECM is more powerful than the DF test, when there are no common factor restrictions in the DGP. However, under the occurrence of both a break and a unit root only the marginal process is considered. Gregory et al. (1996) employ a 'linear quadratic model' to evaluate, using Monte Carlo simulations, Hansen's (1992) tests for structural breaks in the cointegrating relations. They also use Monte Carlo experiments to examine the behaviour of the ADF test for cointegration, when breaks are present in the cointegrating relation and they find considerable reduction in the rejection frequency of the test.

For the single equation framework Hansen (1992) and Quintos and Phillips (1993) provide Lagrange multiplier (LM)-type tests for parameter stability in cointegrating regressions. The tests proposed by Hansen (1992) refer to all the coefficients of the cointegrating regression (full cointegrating vector), whereas the tests of Quintos and Phillips (1993) can also be applied to a subset of the cointegrating coefficients. Some of the statistics they propose can be used to test the null hypothesis of cointegration against the alternative of no cointegration, which is equivalent to testing parameter constancy against a random walk alternative for the intercept coefficient. Hao (1996) compares the various statistics suggested by Hansen (1992) and Quintos and Phillips (1993) analytically and via a Monte Carlo study. He also suggests a test for cointegration which is robust to a discrete jump in the intercept. Gregory and Hansen (1996a, 1996b) provide ADF, Z_α and Z_t -type tests for testing the null of no cointegration against the alternative of cointegration. The intercept and/or cointegrating slope coefficients are allowed to change at an unknown date only under the alternative.

For the multivariate framework Quintos (1997) proposes tests for rank stability and tests for the stability of the long-run matrix in an ECM, under the assumption of correctly specified cointegrating rank. Kuo (1998) proposes LM-type tests for non-constancy in subsets of the cointegrating coefficients when the non-constancy of the parameters is either in the form of random walk coefficients or single jump at an unknown date. Seo (1998) suggests LM statistics for structural change in the cointegrating vector and/or adjustment coefficient vector at an unknown change point, under known cointegrating rank and a normalisation of the cointegrating vector. His tests are based on maximum likelihood estimation of the ECM and do not require sequential estimation, unlike some of the tests mentioned above, which in addition require fully modified estimation (e.g. Hansen (1992), Kuo (1998), Quintos and Phillips (1993)). Hansen and Johansen (1999) propose methods of testing parameter constancy in a cointegrated VAR model based on recursive estimation of the model and suggest tests for constancy of the long-run parameters in an ECM.

Saikkonen and Lütkepohl (1998) propose LM and LR-type tests for the cointegrating rank of a VAR process, when some of the variables have a shift in the mean (modelled by step or impulse dummies) at known date. The first stage of their procedure involves the estimation and subtraction of the deterministic parts (including the impulse and step dummy variables) of the model and at the second stage the cointegration rank of the adjusted series is tested. They find that the inclusion of step and impulse dummies in the model and the estimation of their coefficients do not affect the asymptotic distribution of the tests. Johansen et al. (2000) propose a LR test for the cointegrating rank in a model with piecewise linear trend and known breakpoints. The asymptotic distribution of the test statistic depends on the relative length of the regimes induced by the breaks. Inoue (1999) proposes tests of the cointegrating rank in the presence of a trend break, at an unknown date under the alternative hypothesis.

1.2 Irrelevant or omitted variables

One of the motivations for studying the effects of irrelevant or omitted variables is the well-known results from the standard regression analysis, namely that (i) the ordinary least squares estimators (of the regression coefficients and the variance of the errors) are unbiased but inefficient when irrelevant variables have been included in the SM and (ii) the ordinary least squares estimators are biased when relevant variables have been omitted from the SM.

For this type of misspecification we investigate analytically and via Monte Carlo simulations the effects of irrelevant $I(1)$ variables in the SM and omitted $I(1)$ variables from the SM, on the inference about the cointegrating rank. The consistency of the estimators of the parameters of the ECM under this form of misspecification is also considered. We show that the inclusion of irrelevant variables does not affect the inference about the cointegrating rank or the consistency of the estimators of the cointegrating vectors and the adjustment coefficients. However, simulations show that the inclusion of irrelevant variables leads to reduction in the power of the tests. We also show that omission of relevant variables from the SM leads to either failure in detecting cointegration or underestimation of the cointegrating rank. Moreover, in the omitted variables case, we show that although the estimator of the detected cointegrating vectors is consistent, this is not the case for the estimator of the adjustment coefficient matrix.

This second type of misspecification under consideration can be seen as overspecification or underspecification of the statistical model used for cointegration testing. This means that with respect to the DGP, either some variables have been omitted from the SM (underspecification) or some of the variables included in the SM are irrelevant (overspecification).

Podivinsky (1998) investigates the performance of the LR tests for cointegration when there is a mismatch between the variables used in the SM (used for the cointegration tests) and the variables entering the true cointegrating vectors. Using Monte Carlo simulations he finds that the LR tests performed on an overspecified SM detect at least the true number of cointegrating vectors. He also finds that LR tests based on only two variables: (i) have low power when there are in fact two cointegrating vectors among three variables, and (ii) may not detect the cointegrating vector if there is only one cointegrating vector among three variables.

The issue of omitted variables in relation to the LR tests for cointegration was also considered in the applied econometrics literature (see e.g. DeLoach (2001)).

Finally, other works in the literature relate to the effect of including an irrelevant random walk in the SM, on the test for 'Granger non-causality' using the Wald statistic, see Ohanian (1988) and Toda and Phillips (1993).

1.3 Organisation

The organisation of the subsequent chapters is as follows. Chapter 2 gives an overview of Johansen's maximum likelihood estimation method of cointegrated models and LR tests for the cointegrating rank. Chapter 3 considers the algorithm used in the Monte Carlo simulations for computing the trace and the maximal eigenvalue statistics as well as the estimates of the cointegrating vectors and adjustment coefficients. Chapter 4 employs asymptotic theory to examine the effects of the presence of intercept shifts (occurring at either different dates or a common date) in the DGP (given by a cointegrated VAR process) on the inference about the cointegrating rank, when the SM does not account for those shifts. Chapter 5 provides an extensive Monte Carlo investigation of the performance of the LR tests for cointegration using the trace and the maximal eigenvalue statistics, in the presence of intercept shifts. Chapter 6 considers the effects of including irrelevant $I(1)$ variables in or omitting relevant $I(1)$ variables from the SM, on the inference about the cointegrating rank and the consistency of the estimators of the parameters in the ECM. Chapter 7 concludes.

1.4 Notation

The symbol '$\longmapsto$' denotes a mapping or function. $[n]$ gives the largest integer that is less than or equal to n, L is the lag operator and $\Delta \equiv 1 - L$. $tr(M)$ and $|M|$ denote the trace and determinant respectively of a square matrix M, $rank(M)$ denotes the rank of the matrix M, $sp(M)$ denotes the space spanned by the columns of the matrix M, I_n denotes the identity matrix of dimension n and $diag(m_1, \dots, m_n)$ is a diagonal matrix with $(m_1, \dots, m_n)$ the elements on the main diagonal. 0 is used to denote both the number (scalar) zero and the null matrix or vector, and its dimensions can be inferred from the context. The symbols '$\rightarrow$', '$\xrightarrow{p}$' and '$\xrightarrow{d}$'[1] denote deterministic convergence, convergence in probability and convergence in distribution respectively, as the sample size, T, tends to infinity. The symbols O, o and O_p, o_p denote the order of magnitude of approximations of deterministic and stochastic sequences respectively. Moreover, $E(\cdot)$, $Var(\cdot)$, $Cov(\cdot)$ and $p\text{lim}(\cdot)$ denote the expected value, variance, covariance and probability limit of the argument random quantity respectively. The notation '$Y_t \sim i.i.d.(M, V)$' states that the random variable/vector, Y_t is independent and identically distributed with mean M and variance V, $N_n(M, V)$ stands for the n-dimensional normal distribution with mean M and

---

[1] The convergence in distribution of X_n to X is equivalent to the weak convergence of F_n to F, usually denoted by $F_n \Rightarrow F$, where F_n and F are the distribution functions of X_n and X respectively. Moreover, since $\mu \equiv \mu((-\infty, x])$ and F (similarly $\mu_n \equiv \mu_n((-\infty, x])$ and F_n) for each $x \in \mathbb{R}$, represent the same probability measure, $\mu_n \Rightarrow \mu$ is equivalent to $F_n \Rightarrow F$. However, the weak convergence of distribution functions is linked to $\mathbb{R}$ whereas the weak convergence of probability measures ($\mu_n \Rightarrow \mu$) can be used for any metric space (see Billingsley (1968)). In this thesis we use the term 'convergence in distribution' denoted by '$\xrightarrow{d}$' to mean that the sequence of probability measures associated with a sequence of random variables/vectors, converges weakly to the Wiener measure (Brownian motion) or a funtional of it. A special case of this is the central limit theorem.

variance V and '$Y_t \sim I(d)$' states that the random variable/vector Y_t is integrated of order d where $d = 0, 1$.

Chapter 2
Johansen's procedure

This chapter describes the maximum likelihood method, proposed by Johansen (1988, 1996), for the estimation of the parameters of an ECM under the assumption of cointegration, and the LR statistics for the determination of the cointegrating rank. In addition it provides an outline of the procedure for deriving the asymptotic distribution of the LR test for the cointegrating rank. This procedure will be followed in some of the derivations in Chapters 4 and 6.

2.1 Definitions

We give some definitions that will be used throughout the thesis. The definitions were taken from Johansen (1996, Chapter 3).

Let $\{\varepsilon_t\}$ be a $p \times 1$ sequence and $\varepsilon_t \sim i.i.d.(0, \Omega)$ for all t.

Definition 2.1. A stochastic process X_t is said to be integrated of order zero, $I(0)$, if $X_t - E(X_t) = C(L)\varepsilon_t$, where $C(L) = \sum_{i=0}^{\infty} C_i L^i$, $C(1) \neq 0$ and $C(y)$ is convergent for $|y| \leq 1 + \omega$ and $\omega > 0$.

Definition 2.2. A stochastic process X_t is said to be integrated of order d, $I(d)$, $d = 0, 1, 2, \dots$, if $\Delta^d(X_t - E(X_t))$ is $I(0)$.

Definition 2.3. A p-dimensional stochastic process X_t is said to be cointegrated of order d, b, CI(d, b), with cointegrating vector $\beta \neq 0$, if (i) X_t is $I(d)$ and (ii) $\beta' X_t$ is $I(d-b)$, $d = 1, 2, \ldots; b = 1, 2, \ldots d$.

In this thesis we deal with the case $CI(1,1)$ that is $I(1)$ processes, linear combinations of which are $I(0)$.

2.2 The model

Consider a p-dimensional process X_t generated by a k-th order VAR,

$$X_t = \Pi_1 X_{t-1} + \cdots + \Pi_k X_{t-k} + \Phi D_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T \tag{2.1}$$

for fixed values of $X_{-k+1}, \ldots, X_0$ and $\varepsilon_t \sim i.i.d. N_p(0, \Omega)$. D_t is a $q \times 1$ vector of deterministic terms such as constant, linear trend, seasonal dummies, intervention dummies or other non-stochastic and fixed regressors.

The characteristic polynomial of (2.1), $A(y) = I_p - \sum_{i=1}^{k} \Pi_i y^i$, satisfies the condition that if $|A(y)| = 0$ then either $|y| > 1$ or $y = 1$, which ensures that X_t can be made stationary by differencing.

Equation (2.1) can equivalently be written in an error correction form,

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T \tag{2.2}$$

where $\Pi = \sum_{i=1}^{k} \Pi_i - I_p$ and $\Gamma_i = -\sum_{j=i+1}^{k} \Pi_j$.

As far as the rank of the matrix Π is concerned, three cases might arise: (i) $rank(\Pi) = p$, that is Π has full rank, which means that X_t is $I(0)$; (ii) $rank(\Pi) = 0$, therefore $\Pi = 0$

and the VAR model can be expressed in first differences; (iii) $0 < rank(\Pi) < p$, that is Π has reduced rank, r say, $0 < r < p$, so Π can be expressed as the product of two $p \times r$ matrices α and β of rank r, i.e. $\Pi = \alpha\beta'$. Without any a priori information α and β are not unique because $\Pi = \alpha\beta' = \alpha PP^{-1}\beta' = \alpha^*\beta^{*'}$ (with $\alpha^* = \alpha P$ and $\beta^{*'} = P^{-1}\beta'$) for all invertible $r \times r$ matrices P. Thus, one can only estimate the space spanned by the columns of β (the cointegrating space) and the space spanned by the columns of α. The matrices α and β correspond to the adjustment coefficients and cointegrating vectors respectively. $\beta' X_t$ are the cointegrating relations which are stationary although X_t is not.

The characteristic polynomial derived from (2.2) is given by

$$A(y) = (1-y)I_p - \Pi y - \sum_{i=1}^{k-1} \Gamma_i (1-y) y^i$$

and for case (iii) the characteristic equation $|A(y)| = 0$ has $(p-r)$ unit roots ($y = 1$) and r roots with modulus strictly greater than unity ($|y| > 1$).

2.3 Maximum likelihood estimation

In what follows we consider the model given by (2.2) under case (iii) (see section 2.2). Thus, $X_t \sim I(1)$, $\beta' X_t \sim I(0)$ and $\Delta X_t \sim I(0)$. The detailed properties of the process X_t are given in Theorem 2.1.

The purpose of Johansen's procedure is to derive an estimator for the unrestricted cointegrating vectors and test statistics for the hypothesis of the cointegrating rank under the assumption that there are at most r cointegrating vectors i.e.

$$H(r) : rank(\Pi) \leq r \text{ or } \Pi = \alpha\beta'. \tag{2.3}$$

Let $Y_{0t} = \Delta X_t, Y_{1t} = X_{t-1}, Y_{2t} = (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1}, D'_t)'$ and $\Psi = (\Gamma_1, \ldots, \Gamma_{k-1}, \Phi)$, where Y_{2t} is $[p(k-1)+q] \times 1$ and Ψ is $p \times [p(k-1)+q]$. Using the above notation together with the rank restriction (2.3), (2.2) can be written as

$$Y_{0t} = \alpha\beta' Y_{1t} + \Psi Y_{2t} + \varepsilon_t, \; t = 1, 2, \ldots, T \tag{2.4}$$

where Ψ is unrestricted. Reinsel and Ahn (1992) analyse the case where constraints can be imposed on the coefficient matrices of the first differences of the variables (i.e. restrictions on $\Gamma_1, \ldots, \Gamma_{k-1}$). Since the coefficient of Y_{1t} has reduced rank the technique of reduced rank regression has to be employed, see Anderson (1951). The log-likelihood function of the problem is

$$\begin{aligned} \log \mathcal{L}(\alpha, \beta, \Psi, \Omega) &= -\frac{Tp}{2}\log(2\pi) - \frac{T}{2}\log|\Omega| \qquad (2.5) \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(Y_{0t} - \alpha\beta' Y_{1t} - \Psi Y_{2t})'\Omega^{-1}(Y_{0t} - \alpha\beta' Y_{1t} - \Psi Y_{2t}). \end{aligned}$$

Concentrating (2.5) with respect to Ψ we obtain the following first order condition,

$$\sum_{t=1}^{T}(Y_{0t} - \alpha\beta' Y_{1t} - \hat{\Psi} Y_{2t})Y'_{2t} = 0 \tag{2.6}$$

and $\hat{\Psi}$ denotes the maximum likelihood estimator of Ψ, see equation (2.8).

Define the product moment matrices as

$$M_{ij} = T^{-1}\sum_{t=1}^{T} Y_{it}Y'_{jt}, \; i, j = 0, 1, 2. \tag{2.7}$$

Then from (2.6) we have,

$$\hat{\Psi}(\alpha, \beta) = M_{02}M_{22}^{-1} - \alpha\beta' M_{12}M_{22}^{-1} \tag{2.8}$$

which is the unrestricted estimator of Ψ for fixed α, β and Ω. By substituting (2.8) into (2.4) we get the residuals

$$\hat{\varepsilon}_t = R_{0t} - \alpha\beta' R_{1t} \tag{2.9}$$

where $R_{0t} = Y_{0t} - M_{02}M_{22}^{-1}Y_{2t}$ and $R_{1t} = Y_{1t} - M_{12}M_{22}^{-1}Y_{2t}$. Thus, R_{0t} are the residuals we obtain from the regression of Y_{0t} (or ΔX_t) on Y_{2t} (or $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}, D_t$) and R_{1t} are the residuals from the regression of Y_{1t} (or X_{t-1}) on Y_{2t} (or $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}, D_t$), by application of Frisch-Waugh Theorem (see Davidson (2000, p. 8)). Therefore, (2.9) can take the form of a reduced rank regression in the residuals,

$$R_{0t} = \alpha\beta' R_{1t} + \hat{\varepsilon}_t \tag{2.10}$$

with the following log-likelihood function,

$$\begin{aligned} \log \mathcal{L}(\alpha, \beta, \Psi, \Omega) &= -\frac{Tp}{2}\log(2\pi) - \frac{T}{2}\log|\Omega| \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}(R_{0t} - \alpha\beta' R_{1t})'\Omega^{-1}(R_{0t} - \alpha\beta' R_{1t}) \end{aligned} \tag{2.11}$$

which is the log-likelihood function concentrated with respect to Ψ.

Define the residual sums of squares as

$$S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}' = M_{ij} - M_{i2}M_{22}^{-1}M_{2j},\ i, j = 0, 1. \tag{2.12}$$

Then the estimators of α and Ω obtained by regression of R_{0t} on $\beta' R_{1t}$ for fixed β are given by

$$\hat{\alpha}(\beta) = S_{01}\beta(\beta' S_{11}\beta)^{-1} \tag{2.13}$$

$$\hat{\Omega}(\beta) = S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}. \tag{2.14}$$

By inserting (2.13) and (2.14) into (2.11) we get the maximised likelihood for fixed β,

$$\begin{aligned} \mathcal{L}^{-2/T} &= (2\pi e)^p |\hat{\Omega}(\beta)| \qquad (2.15) \\ &= (2\pi e)^p |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}| \\ &= (2\pi e)^p |S_{00}||\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta| / |\beta' S_{11}\beta| \end{aligned}$$

where the third equality follows from the expansion of the determinant

$$\begin{aligned} \begin{vmatrix} S_{00} & S_{01}\beta \\ \beta' S_{10} & \beta' S_{11}\beta \end{vmatrix} &= |S_{00}||\beta'(S_{11} - S_{10}S_{00}^{-1}S_{01})\beta| \\ &= |\beta' S_{11}\beta||S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}|. \end{aligned}$$

$\mathcal{L}^{-2/T}$ ($\mathcal{L}$) is minimised (maximised) among all $p \times r$ matrices β by solving the eigenvalue problem

$$|\rho S_{11} - (S_{11} - S_{10}S_{00}^{-1}S_{01})| = 0$$

or for $\zeta = (1 - \rho)$ by solving

$$|\zeta S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0 \qquad (2.16)$$

with eigenvalues $1 > \hat{\zeta}_1 > \ldots > \hat{\zeta}_p > 0$ and eigenvectors $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_p)$ normalised by $\hat{V}' S_{11}\hat{V} = I_p$.

The estimates of the cointegrating vectors are given by $\hat{\beta} = (\hat{v}_1, \ldots, \hat{v}_r)$, that is the eigenvectors that correspond to the r largest eigenvalues. Thus, the eigenvalues found by the reduced rank regression technique are the squared sample canonical correlations between R_{0t} and R_{1t}, in other words they are the squared sample canonical correlations between ΔX_t and X_{t-1} after removing the effects of lagged differences and deterministic terms.

Having found $\hat{\beta}$ we substitute back into (2.13) and (2.14) to find $\hat{\alpha}$ and $\hat{\Omega}$.

The maximised likelihood function is

$$\mathcal{L}^{-2/T}(H(r)) \equiv \mathcal{L}^{-2/T} = |S_{00}| \prod_{i=1}^{r} (1 - \hat{\zeta}_i) \tag{2.17}$$

because the normalisation of the eigenvectors leads to $\hat{\beta}' S_{11} \hat{\beta} = I_r$ and $\hat{\beta}' S_{10} S_{00}^{-1} S_{01} \hat{\beta} = diag(\hat{\zeta}_1, \ldots, \hat{\zeta}_r)$. Using (2.17) we can derive the LR test statistics for the hypotheses (i) $H(r)$ against $H(p)$ and (ii) $H(r)$ against $H(r+1)$. The trace statistic

$$-2 \log Q(H(r)|H(p)) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\zeta}_i) \tag{2.18}$$

corresponds to (i) and the maximal eigenvalue statistic

$$-2 \log Q(H(r)|H(r+1)) = -T \log(1 - \hat{\zeta}_{r+1}) \tag{2.19}$$

corresponds to (ii).

The asymptotic distribution of (2.18) is given by

$$-2 \log Q(H(r)|H(p)) \xrightarrow{d} tr\{\int_0^1 (dB)F' [\int_0^1 FF' du]^{-1} \int_0^1 F(dB)'\}. \tag{2.20}$$

B is a $(p-r)$-dimensional Brownian motion, and the elements of F depend on the elements of B and on the deterministic terms in the model. (2.20) was tabulated, for alternative specifications of the deterministic components, by Johansen (1988, 1996), Johansen and Juselius (1990), Osterwald-Lenum (1992) and MacKinnon et al. (1999) using response surfaces.

2.4 Asymptotic distribution

The asymptotic distribution (2.20) is derived in Johansen (1988, 1996) and the limiting result holds for independent and identically distributed errors, without assuming normality. The basic tool needed for the derivation of (2.20) and for various results in this thesis is the Granger Representation Theorem, which is given below.

Theorem 2.1[2] *(Granger Representation Theorem). Let* X_t *be defined by* (2.2) *for* $t = 1, 2, \dots,$ *and let* $\Pi = \alpha\beta'$ *for* α *and* β *defined as above. Assume* $|\alpha'_{\perp}\Gamma\beta_{\perp}| \neq 0$, *where* $\alpha_{\perp}$ *and* $\beta_{\perp}$ *are* $p \times (p-r)$ *matrices of rank* $(p-r)$, *orthogonal to* α *and* β *respectively (i.e.* $\alpha'_{\perp}\alpha = \beta'_{\perp}\beta = 0$*), and* $\Gamma = I_p - \sum_{i=1}^{k-1}\Gamma_i$. *Define* $C = \beta_{\perp}(\alpha'_{\perp}\Gamma\beta_{\perp})^{-1}\alpha'_{\perp}$. *Then,* $\Delta X_t - E(\Delta X_t)$ *and* $\beta' X_t - E(\beta' X_t)$ *can be given initial distributions such that* $\Delta X_t \sim I(0)$, $\beta' X_t \sim I(0)$ *and* $X_t \sim I(1)$. *In addition* ΔX_t *and* X_t *have the following representations*

$$\Delta X_t = C(L)(\varepsilon_t + \Phi D_t) \tag{2.21}$$

$$X_t = C\sum_{i=1}^{t}(\varepsilon_i + \Phi D_i) + C_1(L)(\varepsilon_t + \Phi D_t) + A \tag{2.22}$$

where $A = \beta_{\perp}(\beta'_{\perp}\beta_{\perp})^{-1}(\beta'_{\perp}X_0 - Y_0)$ *i.e. it is such that* $\beta' A = 0$ *and* Y_0 *depends on* α, $\alpha_{\perp}$ *and* ε_0. $C(L) = C(1) + (1-L)C_1(L)$ *and* $C(1) = C$. $C(y) = \sum_{i=0}^{\infty} y^i C_i$ *is convergent for* $|y| < 1 + \omega, \omega > 0$. $C_1(y) = \sum_{i=0}^{\infty} y^i C_{1i}$, *where* $C_{1i} = -\sum_{j=i+1}^{\infty} C_j$, $i = 0, 1, \dots,$ *is also convergent*[3] *for* $|y| < 1 + \omega, \omega > 0$. *Furthermore,* $C_1(1) = -dC(y)/dy|_{y=1} = -\sum_{i=1}^{\infty} iC_i$.

Proof. See Johansen (1991a, 1996).

---

[2] The theorem is adapted from Johansen (1991a, Theorem 4.1) and Johansen (1996, Theorem 4.2).

[3] This follows by Lemma 4.1 in Johansen (1996).

The assumption that $|\alpha_{\perp}^{'}\Gamma\beta_{\perp}| \neq 0$ is a necessary and sufficient condition for $\Delta X_t - E(\Delta X_t)$ and $\beta^{'}X_t - E(\beta^{'}X_t)$ to be given initial distributions such that they become $I(0)$.

Next we give an outline of the derivation of (2.20). Detailed derivations can be found in Johansen (1988, 1996). We present the results for the case where $\Phi D_t = \mu$, i.e. there is an unrestricted constant in the model (2.2).

From the representation (2.22) we can see that when $\Phi D_t = \mu$ the process X_t is the sum of a random walk, a linear trend, an infinite moving average process (stationary) and a constant. Therefore, the process X_t behaves differently in different directions, depending on which linear combination of the process we consider. To see this, let γ and $\tau = C\mu$ be $p \times (p - r - 1)$ and $p \times 1$ respectively such that β, γ, and τ are mutually orthogonal and (β, γ, τ) is $p \times p$ and has rank p. In the β direction, the matrix of cointegrating vectors β, eliminates the non-stationary component since $\beta^{'}C = 0$ and $\beta^{'}X_t$ is a stationary process. If we consider the linear combination $\gamma^{'}X_t$, then the random walk component dominates, since $\gamma^{'}\tau = 0$. In the τ direction, the process $\tau^{'}X_t$ is dominated by a linear trend.

Below we give some results concerning the asymptotic behaviour of S_{ij} $i, j = 0, 1$ in various directions, which are used in the derivation of (2.20),

$$S_{00} \xrightarrow{p} \Sigma_{00} \tag{2.23}$$

$$\beta^{'}S_{11}\beta \xrightarrow{p} \Sigma_{\beta\beta} \tag{2.24}$$

$$\beta^{'}S_{10} \xrightarrow{p} \Sigma_{\beta 0} \tag{2.25}$$

$$T^{-1}B_T^{'}S_{11}B_T \xrightarrow{d} \int_0^1 GG^{'}du \tag{2.26}$$

$$B_T'(S_{10} - S_{11}\beta\alpha') \xrightarrow{d} \int_0^1 G(dW)' \tag{2.27}$$

$$B_T' S_{11}\beta \in O_p(1) \tag{2.28}$$

where $Var(\begin{bmatrix} \Delta X_t \\ \beta' X_{t-1} \end{bmatrix} \mid \Delta X_{t-1} \dots \Delta X_{t-k+1}) = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}$; $B_T = (\bar{\gamma}, T^{-1/2}\bar{\tau})$ with $\bar{\gamma} = \gamma(\gamma'\gamma)^{-1}$, $\bar{\tau} = \tau(\tau'\tau)^{-1}$; $G = \begin{bmatrix} \bar{\gamma}'C(W(u) - \int_0^1 W(u)du) \\ u - 1/2 \end{bmatrix}$, with $T^{-1/2}\bar{\gamma}' X_{[Tu]} \xrightarrow{d} \bar{\gamma}' CW(u)$, $W(u) \equiv W$ is a Brownian motion with variance matrix Ω and $T^{-1}\bar{\tau}' X_{[Tu]} \xrightarrow{d} u$, $u \in [0, 1]$. For the proofs of the above results see Johansen (1996, pp. 146-148).

Using the above results and the fact that the ordered solutions of (2.16) are continuous functions of the elements of S_{ij} $i, j = 0, 1$, we can show that the r largest eigenvalues of (2.16) converge to the eigenvalues of $|\zeta\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}| = 0$ and the $(p - r)$ smallest eigenvalues of (2.16) converge to zero at rate T.

Let $A_T = (\beta, T^{-1/2}B_T)$ and $S(\zeta) = \zeta S_{11} - S_{10}S_{00}^{-1}S_{01}$, then by (2.23)-(2.28),

$$\begin{aligned} & |A_T' S(\zeta) A_T| \xrightarrow{d} \\ & \left| \begin{bmatrix} \zeta\Sigma_{\beta\beta} & 0 \\ 0 & \zeta\int_0^1 GG'du \end{bmatrix} - \begin{bmatrix} \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix} \right| \\ = \; & |\zeta\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}||\zeta \int_0^1 GG'du| \end{aligned}$$

which shows that there are $(p - r)$ zero eigenvalues since the stochastic matrix $\int_0^1 GG'du$ is assumed to be positive definite almost surely.

The asymptotic distribution is derived under the null hypothesis that the cointegrating rank is r. Let $\hat{\zeta}_i$ be an eigenvalue of (2.16) such that $|S(\hat{\zeta}_i)| = 0$. We then consider the asymptotic behaviour of $\left|S(\hat{\zeta}_i)\right|$ in the stationary and non-stationary directions using the following scaling,

$$
\begin{aligned}
|(\beta, B_T)' S(\hat{\zeta}_i)(\beta, B_T)| &= \begin{vmatrix} \beta' S(\hat{\zeta}_i)\beta & \beta' S(\hat{\zeta}_i)B_T \\ B_T' S(\hat{\zeta}_i)\beta & B_T' S(\hat{\zeta}_i)B_T \end{vmatrix} \\
&= |\beta' S(\hat{\zeta}_i)\beta||B_T'[S(\hat{\zeta}_i) - S(\hat{\zeta}_i)\beta(\beta' S(\hat{\zeta}_i)\beta)^{-1}\beta' S(\hat{\zeta}_i)]B_T| = 0.
\end{aligned} \tag{2.29}
$$

For any of the $(p-r)$ smallest eigenvalues we have

$$T\hat{\zeta}_i \xrightarrow{d} \kappa_i, \text{ for } r+1 \leq i \leq p$$

assuming that the asymptotic distribution of κ_i exists, see Davidson (2000, Chapter 16). Using again the results (2.23)-(2.28) we have

$$\beta' S(\hat{\zeta}_i)\beta = (T\hat{\zeta}_i)(T^{-1}\beta' S_{11}\beta) - \beta' S_{10} S_{00}^{-1} S_{01}\beta \xrightarrow{p} -\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}$$

and

$$
\begin{aligned}
& B_T'[S(\hat{\zeta}_i) - S(\hat{\zeta}_i)\beta(\beta' S(\hat{\zeta}_i)\beta)^{-1}\beta' S(\hat{\zeta}_i)]B_T \\
=\; & (T\hat{\zeta}_i)(T^{-1}B_T' S_{11} B_T) - B_T' S_{10}\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' S_{10} B_T + o_p(1) \\
\xrightarrow{d}\; & \kappa_i \int_0^1 GG' du - \int_0^1 G(dW)'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' \int_0^1 (dW)G'.
\end{aligned}
$$

In the second equality the identities $\Sigma_{0\beta} = \alpha\Sigma_{\beta\beta}$ and $\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\alpha(\alpha'\Sigma_{00}^{-1}\alpha)^{-1}\alpha'\Sigma_{00}^{-1} = \alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'$ are used, see Johansen (1996, p. 142).

Since $\beta' S(\hat{\zeta}_i)\beta$ has full rank asymptotically ($\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}$ is $r \times r$ of rank r) the $(p-r)$ smallest eigenvalues of (2.16) scaled by T converge to the solutions of

$$|\kappa_i \int_0^1 GG' du - \int_0^1 G(dW)'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' \int_0^1 (dW)G'| = 0. \tag{2.30}$$

We define the standard Brownian motion $B = (B_1', B_2')'$ where $B_1 = (\bar{\gamma}' C\Omega C'\bar{\gamma})^{-1/2}\bar{\gamma}' CW$ and $B_2 = (\mu'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\mu)^{-1/2}\mu'\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' W$. $F = (F_1', F_2')'$, where $F_1 =$

$B_1 - \int_0^1 B_1 du$, $F_2 = u - 1/2$. Then (2.30) becomes

$$|\kappa_i \int_0^1 FF' du - \int_0^1 F(dB)' \int_0^1 (dB)F'| = 0.$$

From (2.18) we have

$$\begin{aligned} -2\log Q(H(r)|H(p)) &= -T\sum_{i=r+1}^{p} \log(1-\hat{\zeta}_i) = T\sum_{i=r+1}^{p} \hat{\zeta}_i + o_p(1) \\ \xrightarrow{d} \sum_{i=r+1}^{p} \kappa_i &= tr\{\int_0^1 (dB)F' [\int_0^1 FF' du]^{-1} \int_0^1 F(dB)'\}. \end{aligned}$$

Chapter 3
Solving the eigenvalue problem: an algorithm

This chapter describes the algorithm used in the simulation experiments, that appear in Chapters 4, 5 and 6, for the calculation of the trace and maximal eigenvalue statistics as well as the estimators of the cointegrating vectors and adjustment coefficients. The algorithm was programmed in Ox 3.00 (see Doornik (1999)) in the form of a 'function' to produce the simulation results discussed in the following chapters.

Consider the ECM

$$\Delta X_t = \Pi^* X_{t-1}^* + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi D_t + \varepsilon_t, \ t = 1, 2, \dots, T \tag{3.1}$$

where X_t is a p-dimensional, $I(1)$, random vector, the initial values $(X_{-k+1}, \dots, X_0)$ are fixed and $\varepsilon_t \sim i.i.d.N_p(0, \Omega)$. $X_{t-1}^* = (X_{t-1}', d_t')'$ is $p_1 \times 1$, where $p_1 = p + m$, d_t and D_t are vectors of deterministic variables of dimensions $m \times 1$ and $q \times 1$ respectively.

Under the assumption of at most r cointegrating vectors the matrix Π^* can be factorised into $\Pi^* = \alpha\beta'$ where α and β are $p \times r$ and $p_1 \times r$ respectively, of rank r. We assume that m deterministic variables (such as constant or linear trend) lie in the cointegrating space whereas the lagged differences and the deterministic variables held by D_t are unrestricted (i.e. they lie outside the cointegrating space).

The maximum likelihood method proposed by Johansen (1988, 1996), which is described in Chapter 2, amounts to solving the eigenvalue problem

$$|\zeta S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0 \tag{3.2}$$

where $S_{ij} = T^{-1}r_i'r_j$, $i, j = 0, 1$, $r_0' = (r_{01}, \ldots, r_{0T})$ and $r_1' = (r_{11}, \ldots, r_{1T})$. r_{0t} and r_{1t}, $t = 1, 2, \ldots, T$ are the residuals obtained from the regression of ΔX_t and X_{t-1}^* respectively on the lagged differences and unrestricted deterministic terms. The solution to (3.2) gives $1 > \hat{\zeta}_1 > \cdots > \hat{\zeta}_p > \hat{\zeta}_{p+1} = \hat{\zeta}_{p+2} = \cdots = 0$.

The trace statistic for the hypothesis $H(r) : rank(\Pi^*) \leq r$ against $H(p) : rank(\Pi^*) \leq p$ is given by

$$-2 \log Q(H(r)|H(p)) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\zeta}_i) \tag{3.3}$$

and the maximal eigenvalue statistic for $H(r)$ against $H(r+1) : rank(\Pi^*) \leq r + 1$ is given by

$$-2 \log Q(H(r)|H(r+1)) = -T \log(1 - \hat{\zeta}_{r+1}). \tag{3.4}$$

Define $w_t' = (\Delta X_{t-1}', \ldots, \Delta X_{t-k+1}', D_t', X_{t-1}^{*'}, \Delta X_t')$, which is the t-th row of the $T \times (pk + p_1 + q)$ matrix W, where $W' = (w_1, \ldots, w_T)$. In order to implement Johansen's procedure we employ the following algorithm.

Proposition 3.1. (Doornik and O'Brien (2001), O'Brien (1996)). If W has full column rank then the estimates of ζ, α and β can be calculated by:

(a) QR decomposition of W :

$$W = QR$$

where Q is $T \times (pk+p_1+q)$ such that $Q'Q = I_{pk+p_1+q}$ and R is $(pk+p_1+q) \times (pk+p_1+q)$, upper triangular with positive diagonal elements, i.e.

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

where R_{11} *is* $[p(k-1)+q] \times [p(k-1)+q]$, R_{22} *is* $p_1 \times p_1$ *and* R_{33} *is* $p \times p$.

(b) Singular value decomposition of $R_{23}R_{33}^{-1}$:

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

where U *is* $p_1 \times p$ *and* V *is* $p \times p$ *such that* $U'U = V'V = I_p$ *and* $\Sigma_R = diag(\sigma_1, \ldots, \sigma_p)$. σ_i, $i = 1, \ldots, p$ *are the singular values of* $R_{23}R_{33}^{-1}$, *which are real, non-negative and ordered* $\sigma_1 \geq \ldots \geq \sigma_p$. *The number of non-zero singular values corresponds to the rank of* $R_{23}R_{33}^{-1}$.

The estimates of interest are given by

$$\hat{\zeta}_i = \sigma_i^2(1+\sigma_i^2)^{-1}, \quad i = 1, \ldots, p$$

$$\hat{\beta} = T^{1/2}R_{22}^{-1}US_r$$

$$\hat{\alpha} = T^{-1/2}R_{23}'US_r$$

where $S_r = \begin{bmatrix} I_r \\ 0 \end{bmatrix}$, *a* $p \times r$ *matrix.*

Proof. We partition W into $W = (\underset{T\times[p(k-1)+q]}{W_1} \; \underset{T\times p_1}{W_2} \; \underset{T\times p}{W_3})$ where W_1 holds the lagged differences of X_t and the unrestricted deterministic terms. The t-th rows of W_2 and W_3 are given by $X_{t-1}^{*'}$ and $\Delta X_t'$ respectively.

From the QR decomposition we have

$$W'W = R'Q'QR = R'R$$

or

$$\begin{bmatrix} W_1'W_1 & W_1'W_2 & W_1'W_3 \\ W_2'W_1 & W_2'W_2 & W_2'W_3 \\ W_3'W_1 & W_2'W_3 & W_3'W_3 \end{bmatrix} =$$

$$\begin{bmatrix} R_{11}^{'}R_{11} & R_{11}^{'}R_{12} & R_{11}^{'}R_{13} \\ R_{12}^{'}R_{11} & R_{12}^{'}R_{12}+R_{22}^{'}R_{22} & R_{12}^{'}R_{13}+R_{22}^{'}R_{23} \\ R_{13}^{'}R_{11} & R_{13}^{'}R_{12}+R_{23}^{'}R_{22} & R_{13}^{'}R_{13}+R_{23}^{'}R_{23}+R_{33}^{'}R_{33} \end{bmatrix}.$$

Using the equality of the above matrices we get

$$R_{11}^{'}R_{11} = W_1^{'}W_1 \tag{3.5}$$

$$R_{12} = (R_{11}^{'})^{-1}W_1^{'}W_2 \tag{3.6}$$

$$R_{13} = (R_{11}^{'})^{-1}W_1^{'}W_3 \tag{3.7}$$

$$R_{22}^{'}R_{22} = W_2^{'}W_2 - R_{12}^{'}R_{12} = W_2^{'}[I_T - W_1(W_1^{'}W_1)^{-1}W_1^{'}]W_2 = TS_{11} \tag{3.8}$$

$$R_{22}^{'}R_{23} = W_2^{'}W_3 - R_{12}^{'}R_{13} = W_2^{'}[I_T - W_1(W_1^{'}W_1)^{-1}W_1^{'}]W_3 = TS_{10} \tag{3.9}$$

$$R_{23}^{'}R_{23} + R_{33}^{'}R_{33} = W_3^{'}W_3 - R_{13}^{'}R_{13} = W_3^{'}[I_T - W_1(W_1^{'}W_1)^{-1}W_1^{'}]W_3 = TS_{00}. \tag{3.10}$$

Substituting (3.8)-(3.10) into (3.2) gives an alternative eigenvalue problem,

$$0 = |\zeta S_{11} - S_{10}S_{00}^{-1}S_{01}|$$

$$= |T^{-1/2}R_{22}^{'}||\zeta I_{p_1} - R_{23}(R_{23}^{'}R_{23} + R_{33}^{'}R_{33})^{-1}R_{23}^{'}||T^{-1/2}R_{22}|$$

$$= |T^{-1/2}R_{22}^{'}||(\zeta-1)I_{p_1} + [I_{p_1} - R_{23}(R_{23}^{'}R_{23} + R_{33}^{'}R_{33})^{-1}R_{23}^{'}]||T^{-1/2}R_{22}|$$

$$= |T^{-1/2}R_{22}^{'}||(\zeta-1)I_{p_1} + [I_{p_1} + R_{23}(R_{33}^{'}R_{33})^{-1}R_{23}^{'}]^{-1}||T^{-1/2}R_{22}| \tag{3.11}$$

and (3.11) follows from the equality

$$[I_{p_1} + R_{23}(R_{33}'R_{33})^{-1}R_{23}']^{-1} = I_{p_1} - R_{23}(R_{33}'R_{33} + R_{23}'R_{23})^{-1}R_{23}'$$

by applying the formula for the partitioned inverse to the matrix $\begin{bmatrix} I_{p_1} & R_{23} \\ R_{23}' & -R_{33}'R_{33} \end{bmatrix}$.

Then, (3.11) implies

$$|R_{23}(R_{33}'R_{33})^{-1}R_{23}' - (1-\zeta)^{-1}\zeta I_{p_1}| = 0 \tag{3.12}$$

and (3.12) follows from the fact that the eigenvalues of (3.11) are the reciprocals of the eigenvalues of (3.12).

From the singular value decomposition of $R_{23}R_{33}^{-1}$ we have

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

The singular values of $R_{23}R_{33}^{-1}$ are the non-negative square roots of the eigenvalues of $R_{23}R_{33}^{-1}(R_{33}{}^{-1})'R_{23}'$ i.e. $\sigma_i = \sqrt{\hat{\zeta}_i(1-\hat{\zeta}_i)^{-1}}$, $i = 1, \ldots, p$ from which we have $\hat{\zeta}_i = \sigma_i^2/(1+\sigma_i^2)$, $i = 1, \ldots, p$. The eigenvectors of (3.2) which correspond to $\hat{\zeta}_i$, $i = 1, \ldots, p_1$, are $\hat{E} = (\hat{e}_1, \ldots, \hat{e}_{p_1})$, say, and are normalised by $\hat{E}'S_{11}\hat{E} = I_{p_1}$. The eigenvectors of (3.11) which correspond to the symmetric eigenvalue problem

$$|\zeta I_{p_1} - S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}| = 0$$

are given by U and $\hat{E} = S_{11}^{-1/2}U$. So,

$$\hat{\beta} = (\hat{e}_1, \ldots, \hat{e}_r) = S_{11}^{-1/2}US_r = T^{1/2}R_{22}^{-1}US_r,$$

by (3.8) and

$$\hat{\alpha} = S_{01}\hat{\beta} = T^{-1/2}R_{23}'US_r$$

by (3.9) and the above expression for $\hat{\beta}$. ■

Having calculated the $\hat{\zeta}_i$'s we can then compute (3.3) and (3.4) i.e. the LR test statistics for the test for the cointegrating rank.

Chapter 4
LR tests for cointegration and intercept shifts: an asymptotic analysis

In this chapter we use asymptotic theory to investigate the effects of intercept shifts on the inference about the cointegrating rank. We consider shifts occurring at different dates as well as at a common date. We also discuss the effects under alternative specifications of the deterministic term. The asymptotic findings are checked via Monte Carlo simulations.

4.1 The model and preliminary results

The DGP is given by the following VAR(1) model in error correction form

$$\Delta X_t = \Pi X_{t-1} + \Phi D_t + \varepsilon_t, \quad t = 1, 2, \ldots, T \tag{4.1}$$

where X_t is a p-dimensional vector of $I(1)$ variables, which is partitioned into $[X_{1t}^{'}\ X_{2t}^{'}]^{'}$. X_{1t} is $p_1 \times 1$ and contains $I(1)$ variables with intercept shifts and X_{2t} is $p_2 \times 1$ and contains $I(1)$ variables with a drift. The error process ε_t is $i.i.d.$ with mean zero, variance Ω and finite fourth moments. $\Pi = \alpha\beta^{'}$ where α and β are $p \times r$ matrices of rank r and $0 < r < p$, i.e. the variables in the DGP are cointegrated with cointegrating rank r. D_t is a $q \times 1$ vector of deterministic terms partitioned into $[z_t^{'}\ 1]$ and z_t is $q_1 \times 1$ subvector of intercept shifts thus, $q = q_1 + 1$. We consider the case of distinct, non-coincident shifts thus $p_1 = q_1$ and the $p \times q$ coefficient matrix Φ when partitioned conformably with D_t takes the following form,

$$\Phi = \begin{bmatrix} \underset{p_1 \times p_1}{I} & \underset{p_1 \times 1}{0} \\ \underset{p_2 \times p_1}{0} & \underset{p_2 \times 1}{\varphi} \end{bmatrix}. \tag{4.2}$$

For any arbitrary breakpoint t_0, where $t_0 = [T\lambda]$, $\lambda \in (0,1)$, a typical, e.g. the j-th, ($j = 1, 2, \ldots, q_1$) step dummy variable (shift) is given by

$$z_{jt}^* = \begin{cases} 0, & 1 \le t \le t_0 \\ \delta_j, & t_0 + 1 \le t \le T \end{cases}.$$

Following O'Brien (1997, 1999) to simplify the algebra we use the de-meaned shift $z_{jt} = z_{jt}^* - \bar{z}_j^*$, where $\bar{z}_j^* = T^{-1} \sum_{t=1}^{T} z_{jt}^*$. So,

$$z_{jt} = \begin{cases} \delta_j(\lambda - 1), & 1 \le t \le t_0 \\ \delta_j \lambda, & t_0 + 1 \le t \le T \end{cases}. \tag{4.3}$$

In addition,

$$z_{jt} = 0, \quad t \le 0 \tag{4.4}$$

$$Z_{jt} = \sum_{s=1}^{t} z_{js}, \text{ so } Z_{jT} = 0 \tag{4.5}$$

and by rescaling the time axis,

$$z_j(u) = z_{j[Tu]}, \quad u \in [0,1] \tag{4.6}$$

$$Z_j(u) = Z_{j[Tu]}/T, \quad u \in [0,1]. \tag{4.7}$$

Then, when we collect all the step dummies and cumulative step dummies in $q_1 (= p_1) \times 1$ vectors we have $z(u)$, Z_t and $Z(u)$ with their j-th element given by (4.6), (4.5) and (4.7) respectively. The detailed algebraic properties of z_t and Z_t are given in O'Brien (1996, 1997, 1999).

The SM fitted to the data is

$$\Delta X_t = K_1 X_{t-1} + K_2 D_t + e_t, \;\; t = 1, 2, \ldots, T \tag{4.8}$$

with $D_t = 1$ so only an intercept is included in the model.

As described in Chapter 2 the LR test statistics for the hypothesis $H(r)$ are given by

$$-2\log Q(H(r)|H(p)) = -T \sum_{i=r+1}^{p} \log(1 - \hat{\zeta}_i) \simeq T \sum_{i=r+1}^{p} \hat{\zeta}_i \tag{4.9a}$$

$$-2\log Q(H(r)|H(r+1)) = -T\log(1 - \hat{\zeta}_{r+1}) \simeq T\hat{\zeta}_{r+1} \tag{4.9b}$$

and $\hat{\zeta}_{r+1}, \ldots, \hat{\zeta}_p$ correspond to the smallest eigenvalues of $|\zeta S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$. So, the asymptotic behaviour of (4.9a) and (4.9b) calculated from the misspecified model (4.8) depends on the asymptotic properties of the residual sums of squares matrices, S_{ij} $i, j = 0, 1$, also calculated from (4.8), which in turn depend on the DGP (4.1).

For the particular SM under examination, $S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it}R_{jt}^{'}$, $i, j = 0, 1$, (where R_{0t} and R_{1t} are the residuals from the regressions of ΔX_t and X_{t-1} respectively on a vector of ones) can be obtained by applying Frisch-Waugh Theorem as follows. Define the $p \times T$ matrices $\Delta X^{'} = [\Delta X_1 \; \Delta X_2 \ldots \Delta X_T]$ and $X^{'} = [X_0 \; X_2 \ldots X_{T-1}]$ with ΔX_t and X_{t-1}, $t = 1, 2, \ldots, T$ of dimensions $p \times 1$. Let $P = \mathbf{i}(\mathbf{i}^{'}\mathbf{i})^{-1}\mathbf{i}^{'}$, where $\mathbf{i}$ is a $T \times 1$ vector of ones, $\mathbf{i}^{'} = [1, 1, \ldots, 1]$ so that P is a $T \times T$ matrix. Let $M = I_T - P$. Then

$$S_{00} = T^{-1}\Delta X^{'} M \Delta X = T^{-1} \sum_{t=1}^{T} \Delta X_t \Delta X_t^{'} - \bar{\Delta X}\bar{\Delta X}^{'} \tag{4.10}$$

where $\bar{\Delta X} = T^{-1} \sum_{t=1}^{T} \Delta X_t$. Similarly,

$$S_{11} = T^{-1}X^{'}MX = T^{-1} \sum_{t=1}^{T} X_{t-1}X_{t-1}^{'} - \bar{X}\bar{X}^{'} \tag{4.11}$$

where $\bar{X} = T^{-1}\sum_{t=1}^{T} X_{t-1}$ and

$$S_{10} = T^{-1}X'M\Delta X = T^{-1}\sum_{t=1}^{T} X_{t-1}\Delta X_t' - \bar{X}\bar{\Delta}X'. \tag{4.12}$$

From (4.10)-(4.12) it is apparent that S_{ij}, $i, j = 0, 1$, are functions of the original vector process X_t. The structure of X_t, which in turn will determine the limiting properties of S_{ij}, $i, j = 0, 1$, can be analysed using the Granger Representation Theorem. Since we use Johansen's (1996, Chapters 10, 11) methodology in deriving the asymptotic results in this chapter, we apply Johansen's (1991a, 1996) version of the Granger Representation Theorem given in Chapter 2 (Theorem 2.1).

In what follows we give various preliminary asymptotic results necessary to establish the final result. The asymptotic properties of X_t are different in the various directions as shown below.

Lemma 4.1. Let X_t be given by (2.22) *with $D_t' \equiv [z_t' \; 1]$ and Φ as in* (4.2) *and $\varphi \neq 0$. Let $\tau = C\Phi$ be of dimension $p \times q$. Let γ, of dimension $p \times (p - r - q)$, $p > (r + q)$, be chosen orthogonal to β and τ such that (β, γ, τ) are mutually orthogonal and span $\mathbb{R}^p$. Then, when $T \to \infty$ and $u \in [0, 1]$*

$$T^{-1/2}\bar{\gamma}' X_{[Tu]} \xrightarrow{d} \bar{\gamma}' CW(u) \tag{4.13}$$

$$T^{-1}\bar{\tau}' X_{[Tu]} \xrightarrow{d} \begin{bmatrix} Z(u) \\ u \end{bmatrix} \tag{4.14}$$

where $\bar{\gamma} = \gamma(\gamma'\gamma)^{-1}$, $\bar{\tau} = \tau(\tau'\tau)^{-1}$ and $W(u)$ is a p-dimensional Brownian motion, with variance matrix Ω, on the space of continuous functions on $[0, 1]$ denoted by $C[0, 1]$. Define *the $p \times (p-r)$ matrix B_T as $B_T = (\bar{\gamma}, T^{-1/2}\bar{\tau})$ then*

$$T^{-1/2} B_T' X_{[Tu]} \xrightarrow{d} G_0(u) = \begin{bmatrix} \bar{\gamma}' C W(u) \\ Z(u) \\ u \end{bmatrix}. \tag{4.15}$$

Moreover,

$$T^{-1/2} B_T' (X_{[Tu]} - \bar{X}) \xrightarrow{d} G_0(u) - \bar{G}_0 = G = \begin{bmatrix} \bar{\gamma}' C (W(u) - \bar{W}) \\ Z(u) - \bar{Z} \\ u - 1/2 \end{bmatrix} \tag{4.16}$$

where $\bar{G}_0 = \int_0^1 G_0(u) du$.

Proof. See Appendix B.

Another set of preliminary results concerning the asymptotic behaviour of the residual sums of squares is given in the following lemma. In fact the following results are similar to those in section 2.4 (see results (2.23)-(2.28)) but the former account for the presence of step dummy variables.

Lemma 4.2. Under the assumptions of Lemma 4.1

$$S_{00} \xrightarrow{p} \Sigma_{00} + C\Phi R \Phi' C' \equiv \Sigma_{00}^* \tag{4.17}$$

$$\beta' S_{11} \beta \xrightarrow{p} \Sigma_{\beta\beta} + \beta' C_1^1(1) g C_1^1(1)' \beta \equiv \Sigma_{\beta\beta}^* \tag{4.18}$$

$$\beta' S_{10} \xrightarrow{p} \Sigma_{\beta 0} + \beta' C_1^1(1) g C_1' \equiv \Sigma_{\beta 0}^* \tag{4.19}$$

$$T^{-1} B_T' S_{11} B_T \xrightarrow{d} \int_0^1 G G' du. \tag{4.20}$$

$$T^{-1/2} B_T' S_{11} \beta \xrightarrow{d} V C_1^1(1)' \beta \tag{4.21}$$

$$T^{-1/2} B_T' S_{10} \xrightarrow{d} \begin{bmatrix} V & 0 \end{bmatrix} C' = V C_1' \tag{4.22}$$

where $R = \begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix}$ *and* $T^{-1}\sum_{t=1}^{T} z_t z_t' \rightarrow g$ *with* (i,j)*-th element* $\delta_i\delta_j\lambda_l(1-\lambda_m)$ *and* $\lambda_l = \min(\lambda_i, \lambda_j)$, $\lambda_m = \max(\lambda_i, \lambda_j)$. $C_1^1(1)$ *and* C_1 *are the first* p_1 *columns of* $C_1(1)$ *and* C *respectively (see Theorem 2.1) and* $V = \int_0^1 G_0(u)z(u)'du$.

Proof. See Appendix B.

The results of Lemma 4.1 are similar to those in Johansen (1996, Lemma 10.2) but Lemma 4.1 allows p_1 dimensions in $\mathbb{R}^p$ for the step dummies. Comparing the results of Lemma 4.2 with those for the standard case, i.e. no intercept shifts (see Johansen (1996, Lemma 10.3)) we notice that the presence of step dummies in the model increases the (conditional) variance of the stationary components, ΔX_t (4.17) and $\beta' X_t$ (4.18) and alters the covariance (4.19) between them. The stochastic order of magnitude of $S_{11}\beta$ and S_{10} in the non-stationary directions increases by $T^{1/2}$ (4.21, 4.22) compared with the case that no step dummies are present in the DGP. So these terms turn out to be $O_p(T^{1/2})$ instead of $O_p(1)$ because the contribution of the vector with the step dummies (z_t) dominates the asymptotic behaviour of the product moment matrices of the residuals.

If no shifts are present, i.e. $\delta_j = 0$ for $j = 1, 2, \ldots, p_1$, the results of Lemma 4.2 reduce to those for the standard case.

Comparing the results of Lemma 4.1 and 4.2 with the null case (see O'Brien (1996, 1997, 1999)), we observe that in the null case the problem with the order of magnitude appears when the random walk processes (with or without step dummies) in the model interact with the stationary analogue (first differences) of the random walks with the step dummies. In the cointegrated case the problem with the order of magnitude also arises from the interaction of the non-stationary components with the stationary, which in this case are the first differences and the cointegrating relations ($\beta' X_t$).

4.2 The effects of step dummy variables

Using the preliminary results given in section 4.1 we analyse the effects of step dummy variables on the LR tests for cointegration. The null hypothesis is that there are r cointegrating vectors, $H(r)$, against the alternative of stationarity, $H(p)$, (i.e. the p components of the VAR process are stationary) with test statistic $-2\log Q(H(r)|H(p)) = -T\sum_{i=r+1}^{p}\log(1-\hat{\zeta}_i)$, or against the alternative of $(r+1)$ cointegrating vectors, $H(r+1)$ with test statistic $-2\log Q(H(r)|H(r+1)) = -T\log(1-\hat{\zeta}_{r+1})$. $\hat{\zeta}_{r+1},\dots,\hat{\zeta}_p$ are the smallest solutions of

$$|S(\zeta)| = 0 \tag{4.23}$$

where $S(\zeta) = \zeta S_{11} - S_{10}S_{00}^{-1}S_{01}$. In the standard case (no intercept shifts), under the null hypothesis of r cointegrating vectors the ordered eigenvalues of (4.23) converge in probability to $(\zeta_1,\dots,\zeta_r,0,\dots,0)$, where $\zeta_1,\dots,\zeta_r$ are the ordered (positive) eigenvalues of $|\zeta\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}| = 0$, see Johansen (1988, Lemma 4).

Next we analyse the asymptotic properties of $S(\zeta)$ in the presence of intercept shifts, bearing in mind that the ordered eigenvalues of (4.23) are continuous functions of the elements of S_{ij}, $i,j = 0,1$, see Andersson et al. (1983, p. 395).

Let $A_T \equiv (\beta, T^{-1/2}B_T) = (\beta, T^{-1/2}\bar{\gamma}, T^{-1}\bar{\tau})$ then A_T is a $p \times p$ non-singular matrix since β, γ and τ span $\mathbb{R}^p$ (see section 4.1). Hence the eigenvalues of (4.23) also satisfy

$$|A_T' S(\zeta) A_T| = 0 \tag{4.24}$$

see Anderson (1984, p. 589). Partitioning (4.24) we get

$$|A_T' S(\zeta) A_T| = \begin{vmatrix} \beta' S(\zeta)\beta & T^{-1/2}\beta' S(\zeta) B_T \\ T^{-1/2} B_T' S(\zeta)\beta & T^{-1} B_T' S(\zeta) B_T \end{vmatrix} =$$

$$\begin{vmatrix} \zeta\beta' S_{11}\beta - \beta' S_{10} S_{00}^{-1} S_{01}\beta & T^{-1/2}\zeta\beta' S_{11} B_T - T^{-1/2}\beta' S_{10} S_{00}^{-1} S_{01} B_T \\ T^{-1/2}\zeta B_T' S_{11}\beta - T^{-1/2} B_T' S_{10} S_{00}^{-1} S_{01}\beta & T^{-1}\zeta B_T' S_{11} B_T - T^{-1} B_T' S_{10} S_{00}^{-1} S_{01} B_T \end{vmatrix} = 0. \tag{4.25}$$

Using the results of Lemma 4.2 and (4.25) we find that

$$|A_T' S(\zeta) A_T| \xrightarrow{d}$$

$$\begin{vmatrix} \zeta\Sigma_{\beta\beta}^* - \Sigma_{\beta 0}^* \Sigma_{00}^{*-1} \Sigma_{0\beta}^* & \zeta\beta' C_1^1(1) V' - \Sigma_{\beta 0}^* \Sigma_{00}^{*-1} C_1 V' \\ \zeta V C_1^1(1)'\beta - V C_1' \Sigma_{00}^{*-1} \Sigma_{0\beta}^* & \zeta \int_0^1 G G' du - V C_1' \Sigma_{00}^{*-1} C_1 V' \end{vmatrix}$$

$$= |\zeta M_1 - M_2| = 0 \tag{4.26}$$

where $M_1 = \begin{bmatrix} \Sigma_{\beta\beta}^* & \beta' C_1^1(1) V' \\ V C_1^1(1)'\beta & \int_0^1 G G' du \end{bmatrix}$ and $M_2 = \begin{bmatrix} \Sigma_{\beta 0}^* \Sigma_{00}^{*-1} \Sigma_{0\beta}^* & \Sigma_{\beta 0}^* \Sigma_{00}^{*-1} C_1 V' \\ V C_1' \Sigma_{00}^{*-1} \Sigma_{0\beta}^* & V C_1' \Sigma_{00}^{*-1} C_1 V' \end{bmatrix}$.
M_1 and M_2 are symmetric and M_1 is the probability limit of S_{11} which is by assumption non-singular.

Unlike the standard case, (4.26) is not the determinant of a diagonal matrix so it is not obvious how many of the roots of (4.23) converge in probability to positive and how many to zero eigenvalues of (4.26).

Let $F = M_1^{-1/2} M_2 M_1^{-1/2}$, then F is positive semi-definite and symmetric and

$$|\zeta M_1 - M_2| = |M_1||\zeta I_p - F| = 0.$$

Thus, the rank of F, which is the number of non-zero eigenvalues of F and equivalently the number of non-zero roots of (4.26), is informative in establishing an upper (lower) bound for the number of positive (zero) eigenvalues in the limit and consequently an upper bound for the number of spurious cointegrating vectors that might arise asymptotically.

Define

$$Q = \begin{bmatrix} I_r & 0 \\ -VC_1'\Sigma_{00}^{*-1}\Sigma_{0\beta}^{*}(\Sigma_{\beta 0}^{*}\Sigma_{00}^{*-1}\Sigma_{0\beta}^{*})^{-1} & I_{p-r} \end{bmatrix},$$

a non-singular matrix ($|Q| = 1$) and

$$D = \begin{bmatrix} \Sigma_{\beta 0}^{*}\Sigma_{00}^{*-1}\Sigma_{0\beta}^{*} & 0 \\ 0 & VC_1'N^*C_1V' \end{bmatrix}$$

where $N^* = \Sigma_{00}^{*-1} - \Sigma_{00}^{*-1}\Sigma_{0\beta}^{*}(\Sigma_{\beta 0}^{*}\Sigma_{00}^{*-1}\Sigma_{0\beta}^{*})^{-1}\Sigma_{\beta 0}^{*}\Sigma_{00}^{*-1}$.

Then $M_2 = QDQ'$ and

$$\begin{aligned} rank(F) &= rank(M_2) = rank(D) \\ &= rank(\Sigma_{\beta 0}^{*}\Sigma_{00}^{*-1}\Sigma_{0\beta}^{*}) + rank(VC_1'N^*C_1V') \\ &= r + rank(VC_1'N^*C_1V') \end{aligned}$$

since $(\Sigma_{\beta 0}^{*}\Sigma_{00}^{*-1}\Sigma_{0\beta}^{*})$ is assumed to be non-singular of rank r. Hence in the limit there are more than r positive eigenvalues given that $(VC_1'N^*C_1V')$ is not the null matrix and $rank(VC_1'N^*C_1V')$ is positive. Thus, asymptotically there appear to be more cointegrating vectors (cointegrating/stationary relationships) than in the DGP.

Proposition 4.1. The rank of VC_1', b say, which is at most p_1 (i.e. $b \leq p_1$) *gives an upper bound for the number of non-zero eigenvalues, that is the number of spurious cointegrating relationships that arise as $T \to \infty$.*

Proof. N^* is a $p \times p$ matrix of rank $(p-r)$ since $(\Sigma^*_{0\beta})' N^* = 0$ and therefore N^* lies in the null space spanned by the columns of $\Sigma^*_{0\beta}$, which has rank r. So N^* can be decomposed into $N^* = P^* P^{*'}$ where P^* is a $p \times (p-r)$ matrix of rank $(p-r)$ and $P^{*'}\Sigma^*_{0\beta} = 0$. Then

$$rank(VC_1' N^* C_1 V') = rank[(VC_1'P^*)(VC_1'P^*)'] = rank(VC_1'P^*) \leq$$

$$\min[rank(VC_1'), (p-r)] \leq \min[p_1,\ (p-r)] = p_1$$

because V is $(p-r) \times p_1$, C_1 is $p \times p_1$ and of the restriction $(p-r-q) > 0$ indicating the existence of the γ direction. ■

Proposition 4.2. The rank of $VC_1' = \begin{bmatrix} V & 0 \end{bmatrix} C'$ *cannot exceed* l_1, *the number of variables with intercept shifts or linear combinations thereof that are weakly exogenous*[4].

Proof. $VC_1' = \begin{bmatrix} V & 0 \end{bmatrix} C' = \begin{bmatrix} \bar{\gamma}' C \int_0^1 W(u)z(u)'du & 0 \\ \int_0^1 Z(u)z(u)'du & 0 \\ \int_0^1 uz(u)'du & 0 \end{bmatrix} \alpha_\perp H$, where $H = (\beta_\perp' \alpha_\perp)^{-1}\beta_\perp'$ is $(p-r) \times p$ and has rank $(p-r)$, by definition. Partitioning $\alpha_\perp$ conformably with $\begin{bmatrix} V & 0 \end{bmatrix}$,

$$\begin{bmatrix} \bar{\gamma}' C \int_0^1 W(u)z(u)'du & 0 \\ \int_0^1 Z(u)z(u)'du & 0 \\ \int_0^1 uz(u)'du & 0 \end{bmatrix} \begin{bmatrix} \underset{p_1 \times (p-r)}{\alpha_\perp^{(1)}} \\ \underset{p_2 \times (p-r)}{\alpha_\perp^{(2)}} \end{bmatrix} H$$

$$= \begin{bmatrix} \bar{\gamma}' C \int_0^1 W(u)z(u)'du\alpha_\perp^{(1)} \\ \int_0^1 Z(u)z(u)'du\alpha_\perp^{(1)} \\ \int_0^1 uz(u)'du\alpha_\perp^{(1)} \end{bmatrix} H = V\alpha_\perp^{(1)}H.$$

Thus, $rank(VC_1') = rank(V\alpha_\perp^{(1)}H) \leq rank(\alpha_\perp^{(1)})$ and the rank of $\alpha_\perp^{(1)}$ depends on the number of variables with intercept shifts that are weakly exogenous. Partitioning α into $\alpha = \begin{bmatrix} \underset{p_1 \times r}{\alpha^{(1)}} \\ \underset{p_2 \times r}{\alpha^{(2)}} \end{bmatrix}$ and noting that the matrix $\begin{bmatrix} \alpha & \alpha_\perp \end{bmatrix} = \begin{bmatrix} \alpha^{(1)} & \alpha_\perp^{(1)} \\ \alpha^{(2)} & \alpha_\perp^{(2)} \end{bmatrix}$ has full rank, p, we must have $rank(\alpha^{(1)}) + rank(\alpha_\perp^{(1)}) = p_1$. Suppose that $l_1 \leq p_1$ rows of $\alpha^{(1)}$ are zero, which

---

[4] We refer to the notion of weak exogeneity as it is given in Theorem 8.1 in Johansen (1996).

means that l_1 variables with intercept shifts are weakly exogenous, and that the remaining $(p_1 - l_1)$ rows are linearly independent giving $rank(\alpha^{(1)}) = (p_1 - l_1)$. Then, l_1 rows of $\alpha_\perp^{(1)}$ must be non-zero and linearly independent with $rank(\alpha_\perp^{(1)}) = l_1$. Hence, $rank(VC_1') \leq l_1$ and Proposition 4.2 follows. Second, suppose $rank(\alpha^{(1)}) = (p_1 - l_1)$, but $\alpha^{(1)}$ contains no zero rows. If we take appropriate linear combinations of the p_1 variables involved, we can generate a transformed $\alpha^{(1)}$ with l_1 zero rows corresponding to weakly exogenous linear combinations of the shifted variables. Finally, note that we can transform an ECM such that the adjustment coefficient matrix α has r linearly independent rows and $(p - r)$ zero rows. ■

Corollary 4.1. If $\alpha_\perp^{(1)} = 0$ *which together with the restriction* $rank(\alpha^{(1)}) + rank(\alpha_\perp^{(1)}) = p_1$ *imply that* $\alpha^{(1)} \neq 0$ *and* $rank(\alpha^{(1)}) = p_1$ *i.e. none of the variables with intercept shifts is weakly exogenous, then* $rank(VC_1') = 0$.

Corollary 4.2. If the variables without intercept shifts are all weakly exogenous i.e. $\alpha^{(2)} = 0$ *(which implies* $rank(\alpha_\perp^{(2)}) = p_2$*) and the cointegrating rank equals the number of variables with intercept shifts (*$r = p_1$*) then* $rank(\alpha_\perp^{(1)}) = 0$.

Corollaries 4.1 and 4.2 describe special cases, where spurious cointegration does not arise, despite the presence of intercept shifts in the DGP.

Corollary 4.2 uses the fact that $rank(\alpha_\perp) = (p - r)$ and if $r = p_1$, $rank(\alpha_\perp) = p_2 = rank(\alpha_\perp^{(2)})$ since all the linearly independent rows of $\alpha_\perp$ are given by the p_2 rows of $\alpha_\perp^{(2)}$.

If we have r cointegrating relations loading on the first r variables, with the remaining $(p - r)$ weakly exogenous and the shifted variables are a subset of the former, then intuitively the system is driven by unshifted stochastic trends, and the shifts do not affect its long-run behaviour.

The above analysis suggests that there are more than r (less than $p-r$) eigenvalues of (4.23) which are $O_p(1)$ ($O_p(T^{-1})$). An intuitive explanation of this finding is given below.

Let $\hat{\zeta}_i$, $i = 1, 2, \ldots, p$ be an eigenvalue of (4.23) and the corresponding eigenvectors are given by the columns of $\hat{V} = (\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_p)$, then

$$S_{11}\hat{v}_i\hat{\zeta}_i = S_{10}S_{00}^{-1}S_{01}\hat{v}_i, \;\; i = 1, 2, \ldots, p. \tag{4.27}$$

Since the eigenvectors are normalised by $\hat{V}'S_{11}\hat{V} = I_p$ (see section 2.3), pre-multiplying (4.27) by $\hat{v}_i'$ we get

$$\hat{\zeta}_i = \hat{v}_i'S_{10}S_{00}^{-1}S_{01}\hat{v}_i, \;\; i = 1, 2, \ldots, p. \tag{4.28}$$

However, the normalisation $\hat{V}'S_{11}\hat{V} = I_p$ eliminates S_{11} from the expression for the eigenvalue and (4.28) is not informative about the stochastic order of magnitude of the eigenvalue. Thus, a re-normalisation of the eigenvectors is required, see Davidson (2000, p. 395). Let $\hat{e}_i = (\hat{v}_i'\hat{v}_i)^{-1/2}\hat{v}_i$ be a vector of unit length and $\hat{e}_i'S_{11}\hat{e}_i = (\hat{v}_i'\hat{v}_i)^{-1}$, by $\hat{V}'S_{11}\hat{V} = I_p$, then (4.28) becomes

$$\hat{\zeta}_i = \frac{\hat{e}_i'S_{10}S_{00}^{-1}S_{01}\hat{e}_i}{\hat{e}_i'S_{11}\hat{e}_i}, \;\; i = 1, 2, \ldots, p. \tag{4.29}$$

The stochastic order of the various terms in (4.29) depends on whether $\hat{e}_i$ and hence $\hat{v}_i$ converge to a point in the cointegrating space. The following argument under correct specification (absence of intercept shifts) assumes convergence of the r eigenvectors, that correspond to the r largest eigenvalues, to points in the cointegrating space, see Davidson (2000, p. 395). We assume that this remains true in the presence of intercept shifts. Therefore, we expect the eigenvectors which correspond to the r largest eigenvalues of (4.23) to converge to points in the cointegrating space, so that $\hat{e}_i' S_{10}$ and $\hat{e}_i' S_{11} \hat{e}_i$ are $O_p(1)$ and therefore $\hat{\zeta}_i = O_p(1)$, for $i = 1, \ldots, r$ (by (4.17)-(4.19)). For the remaining $(p - r)$ eigenvectors which correspond to the $(p - r)$ smallest eigenvalues of (4.23), in the presence of intercept shifts, we find that $\hat{e}_i' S_{10}$ is $O_p(T^{1/2})$ for $i = r + 1, \ldots, p$, by (4.22), instead of $O_p(1)$, but $\hat{e}_i' S_{11} \hat{e}_i$, for $i = r + 1, \ldots, p$ behaves as in the standard case, i.e. it is $O_p(T)$ by (4.20). Thus, the r largest eigenvalues are $O_p(1)$ and therefore 'well behaved' asymptotically. However, we find that some of the remaining $(p - r)$ eigenvalues are spuriously $O_p(1)$ (at most p_1 have positive probability limits, see Proposition 4.1) instead of $O_p(T^{-1})$.

Given that in the presence of 'effective' intercept shifts (i.e. $b > 0$) there are more than r eigenvalues which are $O_p(1)$, using the test statistics designed for the standard case (correctly specified model) we find that both tests reject the null hypothesis of r cointegrating vectors with probability one as the sample size tends to infinity. The maximal eigenvalue statistic uses the largest of the $(p - r)$ smallest eigenvalues of (4.23) which according to the analysis of (4.26) seems to be non-zero in the limit and therefore $O_p(1)$. Thus,

$$-2 \log Q(H(r)|H(r + 1)) = -T \log(1 - \hat{\zeta}_{r+1}) \geq T\hat{\zeta}_{r+1} = TO_p(1) \to \infty$$

as $T \to \infty$ and $H(r)$ is rejected with probability one. For the trace statistic which is the sum of the $(p-r)$ smallest eigenvalues of (4.23) we have

$$\begin{aligned} -2\log Q(H(r)|H(p)) &= -T\sum_{i=r+1}^{p}\log(1-\hat{\zeta}_i) \geq T\sum_{i=r+1}^{p}\hat{\zeta}_i \\ &= T\sum_{i=r+1}^{b}\hat{\zeta}_i + T\sum_{i=b+1}^{p}\hat{\zeta}_i = TO_p(1) + TO_p(T^{-1}) \to \infty \end{aligned}$$

as $T \to \infty$ and $H(r)$ is rejected with probability one asymptotically.

Below we explore the procedure followed by Johansen (1996, Chapter 11) in deriving the asymptotic distribution of the test statistics. We find that this procedure fails in the presence of intercept shifts since they change the stochastic order of certain residual product moment matrices and consequently the stochastic order of the eigenvalues used in the tests.

The asymptotic distributions of the trace and the maximal eigenvalue statistics, as outlined in Chapter 2, are derived by first finding the limit of the eigenvalue equation (4.23), under the maintained hypothesis that the cointegrating rank is r and therefore that the $(p-r)$ smallest eigenvalues (associated with the non-stationary directions) are $O_p(T^{-1})$. Then the limiting expression is a stochastic $(p-r)\times(p-r)$ matrix whose eigenvalues are used in the tabulation of the asymptotic distributions of the trace and maximal eigenvalue statistics.

Initially we consider the eigenvalue equation (4.23) in the stationary and non-stationary directions.

$$|(\beta, B_T)'S(\zeta)(\beta, B_T)| =$$

$$|\zeta\begin{bmatrix} \beta' S_{11}\beta & \beta' S_{11}B_T \\ B_T' S_{11}\beta & B_T' S_{11}B_T \end{bmatrix} - \begin{bmatrix} \beta' S_{10}S_{00}^{-1}S_{01}\beta & \beta' S_{10}S_{00}^{-1}S_{01}B_T \\ B_T' S_{10}S_{00}^{-1}S_{01}\beta & B_T' S_{10}S_{00}^{-1}S_{01}B_T \end{bmatrix}| = 0. \qquad (4.30)$$

The procedure assumes that the $(p-r)$ roots of (4.30) are $O_p(T^{-1})$ so as $T \to \infty$, for any of the smallest $(p-r)$ eigenvalues, $\hat{\zeta}$ say, we have $T\hat{\zeta} \to \kappa$. Multiplying and dividing the first matrix in (4.30) by T and using the stochastic order of magnitude of the matrices in (4.30) (see Lemma 4.2) we have

$$|\kappa \begin{bmatrix} o_p(1) & o_p(1) \\ o_p(1) & O_p(1) \end{bmatrix} - \begin{bmatrix} O_p(1) & O_p(T^{1/2}) \\ O_p(T^{1/2}) & O_p(T) \end{bmatrix}| = 0$$

or, after expanding (4.30),

$$|O_p(1)| \times |\kappa O_p(1) - O_p(T)| = 0$$

which is not defined as $T \to \infty$. Note that in the absence of intercept shifts for (4.30) we have $|O_p(1)| \times |\kappa O_p(1) - O_p(1)| = 0$ (see Johansen (1996, pp. 159-160)) and therefore an asymptotic distribution. It would seem necessary to define a new direction of dimension b and scale accordingly, but this does not seem possible.

4.3 Monte Carlo simulations

Monte Carlo simulations can be regarded as having (at least) two functions. One is to measure the size of small sample effects, and this is deferred until Chapter 5. Another is to verify the correctness of asymptotic results. Below we illustrate the asymptotic findings using the results from Monte Carlo simulations. All simulation experiments are based on 10,000 replications and were programmed in Ox 3.00 (see Doornik (1999)). A detailed presentation of the DGPs used for the simulations appears in Appendix C.

Figures 4.1 and 4.2 show the rejection frequency (abbreviated as rf on the vertical axis) of the null hypotheses of one and two cointegrating vectors using the trace and maximal eigenvalue statistics. The DGP includes four random walks, three of which have drifts, one has an intercept shift of magnitude 0.5 (i.e. $\delta = 0.5$), at $T/2$, and there is one cointegrating vector. The true null hypothesis is rejected with frequency that tends to one as the sample size increases, implying the incorrect acceptance of at least two cointegrating vectors. The false null hypothesis of two cointegrating vectors is rejected with frequency that does not exceed 0.11 which reflects the distortion occurring in the size of the test due to misspecification caused by the unmodelled intercept shift.

Figures 4.3 and 4.4 show the frequency of rejecting the null hypotheses of one, two and three cointegrating vectors when the DGP contains five random walks, two of which have intercept shifts (of magnitude 0.5) at two different dates, $T/3$ and $2T/3$. The remaining three random walks have drifts. The frequency of rejecting the true null hypothesis of one cointegrating vector tends to one as the sample size increases. The false null of two cointegrating vectors is rejected with frequency that increases with the sample size and which is much higher than the asymptotic size of the test. So as the sample size gets larger the tests indicate (quite often) that there are three cointegrating relations, as the asymptotic analysis suggests.

Figure 4.5 shows the rejection frequency of the true null hypothesis of one cointegrating vector when the DGP consists of four random walks, one of which has an intercept shift (of magnitude 0.5 at $T/2$) and is not weakly exogenous with respect to the single cointegrating relation. The other three random walks contain drifts. In accordance with the asymptotic result, the rejection frequency is very close to the size of the tests hence systematic acceptance of spurious cointegrating relations is not expected.

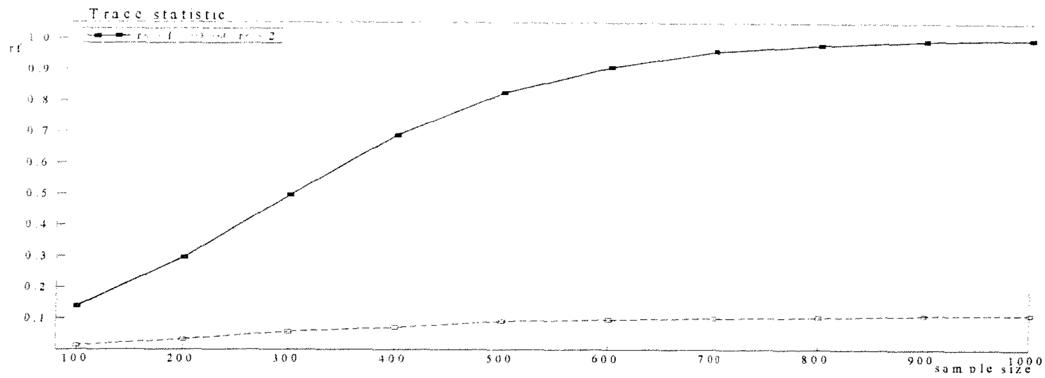

Figure 4.1. Frequency of rejecting the null hypotheses $r \leq 1$ (true) and $r \leq 2$ using the trace statistic.

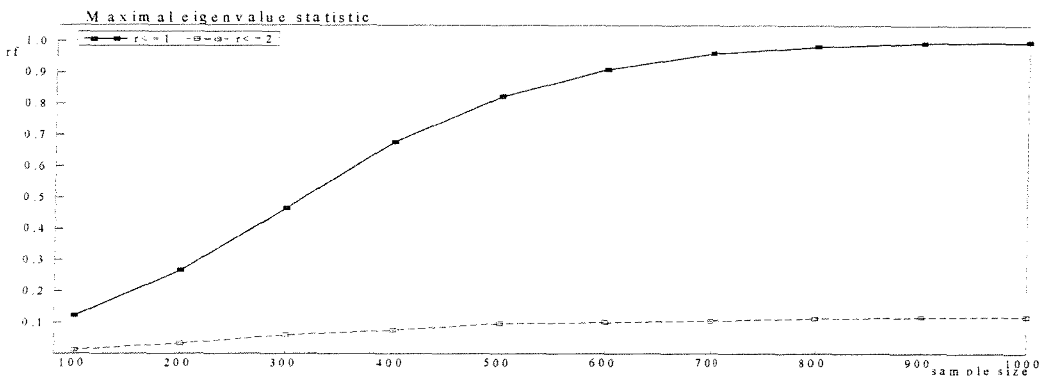

Figure 4.2. Frequency of rejecting the null hypotheses $r \leq 1$ (true) and $r \leq 2$ using the maximal eigenvalue statistic.

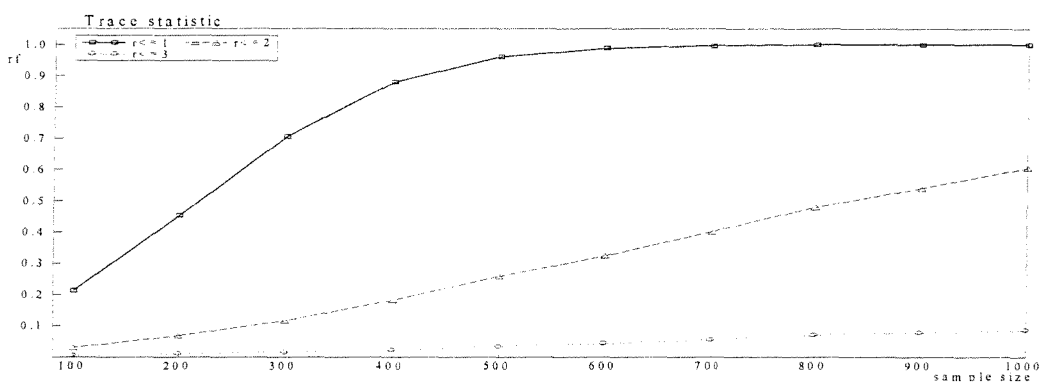

Figure 4.3. Frequency of rejecting the null hypotheses $r \leq 1$ (true), $r \leq 2$ and $r \leq 3$ using the trace statistic.

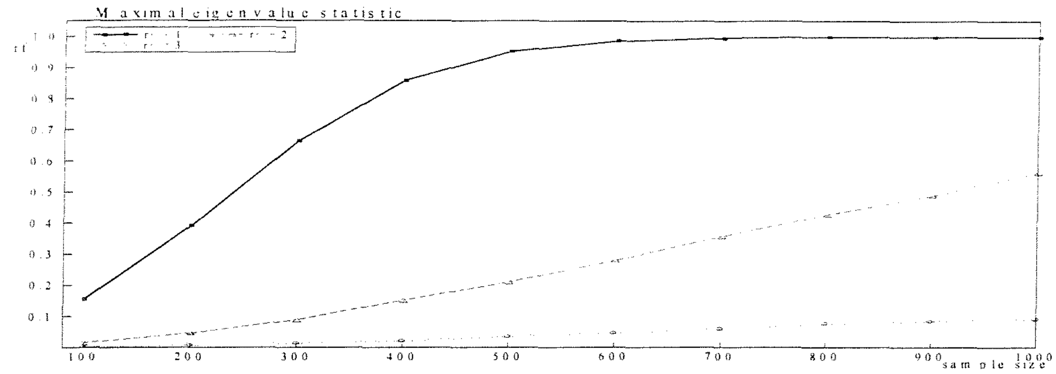

Figure 4.4. Frequency of rejecting the null hypotheses $r \leq 1$ (true), $r \leq 2$ and $r \leq 3$ using the maximal eigenvalue statistic.

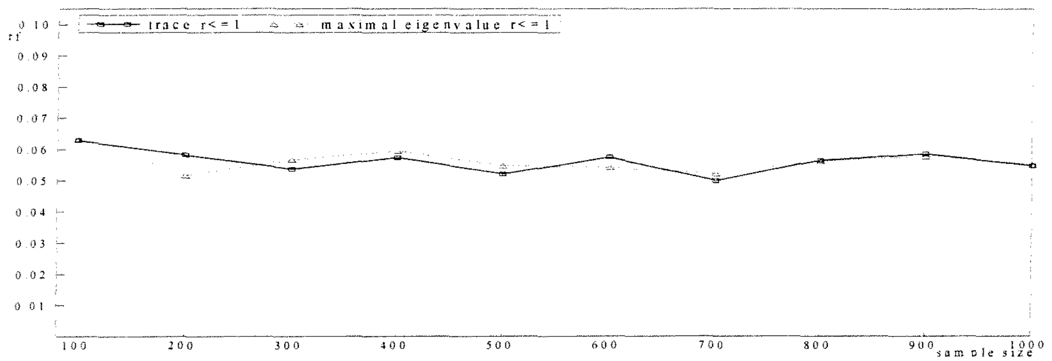

Figure 4.5. Frequency of rejecting the null hypothesis $r \leq 1$ (true) using the trace and the maximal eigenvalue statistics.

4.4 The null case

Below we investigate how the results reduce in the null case ($r = 0$), without a constant term in the DGP, which is the case analysed by O'Brien (1996, 1997, 1999) where a different approach is followed in deriving the asymptotic results.

The directions of the process X_t under consideration are γ ($p \times p - p_1$ and $p - p_1 = p_2$) and τ ($p \times p_1$) which correspond to the stochastic trends and the non-stationary part of the process due to intercept shifts respectively. In addition, when $r = 0$, $C = I_p$.

The roots of

$$|\zeta I_p - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}| = 0 \tag{4.31}$$

are the same as those of (4.23). We use the fact that the eigenvalues are continuous functions of the elements of S_{ij}, $i, j = 0, 1$, to investigate the stochastic order of the eigenvalues.

Let $B_T = (\bar{\gamma}\,, T^{-1/2}\bar{\tau})$ be a $p \times p$ non-singular matrix, then

$$|\zeta I_p - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}| =$$

$$|\zeta I_p - (T^{-1}B_T'S_{11}B_T)^{-1}(T^{-1/2}B_T'S_{10})S_{00}^{-1}(S_{01}B_T T^{-1/2})| = 0$$

and $(T^{-1}B_T'S_{11}B_T)^{-1}(T^{-1/2}B_T'S_{10})S_{00}^{-1}(S_{01}B_T T^{-1/2}) = O_p(1)$ (by (4.20) and (4.22) using B_T defined in this section) which makes the roots of (4.23) and (4.31) $O_p(1)$.

Thus, an expression for the asymptotic distribution can be found,

$$\sum_{i=1}^{p}\hat{\zeta}_i = tr\{(T^{-1}B_T'S_{11}B_T)^{-1}(T^{-1/2}B_T'S_{10})S_{00}^{-1}(S_{01}B_T T^{-1/2})\} \xrightarrow{d}$$

$$tr\{(\int_0^1 GG'du)^{-1}\begin{bmatrix} \int_0^1 G_0(u)z(u)'du & 0 \end{bmatrix}\Sigma_{00}^{*-1}\begin{bmatrix} \int_0^1 G_0(u)z(u)'du & 0 \end{bmatrix}'\}$$

$$= tr\{\begin{bmatrix} (\int_0^1 G_0(u)z(u)'du)'(\int_0^1 GG'du)^{-1}(\int_0^1 G_0(u)z(u)'du) & 0 \\ 0 & 0 \end{bmatrix}\Sigma_{00}^{*-1}\}$$

$$= tr\{(\int_0^1 G_0(u)z(u)'du)'(\int_0^1 GG'du)^{-1}(\int_0^1 G_0(u)z(u)'du)\Sigma_1^*\} \tag{4.32}$$

where now $G_0(u) = \begin{bmatrix} \bar{\gamma}'W(u) \\ Z(u) \end{bmatrix}$, $G = \begin{bmatrix} \bar{\gamma}'(W(u) - \bar{W}) \\ Z(u) - \bar{Z} \end{bmatrix}$, $W(u)$ is $p_2 \times 1$ and $\Sigma_{00}^{*-1} = \begin{bmatrix} \Sigma_1^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\prime} & \Sigma_2^* \end{bmatrix}$.

We can set δ_i's to unity by writing $G_0(u) = \Upsilon_\delta G_0^*(u)$ where $\Upsilon_\delta = \begin{bmatrix} I_{p-p_1} & 0 \\ 0 & \tilde{\Delta} \end{bmatrix}$, $\tilde{\Delta} = diag(\delta_1 \ldots \delta_{p_1})$ and $G_0^*(u) = \begin{bmatrix} \bar{\gamma}'W(u) \\ Z^*(u) \end{bmatrix}$ with $Z^*(u) = \tilde{\Delta}^{-1}Z(u)$. Similarly $G =$

$\Upsilon_\delta G^*$. Then (4.32) becomes

$$tr\{(\int_0^1 G_0^*(u)z^*(u)'du)'(\int_0^1 G^*G^{*'}du)^{-1}(\int_0^1 G_0^*(u)z^*(u)'du)\Sigma_1^*\}. \tag{4.33}$$

To simplify even further we consider the case $p_1 = 1$, where $\int_0^1 z^*(u)Z^*(u)du = 0$, see O'Brien (1996). Let $\bar{\gamma}'W(u) = W_2(u)$ then dropping u argument, (4.33) is written as

$$tr\{\begin{bmatrix} \int_0^1 z^*W_2'du & 0 \end{bmatrix}[F(W_2, Z^*)]^{-1}\begin{bmatrix} \int_0^1 W_2z^*du \\ 0 \end{bmatrix}\}$$

where $F(W_2, Z^*) = \begin{bmatrix} \int_0^1(W_2 - \bar{W}_2)(W_2 - \bar{W}_2)'du & \int_0^1(W_2 - \bar{W}_2)(Z^* - \bar{Z}^*)du \\ \int_0^1(Z^* - \bar{Z}^*)(W_2 - \bar{W}_2)'du & \int_0^1(Z^* - \bar{Z}^*)^2du \end{bmatrix}$.

Using the formula for the partitioned inverse and the fact that $\int_0^1(Z^* - \bar{Z}^*)^2du = \lambda^2(\lambda-1)^2/12$ (see O'Brien (1999, equations (1) and (3), p. 25)) the limit of the trace statistic found above becomes,

$$tr\{\int_0^1 z^*W_2'du[f(W_2, Z^*)]^{-1}\int_0^1 W_2z^*du\Sigma_1^*\}$$

where

$$f(W_2, Z^*) =$$

$$\int_0^1(W_2-\bar{W}_2)(W_2-\bar{W}_2)'du - \frac{\lambda^2(\lambda-1)^2}{12}\int_0^1(W_2-\bar{W}_2)(Z^*-\bar{Z}^*)du\int_0^1(Z^*-\bar{Z}^*)(W_2-\bar{W}_2)'du$$

which coincides with the result in O'Brien (1999, p. 12).

4.5 Occurrence of intercept shifts at a common date

In this section we consider the case that the intercept shifts, concerning the last m, $2 \le m \le p_1$, elements of the vector process X_t, occur at a common date. The effect of this assumption is to 'redistribute' the number of variables between the non-stationary stochastic (γ) and deterministic (τ) directions of $\mathbb{R}^p$.

For simplicity we use the following form of step dummy variable,

$$z_{it} = \begin{cases} 0, & 1 \le t \le t_{0i} \\ \delta_i, & t_{0i} + 1 \le t \le T \end{cases},$$

and for shift dates t_{0i}, $i = m, m+1, \ldots, p_1$ we assume that $t_{0m} = t_{0(m+1)} = \cdots = t_{0p_1}$. The error correction form of the model

$$\Delta X_t = \alpha\beta' X_{t-1} + \Phi D_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T \tag{4.34}$$

with all its components defined as previously, can be written as

$$X_t = \Upsilon X_{t-1} + d_t + \varepsilon_t \tag{4.35}$$

where $\Upsilon = I_p + \alpha\beta'$, $d_t' = \begin{bmatrix} z_t' & \varphi' \end{bmatrix}$, z_t is a $p_1 \times 1$ vector of step dummy variables and φ is a $p_2 \times 1$ vector of constants.

Let

$$H = \begin{bmatrix} I_{p_1-m} & 0 & 0 \\ 0 & H_{22} & 0 \\ 0 & 0 & I_{p_2} \end{bmatrix},$$

where the submatrix H_{22} is $m \times m$ defined as

$$H_{22} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -h_1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -h_2 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -h_{m-1} & 1 \end{bmatrix}$$

and $h_i = \delta_{(p_1-m+1+i)}/\delta_{(p_1-m+i)}$, $i = 1, 2, \ldots, m-1$. We can then transform (4.35), by pre-multiplying by H, into

$$HX_t = (H\Upsilon H^{-1})HX_{t-1} + Hd_t + H\varepsilon_t$$

or more compactly,

$$X_t^* = \Upsilon^* X_{t-1}^* + d_t^* + \varepsilon_t^* \tag{4.36}$$

where $X_t^* = HX_t$, $\Upsilon^* = H\Upsilon H^{-1}$, $d_t^{*'} = \left[\begin{array}{cc} z_t^{*'} & \varphi^{*'} \end{array}\right]$, z_t^* is $(p_1 - (m-1)) \times 1$, φ^* is $(p_2 + (m-1)) \times 1$ with the first $(m-1)$ elements equal to zero. Then (4.36) can be written in an error correction form as

$$\Delta X_t^* = \alpha^* \beta^{*'} X_{t-1}^* + d_t^* + \varepsilon_t^* \tag{4.37}$$

since $\Upsilon^* - I_p = H(\Upsilon - I_p)H^{-1} = H\alpha\beta' H^{-1} = \alpha^*\beta^{*'}$ where $\alpha^* = H\alpha$, $\beta^* = H^{-1}\beta$ and $rank(\alpha^*) = rank(\beta^*) = r$.

The transformed system (4.37) has the same properties as (4.34), since H is non-singular, but the dimensions of the non-stationary directions change. The dimension of the non-stationary deterministic direction (τ) is reduced by $(m-1)$ and becomes $p_1 - (m-1) + 1, = q^*$ say. The non-stationary stochastic direction (γ) is increased by $(m-1)$ and becomes $[p_1 - (m-1)] + [p_2 + (m-1)] - r - (p_1 - (m-1) + 1) = p - r - q^*$, since the number of variables with the intercept shifts is reduced by $(m-1)$ and the set of non-shifted variables now involves p_2 random walks with drifts, as before, and $(m-1)$ random walks without drift.

The method of analysis presented in section 4.2 can be similarly applied using (4.37) and appropriately modifying the dimensions of γ and τ. In this case the matrix V is $(p - r) \times [p_1 - (m-1)]$ and C_1 is $p \times [p_1 - (m-1)]$. Therefore an upper bound for the number of spurious cointegrating relations that arise as $T \to \infty$ is given by

$$rank(VC_1') \leq \min(p - r,\ p_1 - (m-1)) = p_1 - (m-1)$$

since $p - r - (p_1 - (m-1)) > 1$, assuming that the γ direction exists.

Thus, when there are m variables with intercept shifts at a common date the upper bound for the number of spurious cointegrating relations is given by the number of distinct shifts in the DGP.

Figures 4.6 and 4.7 show the rejection frequency for the null hypotheses of one and two cointegrating vectors, using the trace and maximal eigenvalue statistics, when the DGP consists of five random walks and one cointegrating vector. Three random walks have drifts, and two have intercept shifts of magnitude 0.5, occurring at a common date, $T/2$ (detailed description of the DGP can be found in Appendix C). For both statistics the frequency of rejecting the true null hypothesis of one cointegrating vector tends to unity as the sample size increases. The rejection frequency of the false null hypothesis of two cointegrating vectors does not seem to indicate acceptance of a third cointegrating vector too often, which agrees with the asymptotic result.

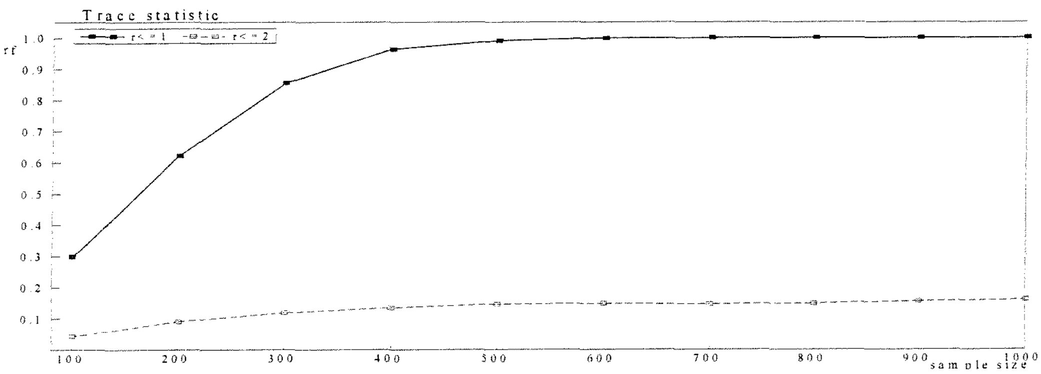

Figure 4.6. Frequency of rejecting the null hypotheses $r \leq 1$ (true) and $r \leq 2$ using the trace statistic.

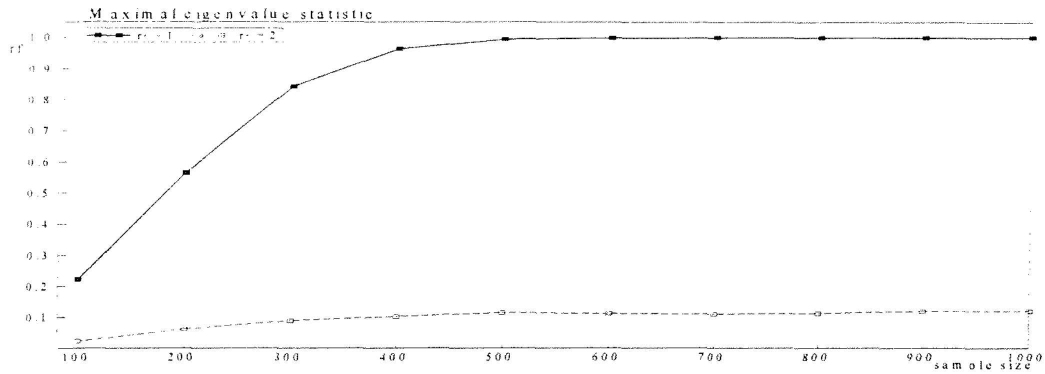

Figure 4.7. Frequency of rejecting the null hypotheses $r \leq 1$ (true) and $r \leq 2$ using the maximal eigenvalue statistic.

4.6 A generalisation

In this section we show that the result concerning the overestimation of the cointegrating rank follows unchanged when we allow for a more general model. However, in the general form of the model presented in this section it is not obvious how the weak exogeneity status of the variables affects the results about the overfit of the cointegrating rank.

A generalised version of (4.1) would require changing the original parameter Φ, $\Phi_o = \begin{bmatrix} I_{p_1} & 0 \\ 0 & \varphi \end{bmatrix}$ say, into $\Phi_n = \begin{bmatrix} \underset{p\times p_1}{\tilde{\Phi}} & \underset{p\times 1}{\tilde{\varphi}} \end{bmatrix}$, with $rank(\tilde{\Phi}) = p_1$ to maintain the assumption of distinct intercept shifts.

Let $M = \begin{bmatrix} \underset{p\times p_1}{\tilde{\Phi}} & \underset{p\times(p-p_1)}{\tilde{\Phi}_\perp} \end{bmatrix}$ be a $p \times p$ full rank matrix with $\tilde{\Phi}'\tilde{\Phi}_\perp = 0$, then pre-multiplying (4.1) by M, we get the generalised specification,

$$\Delta X_t^* = \alpha^* \beta^{*'} X_{t-1}^* + \Phi_n D_t + \varepsilon_t^* \tag{4.38}$$

where $X_t^* = MX_t$, $\alpha^* = M\alpha$, $\beta^* = M^{-1'}\beta$, $\Phi_n = M\Phi_o = \begin{bmatrix} \tilde{\Phi} & \tilde{\Phi}_\perp\varphi \end{bmatrix}$ and $\varepsilon_t^* = M\varepsilon_t$ with mean zero and variance $M\Omega M'$.

The representations given below follow from the Granger Representation Theorem,

$$X_t^* = C^* \sum_{i=1}^{t} (\varepsilon_i^* + \Phi_n D_i) + C_1^*(L)(\varepsilon_t^* + \Phi_n D_t) + A^* \tag{4.39}$$

$$\Delta X_t^* = C^*(L)(\varepsilon_t^* + \Phi_n D_t). \tag{4.40}$$

where $C_1^*(L) = MC_1(L)M^{-1}$, $C^*(L) = MC(L)M^{-1}$, $C^*(1) = C^* = MCM^{-1}$, $A^* = MA$ and therefore $\beta^{*'}C^* = \beta' CM^{-1} = 0$, $\beta^{*'}A^* = \beta' A = 0$.

Let $\tau^* = C^*\Phi_n$ and γ^* be chosen such that $(\beta^*, \gamma^*, \tau^*)$ are mutually orthogonal and span $\mathbb{R}^p$, then the preliminary results given in section 4.1 can be derived in the same way using (4.39) and (4.40), and can be restated as follows:

$$S_{00}^* \xrightarrow{p} M\Sigma_{00}^* M' \tag{4.41}$$

$$\beta^{*'} S_{11}^* \beta^* \xrightarrow{p} \Sigma_{\beta\beta}^* \tag{4.42}$$

$$\beta^{*'} S_{10}^* \xrightarrow{p} \Sigma_{\beta 0}^* M' \tag{4.43}$$

$$T^{-1} B_T^{*'} S_{11}^* B_T^* \xrightarrow{d} \int_0^1 G^* G^{*'} du \tag{4.44}$$

where $B_T^* = (\bar{\gamma}^*, T^{-1/2}\bar{\tau}^*)$, $\bar{\gamma}^* = \gamma^*(\gamma^{*'}\gamma^*)^{-1}$, $\bar{\tau}^* = \tau^*(\tau^{*'}\tau^*)^{-1}$, $G^* = G_0^* - \int_0^1 G_0^* du$, $G_0^* = \begin{bmatrix} \bar{\gamma}^{*'} C^* W^*(u) \\ Z(u) \\ u \end{bmatrix}$ and $W^*(u) = MW(u)$;

$$T^{-1/2} B_T^{*'} S_{11}^* \beta^* \xrightarrow{d} V^* \tilde{\Phi}' C_1^*(1)' \beta^* \tag{4.45}$$

$$T^{-1/2} B_T^{*'} S_{10}^* \xrightarrow{d} V^* \tilde{\Phi}' C^{*'} \tag{4.46}$$

where $V^* = \int_0^1 G_0^* z(u)' du$.

S_{ij}, $i,j = 0,1$ are the residual product moment matrices calculated from (4.38) and Σ^*_{ij}, $i,j = 0,\beta$ are just the expressions defined in section 4.1.

Let $A^*_T = (\beta^*, T^{-1/2}B^*_T)$ and $S^*(\zeta) = \zeta S^*_{11} - S^*_{10}S^{*-1}_{00}S^*_{01}$, then the limit of the scaled form of the eigenvalue equation $|A^{*'}_T S^*(\zeta)A^*_T| = 0$ is given by

$$|A^{*'}_T S^*(\zeta)A^*_T| \xrightarrow{d} |\zeta M^*_1 - M^*_2| = 0$$

where $M^*_1 = \begin{bmatrix} \Sigma^*_{\beta\beta} & \beta^{*'}C^*_1(1)\tilde{\Phi}V^{*'} \\ V^*\tilde{\Phi}'C^*_1(1)'\beta^* & \int_0^1 G^*G^{*'}du \end{bmatrix}$ and

$$M^*_2 = \begin{bmatrix} \Sigma^*_{\beta 0}\Sigma^{*-1}_{00}\Sigma^*_{0\beta} & \Sigma^*_{\beta 0}\Sigma^{*-1}_{00}M^{-1}C^*\tilde{\Phi}V^{*'} \\ V^*\tilde{\Phi}'C^{*'}M'^{-1}\Sigma^{*-1}_{00}\Sigma^*_{0\beta} & V^*\tilde{\Phi}'C^{*'}(M\Sigma^*_{00}M')^{-1}C^*\tilde{\Phi}V^{*'} \end{bmatrix}.$$

As M^*_1 and M^*_2 are symmetric matrices and M^*_1 is non-singular

$$|\zeta M^*_1 - M^*_2| = |\zeta I_p - F^*| = 0$$

where $F^* = M^{*-1/2}_1 M^*_2 M^{*-1/2}_1$. M^*_2 can be decomposed into $M^*_2 = Q^*D^*Q^{*'}$ where

$$Q^* = \begin{bmatrix} I_r & 0 \\ -V^*\tilde{\Phi}'C^{*'}M'\Sigma^{*-1}_{00}\Sigma^*_{0\beta}(\Sigma^*_{\beta 0}\Sigma^{*-1}_{00}\Sigma^*_{0\beta})^{-1} & I_{p-r} \end{bmatrix} \text{ and}$$

$$D^* = \begin{bmatrix} \Sigma^*_{\beta 0}\Sigma^{*-1}_{00}\Sigma^*_{0\beta} & 0 \\ 0 & V^*\tilde{\Phi}'C^{*'}M'^{-1}N^*M^{-1}C^*\tilde{\Phi}V^{*'} \end{bmatrix}.$$

The rank of F^* gives the number of non-zero eigenvalues in the limit and

$$rank(F^*) = rank(M^*_2) = rank(D^*)$$

$$= rank(\Sigma^*_{\beta 0}\Sigma^{*-1}_{00}\Sigma^*_{0\beta}) + rank(V^*\tilde{\Phi}'C^{*'}M'^{-1}N^*M^{-1}C^*\tilde{\Phi}V^{*'}).$$

Decomposing N^*, as in section 4.2, into $N^* = P^*P^{*'}$, where P^* is a $p \times (p-r)$ matrix of rank $(p-r)$, we can find an upper bound for the number of spurious cointegrating relations that is given by

$$rank(V^*\tilde{\Phi}'M^{-1}C'N^*CM^{-1}\tilde{\Phi}V^{*'}) = rank(V^*\tilde{\Phi}'M^{-1}C'P^*) \leq$$

$$\min[rank(V^*\tilde{\Phi}'),\ p-r] = \min(p_1,\ p-r) = p_1$$

as in section 4.2, because V^* is $(p-r) \times p_1$, $\tilde{\Phi}$ is $p \times p_1$ and has rank p_1, and of the assumption $(p-r-q) > 0$ which assures the existence of the direction γ^*.

The key expression in this case is $V^*\tilde{\Phi}'$ and because of the presence of the matrix $\tilde{\Phi}$ the exogeneity status of the variables does not seem to affect the upper bound for the number of spurious cointegrating vectors. When $V^*\tilde{\Phi}'C^{*'} = 0$ spurious cointegration does not occur and this happens if $\alpha_{\perp}^{*'}\tilde{\Phi} = 0$ i.e. when the cumulative shift does not enter the level of the process (see (4.39)).

When m of the intercept shifts occur at the same date, there are $p_1 - (m-1)$ distinct intercept shifts and $\tilde{\Phi}$ does not have full column rank, in fact $rank(\tilde{\Phi}) = (p_1 - m + 1)$ and then the upper bound becomes $(p_1 - m + 1)$ i.e. the number of distinct shifts in the model (see also section 4.5).

4.7 Co-breaking

Co-breaking refers to the elimination of deterministic shifts using linear combinations of variables, either at the same point in time (contemporaneous co-breaking) or at different points in time (intertemporal co-breaking). Co-breaking is defined for processes with well-defined unconditional expectations.

Let $\{Y_t\}$, $t = 1, 2, \ldots$, be a p-dimensional stochastic process, whose unconditional expectation around an initial parameter ψ at $t = 0$ is given by

$$E(Y_t - \psi) = \mu_t,\ |\mu_t| < \infty$$

then we have the following definitions (Clements and Hendry (1999, pp. 249-252)):

Definition 4.1. The $p \times s$ *matrix* F *of rank* s $(p > s > 0)$ *is said to be contemporaneous mean co-breaking of order* s *for* $\{Y_t\}$ *if* $F'\mu_t = 0$, $t = 1, 2, \ldots, T$.

Definition 4.2. The $p \times s$ *polynomial matrix* $F(L) = \sum_{i=0}^{m} F_i L^i$ *of degree* $m > 0$ *with* $rank[(F_0' \, F_1' \ldots F_{m-1}' \, F_m')] = s$, $(p \geq s > 0)$ *is said to be intertemporal mean co-breaking of order* s *for* $\{Y_t\}$ *if* $F(L)'\mu_t = 0$, $t = 1, 2, \ldots, T$ *and no* $p \times s$ *matrix polynomial of degree* $(m-1)$ *and rank* s *annihilates* μ_t, $t = 1, 2, \ldots, T$.

When a process is non-stationary co-breaking can be considered in terms of functions of the process which have well-defined unconditional expectations such as the first differences or the cointegrating relations.

In order to consider co-breaking we use the general form of (4.1) given by (4.38), where the deterministic term of each equation of the VAR model involves a linear combination of the step dummy variables and a constant.

Taking the expectations of the stationary components in (4.39) and (4.40) we obtain

$$E(\beta^{*'} X_t^*) = \beta^{*'} C_1^*(L)\Phi_n D_t = \beta^{*'}(C_1^*(L)\tilde{\Phi} z_t + C_1^*(1)\tilde{\varphi}) = \mu_t^{\beta}$$

$$E(\Delta X_t^*) = C^*(L)\Phi_n D_t = C^*(L)\tilde{\Phi} z_t + C^* \tilde{\varphi} = \mu_t^{\Delta}.$$

Below we examine whether the transformed cointegrating vectors β^* are co-breaking for μ_t^{Δ}, i.e. whether they eliminate changes in μ_t^{Δ}.

(4.38) in mean deviation form becomes

$$\Delta X_t^* = \mu_t^{\Delta} + \alpha^*(\beta^{*'} X_{t-1}^* - \mu_{t-1}^{\beta}) + \varepsilon_t^*. \tag{4.47}$$

Pre-multiplying (4.47) by $\beta^{*'}$ and taking expectations we have

$$E(\beta^{*'}\Delta X_t) = \beta^{*'}\mu_t^{\Delta} \text{ or}$$

$$\Delta\mu_t^{\beta} = \beta^{*'}\mu_t^{\Delta}.$$

If $\Delta\mu_t^{\beta} = \beta^{*'}C_1^*(L)\tilde{\Phi}\Delta z_t = 0$, for all t then the mean of the cointegrating relations does not change and the transformed cointegrating vectors, β^* are co-breaking for ΔX_t. However, in general $\beta^{*'}C_1^*(L)\tilde{\Phi}\Delta z_t$ is not zero for all t, regardless of whether the intercept shifts occur at different or at the same date. Δz_t is not equal to zero for all values of t and $\beta^{*'}C_1^*(1) \neq 0$ (since $\beta^{*'}X_t^*$ is $I(0)$, see Definition 2.1). Thus, the transformed cointegrating vectors, β^*, do not seem to induce contemporaneous co-breaking for the impact of intercept shifts on ΔX_t.

For example when $p_1 = 2$ and the intercept shifts occur at two different dates, t_{01} and t_{02}, we have for $t = 2, \ldots, T$

$$\Delta z_{t-k} = \begin{cases} \begin{bmatrix} \delta_1 \\ 0 \end{bmatrix}, t = t_{01} + k + 1 \\ \begin{bmatrix} 0 \\ \delta_2 \end{bmatrix}, t = t_{02} + k + 1 \\ 0, \text{otherwise} \end{cases},$$

and therefore

$$\beta^{*'}C_1^*(L)\tilde{\Phi}\Delta z_t = \begin{cases} \beta^{*'}C_1^*(1)\tilde{\Phi}\begin{bmatrix} \delta_1 \\ 0 \end{bmatrix}, t = t_{01} + k + 1 \\ \beta^{*'}C_1^*(1)\tilde{\Phi}\begin{bmatrix} 0 \\ \delta_2 \end{bmatrix}, t = t_{02} + k + 1 \\ 0, \text{otherwise} \end{cases}.$$

If on the other hand the intercept shifts occur at the same date, t_0,

$$\Delta z_{t-k} = \begin{cases} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}, t = t_0 + k + 1 \\ 0, \text{otherwise} \end{cases}$$

and

$$\beta^{*'}C_1^*(L)\tilde{\Phi}\Delta z_t = \begin{cases} \beta^{*'}C_1^*(1)\tilde{\Phi}\begin{bmatrix}\delta_1\\ \delta_2\end{bmatrix}, t = t_0 + k + 1 \\ 0, \text{otherwise} \end{cases}.$$

Nevertheless, the matrix $\alpha_{\perp}^*$ (such that $\alpha_{\perp}^{*'}\alpha^* = 0$) which characterises common trends is co-breaking (contemporaneously) for changes in the mean of the cointegrating relations but the effects of intercept shifts are not eliminated,

$$\alpha_{\perp}^{*'}\Delta X_t^* = \alpha_{\perp}^{*'}\mu_t^{\Delta} + \alpha_{\perp}^{*'}\varepsilon_t^*$$

since the transformed system still depends on the intercept shifts affecting the mean of ΔX_t^*. Note that $\alpha_{\perp}^{*'}$ eliminates both shifts in the mean of the cointegrating relations and the cointegrating relations themselves.

Next we show that the cointegrating vectors in the transformed model (4.38) can induce intertemporal co-breaking for changes in the mean of the cointegrating relations under certain restrictions.

After pre-multiplying (4.38) by $\beta^{*'}$ and taking expectations we have,

$$F(L)'E(\beta^{*'}X_t) = \beta^{*'}\Phi_n D_t \quad \text{or}$$

$$F(L)'\mu_t^{\beta} = \beta^{*'}\Phi_n D_t$$

where $F(L)' = I_r - (I_r + \beta^{*'}\alpha^*)L$. Thus, intertemporal co-breaking of order r occurs if $\beta^{*'}\Phi_n = 0$ so that $F(L)'\mu_t^{\beta} = 0$ and $F(L)$ is a matrix polynomial of degree one.

Below we give an example in which the cointegrating vectors are co-breaking in the intertemporal sense when there is a common shift.

Let $p = 2$, $\Phi = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$, $D_t = z_t$, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ and $\beta_1 = \phi_2$, $\beta_2 = -\phi_1$ such that $\beta^{'}\Phi = 0$ then the ECM takes the form

$$\Delta X_t = \alpha\beta^{'}X_{t-1} + \Phi z_t + \varepsilon_t. \tag{4.48}$$

Pre-multiplying (4.48) by $\beta^{'}$ and taking expectations we have

$$[1 - (1 + \beta^{'}\alpha)L]E(\beta^{'}X_t) = 0 \text{ or}$$

$$[1 - (1 + \beta^{'}\alpha)L]\mu_t^{\beta} = 0. \tag{4.49}$$

(4.49) can be written in the form

$$F(L)\mu_t^{\beta} = 0 \tag{4.50}$$

where $F(L) = 1 - (1 + \beta^{'}\alpha)L$. (4.50) coincides with the definition of intertemporal co-breaking of order 1, where the matrix polynomial $F(L)$ is of degree one.

4.8 A digression: alternative specifications of the deterministic term

Consider again the model (DGP) in error correction form

$$\Delta X_t = \alpha\beta^{'}X_{t-1} + \Phi D_t + \varepsilon_t, \ \ t = 1, 2, \ldots, T \tag{4.51}$$

with all its components except ΦD_t defined as in section 4.1.

Since the asymptotic distribution of the LR tests for cointegration depends on the deterministic terms in the model, we examine how alternative specifications of the deterministic term, ΦD_t, affect the analysis in the presence of intercept shifts. There are many cases regarding the deterministic terms depending on whether the constant and/or the linear trend in the DGP lie in the cointegrating space (see Johansen (1996, p. 81)) and whether the SM coincides with the DGP. Below we analyse three cases, which are by no means exhaustive. These cases are:

Case (i): unrestricted constant in the SM but the linear trend is absent from the level of the process, X_t in the DGP.

Case (ii): constant restricted to lie in the cointegrating space in the DGP and SM.

Case (iii): constant and step dummy variables restricted to lie in the cointegrating space in the DGP and restricted constant in the SM.

Given the different directions in which the process X_t behaves differently, and the representation of X_t and ΔX_t given by the Granger Representation Theorem, the asymptotic results for the three cases can be derived. The proofs parallel those for the main case analysed in sections 4.1, 4.2 and in the Appendices A and B, so detailed derivations are omitted.

Case (i)

We assume that $C_2\varphi = 0$, where $C = \begin{bmatrix} \underset{p\times p_1}{C_1} & \underset{p\times p_2}{C_2} \end{bmatrix}$; $\Phi = \begin{bmatrix} I_{p_1} & 0 \\ 0 & \underset{p_2\times 1}{\varphi} \end{bmatrix}$ and $D_t = \begin{bmatrix} \underset{p_1\times 1}{z_t} \\ 1 \end{bmatrix}$.

In this case the SM is estimated as in the case where the constant is unrestricted (see section 4.1). The linear trend is absent from the level of the process X_t. It follows from the Granger Representation Theorem that

$$X_t = C\sum_{i=1}^{t}\varepsilon_i + C_1 Z_t + C_1(L)(\varepsilon_t + \Phi D_t) + A$$

and

$$\Delta X_t = C(L)(\varepsilon_t + \Phi D_t) = C(L)\varepsilon_t + C_1 z_t + C_1(L)(1-L)\Phi D_t.$$

The asymptotic properties of the process are considered in three different directions; the stationary given by β, $p \times r$, the non-stationary that annihilates the deterministic terms given by γ, $p \times (p-r-p_1)$ and the direction that annihilates the stochastic trends, given by $\tau = C_1$, $p \times p_1$. (β, γ, τ) span $\mathbb{R}^p$. The residual product moment matrices, S_{ij}, $i, j = 0, 1$ are defined as in section 4.1 and $B_T = (\bar{\gamma}, T^{-1/2}\bar{\tau})$, with $\bar{\gamma} = \gamma(\gamma'\gamma)^{-1}$ and $\bar{\tau} = \tau(\tau'\tau)^{-1}$.

Then (4.17)-(4.19) follow unchanged. We now have

$$T^{-1}B_T' S_{11} B_T \xrightarrow{d} \int_0^1 GG' du$$

where $G = \begin{bmatrix} \bar{\gamma}' C(W(u) - \bar{W}) \\ Z(u) - \bar{Z} \end{bmatrix}$;

$$T^{-1/2}B_T' S_{11}\beta \xrightarrow{d} \int_0^1 G_0 z(u)' du C_1(1)'\beta$$

where $G_0 = \begin{bmatrix} \bar{\gamma}' CW(u) \\ Z(u) \end{bmatrix}$ and

$$T^{-1/2}B_T' S_{10} \xrightarrow{d} \int_0^1 G_0 z(u)' du C_1'. \tag{4.52}$$

Adapting Proposition 4.1 we find that the rank of the limiting matrix in (4.52), $rank(V_{(i)}C_1')$ $= b$, say where $b \leq p_1$ and $V_{(i)} = \int_0^1 G_0 z(u)' du$ a $(p-r) \times p_1$ matrix, gives an upper bound for the number of spurious cointegrating relationships as $T \to \infty$.

Case (ii)

We assume that $\Phi = \begin{bmatrix} \underset{p \times p_1}{\tilde{\Phi}} & \underset{p \times 1}{\varphi} \end{bmatrix}$, $\varphi = \alpha\varphi_0$ so that (4.51) can be written as

$$\Delta X_t = \alpha\beta^{*\prime} X_{t-1}^* + \tilde{\Phi} z_t + \varepsilon_t$$

where $\beta^{*'} = \begin{bmatrix} \beta' & \varphi_0 \end{bmatrix}$ and $X^{*'}_{t-1} = \begin{bmatrix} X'_{t-1} & 1 \end{bmatrix}$. Then,

$$X_t = C(\sum_{i=1}^{t} \varepsilon_i + \tilde{\Phi} Z_t) + C_1(L)(\varepsilon_t + \Phi D_t) + A$$

and

$$\Delta X_t = C(L)(\varepsilon_t + \Phi D_t) = C(L)(\varepsilon_t + \tilde{\Phi} z_t),$$

where $D'_t = \begin{bmatrix} z'_t & 1 \end{bmatrix}$. In this case the eigenvalues used in the LR tests are obtained by solving

$$|\zeta^* S^*_{11} - S^*_{10} S^{*-1}_{00} S^*_{01}| = 0$$

where S^*_{ij}, $i, j = 0, 1$ are the product moment matrices of ΔX_t and $(X'_{t-1}, 1)$.

In deriving the asymptotic results we consider the behaviour of the process X_t in three directions. The stationary direction given by $\beta^+ = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$, $(p+1) \times r$, the non-stationary with the restricted constant given by $\gamma^+ = \begin{bmatrix} \bar{\gamma} & 0 \\ 0 & T^{1/2} \end{bmatrix}$, $(p+1) \times (p - r - p_1 + 1)$ and the direction where the step dummy variables dominate given by $\tau^+ = \begin{bmatrix} \bar{\tau} \\ 0 \end{bmatrix}$, $(p+1) \times p_1$. $\bar{\gamma}$ and $\bar{\tau}$ are $p \times (p - r - p_1)$ and $p \times p_1$ respectively and mutually orthogonal. In this setup the asymptotic results are modified as follows,

$$S^*_{00} \xrightarrow{p} \Sigma_{00} + C\Phi R \Phi' C'$$

$$\beta^{+'} S^*_{10} \xrightarrow{p} \Sigma_{\beta 0} + \beta' C_1(1) \Phi \bar{R} \Phi' C'$$

$$\beta^{+'} S^*_{11} \beta^+ \xrightarrow{p} \Sigma_{\beta\beta} + \beta' C_1(1) \Phi \bar{R} \Phi' C_1(1)' \beta$$

where $R = \begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix}$ as before, $\bar{R} = \begin{bmatrix} g & 0 \\ 0 & 1 \end{bmatrix}$, g is defined as in section 4.1 and Φ is given above. Let $B_T^+ = \begin{bmatrix} \gamma^+ & T^{-1/2}\tau^+ \end{bmatrix}$ then

$$T^{-1}B_T^{+'}S_{11}^*B_T^+ \xrightarrow{d} \int_0^1 G^*G^{*'}du$$

where $G^* = \begin{bmatrix} \bar{\gamma}'CW(u) \\ 1 \\ Z(u) \end{bmatrix}$;

$$T^{-1/2}B_T^{+'}S_{11}^*\beta^+ \xrightarrow{d} \int_0^1 \begin{bmatrix} \tilde{G}z(u)' & G^* \end{bmatrix} du\Phi'C_1(1)'\beta$$

where $\tilde{G} = \begin{bmatrix} \bar{\gamma}'CW(u) \\ 0 \\ Z(u) \end{bmatrix}$ and

$$T^{-1/2}B_T^{+'}S_{10}^* \xrightarrow{d} \int_0^1 \tilde{G}z(u)'du\tilde{\Phi}'C'. \tag{4.53}$$

The rank of the limiting matrix in (4.53) $rank(V_{(ii)}\tilde{\Phi}'C') = b$, where $V_{(ii)} = \int_0^1 \tilde{G}z(u)'du$ a $(p-r+1)\times p_1$, does not exceed $\min(p_1, p-r) = p_1$ (i.e. $b \leq p_1$) which is the upper bound for the number of spurious cointegrating vectors. Note that in this case the restriction for the existence of γ^+ direction is $(p-r) > (p_1 - 1)$, which implies $(p-r) \geq p_1$ therefore p_1 is still an upper bound.

Case (iii)

We assume that there are $k \leq r$ step dummy variables that lie, together with the constant term, in the cointegrating space. Thus, $\Phi = \alpha\Phi_0$, where Φ_0 is $r \times (k+1)$, $\Phi_0 = \begin{bmatrix} \underset{r\times k}{\tilde{\Phi}} & \underset{r\times 1}{\varphi_0} \end{bmatrix}$ and $D_t' = \begin{bmatrix} \underset{1\times k}{z_t'} & 1 \end{bmatrix}$. Then (4.51) can be expressed as

$$\Delta X_t = \alpha\beta^{*'}X_{t-1}^* + \varepsilon_t$$

where $\beta^{*'} = \left[\begin{array}{ccc} \beta' & \tilde{\Phi} & \varphi_0 \end{array} \right]$ and $X^{*'}_{t-1} = \left[\begin{array}{ccc} X'_{t-1} & z'_t & 1 \end{array} \right]$. From the Granger Representation Theorem we obtain

$$X_t = C \sum_{i=1}^{t} \varepsilon_i + C_1(L)(\varepsilon_t + \Phi D_t) + A$$

and

$$\Delta X_t = C(L)(\varepsilon_t + \Phi D_t) = C(L)\varepsilon_t + C_1(L)(1 - L)\alpha\tilde{\Phi} z_t.$$

The residual product moment matrices S^*_{ij}, $i, j = 0, 1$ are computed as in the restricted constant case (case (ii)), assuming that a restricted constant is included in the SM but the presence of shifts is (again) ignored. The properties of the process X_t differ in the direction $\beta^+ = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$ and $B_T = \begin{bmatrix} \bar{\beta}_\perp & 0 \\ 0 & T^{1/2} \end{bmatrix}$ (where $\bar{\beta}_\perp = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}$) which correspond to the stationary and non-stationary (stochastic trends) directions respectively.

The two problematic terms $B'_T S^*_{11}\beta$ and $B'_T S^*_{10}$ (that need to be rescaled in the presence of intercept shifts) are $O_p(1)$ in this case (as in the standard case i.e. no intercept shifts), thus if scaled by $T^{-1/2}$ they converge in probability to zero. This is due to the fact that the cumulative step dummy variables do not appear in the representation of the level of X_t, because of the way Φ is defined. Only the variances/covariances of the stationary components seem to be affected by the presence of intercept shifts, but not the inference for the cointegrating rank. In this case the impact of the intercept shifts is only in the form of 'smoothed' dummy variables, which enter the representation of X_t as an infinite-lag polynomial for z_t.

To sum up, the scaled version of the matrix, which has the same eigenvalues as the roots of (4.23), i.e. the eigenvalues used in the LR tests, for the case analysed in sections

4.1 and 4.2 and case (i) of this section is

$$(T^{-1}B_T^{'}S_{11}B_T)^{-1}(T^{-1/2}B_T^{'}S_{10})S_{00}^{-1}(S_{01}B_TT^{-1/2})$$

$$= (B_T^{'}S_{11}B_T)^{-1}(B_T^{'}S_{10})S_{00}^{-1}(S_{01}B_T) = O_p(1)$$

(with B_T appropriately defined for each case) and for case (ii)

$$(T^{-1}B_T^{+'}S_{11}^{*}B_T^{+})^{-1}(T^{-1/2}B_T^{+'}S_{10}^{*})S_{00}^{-1}(S_{01}^{*}B_T^{+}T^{-1/2})$$

$$= (B_T^{+'}S_{11}^{*}B_T^{+})^{-1}(B_T^{+'}S_{10}^{*})S_{00}^{-1}(S_{01}^{*}B_T^{+}) = O_p(1)$$

instead of $O_p(T^{-1})$ as in the absence of intercept shifts. For case (iii) we have

$$(T^{-1}B_T^{'}S_{11}^{*}B_T)^{-1}(B_T^{'}S_{10}^{*})S_{00}^{-1}(S_{01}^{*}B_T)$$

$$T(B_T^{'}S_{11}^{*}B_T)^{-1}(B_T^{'}S_{10}^{*})S_{00}^{-1}(S_{01}^{*}B_T) = O_p(1)$$

therefore

$$(B_T^{'}S_{11}^{*}B_T)^{-1}(B_T^{'}S_{10}^{*})S_{00}^{-1}(S_{01}^{*}B_T) = O_p(T^{-1}).$$

Thus, applying the trace or the maximal eigenvalue statistic, which require scaling by T, in the case analysed in sections 4.1 and 4.2 and cases (i) and (ii) will lead to rejection of the null hypothesis of cointegrating rank r with probability one as $T \to \infty$. On the other hand, intercept shifts that lie in the cointegrating space but are omitted from the SM will not affect the inference about the cointegrating rank.

4.9 Concluding remarks

This chapter has considered the effects of intercept shifts on the trace and maximal eigenvalue statistics used for cointegration testing. It was shown that when step dummy variables, which capture the impacts of intercept shifts, are present in the DGP but not in the statistical model used for cointegration testing, these statistics reject the null hypothesis of r cointegrating vectors with probability one as $T \to \infty$. As a result, the cointegrating rank is overestimated. The extent of the overestimation depends on the number of distinct intercept shifts in the DGP and on the weak exogeneity status of the variables.

The model under examination is quite simple, being a VAR(1). A restrictive assumption in the analysis is the one about the existence of the non-stationary stochastic direction (γ), given by $(p-r) > q$. When $(p-r) = q$ the asymptotic results do not involve stochastic terms (Brownian motions).

A possible extension of this investigation is the derivation of the asymptotic distribution of the test statistics considered, in the presence of intercept shifts. Although for the null case ($r = 0$) the procedure of deriving the asymptotic distribution is tractable (see O'Brien (1999)) this does not seem to be the case under the assumption of cointegration ($r > 0$).

However, it seems that under certain circumstances, ignoring the presence of intercept shifts leads to misleading inference about the cointegrating rank. So a priori testing for the presence of shifts applied on the univariate representation of the processes involved in the VAR (see e.g. Perron (1989), Perron and Vogelsang (1992), Zivot and Andrews (1992)) and/or application of cointegration tests that allow for shifts in the mean of the process (see e.g. Gregory and Hansen (1996a,b), Inoue (1999), Johansen et al. (2000), Saikkonen and Lütkepohl (1998)) appear to be a 'safer' strategy to follow.

Chapter 5
LR tests for cointegration and intercept shifts: a finite sample analysis

In this chapter we use Monte Carlo simulations to investigate the finite sample performance of the LR tests for cointegration in the presence of intercept shifts, implemented using the trace or the maximal eigenvalue statistic. The investigation is carried out using alternative specifications for the constant term (absence of constant term, constant restricted to lie in the cointegrating space and unrestricted constant term) in the SM, in conjunction with alternative designs concerning the variables (shifted, non-shifted) entering the cointegrating vectors. The setup of the analysis allows for some degree of control over the local power of the tests. All simulation experiments were programmed in Ox 3.00 (see Doornik (1999)).

5.1 Local power

Since the aim of this chapter is the investigation of the finite sample performance of the LR tests for cointegration under misspecification, it is necessary to have some degree of control on the power of the tests so that the conclusions drawn will be conditional on a certain power level. The motivation for this is as follows. If the power of the tests for the cointegrating rank is low, and correct inferences are difficult, mistakes caused by spurious cointegration may be less important. In such situations inference will tend to be imprecise in any case. If, however, the tests for the cointegrating rank are very likely to detect the correct cointegrating rank in the absence of intercept shifts, but spurious cointegration appears with appreciable probability, then it may be taken more seriously, as distorting an otherwise clear picture. In the absence of control over the exact power, we approximate by endeavouring to control local asymptotic power.

Below we present the theoretical framework, which is the simplest possible one, without deterministic terms and short-term dynamics. For detailed treatments of the model without deterministic terms see Johansen (1991b; 1996, Chapter 14). Cases involving deterministic terms are analysed by Rahbek (1994) and Saikkonen and Lütkepohl (1999).

The model is given by

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \;\; t = 1, 2, \ldots, T$$

where $\varepsilon_t \sim i.i.d.(0, \Omega)$, $X_0 = 0$.

The null hypothesis is

$$H(r) : \Pi = \alpha\beta'$$

and the local alternative is

$$H_T(r, s) : \Pi_T = \alpha\beta' + T^{-1}\alpha_1\beta_1'$$

where α, β are $p \times r$ and α_1, β_1 are $p \times s$, so under the local alternative there are $(r + s)$ cointegrating vectors, s of which are attached adjustment coefficients whose magnitude is inversely proportional to the sample size and therefore small. Thus, the s cointegrating vectors cannot be easily detected by the cointegration tests.

Assuming that the eigenvalues of the matrix $(I_r + \beta^{'}\alpha)$ lie inside the unit circle i.e. the process X_t is $I(1)$ under $H(r)$, then the asymptotic distribution of the trace statistic[5] (2.18) (for the hypothesis $H(r) : \Pi = \alpha\beta^{'}$) under the local alternative $H_T(r,s) : \Pi_T = \alpha\beta^{'} + T^{-1}\alpha_1\beta_1^{'}$ is given by

$$tr\{\int_0^1 (dK)K^{'}(\int_0^1 KK^{'}du)^{-1}\int_0^1 K(dK)^{'}\} \tag{5.1}$$

where K is a $(p-r)$ Ornstein-Uhlenbeck process which is defined by the stochastic differential equation

$$-ab^{'}K(u) + dK(u) = dB(u), \;\; u \in [0,1]$$

or equivalently by

$$-ab^{'}\int_0^u K(s)ds + K(u) = B(u), \;\; u \in [0,1] \tag{5.2}$$

and $B(u)$ is a $(p-r)$-dimensional standard Brownian motion. The asymptotic distribution under the local alternative depends on the parameters of the model through

$$a = (\alpha_{\perp}^{'}\Omega\alpha_{\perp})^{-1/2}\alpha_{\perp}^{'}\alpha_1$$

and

$$b = (\alpha_{\perp}^{'}\Omega\alpha_{\perp})^{1/2}(\beta_{\perp}^{'}\alpha_{\perp})^{-1}\beta_{\perp}^{'}\beta_1$$

where $\alpha_{\perp}$ and $\beta_{\perp}$ are $p \times (p-r)$ matrices orthogonal to α and β respectively.

---

[5] For a discussion on the local power of the LR test based on the maximal eigenvalue statistic see Paruolo (2001).

When $s = 1$ so that α_1 and β_1 are $p \times 1$ vectors the asymptotic power function depends on

$$f = b^{'}a = \beta_1^{'}C\alpha_1 < 0$$

and

$$\begin{aligned} g^2 &= a^{'}ab^{'}b \\ &= (\alpha_1^{'}\alpha_\perp(\alpha_\perp^{'}\Omega\alpha_\perp)^{-1}\alpha_\perp^{'}\alpha_1)(\beta_1^{'}C\Omega C^{'}\beta_1) - (\beta_1^{'}C\alpha_1)^2 \end{aligned}$$

where $C = \beta_\perp(\alpha_\perp^{'}\beta_\perp)^{-1}\alpha_\perp^{'}$, furthermore when $(p-r) = 1$ the asymptotic power depends only on f, see Johansen (1991b, p. 327).

The process (5.2) can be decomposed into three orthogonal directions so that it depends only on $a^{'}a$, $b^{'}b$ and $b^{'}a$ (see equations (14.16)-(14.18) in Johansen (1996)) and for $s = 1$ these are given by

$$-f\int_0^u K_1(s)ds + K_1(u) = B_1(u) \tag{5.3}$$

$$-g\int_0^u K_1(s)ds + K_2(u) = B_2(u) \tag{5.4}$$

$$K_3(u) = B_3(u) \tag{5.5}$$

where the first two equations are one-dimensional and the third is $(p-r-2)$-dimensional.

The DGPs used in the Monte Carlo analysis in section 5.2 are four-dimensional VAR processes ($p = 4$) with one cointegrating vector. So we simulate the local power of the LR tests for $(p-r) = 4$ to investigate the probability with which the single stationary (or near-integrated) relation can be detected (by rejecting $r = 0$) for different values of f and g.

Since the main quantity in the expression for the asymptotic power (5.1) is the process K, it is required to simulate the discrete analogues of (5.3)-(5.5) for $(p-r)=4$,

$$K_{1t} = (1 + T^{-1}f)K_{1(t-1)} + u_{1t} \tag{5.6}$$

$$K_{2t} = K_{2(t-1)} + (T^{-1}g)K_{1(t-1)} + u_{2t} \tag{5.7}$$

$$K_{3t} = K_{3(t-1)} + u_{3t} \tag{5.8}$$

$$K_{4t} = K_{4(t-1)} + u_{4t} \tag{5.9}$$

$t = 1, 2, \ldots, T, T = 400$, $K_{i0} = 0$, $u_{it} \sim i.i.d.N(0,1)$, $i = 1, 2, 3, 4$.

We can then calculate

$$tr\{\sum_{t=1}^{T} \Delta K_t K_t^{*'} (\sum_{t=1}^{T} K_t^* K_t^{*'})^{-1} \sum_{t=1}^{T} K_t^* \Delta K_t^{'}\} \tag{5.10}$$

and

$$\max eig\{\sum_{t=1}^{T} \Delta K_t K_t^{*'} (\sum_{t=1}^{T} K_t^* K_t^{*'})^{-1} \sum_{t=1}^{T} K_t^* \Delta K_t^{'}\} \tag{5.11}$$

where $K_t^{'} = (K_{1t}, K_{2t}, K_{3t}, K_{4t})$ and $\max eig\{\cdot\}$ denotes the largest eigenvalue of the argument matrix. (5.10) and (5.11) were computed for three alternative specifications of the deterministic term in the DGP, which give rise to three alternative definitions for K_t^*:

(i) no deterministic term, where $K_t^* = K_{t-1}$

(ii) constant restricted to lie in the cointegrating space, where $K_t^{*'} = (K_{t-1}^{'} : 1)$

(iii) in the DGP the constant is restricted to lie in the cointegrating space but the SM allows for unrestricted constant, where $K_t^* = K_{t-1} - T^{-1} \sum_{t=1}^{T} K_{t-1}$.

The asymptotic power for different values of f and g is computed as the rejection frequency of the null hypothesis $H(0)$, by comparing (5.10) and (5.11) with the appropriate 95% critical values (under the null) given in Tables 0 and 1* in Osterwald-Lenum (1992), for cases (i) and (ii) respectively, and in Table A2 in Johansen and Juselius (1990) for case (iii). The number of replications is 5,000.

The tables for the simulated power function appear in Appendix D. For all cases the probability associated with $f = g = 0$ corresponds to the asymptotic size of the tests, which is 5%. For the trace test in case (i) (without deterministic terms) the power is lower for $(p-r) = 4$ compared to the cases where $(p-r) = 1, 2, 3$, which appear in Johansen (1991b, 1996). This is a manifestation of the results stated in Johansen (1991b, 1996) and Saikkonen and Lütkepohl (1999), namely that the power decreases as the number of common trends, $(p-r)$, increases, which makes it more difficult for the test to distinguish the near-integrated process from the integrated ones. It is also observed that for both test statistics the power is higher when there are no deterministic terms (in either the DGP or the SM). The same result was found by Saikkonen and Lütkepohl (1999) for the trace test and $(p-r) = 1, 2, 3$. Moreover, looking at the tabulated values of the local power, neither test seems to dominate uniformly the other (in terms of local power). The trace test appears to have higher local power for moderate values of f and g, whereas the maximal eigenvalue test tends to be more powerful (locally) for extreme values of f and g. This observation is in agreement with the findings of Paruolo (2001) for $(p-r) = 1, 2, 3$.

5.2 Monte Carlo experiments

The general form of the DGPs written in error correction form is given by

$$\Delta X_t = \alpha\beta^{*'} X^*_{t-1} + \Phi_0 z_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T \tag{5.12}$$

where X_t is a four-dimensional $I(1)$ process, $\alpha' = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \end{bmatrix}$, $\beta^{*'} = \begin{bmatrix} \beta' & \mu' \end{bmatrix}$, $\beta' = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{bmatrix}$, $X^{*'}_{t-1} = \begin{bmatrix} X'_{t-1} & 1 \end{bmatrix}$, $\Phi'_0 = \begin{bmatrix} \underset{2\times2}{I} & \underset{2\times2}{0} \end{bmatrix}$ and z_t is a 2×1 vector of de-meaned step dummy variables (they sum to zero over the sample period) used to model intercept shifts. For any arbitrary date, t_{0i}, $1 \leq t_{0i} \leq T$; $t_{0i} = [T\lambda_i]$, $\lambda_i \in (0, 1)$, a typical (i-th) element of z_t, z_{it} is defined by

$$z_{it} = \begin{cases} \delta_i(\lambda_i - 1), & 1 \leq t \leq t_{0i} \\ \delta_i\lambda_i, & t_{0i} + 1 \leq t \leq T \end{cases}.$$

$\varepsilon_{it} \sim i.i.d.N(0, 1)$, $i = 1, 2, 3, 4$, $t = 1, 2, \ldots, T$, where $T = 50, 100, 200$. So we consider four-variable models with two variables having equations in the ECM that contain intercept shifts and one cointegrating vector. The shifts $z'_t = \begin{bmatrix} z_{1t} & z_{2t} \end{bmatrix}$ are either at two different dates, $T/3$ and $2T/3$ (distinct shifts) with $\lambda_1 = 1/3$, $\lambda_2 = 2/3$ and $z_{1t} \neq z_{2t}$, or at a common date, $T/2$, with $\lambda_1 = \lambda_2 = 1/2$ and $z_{1t} = z_{2t}$. In addition the impact of the magnitude of the shift on the LR tests is also examined by allowing different values for δ_i; in particular $\delta_i = 0, 0.5, 0.6, 0.7, 0.85, 1$. These values were also used in studying the effects of intercept shifts in the null case in O'Brien (1996, 1997, 1999).

Since the local power of the LR tests for cases (ii) and (iii) (see section 5.1) is not affected by the actual value of μ (see Saikkonen and Lütkepohl (1999)) we set $\mu = 0$ in generating the data (in the DGPs). However, the value of μ has noticeable effect on the speed with which the small sample behaviour tends to the asymptotic distribution under the local alternatives as can be discovered by simulations with different values for μ (see O'Brien (2001)). It would be possible to investigate the small sample effects of varying μ, but this would substantially increase the number of experiments, and possibly also the range of T values to be considered, given the slower convergence to the asymptotic distribution associated with $\mu \neq 0$.

As far as the specification of the constant term in (5.12) is concerned μ is set to zero in all DGPs, as mentioned above. In the SM we employ the three cases described in section 5.1:

Case (i) no deterministic term; $\mu = 0$, $X_{t-1}^{*} = X_{t-1}$.

Case (ii) constant restricted to lie in the cointegrating space; $\alpha\mu' \neq 0$.

Case (iii) constant restricted to lie in the cointegrating space in the DGP (i.e. $\alpha\mu' \neq 0$) but the SM is estimated unrestrictedly.

Furthermore, we use three alternative designs which give rise to parametrisations which are compatible with those of the system in (5.6)-(5.9). Hence, the parameters of the DGPs can be expressed in terms of f and g (the parameters upon which the local power of the tests depends) and for given values of f and g a certain level of asymptotic local power can be attained. Accordingly we gain a certain degree of control over the asymptotic power and we can investigate the effects of the misspecification (intercept shifts in the DGP but not in the SM) for cases that the LR tests have low, medium and high local power.

Having set $\mu = 0$ so that

$$\Delta X_t = \alpha\beta' X_{t-1} + \Phi_0 z_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T \tag{5.13}$$

the three designs used are:

(D1) $\alpha_3 \neq 0$, $\alpha_4 \neq 0$, $\beta_3 = 1$, $\beta_4 = -1$, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0$

(D2) $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, $\beta_1 = 1$, $\beta_2 = -1$, $\alpha_3 = \alpha_4 = \beta_3 = \beta_4 = 0$

(D3) $\alpha_2 \neq 0$, $\alpha_3 \neq 0$, $\beta_2 = 1$, $\beta_3 = -1$, $\alpha_1 = \alpha_4 = \beta_1 = \beta_4 = 0$

and in matrix form:

(D1) $\alpha' = \begin{bmatrix} 0 & 0 & \alpha_3 & \alpha_4 \end{bmatrix}, \beta' = \begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}$

(D2) $\alpha' = \begin{bmatrix} \alpha_1 & \alpha_2 & 0 & 0 \end{bmatrix}, \beta' = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix}$

(D3) $\alpha' = \begin{bmatrix} 0 & \alpha_2 & \alpha_3 & 0 \end{bmatrix}, \beta' = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}$.

The cointegrating vector involves only non-shifted variables in (D1), only shifted variables in (D2) and a mixture of shifted and non-shifted variables in (D3). The only undetermined parameters in the designs are the adjustment coefficients and their relation with the parameters f and g is given in the following proposition.

Proposition 5.1. $\alpha_i = (f+g)/2T$ *and* $\alpha_j = (g-f)/2T$ *for* $i = 3, 1, 2$ *and* $j = 4, 2, 3$ *in* (D1), (D2), (D3) *respectively.*

The proof is given only for (D1) since it is similar for the rest of the designs.

Proof. We are interested in situations where inferences are clear in the absence of step dummy variables. Thus, we set $\Phi_0 = 0$ in (5.13) i.e.

$$\Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t, \ \ t = 1, 2, \ldots, T \tag{5.14}$$

or

$$X_t = (I_4 + \alpha\beta')X_{t-1} + \varepsilon_t, \ \ t = 1, 2, \ldots, T \tag{5.15}$$

in order to relate the parameters of (5.15) to those in (5.6)-(5.9).

Let $B = \begin{bmatrix} \beta & \beta_\perp \end{bmatrix}$, a $p \times p$ full rank matrix. Given β, a possible choice for $\beta_\perp$ is

$$\beta_\perp = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ so that } B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}.$$

By pre-multiplying (5.15) by B' we obtain

$$X_{3t} - X_{4t} = (1 + \alpha_3 - \alpha_4)(X_{3(t-1)} - X_{4(t-1)}) + (\varepsilon_{3t} - \varepsilon_{4t}) \tag{5.16}$$

$$X_{1t} = X_{1(t-1)} + \varepsilon_{1t} \tag{5.17}$$

$$X_{2t} = X_{2(t-1)} + \varepsilon_{2t} \tag{5.18}$$

$$X_{3t} + X_{4t} = X_{3(t-1)} + X_{4(t-1)} + (\alpha_3 + \alpha_4)(X_{3(t-1)} - X_{4(t-1)}) + (\varepsilon_{3t} + \varepsilon_{4t}). \tag{5.19}$$

Let $K_{1t} = X_{3t} - X_{4t}$, a stationary or near-integrated process, $K_{2t} = X_{3t} + X_{4t}$, $K_{3t} = X_{1t}$, $K_{4t} = X_{2t}$, $I(1)$ processes, by definition. Then the system (5.16)-(5.19) can be expressed as

$$K_{1t} = (1 + \alpha_3 - \alpha_4)K_{1(t-1)} + u_{1t} \tag{5.20}$$

$$K_{2t} = K_{2(t-1)} + (\alpha_3 + \alpha_4)K_{1(t-1)} + u_{2t} \tag{5.21}$$

$$K_{3t} = K_{3(t-1)} + u_{3t} \tag{5.22}$$

$$K_{4t} = K_{4(t-1)} + u_{4t} \tag{5.23}$$

where $u_{1t} = \varepsilon_{3t} - \varepsilon_{4t}$, $u_{2t} = \varepsilon_{3t} + \varepsilon_{4t}$, $u_{3t} = \varepsilon_{1t}$, $u_{4t} = \varepsilon_{2t}$. Comparing the coefficients in (5.20)-(5.23) to those in (5.6)-(5.9) we obtain,

$$\alpha_3 - \alpha_4 = T^{-1}f \tag{5.24}$$

$$\alpha_3 + \alpha_4 = T^{-1}g. \tag{5.25}$$

Proposition 5.1 follows by solving (5.24) and (5.25) simultaneously with respect to α_3 and α_4. ■

Note that the LR test statistics calculated from (5.14) and (5.20)-(5.23) are algebraically equivalent due to the invariance of the eigenvalues to linear transformations. The eigenvalues associated with (5.14) are calculated from

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0$$

(see Chapters 2 and 3), then the corresponding eigenvalue equation for (5.20)-(5.23) is

$$|\zeta B^{'}S_{11}B - B^{'}S_{10}B(B^{'}S_{00}B)^{-1}B^{'}S_{01}B| = |B^{'}||\zeta S_{11} - S_{10}S_{00}^{-1}S_{01}||B| = 0$$

and thus $\lambda = \zeta$.

An alternative way to express the parameters of (5.14) in terms of f and g is to apply the formulae for f and g given in section 5.1. Since we consider the case of just one extra cointegrating vector, i.e. $r = 0$ and $s = 1$, under the null $C = \Omega = I_4$, $\alpha_\perp$ and $\beta_\perp$ are $p \times p$ matrices with full rank and α_1, β_1 in section 5.1 correspond to α, β respectively used in this section. Then,

$$T^{-1}f = \beta^{'}\alpha = \alpha_3 - \alpha_4$$

$$(T^{-1}g)^2 = \alpha^{'}\alpha\beta^{'}\beta - (\beta^{'}\alpha)^2 = (\alpha_3 + \alpha_4)^2$$

which coincide with (5.24) and (5.25), using again α, β as defined in (D1).

In Johansen's (1996, equation 14.2) notation, the deviation from the null is $T^{-1}\alpha_1\beta_1^{'}$, corresponding to $\alpha\beta^{'}$ used in this section. Thus, $f = \beta_1^{'}\alpha_1$ in Johansen (1996), after simplification, corresponds to $T\beta^{'}\alpha$ as shown above. A similar adjustment is required for g.

Consequently the presence of the cointegrating vector can be detected with different probabilities, in other words with different local power, depending on the values taken by f and g. For each of the cases (i)-(iii) concerning the deterministic term, we use six pairs of (f, g) values, two pairs for each power level: low (about 0.2), medium (about 0.55) and high (about 0.85). The values of (f, g) are: $\{$(-21, 0), (-21, 6), (-18, 18), (-30, 12), (-15, 24), (-48, 6)$\}$ for case (i), $\{$(-15, 12), (-24, 6), (-27, 18), (-30, 18), (-42, 24), (-54, 6)$\}$ for case (ii) and $\{$(-9, 12), (-21, 6), (-24, 18), (-36, 6), (-9, 24), (-48, 12)$\}$ for case (iii). The exact value of the asymptotic local power which corresponds to each pair appears underlined in the appropriate table (Tables D.1-D.6) in Appendix D. For some pairs involving extreme values of f and g the power for the maximal eigenvalue test is slightly higher and for pairs with moderate values of f and g the trace test is slightly more powerful (see also section 5.1).

To verify that the designs under consideration conform with the chosen power levels, we computed the frequencies of rejecting the null of $r = 0$, given that the DGPs contain one cointegrating vector, with adjustment coefficients given by the relevant f and g values that are associated with the chosen power level. Setting $\delta_i = 0$ for this experiment, and using $T = 50, 100, 200$ and 10,000 replications the rejection frequencies seem to converge (though slowly in some cases) to the predetermined power levels indicated by the choice of f and g. The tables are omitted for the sake of brevity.

The condition for the presence of one cointegrating vector, i.e. the stability condition for the cointegrating relation (stationary process) is that the eigenvalues of the matrix $(I_r + \beta^{'}\alpha)$ lie inside the unit circle, which reduces to

$$-2 < \alpha_i - \alpha_j < 0$$

for $i = 3, 1, 2$ and $j = 4, 2, 3$ in (D1), (D2), (D3) respectively, or equivalently

$$-2 < T^{-1}f < 0 \tag{5.26}$$

for all designs. f is always negative, by definition (see section 5.1), and given the choices of f and T values, the stability condition (5.26) is satisfied.

Finally a word on the taxonomy of the Monte Carlo experiments. Each of the designs is used together with each of the three cases regarding the specification of the deterministic term, which generates nine experiments as shown in Table 5.1. The nine experiments shown in Table 5.1 were conducted both under the assumption of distinct and common shifts resulting in eighteen experiments in all.

Table 5.1. The taxonomy of the experiments

Design / Specification of deterministic term	D1	D2	D3
case (i)	D1(i)	D2(i)	D3(i)
case (ii)	D1(ii)	D2(ii)	D3(ii)
case (iii)	D1(iii)	D2(iii)	D3(iii)

5.3 Monte Carlo results

The results of the simulations are presented graphically in Appendix E. The graphs show the rejection frequency (abbreviated as rf on the vertical axis) of the null hypothesis of one cointegrating vector (detected with various power levels) for different magnitudes of the shift (δ) and different sample sizes, using the experiments shown in Table 5.1.

In each set of six graphs, the left hand column shows rejection frequencies for the trace test; the right hand column shows rejection frequencies for the maximal eigenvalue test. The three rows of graphs represent low, medium and high power respectively, as one works down. On each graph there are six 'curves', two for each of three sample sizes ($T = 50, 100, 200$). First we discuss Figures E.1-E.18.

The rejection frequencies increase as the sample size increases (which is in agreement with the asymptotic analysis) and as the magnitude of the shift grows larger. This pattern persists for Figures E.1-E.18.

The higher the power with which the cointegrating vector can be detected, the higher the rejection frequency of the true null. Thus, as one works down through a sextet of graphs, the rejection frequency curves rise. Again, the pattern persists for Figures E.1-E.18. So in the cases where genuine cointegration cannot be easily detected, spurious cointegration arises less frequently.

However, for many cases the frequency of rejecting the true null is approximately equal to or exceeds the probability of finding the true cointegrating vector (i.e. the empirical size exceeds the asymptotic local power). This occurs when the rejection frequency exceeds 0.2 in the top graphs in a sextet (Figures E.1, E.3, E.4, E.6, E.7, E.9, E.10, E.11, E.13, E.14, E.16, E.17, E.18), or 0.5 in the middle graphs (Figures E.1, E.4, E.6, E.7, E.9, E.10, E.11, E.13, E.16, E.17), or 0.8 in the bottom graphs (Figures E.4, E.7, E.9, E.10, E.11, E.13, E.14, E.16, E.17).

When there are two distinct intercept shifts in variables in the cointegrating vector, D2 (Figures E.2, E.5, E.8) yields lower rejection frequencies than D1 (Figures E.1, E.4, E.7), where the variables in the cointegrating vector do not have shifts and D3 (Figures E.3, E.6, E.9), the mixed case, especially at low and medium power levels. When there is a common shift D3 (Figures E.12, E.15, E.18) generates lower rejection frequencies.

For both the distinct and common shifts specifications the rejection frequencies appear to be higher when a constant term (restricted or unrestricted) is introduced in the model (compare Figures E.1-E.3 and E.10-E.12 with Figures E.4-E.9 and E.13-E.18).

For most cases the rejection frequencies for the same sample size evolve similarly along the δ-axis for a given power level. An apparent exception is the case of D2 for large T and medium/high power level (see Figures E.2, E.5, E.8, E.11, E.14, E.17).

In general, D1 seems to produce larger impact of intercept shifts on the rejection frequencies of the LR tests. This relates to Corollary 4.1 (Chapter 4) according to which overfit of the cointegrating rank does not occur when $\alpha_{\perp}^{(1)} = 0$, where $\alpha_{\perp}^{(1)}$ is the $p_1 \times (p-r)$ submatrix of $\alpha_{\perp}$ which corresponds to the variables with the intercept shifts.

Even though $\alpha_{\perp}$ is not uniquely defined a plausible choice (for the sake of argument) could be $\alpha_{\perp} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\alpha_2 \\ 0 & 0 & \alpha_1 \end{bmatrix}$ for D1, $\alpha_{\perp} = \begin{bmatrix} 0 & 0 & -\alpha_2 \\ 0 & 0 & \alpha_1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ for D2 and $\alpha_{\perp} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -\alpha_2 \\ 0 & 0 & \alpha_1 \\ 0 & 1 & 0 \end{bmatrix}$ for D3. Thus, for D1 $\alpha_{\perp}^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is clearly non-null whereas for D2 $\alpha_{\perp}^{(1)} \rightarrow 0$ and for D3 $\alpha_{\perp}^{(1)} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, as $T \rightarrow \infty$, by the definition of α_1 and α_2 (see Proposition 5.1). So the higher rejection frequencies associated with D1, as T becomes large, are somewhat expected since D1 is further from the argument of Corollary 4.1 than the other two designs.

Figures E.19-E.21 show the rejection frequencies of the trace and maximal eigenvalue statistics when $\alpha_{\perp}^{(1)} = 0$ (see also Corollary 4.1). The DGP for this experiment is D2 with $\alpha_2 = 0$, therefore the power levels are chosen for different values of f and g[6] such that $f = -g$. From the formulae for f and g^2 in section 5.1 (scaling appropriately by T^{-1}) we find that $T^{-1}f = \alpha_1 < 0$ and $(T^{-1}g)^2 = \alpha_1^2$ thus $T^{-1}g = -\alpha_1$ and therefore $f = -g$. So, the DGP includes only one variable with intercept shift at $T/2$, which is not weakly exogenous with respect to the cointegrating relation. The rejection frequencies appear to be close to the asymptotic size of the tests (5%) hence systematic acceptance of spurious cointegration does not seem to occur in this case. The Monte Carlo precision with 10,000 replications is $\pm 0.43\%$, so the tests seem marginally oversized for high power levels still at $T = 200$ (last row of the sextets in E.19-E.21).

In the absence of intercept shifts ($\delta_i = 0$) the rejection frequencies correspond to the empirical size, and for most cases they are reasonably close to 5%. Under the assumption $\delta_i = 0$ there is no distinction among designs so the results of the simulations for the different designs are combined for the purposes of Figures E.22 and E.23. Figure E.22 shows the empirical size for different power levels indicated by the numbers 1, 2 and 3, which correspond to low, medium and high power. The empirical size increases with asymptotic power. For the low power level tests are undersized and size corrected tests would give

[6] $f = -g = -12, -18, -24$; $f = -g = -12, -18, -30$; $f = -g = -12, -18, -30$ for cases (i), (ii) and (iii) respectively and each value of f $(= -g)$ corresponds to low, medium and high power (see Tables D.1-D.6).

more rejections. For the medium power level this effect is slight. For the high power level some cases are oversized and size corrected tests would give fewer rejections. This pattern is similar for the trace and the maximal eigenvalue statistic. Figure E.23 shows the empirical size for different sample sizes (T) indicated by the numbers 1, 2 and 3, which correspond to $T = 50$, $T = 100$ and $T = 200$. There is large variation of the empirical size for a given sample size. Even for $T = 200$ the concentration around 5% seems to be low. Again the behaviour of the trace and maximal eigenvalue tests is similar. The scatter diagrams also indicate that the power level and the deterministic term specification dominate the effect of the sample size.

For comparison with O'Brien (1999), Figures E.24-E.25 and E.26-E.27 show the rejection frequency for alternative designs and power levels when there are two different shifts (at $T/3$ and $2T/3$) and a common shift (at $T/2$) respectively. The sample size is 150 which is representative of many econometric applications and the magnitude of the shift is set to 0.5, since a shift of that size can be easily missed in empirical work and not accounted for in the SM. For the majority of cases the frequency of rejecting the true null of one cointegrating vector is above 5% (asymptotic size) regardless of whether the effect of design or power is examined. Particularly there are situations that the rejection frequency exceeds 30%. As noticed before D2 seems to produce lower rejection frequencies when the shifts are distinct and D3 gives lower rejection frequency when there is a common shift in the DGP. In addition we observe that the higher the power level the higher the rejection frequency, as before.

5.4 Concluding remarks

In this chapter we have investigated the finite sample performance of the LR tests for cointegration proposed by Johansen (1988, 1996) when the DGP includes intercept shifts which are not accounted for in the SM used for cointegration testing. The effects of the misspecification are analysed for different levels of local power and different experimental designs with regard to which variables (shifted/non-shifted) enter the cointegrating vector. As the asymptotic analysis predicts (see Chapter 4) the frequency of rejecting the true null hypothesis of one cointegrating vector increases as the sample size becomes larger; therefore we spuriously accept more cointegrating relations than in the DGP. In addition, it is found that the true null hypothesis is rejected more frequently as the magnitude of the shift increases. These patterns arise in the presence of both a common shift and distinct shifts.

Since we opted to have some degree of control on the asymptotic power the analysis is carried out in a rather simplified setup. In considering the local asymptotic power we assume that $r = 0$ under the null and that $s = 1$ i.e. there is a single cointegrating vector, under the local alternative. However, assuming $r = 1$ under the null and carrying out the investigation for $s = 2$ under local alternatives complicates the parametrisation, because $\alpha_\perp$ and $\beta_\perp$ are no longer invertible matrices and C is no longer the identity matrix. Moreover, controlling for the local power in the case of more than one extra cointegrating vector under the alternative ($s > 1$) also complicates the analysis because the parameters upon which the power depends are no longer scalars.

Overall it appears that the intercept shifts have rather noticeable effect on the LR tests for cointegration. For example, for $T = 200$, $\delta = 0.5$ and medium local power, in cases (ii) and (iii) (which are model specifications frequently used in empirical works) the rejection frequencies of the true null are 30%-36% and 21%-29% for the common shift and distinct shifts cases respectively.

Chapter 6
Irrelevant or omitted variables in cointegration analysis

This chapter examines the impact of including irrelevant $I(1)$ variables in, or omitting relevant $I(1)$ variables from the SM used for cointegration analysis, on the inference about the cointegrating rank and the consistency of the estimators of the parameters of the ECM. The analytical findings are supplemented by a Monte Carlo investigation.

6.1 Irrelevant variables

One might think that one could rely on the assertion that Johansen's procedure (see Chapter 2) does not assume any row of β or α is non-zero. Thus, as a zero row of β excludes a variable from the cointegrating relations and a zero row of α excludes the cointegrating relations from the process generating a variable, it may be 'obvious' that such an 'irrelevant' variable will not affect the estimation, and the zero rows will be as efficiently estimated as the other rows of beta. However, to justify this assertion, a careful check of Johansen's procedure is required. This section provides such a check.

By the term 'irrelevant' variables we refer to variables that do not enter the cointegrating relations. We do not assume that the error terms of the relevant variables (i.e. variables that enter the cointegrating relations) are uncorrelated with those of the irrelevant ones as we analyse this as a special case.

6.1.1 The model and some results

The DGP is given by a VAR(1) model in error correction form,

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \ \ t = 1, 2, \dots, T \tag{6.1}$$

where $\varepsilon_t \sim i.i.d.(0, \Omega)$ with finite fourth moments and X_t is a $p \times 1$, $I(1)$ process. X_t is also cointegrated with $r < p$ cointegrating vectors, hence $\Pi = \alpha\beta'$ (α and β are $p \times r$ matrices) and $\beta' X_t \sim I(0)$. The SM used for performing cointegration tests consists of $p^+ > p$ variables so that $l \equiv (p^+ - p)$ $I(1)$ variables are irrelevant (i.e. they do not enter any of the r cointegrating relations). The SM in error correction form is then

$$\Delta X_t^+ = \Pi^+ X_{t-1}^+ + \varepsilon_t^+, \ \ t = 1, 2, \dots, T \tag{6.2}$$

where $\varepsilon_t^+ \sim i.i.d.(0, \Omega^+)$ with finite fourth moments and X_t^+ is $p^+ \times 1$, $I(1)$ process with its first p elements being those in (6.1) i.e.

$$X_t^+ = \begin{bmatrix} X_{1t} \\ \vdots \\ X_{pt} \\ X_{(p+1)t} \\ \vdots \\ X_{p^+t} \end{bmatrix} = \begin{bmatrix} X_t \\ X_t^{(l)} \end{bmatrix} \tag{6.3}$$

where $X_t^{(l)'} = \begin{bmatrix} X_{(p+1)t} & \cdots & X_{p^+t} \end{bmatrix}$. Thus, $X_t^{(l)} \sim I(1)$ and also non-cointegrated. We define the $p^+ \times p$ selection matrix $H = \begin{bmatrix} I_p \\ \underset{l\times p}{0} \end{bmatrix}$, then $H' X_t^+ = X_t$. The eigenvalue equation for the SM (6.2) is

$$|S^+(\zeta)| = 0 \tag{6.4}$$

where $S^+(\zeta) = \zeta S_{11}^+ - S_{10}^+ S_{00}^{+-1} S_{01}^+$. In addition,

$S_{11}^+ = T^{-1} \sum_{t=1}^{T} (X_{t-1}^+ - \bar{X}^+)(X_{t-1}^+ - \bar{X}^+)'$

$$S_{00}^{+} = T^{-1}\sum_{t=1}^{T}(\Delta X_t^{+} - \bar{\Delta} X^{+})(\Delta X_t^{+} - \bar{\Delta} X^{+})'$$

$$S_{10}^{+} = S_{01}^{+'} = T^{-1}\sum_{t=1}^{T}(X_{t-1}^{+} - \bar{X}^{+})(\Delta X_t^{+} - \bar{\Delta} X^{+})'$$

where $\bar{X}^{+} = T^{-1}\sum_{t=1}^{T} X_{t-1}^{+}$ and $\bar{\Delta} X^{+} = T^{-1}\sum_{t=1}^{T}\Delta X_t^{+}$. Using the partition in (6.3) we obtain

$$S_{11}^{+} = \begin{bmatrix} S_{11} & S_{11l} \\ S_{1l1} & S_{1l1l} \end{bmatrix}, S_{00}^{+} = \begin{bmatrix} S_{00} & S_{00l} \\ S_{0l0} & S_{0l0l} \end{bmatrix}, S_{10}^{+} = \begin{bmatrix} S_{10} & S_{10l} \\ S_{1l0} & S_{1l0l} \end{bmatrix} \text{ where}$$

$$S_{11} = T^{-1}\sum_{t=1}^{T}(X_{t-1} - \bar{X})(X_{t-1} - \bar{X})'$$

$$S_{11l} = S_{1l1}' = T^{-1}\sum_{t=1}^{T}(X_{t-1} - \bar{X})(X_{t-1}^{(l)} - \bar{X}^{(l)})'$$

$$S_{1l1l} = T^{-1}\sum_{t=1}^{T}(X_{t-1}^{(l)} - \bar{X}^{(l)})(X_{t-1}^{(l)} - \bar{X}^{(l)})'$$

$$S_{00} = T^{-1}\sum_{t=1}^{T}(\Delta X_t - \bar{\Delta} X)(\Delta X_t - \bar{\Delta} X)'$$

$$S_{00l} = S_{0l0}' = T^{-1}\sum_{t=1}^{T}(\Delta X_t - \bar{\Delta} X)(\Delta X_t^{(l)} - \bar{\Delta} X^{(l)})'$$

$$S_{0l0l} = T^{-1}\sum_{t=1}^{T}(\Delta X_t^{(l)} - \bar{\Delta} X^{(l)})(\Delta X_t^{(l)} - \bar{\Delta} X^{(l)})'$$

$$S_{10} = T^{-1}\sum_{t=1}^{T}(X_{t-1} - \bar{X})(\Delta X_t - \bar{\Delta} X)'$$

$$S_{10l} = S_{10l}' = T^{-1}\sum_{t=1}^{T}(X_{t-1} - \bar{X})(\Delta X_t^{(l)} - \bar{\Delta} X^{(l)})'$$

$$S_{1l0l} = T^{-1}\sum_{t=1}^{T}(X_{t-1}^{(l)} - \bar{X}^{(l)})(\Delta X_t^{(l)} - \bar{\Delta} X^{(l)})'$$

and $\bar{X} = T^{-1}\sum_{t=1}^{T} X_{t-1}$, $\bar{X}^{(l)} = T^{-1}\sum_{t=1}^{T} X_{t-1}^{(l)}$, $\bar{\Delta} X = T^{-1}\sum_{t=1}^{T}\Delta X_t$, $\bar{\Delta} X^{(l)} = T^{-1}\sum_{t=1}^{T}\Delta X_t^{(l)}$.

We then define

$$Var\begin{bmatrix} \Delta X_t^{+} \\ \beta' X_t \end{bmatrix} = \begin{bmatrix} \Sigma_{00}^{+} & \Sigma_{0\beta}^{+} \\ \Sigma_{\beta 0}^{+} & \Sigma_{\beta\beta} \end{bmatrix}$$

which can be expanded into

$$Var\begin{bmatrix} \Delta X_t \\ \Delta X_t^{(l)} \\ \beta' X_t \end{bmatrix} = \begin{bmatrix} \Sigma_{00} & \Sigma_{00l} & \Sigma_{0\beta} \\ \Sigma_{0l0} & \Sigma_{0l0l} & \Sigma_{0l\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta 0l} & \Sigma_{\beta\beta} \end{bmatrix}$$

where $\Sigma_{00}^{+} = \begin{bmatrix} \Sigma_{00} & \Sigma_{00l} \\ \Sigma_{0l0} & \Sigma_{0l0l} \end{bmatrix}$ and $\Sigma_{0\beta}^{+} = \begin{bmatrix} \Sigma_{0\beta} \\ \Sigma_{0l\beta} \end{bmatrix}$.

Note that in this chapter sufficient conditions (assumptions) for the Weak Law of Large Numbers (WLLN) used, are $\varepsilon_t \sim i.i.d.(0, \Omega)$ with finite fourth moments, which can be expressed in terms of the elements of the vector ε_t as $E\left|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}\right| < \infty$ for $i, j, k, l = 1, 2, \ldots, p$. Given that after application of the Granger Representation Theorem (see Theorem 2.1) the first differences (ΔX_t) and cointegrating relations ($\beta' X_t$) have infinite moving average representations, with convergent lag polynomials, the assumptions on ε_t imply (see Hamilton (1994, Propositions 10.2 and 18.1)):

(a) finite fourth moments for the first differences and cointegrating relations i.e. $E\left|\Delta X_{it_1}\Delta X_{jt_2}\Delta X_{kt_3}\Delta X_{lt_4}\right| < \infty$, $E\left|w_{it_1}w_{jt_2}w_{kt_3}w_{lt_4}\right| < \infty$ for $i, j, k, l = 1, 2, \ldots, p$ and for all t_1, t_2, t_3, t_4; where ΔX_{it} and w_{it} are the i-th elements of ΔX_t and $\beta' X_t$ respectively

(b) ergodicity for the second moments of the first differences and cointegrating relations i.e. $T^{-1}\sum_{t=1}^{T}\Delta X_{it}\Delta X_{j(t-s)} \xrightarrow{p} E(\Delta X_{it}\Delta X_{j(t-s)})$ and $T^{-1}\sum_{t=1}^{T} w_{it}w_{j(t-s)} \xrightarrow{p} E(w_{it}w_{j(t-s)})$ for $i, j = 1, 2, \ldots, p$ and for all s.

The implications of (a) and (b) above also hold for ΔX_t^{+} and $\Delta X_t^{(l)}$ used in this section and ΔX_t^{*} and $\beta_{11}' X_t^{*}$ in section 6.2, since they can be written as infinite lag polynomials with error terms that are $i.i.d.$ with finite fourth moments.

Let B_T^{+} be a $p^{+} \times p^{+}$ full rank matrix given by $B_T^{+} = \begin{bmatrix} \underset{p^{+}\times r}{\beta^{+}} & \underset{p^{+}\times(p^{+}-r)}{T^{-1/2}\bar{\beta}_{\perp}^{+}} \end{bmatrix}$ where $\beta^{+} = \begin{bmatrix} \underset{p\times r}{\beta} \\ \underset{l\times r}{0} \end{bmatrix}$, $\bar{\beta}_{\perp}^{+} = \beta_{\perp}^{+}(\beta_{\perp}^{+\prime}\beta_{\perp}^{+})^{-1}$ and $\beta^{+\prime}\beta_{\perp}^{+} = 0$, then

$$|B_T^{+\prime} S^{+}(\zeta) B_T^{+}| =$$

$$\left| \left[\begin{array}{cc} \zeta\beta^{+'}S_{11}^{+}\beta^{+} & T^{-1/2}\zeta\beta^{+'}S_{11}^{+}\bar{\beta}_{\perp}^{+} \\ T^{-1/2}\zeta\bar{\beta}_{\perp}^{+'}S_{11}^{+}\beta^{+} & T^{-1}\zeta\bar{\beta}_{\perp}^{+'}S_{11}^{+}\bar{\beta}_{\perp}^{+} \end{array} \right] \right.$$

$$\left. - \left[\begin{array}{cc} \beta^{+'}S_{10}^{+}S_{00}^{+-1}S_{01}^{+}\beta^{+} & T^{-1/2}\beta^{+'}S_{10}^{+}S_{00}^{+-1}S_{01}^{+}\bar{\beta}_{\perp}^{+} \\ T^{-1/2}\bar{\beta}_{\perp}^{+'}S_{10}^{+}S_{00}^{+-1}S_{01}^{+}\beta^{+} & T^{-1}\bar{\beta}_{\perp}^{+'}S_{10}^{+}S_{00}^{+-1}S_{01}^{+}\bar{\beta}_{\perp}^{+} \end{array} \right] \right| = 0. \tag{6.5}$$

For the asymptotic analysis of (6.5) we need the following results

$$\beta^{+'}S_{11}^{+}\beta^{+} = \beta' S_{11}\beta \xrightarrow{p} \Sigma_{\beta\beta} \tag{6.6}$$

$$\beta^{+'}S_{10}^{+} = \left[\begin{array}{cc} \beta' S_{10} & \beta' S_{10\iota} \end{array} \right] \xrightarrow{p} \left[\begin{array}{cc} \Sigma_{\beta 0} & \Sigma_{\beta 0\iota} \end{array} \right] \tag{6.7}$$

$$S_{00}^{+} = \left[\begin{array}{cc} S_{00} & S_{00\iota} \\ S_{0\iota 0} & S_{0\iota 0\iota} \end{array} \right] \xrightarrow{p} \Sigma_{00}^{+} = \left[\begin{array}{cc} \Sigma_{00} & \Sigma_{00\iota} \\ \Sigma_{0\iota 0} & \Sigma_{0\iota 0\iota} \end{array} \right] \tag{6.8}$$

by the WLLN. For (6.8), by Slutsky's Theorem (see Davidson (2000, pp. 39, 46)) and assuming invertibility in the limit we obtain

$$S_{00}^{+-1} \xrightarrow{p} \left[\begin{array}{cc} \Sigma^{00} & \Sigma^{00\iota} \\ \Sigma^{0\iota 0} & \Sigma^{0\iota 0\iota} \end{array} \right] \tag{6.9}$$

where the probability limit on the right-hand side is the partitioned inverse of Σ_{00}^{+}. Furthermore,

$$\bar{\beta}_{\perp}^{+'}S_{10}^{+} = O_p(1) \tag{6.10}$$

$$\beta^{+'}S_{11}^{+}\bar{\beta}_{\perp}^{+} = \left[\begin{array}{cc} \beta' S_{11} & \beta' S_{11\iota} \end{array} \right] \bar{\beta}_{\perp}^{+} = O_p(1) \tag{6.11}$$

because they are averages of products of an $I(0)$ and an $I(1)$ process.

Along with β^{+} we define $\alpha^{+} = \left[\begin{array}{c} \underset{p\times r}{\alpha} \\ \underset{\iota\times r}{\alpha^{(\iota)}} \end{array} \right]$ such that Π^{+} in (6.2) can be written as $\Pi^{+} = \alpha^{+}\beta^{+'}$ and $\alpha_{\perp}^{+}$ is such that $\alpha^{+'}\alpha_{\perp}^{+} = 0$.

By the Granger Representation Theorem the process X_t^+ has the following representation

$$X_t^+ = C^+ \sum_{i=1}^{t} \varepsilon_i^+ + C_1^+(L)\varepsilon_t^+ \tag{6.12}$$

(see Theorem 2.1) and $\beta^{+'}C^+ = 0$, where $C^+ = \beta_\perp^+(\alpha_\perp^{+'}\beta_\perp^+)^{-1}\alpha_\perp^{+'}$, so that $\beta^{+'}X_t^+ = \beta' X_t = \beta^+ C_1^+(L)\varepsilon_t^+ \sim I(0)$. Then we consider the behaviour of (6.12) in the non-stationary direction $\bar{\beta}_\perp^+$. By application of the Functional Central Limit Theorem on (6.12) and the CMT (see Theorems A.1 and A.3) we have

$$T^{-1/2}\bar{\beta}_\perp^{+'} X_{[Tu]}^+ = T^{-1/2}\bar{\beta}_\perp^{+'}(C^+ \sum_{i=1}^{[Tu]} \varepsilon_i^+ + C_1^+(L)\varepsilon_{[Tu]}^+) \xrightarrow{d} \bar{\beta}_\perp^{+'} C^+ W^+(u)$$

where $W^+(u)$ is a p^+-dimensional Brownian motion with variance Ω^+, $u \in [0,1]$,

$$\bar{\beta}_\perp^{+'} \bar{X}^+ \xrightarrow{d} \bar{\beta}_\perp^{+'} C^+ \int_0^1 W^+(u)du$$

and

$$T^{-1}\bar{\beta}_\perp^{+'} S_{11}^+ \bar{\beta}_\perp^{+'} = T^{-2}\bar{\beta}_\perp^{+'} \sum_{t=1}^{T}(X_{t-1}^+ - \bar{X}^+)(X_{t-1}^+ - \bar{X}^+)'\bar{\beta}_\perp^+$$

$$\xrightarrow{d} \bar{\beta}_\perp^{+'} C^+ \int_0^1 \tilde{W}^+ \tilde{W}^{+'} du C^{+'} \bar{\beta}_\perp^+ \tag{6.13}$$

where $\tilde{W}^+ = W^+(u) - \int_0^1 W^+(u)du$.

Using the results in (6.6), (6.7), (6.9), (6.10), (6.11) and (6.13) the limit of (6.5) becomes,

$$|B_T^{+'} S^+(\varsigma) B_T^+| \xrightarrow{d}$$

$$\begin{vmatrix} \varsigma\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma^{00}\Sigma_{0\beta} - \Psi & 0 \\ 0 & \varsigma\bar{\beta}_\perp^{+'} C^+ \int_0^1 \tilde{W}^+ \tilde{W}^{+'} du C^{+'} \bar{\beta}_\perp^+ \end{vmatrix}$$

$$= |\zeta\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma^{00}\Sigma_{0\beta} - \Psi| \times |\zeta\bar{\beta}_{\perp}^{+'} C^{+} \int_0^1 \tilde{W}^{+}\tilde{W}^{+'} du C^{+'}\bar{\beta}_{\perp}^{+}| = 0 \tag{6.14}$$

where $\Psi = \Sigma_{\beta 0l}\Sigma^{0l0}\Sigma_{0\beta} + \Sigma_{\beta 0}\Sigma^{00l}\Sigma_{0l\beta} + \Sigma_{\beta 0l}\Sigma^{0l0l}\Sigma_{0l\beta}$.

(6.14) has r positive eigenvalues given by the first factor and $(p^{+} - r)$ zero eigenvalues given by the second, since the stochastic matrix $\bar{\beta}_{\perp}^{+'} C^{+} \int_0^1 \tilde{W}^{+}\tilde{W}^{+'} du C^{+'}\bar{\beta}_{\perp}^{+}$ with dimensions $(p^{+} - r) \times (p^{+} - r)$ is positive definite almost surely. Note that

$$\begin{aligned} \lim_{T\to\infty} E[T^{-1}\bar{\beta}_{\perp}^{+'}(X_T^{+} - \bar{X})(X_T^{+} - \bar{X})'\bar{\beta}_{\perp}^{+}] &= \bar{\beta}_{\perp}^{+'} C^{+}\Omega^{+}C^{+'}\bar{\beta}_{\perp}^{+} \\ &= (\alpha_{\perp}^{+'}\beta_{\perp}^{+})^{-1}\alpha_{\perp}^{+'}\Omega^{+}\alpha_{\perp}^{+}(\beta_{\perp}^{+'}\alpha_{\perp}^{+})^{-1}, \end{aligned}$$

which is a $(p^{+} - r) \times (p^{+} - r)$ matrix of rank $(p^{+} - r)$ and corresponds to the long-run covariance matrix of the process X_t^{+} in the non-stationary direction, which is positive definite[7]. For given u, $\tilde{W}^{+}(u) \sim N(0, u\Omega^{+})$ so that $E(\tilde{W}^{+}(u)\tilde{W}^{+}(u)') = u\Omega^{+}$. Thus, $\int_0^1 \tilde{W}^{+}\tilde{W}^{+'} du$ is positive definite almost surely and $\bar{\beta}_{\perp}^{+'} C^{+} \int_0^1 \tilde{W}^{+}\tilde{W}^{+'} du C^{+'}\bar{\beta}_{\perp}^{+}$ has full rank $(p^{+} - r)$ almost surely (see also Davidson (2000, Chapter 15), Hamilton (1994, Chapter 18)).

Therefore when performing LR tests for cointegration on the overspecified model (6.2) we must be able to infer the true cointegrating rank as the sample size becomes large though the number of common trends is overestimated.

In particular the effect of irrelevant variables depends on whether Σ_{00}^{+} is block diagonal and whether the irrelevant variables are weakly exogenous for the parameters α and β that appear in the DGP. Below we state more precisely what we mean by weak exogeneity.

[7] The positive definiteness of $\lim_{T\to\infty} E(T^{-1}\bar{\beta}_{\perp}^{+'} X_T^{+} X_T^{+'}\bar{\beta}_{\perp}^{+})$ is an assumption for the Functional Central Limit Theorem to hold (see Davidson (2000, p. 365); Phillips and Durlauf (1986)).

We use the partition $X_t^{+'} = \left[\begin{array}{cc} X_t' & X_t^{(l)'} \end{array} \right]'$ and the fact that Π^+ can be written as the product of two $p^+ \times r$ matrices α^+ and β^+ of rank r i.e. $\Pi^+ = \alpha^+\beta^{+'}$ with α^+ and β^+ defined as above. Then (6.2) can be written as

$$\left[\begin{array}{c} \Delta X_t \\ \Delta X_t^{(l)} \end{array} \right] = \left[\begin{array}{c} \alpha \\ \alpha^{(l)} \end{array} \right] \left[\begin{array}{cc} \beta' & 0 \end{array} \right] \left[\begin{array}{c} X_{t-1} \\ X_{t-1}^{(l)} \end{array} \right] + \left[\begin{array}{c} \varepsilon_t \\ \varepsilon_t^{(l)} \end{array} \right]$$

or

$$\left[\begin{array}{c} \Delta X_t \\ \Delta X_t^{(l)} \end{array} \right] = \left[\begin{array}{c} \alpha\beta' X_{t-1} \\ \alpha^{(l)}\beta' X_{t-1} \end{array} \right] + \left[\begin{array}{c} \varepsilon_t \\ \varepsilon_t^{(l)} \end{array} \right]$$

where $\left[\begin{array}{cc} \underset{1\times p}{\varepsilon_t'} & \underset{1\times l}{\varepsilon_t^{(l)'}} \end{array} \right]' = \varepsilon_t^{+'}$. Note that with respect to the underlying parameters of the SM $\alpha^+ = \Sigma_{0\beta}^+ \Sigma_{\beta\beta}^{-1} = \left[\begin{array}{c} \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1} \\ \Sigma_{0l\beta}\Sigma_{\beta\beta}^{-1} \end{array} \right]$, see equation (10.3) in Johansen (1996).

If $\alpha^{(l)} = 0$ we say that the vector process $X_t^{(l)}$ is weakly exogenous for the parameters α and β and their maximum likelihood estimators can be calculated from the conditional model (conditional on $\Delta X_t^{(l)}$). In fact this is the definition of weak exogeneity given in Johansen (1996, Theorem 8.1).

Then the four cases that arise are:

(i) Σ_{00}^+ block diagonal and $\Sigma_{0l\beta} = 0$ (implying $\alpha^{(l)} = 0$); the inclusion of irrelevant variables does not affect the probability limit of the r positive eigenvalues ($\Psi = 0$).

(ii) Σ_{00}^+ block diagonal and $\Sigma_{0l\beta} \neq 0$ (implying $\alpha^{(l)} \neq 0$); the inclusion of irrelevant variables changes the magnitude of the probability limit of the r positive eigenvalues ($\Psi \neq 0$).

(iii) Σ_{00}^+ non-block diagonal and $\Sigma_{0l\beta} = 0$ (implying $\alpha^{(l)} = 0$); the inclusion of irrelevant variables does not change the magnitude of the probability limit of the r positive eigenvalues ($\Psi = 0$). However, diagonalising Σ_{00}^{-} changes the exogeneity status of $X_t^{(l)}$ from weakly exogenous to non-weakly exogenous.

(iv) Σ_{00}^{+} non-block diagonal and $\Sigma_{0l\beta} \neq 0$ (implying $\alpha^{(l)} \neq 0$); the inclusion of irrelevant variables changes the magnitude of the probability limit of the r positive eigenvalues ($\Psi \neq 0$) and $X_t^{(l)}$ is not weakly exogenous with respect to β and α in the DGP.

In the case that Σ_{00}^{+} is not block diagonal one can transform (6.2) into

$$\Delta X_t^{++} = \alpha^{++}\beta^{++'}X_{t-1}^{++} + \varepsilon_t^{++}$$

where $\Delta X_t^{++} = P\Delta X_t^{+}$, $X_{t-1}^{++} = PX_{t-1}^{+}$, $\varepsilon_t^{++} = P\varepsilon_t^{+}$, $\alpha^{++} = P\alpha^{+}$, $\beta^{++'} = \beta^{+'}P^{-1}$ and P is such that $Var(\Delta X_t^{+}) = \Sigma_{00}^{+} = P^{-1}P'^{-1}$ and $Var(P\Delta X_t^{+}) = I_{p^+}$. Thus, P might be the Cholesky factor of Σ_{00}^{+-1} and $P = \begin{bmatrix} \underset{p\times p}{P_{11}} & \underset{p\times l}{0} \\ \underset{l\times p}{P_{21}} & \underset{l\times l}{P_{22}} \end{bmatrix}$. Then

$$\beta^{+'}P^{-1} = \begin{bmatrix} \beta' & 0 \end{bmatrix} \begin{bmatrix} P_{11}^{-1} & 0 \\ -P_{22}^{-1}P_{21}P_{11}^{-1} & P_{22}^{-1} \end{bmatrix} = \begin{bmatrix} \beta' P_{11}^{-1} & 0 \end{bmatrix}$$

and

$$P\alpha^{+} = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^{(l)} \end{bmatrix} = \begin{bmatrix} P_{11}\alpha \\ P_{21}\alpha + P_{22}\alpha^{(l)} \end{bmatrix}.$$

However, the transformation to achieve the block diagonality of Σ_{00}^{+} reparametrises α. For case (iii), since $P_{21}\alpha \neq 0$ the irrelevant variables are no longer weakly exogenous even though $\alpha^{(l)} = 0$.

Although we can detect the true number of cointegrating vectors (r), in general, the magnitude of the positive eigenvalues in the limit is altered by the presence of irrelevant variables.

6.1.2 Consistency

Under the assumption of cointegrating rank r, Π^+ in (6.2) can be written as $\Pi^+ = \alpha^+\beta^{+'}$, where α^+ and β^+ are $p^+ \times r$ matrices of rank r. Let $\hat{\beta}^+$, $\hat{\alpha}^+$ and $\hat{\Omega}^+$ be the maximum likelihood estimators of β^+, α^+ and Ω^+ respectively in the SM (6.2). In order to analyse the consistency properties of the estimators we consider a linear transformation of the columns of $\hat{\beta}^+$[8], namely $\tilde{\beta}^+ = \hat{\beta}^+(\bar{\beta}^{+'}\hat{\beta}^+)^{-1}$, where $\bar{\beta}^+ = \beta^+(\beta^{+'}\beta^+)^{-1}$, and $\tilde{\beta}^+$ also maximises the likelihood function. We also consider $\tilde{\alpha}^+ = \hat{\alpha}^+\hat{\beta}^{+'}\bar{\beta}^+ = S_{01}^+\tilde{\beta}^+(\tilde{\beta}^{+'}S_{11}^+\tilde{\beta}^+)^{-1}$, where the second equality follows from the definition of $\tilde{\beta}^+$ and the fact that $\hat{\alpha}^+ = S_{01}^+\hat{\beta}^+(\hat{\beta}^{+'}S_{11}^+\hat{\beta}^+)^{-1}$, see equation (2.13) in section 2.3.

Since β^+ and $\bar{\beta}_\perp^+$ span $\mathbb{R}^{p^+}$ and the inverse of B_T^+ is given by $B_T^{+-1} = \begin{bmatrix} \underset{r\times p^+}{\bar{\beta}^{+'}} \\ T^{1/2}\underset{(p^+-r)\times p^+}{\beta_\perp^{+'}} \end{bmatrix}$, by forming $B_T^+B_T^{+-1}$ the following relation holds

$$\beta^+\bar{\beta}^{+'} + \bar{\beta}_\perp^+\beta_\perp^{+'} = I_{p^+}.$$

Then,

$$\begin{aligned} \tilde{\beta}^+ &= \hat{\beta}^+(\bar{\beta}^{+'}\hat{\beta}^+)^{-1} \\ &= \beta^+\bar{\beta}^{+'}\hat{\beta}^+(\bar{\beta}^{+'}\hat{\beta}^+)^{-1} + \bar{\beta}_\perp^+\beta_\perp^{+'}\hat{\beta}^+(\bar{\beta}^{+'}\hat{\beta}^+)^{-1} \\ &= \beta^+ + \bar{\beta}_\perp^+ b^+ \end{aligned} \tag{6.15}$$

where $b^+ = \beta_\perp^{+'}\tilde{\beta}^+$.

---

[8] In fact for any normalisation c we can define $\hat{\beta}_c = \hat{\beta}(c'\hat{\beta})^{-1} = \tilde{\beta}(c'\tilde{\beta})^{-1}$; expanding around β and normalising β and $\hat{\beta}$ by $c'\beta = c'\hat{\beta} = I_r$, we obtain $\hat{\beta} - \beta = (I_p - \beta c')(\tilde{\beta} - \beta) + O_p(|\tilde{\beta} - \beta|^2)$ (Johansen (1996, p. 180)) therefore the properties of $\hat{\beta}$ follow from those of $\tilde{\beta}$.

Proposition 6.1. (i) The estimator $\tilde{\beta}^{+}$ *associated with the overspecified model* (6.2) *is consistent i.e.* $\tilde{\beta}^{+} \xrightarrow{p} \beta^{+} = \begin{bmatrix} \beta \\ 0 \end{bmatrix}$.

(ii) The first p rows of the estimator $\tilde{\alpha}^{+} = S_{01}^{+}\tilde{\beta}^{+}(\tilde{\beta}^{+'}S_{11}^{+}\tilde{\beta}^{+})^{-1}$ *associated with* (6.2) *are consistent estimators of the adjustment coefficients* (α) *in the DGP* (6.1).

(iii) The top left, $p \times p$ *block of* $\hat{\Omega}^{+} = S_{00}^{+} - S_{01}^{+}\tilde{\beta}^{+}(\tilde{\beta}^{+'}S_{11}^{+}\tilde{\beta}^{+})^{-1}\tilde{\beta}^{+'}S_{10}^{+}$ *is a consistent estimator of* $\Omega = Var(\varepsilon_t)$ *in the DGP.*

Proof. (i) Let $\hat{V}^{+} = \begin{bmatrix} \underset{p^{+}\times r}{\hat{\beta}^{+}} & \underset{p^{+}\times(p^{+}-r)}{\hat{V}_2^{+}} \end{bmatrix}$ be the $p^{+} \times p^{+}$ matrix whose columns correspond to the eigenvectors of (6.4). The eigenvectors that correspond to the r largest eigenvalues of (6.4) define $\hat{\beta}^{+}$. (6.5) has the same eigenvalues as (6.4) but its eigenvectors are given by $B_T^{+-1}\hat{V}^{+}$. The space spanned by the first r eigenvectors of (6.5) is given by $sp(B_T^{+-1}\hat{\beta}^{+}) = sp(B_T^{+-1}\tilde{\beta}^{+})$, since $\tilde{\beta}^{+}$ is a linear transformation of $\hat{\beta}^{+}$, and

$$B_T^{+-1}\tilde{\beta}^{+} = \begin{bmatrix} \bar{\beta}^{+'} \\ T^{1/2}\beta_{\perp}^{+'} \end{bmatrix}\tilde{\beta}^{+} = \begin{bmatrix} \bar{\beta}^{+'}\beta^{+} \\ T^{1/2}\beta_{\perp}^{+'}\bar{\beta}_{\perp}^{+}b^{+} \end{bmatrix} = \begin{bmatrix} I_r \\ T^{1/2}b^{+} \end{bmatrix}. \tag{6.16}$$

The ordered eigenvalues of (6.5) converge to the eigenvalues of (6.14) and therefore the last $(p^{+} - r)$ rows of (6.16) must converge in probability to zero, since in the limit there are only r positive eigenvalues (given by the first factor of (6.14)) which correspond to the eigenvectors given by the first block of (6.16). Hence,

$$T^{1/2}b^{+} \xrightarrow{p} 0 \tag{6.17}$$

and by (6.15)

$$T^{1/2}(\tilde{\beta}^{+} - \beta^{+}) \xrightarrow{p} 0. \tag{6.18}$$

Thus, (6.18) shows the consistency of $\tilde{\beta}^{+}$ and in addition that $(\tilde{\beta}^{+} - \beta^{+}) = o_p(T^{-1/2})$.

(ii) We need to find the limits of $(\tilde{\beta}^{+'} S_{11}^{+} \tilde{\beta}^{+})$ and $(S_{01}^{+} \tilde{\beta}^{+})$. Using (6.15)

$$\tilde{\beta}^{+'} S_{11}^{+} \tilde{\beta}^{+} = (\beta^{+} + \bar{\beta}_{\perp}^{+} b^{+})' S_{11}^{+} (\beta^{+} + \bar{\beta}_{\perp}^{+} b^{+})$$

$$= \beta^{+'} S_{11}^{+} \beta^{+} + T^{-1/2} \beta^{+'} S_{11}^{+} \bar{\beta}_{\perp}^{+} (T^{1/2} b^{+}) + (T^{1/2} b^{+'}) T^{-1/2} \bar{\beta}_{\perp}^{+'} S_{11}^{+} \beta^{+}$$
$$+ (T^{1/2} b^{+'}) (T^{-1} \bar{\beta}_{\perp}^{+'} S_{11}^{+} \bar{\beta}_{\perp}^{+}) (T^{1/2} b^{+})$$

$$\xrightarrow{p} \beta^{+'} S_{11}^{+} \beta^{+} = \beta' S_{11} \beta \xrightarrow{p} \Sigma_{\beta\beta} \tag{6.19}$$

by (6.6), (6.10), (6.13) and (6.17);

$$S_{01}^{+} \tilde{\beta}^{+} = S_{01}^{+} (\beta^{+} + \bar{\beta}_{\perp}^{+} b^{+}) = S_{01}^{+} \beta^{+} + T^{-1/2} S_{01}^{+} \bar{\beta}_{\perp}^{+} (T^{1/2} b^{+})$$

$$\xrightarrow{p} S_{01}^{+} \beta^{+} = \begin{bmatrix} S_{01}\beta \\ S_{0l}\beta \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \Sigma_{0\beta} \\ \Sigma_{0l\beta} \end{bmatrix} \tag{6.20}$$

by (6.7), (6.11) and (6.17). Thus, by (6.19), (6.20) and Slutsky's Theorem,

$$\tilde{\alpha}^{+} \xrightarrow{p} \begin{bmatrix} \Sigma_{0\beta} \Sigma_{\beta\beta}^{-1} \\ \Sigma_{0l\beta} \Sigma_{\beta\beta}^{-1} \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha^{(l)} \end{bmatrix}$$

where $\alpha = \Sigma_{0\beta} \Sigma_{\beta\beta}^{-1}$, which is the definition of α in the DGP (6.1), see equation (10.3) in Johansen (1996).

(iii) First note that the right-hand side of the equality of $\hat{\Omega}^{+}$ follows from the definition of $\tilde{\beta}^{+}$ and the fact that $\hat{\Omega}^{+} = S_{00}^{+} - S_{01}^{+} \hat{\beta}^{+} (\hat{\beta}^{+'} S_{11}^{+} \hat{\beta}^{+})^{-1} \hat{\beta}^{+'} S_{10}^{+}$, see equation (2.14) in section 2.3. Then using (6.8), (6.19), (6.20) and Slutsky's Theorem we find that

$$\hat{\Omega}^{+} \xrightarrow{p} \begin{bmatrix} \Sigma_{00} & \Sigma_{00l} \\ \Sigma_{0l0} & \Sigma_{0l0l} \end{bmatrix} - \begin{bmatrix} \Sigma_{0\beta} \\ \Sigma_{0l\beta} \end{bmatrix} \Sigma_{\beta\beta}^{-1} \begin{bmatrix} \Sigma_{\beta 0} & \Sigma_{\beta 0l} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{00} & \Sigma_{00l} \\ \Sigma_{0l0} & \Sigma_{0l0l} \end{bmatrix} - \begin{bmatrix} \alpha \\ \alpha^{(l)} \end{bmatrix} \Sigma_{\beta\beta} \begin{bmatrix} \alpha' & \alpha^{(l)'} \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{00} - \alpha\Sigma_{\beta\beta}\alpha' & \Sigma_{00l} - \alpha\Sigma_{\beta\beta}\alpha^{(l)'} \\ \Sigma_{0l0} - \alpha^{(l)}\Sigma_{\beta\beta}\alpha' & \Sigma_{0l0l} - \alpha^{(l)}\Sigma_{\beta\beta}\alpha^{(l)'} \end{bmatrix}$$

and $\Sigma_{00} - \alpha\Sigma_{\beta\beta}\alpha' = \Omega$, see equation (10.4) in Johansen (1996). ■

6.2 Omitted variables

Next we investigate the case where relevant $I(1)$ variables have been omitted from the VAR model used for cointegration analysis. The analysis is based on the fact that the cointegrating vectors, β, (as well as the adjustment coefficients, α) are not identified so β (and therefore α) can be replaced by a non-singular transformation e.g. we can replace β' by a row equivalent matrix of β'. To avoid complicating the notation we retain the same symbols for the parameters (and variables) and their non-singular transformations.

6.2.1 The model and some results

The DGP is given by a VAR(1) model in error correction form,

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \ \ t = 1, 2, \ldots T \tag{6.21}$$

where $\varepsilon_t \sim i.i.d.(0, \Omega)$ with finite fourth moments and X_t is a $p \times 1$, $I(1)$ process. In addition X_t is cointegrated so that $\Pi = \alpha\beta'$ (α and β are $p \times r$ matrices) with $r \leq p - 1$ cointegrating vectors β such that $\beta' X_t \sim I(0)$.

The SM used for cointegration testing is assumed to be underspecified i.e. it includes only a subset of the variables of the DGP. More specifically, let $H = \begin{bmatrix} I_{p^*} \\ \underset{k \times p^*}{0} \end{bmatrix}$ be a selection matrix, then the SM includes $p^* < p$ variables given by $X_t^* = H' X_t$ so that $k \equiv (p - p^*)$ relevant variables are omitted.

The misspecified SM takes the form of a multivariate regression of $H'\Delta X_t = \Delta X_t^*$ on $H'X_{t-1}$. The relation between ΔX_t^* and X_{t-1}^* does not have an error correction form as the model

$$\Delta X_t^* = \Pi^* X_{t-1}^* + e_t^*,\ t = 1, 2, \dots, T \tag{6.22}$$

is misspecified. In particular e_t^* is correlated with X_{t-1}^*. We also define $\beta^{(1)} = H'\beta$, and $\alpha^{(1)} = H'\alpha$, but $\Pi^* \neq \alpha^{(1)}\beta^{(1)'}$and $\Pi^* \neq H'\Pi H$ as $HH' \neq I_p$.

Although $\beta' X_t$ is $I(0)$, $\beta^{(1)'}X_t^*$ is not necessarily $I(0)$ since a linear combination of $I(1)$ variables is in general $I(1)$. The nature of $\beta^{(1)'}X_t^*$ is determined by the variables entering the cointegrating relations in the DGP. Since only the space spanned by the columns of β can be estimated, in general, $(r-k)$ cointegrating vectors (stationary relations) can be found by applying elementary row operations on β'. Thus, β' can be transformed so that

$$\beta' = \begin{bmatrix} \beta_{11} & \beta_{21} & \cdots & \beta_{p1} \\ \beta_{12} & \beta_{22} & \cdots & \beta_{p2} \\ \vdots & \vdots & & \vdots \\ \beta_{1r} & \beta_{2r} & \cdots & \beta_{pr} \end{bmatrix} \approx$$

$$\begin{bmatrix} \beta^+_{11} & \beta^+_{21} & \cdots & \beta^+_{(p-r)1} & 1 & 0 & 0 & \cdots & 0 & 0 \\ \beta^+_{12} & \beta^+_{22} & \cdots & \cdots & \beta^+_{(p-(r-1))2} & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta^+_{1r} & \beta^+_{2r} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \beta^+_{(p-1)r} & 1 \end{bmatrix} \tag{6.23}$$

where the symbol $\approx$ denotes the row equivalent matrix of β' given by (6.23) and $p - (r - i) = p^* - (r - k) + i$, $i = 1, 2, \dots, r$ is the number of non-zero elements on the i-th row. Given that only p^* variables are included in the SM, we should be able to recover i cointegrating relations (using the underspecified SM), as long as $p^* - (r - k) + i \leq p^*$. Thus, at most $(i = r - k)$, $(r - k)$ cointegrating relations can be estimated from the SM, by applying the same row operations on $\beta^{(1)'}$ as on β'.

Below we distinguish two cases:

Case (i). $(r-k) \leq 0$, where all the cointegrating relations in the DGP involve at least one of the omitted variables, therefore $\beta^{(1)'} X_t^* \sim I(1)$.

Case (ii). $(r-k) > 0$, where there are $q < r$, $q \geq (r-k)$, cointegrating relations in the DGP which do not involve any of the k omitted variables, accounting also for the event of fortuitous zeros. Therefore, some elements of $\beta^{(1)'} X_t^*$, $\beta_{11}^{'} X_t^*$, say are stationary, where β_{11} is a submatrix of $\beta^{(1)}$ in the following partition, $\beta^{(1)} = \left[\underset{p^* \times q}{\beta_{11}} \quad \underset{p^* \times (r-q)}{\beta_{12}} \right]$. Then $\beta^{(1)'} X_t^* = \left[\begin{array}{c} \beta_{11}^{'} X_t^* \\ \beta_{12}^{'} X_t^* \end{array} \right]$ and $\beta_{11}^{'} X_t^* \sim I(0)$ while $\beta_{12}^{'} X_t^* \sim I(1)$. Here we assume that the actual cointegrating vectors can be found at the first q rows of $\beta^{(1)'}$. Nevertheless, if the above ordering is not satisfied, the cointegrating vectors can be isolated in the first q rows of $\beta^{(1)'}$ using elementary row operations (see above).

The eigenvalue equation that corresponds to (6.22) is

$$|\zeta S_{11}^* - S_{10}^* S_{00}^{*-1} S_{01}^*| = 0 \tag{6.24}$$

where $S_{11}^* = T^{-1} \sum_{t=1}^{T} (X_{t-1}^* - \bar{X}^*)(X_{t-1}^* - \bar{X}^*)'$, $S_{00}^* = T^{-1} \sum_{t=1}^{T} (\Delta X_t^* - \bar{\Delta X}^*)(\Delta X_t^* - \bar{\Delta X}^*)'$, $S_{10}^* = S_{01}^{*'} = T^{-1} \sum_{t=1}^{T} (X_{t-1}^* - \bar{X}^*)(\Delta X_t^* - \bar{\Delta X}^*)'$, $\bar{X}^* = T^{-1} \sum_{t=1}^{T} X_{t-1}^*$ and $\bar{\Delta X}^* = T^{-1} \sum_{t=1}^{T} \Delta X_t^*$.

The eigenvalue equation that corresponds to the DGP is

$$|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$$

with S_{ij}, $i, j = 0, 1$, defined in terms of the process X_t (DGP) similarly as above.

Note that we can partition the stochastic process X_t into $X_t = \left[\begin{array}{c} \underset{p^* \times 1}{X_t^*} \\ \underset{k \times 1}{X_t^{(k)}} \end{array} \right]$ where the upper $(p^* \times 1)$ block holds the variables included in the SM and the lower $(k \times 1)$

block corresponds to the omitted variables. Then, S^*_{ij}, $i, j = 0, 1$, is given by the top left submatrix of the corresponding S_{ij}, $i, j = 0, 1$.

The matrix $S^{*-1}_{11} S^*_{10} S^{*-1}_{00} S^*_{01}$ has the same eigenvalues as the roots of (6.24), which coincide with the non-zero eigenvalues of

$$S^* = (DS_{11}D)^+(DS_{10}D)(DS_{00}D)^+(DS_{01}D)$$

where $D = \begin{bmatrix} I_{p^*} & \underset{p^*\times k}{0} \\ \underset{k\times p^*}{0} & \underset{k\times k}{0} \end{bmatrix}$ and here the superscript + denotes the Moore-Penrose (generalised) inverse.

Let $Q = \begin{bmatrix} S^*_{11} & 0 \\ 0 & I_k \end{bmatrix}$, $|Q| \neq 0$ then,

$$|\zeta I_p - S^*| = |Q^{-1}||Q(\zeta I_p - S^*)| = |Q^{-1}||S^*(\zeta)| = 0,$$

where $S^*(\zeta) = Q(\zeta I_p - S^*)$. Expanding the above equation,

$$|S^*(\zeta)| = \begin{vmatrix} \zeta S^*_{11} - S^*_{10} S^{*-1}_{00} S^*_{01} & 0 \\ 0 & \zeta I_k \end{vmatrix} = |\zeta I_k||\zeta S^*_{11} - S^*_{10} S^{*-1}_{00} S^*_{01}| = 0. \qquad (6.25)$$

As expected, there are k zero eigenvalues which correspond to the omitted variables. The second factor of (6.25) is the characteristic polynomial in (6.24) associated with the SM. If the LR tests are to indicate the existence of cointegration in the underspecified model, the second factor of (6.25) must give some eigenvalues with positive probability limits.

Define $B_T = (\beta, T^{-1/2}\bar{\beta}_\perp)$, where $\bar{\beta}_\perp = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}$, $\beta_\perp$ is $p \times (p-r)$ such that $\beta'\beta_\perp = 0$ and $\beta = \begin{bmatrix} \underset{p^*\times r}{\beta^{(1)}} \\ \underset{k\times r}{\beta^{(2)}} \end{bmatrix}$, $\bar{\beta}_\perp = \begin{bmatrix} \underset{p^*\times(p-r)}{\bar{\beta}^{(1)}_\perp} \\ \underset{k\times(p-r)}{\bar{\beta}^{(2)}_\perp} \end{bmatrix}$ then,

$$|B'_T S^*(\zeta) B_T| = \begin{vmatrix} \beta' S^*(\zeta)\beta & T^{-1/2}\beta' S^*(\zeta)\bar{\beta}_\perp \\ T^{-1/2}\bar{\beta}'_\perp S^*(\zeta)\beta & T^{-1}\bar{\beta}'_\perp S^*(\zeta)\bar{\beta}_\perp \end{vmatrix}$$

$$= \left| \begin{bmatrix} \varsigma(\beta^{(1)'} S_{11}^{*} \beta^{(1)} + \beta^{(2)'} \beta^{(2)}) & T^{-1/2}\varsigma(\beta^{(1)'} S_{11}^{*} \bar{\beta}_{\perp}^{(1)} + \beta^{(2)'} \bar{\beta}_{\perp}^{(2)}) \\ T^{-1/2}\varsigma(\bar{\beta}_{\perp}^{(1)'} S_{11}^{*} \beta^{(1)} + \bar{\beta}_{\perp}^{(2)'} \beta^{(2)}) & T^{-1}\varsigma(\bar{\beta}_{\perp}^{(1)'} S_{11}^{*} \bar{\beta}_{\perp}^{(1)} + \bar{\beta}_{\perp}^{(2)'} \bar{\beta}_{\perp}^{(2)}) \end{bmatrix} \right.$$

$$\left. - \begin{bmatrix} \beta^{(1)'} S_{10}^{*} S_{00}^{*-1} S_{01}^{*} \beta^{(1)} & T^{-1/2}\beta^{(1)'} S_{10}^{*} S_{00}^{*-1} S_{01}^{*} \bar{\beta}_{\perp}^{(1)} \\ T^{-1/2}\bar{\beta}_{\perp}^{(1)'} S_{10}^{*} S_{00}^{*-1} S_{01}^{*} \beta^{(1)} & T^{-1}\bar{\beta}_{\perp}^{(1)'} S_{10}^{*} S_{00}^{*-1} S_{01}^{*} \bar{\beta}_{\perp}^{(1)} \end{bmatrix} \right| = 0. \tag{6.26}$$

In order to analyse the limiting behaviour of (6.26) we resort to the Granger Representation Theorem which gives the following representation for X_t in (6.21)

$$X_t = C\sum_{i=1}^{t} \varepsilon_i + C_1(L)\varepsilon_t \tag{6.27}$$

(see Theorem 2.1). Then for the p^*-dimensional vector of variables X_t^* included in the SM we have the following representation, by using (6.27),

$$X_t^* = C^*\sum_{i=1}^{t} \varepsilon_i + C_1^*(L)\varepsilon_t \tag{6.28}$$

where $C^* = H'C$, $C_1^*(L) = H'C_1(L)$ both of dimensions $p^* \times p$ and $rank(C^*) = \min(p^*, p^* - (r-k))$. Thus, for case (i) $rank(C^*) = p^*$ and for case (ii) $rank(C^*) = (p^* - q)$.

Let the non-stationary direction for the process X_t^* be B^* which is $p^* \times p$ for case (i) and $p^* \times (p-q)$ for case (ii) (for the detailed form of B^* see under the relevant cases below). By application of the Functional Central Limit Theorem on (6.28) and the CMT (see Theorems A.1 and A.3) we have

$$T^{-1/2}B^{*'}X_{[Tu]}^* = T^{-1/2}B^{*'}(C^*\sum_{i=1}^{[Tu]} \varepsilon_{[Tu]} + C_1^*(L)\varepsilon_{[Tu]}) \xrightarrow{d} B^{*'}C^*W(u)$$

where $W(u)$ is a p-dimensional Brownian motion with variance Ω, $u \in [0,1]$

$$B^{*'}\bar{X}^* \xrightarrow{d} B^{*'}C^*\int_0^1 W(u)du$$

and

$$T^{-1}B^{*\prime}S_{11}^{*}B^{*} = T^{-2}B^{*\prime}\sum_{t=1}^{T}(X_{t-1}^{*}-\bar{X}^{*})(X_{t-1}^{*}-\bar{X}^{*})'B^{*} \quad (6.29)$$

$$\xrightarrow{d} B^{*\prime}C^{*}\int_{0}^{1}\tilde{W}\tilde{W}'C^{*\prime}B^{*}du$$

where $\tilde{W} = W(u) - \int_0^1 W(u)du$.

Below we present the asymptotics for the two cases.

Case (i)

Since $\beta^{(1)\prime}X_t^*$ is not $I(0)$, because of the omission of relevant variables, (6.26) is not appropriately scaled for convergence. Pre- and post-multiplying (6.26) by the scaling matrix $\Upsilon_T = \begin{bmatrix} T^{-1/2}I_r & 0 \\ 0 & I_{p-r} \end{bmatrix}$ we obtain,

$$|\Upsilon_T' B_T' S^*(\zeta) B_T \Upsilon_T| =$$

$$\left| \begin{bmatrix} T^{-1}\zeta\beta^{(1)\prime}S_{11}^{*}\beta^{(1)} + o_p(1) & T^{-1}\zeta\beta^{(1)\prime}S_{11}^{*}\bar{\beta}_{\perp}^{(1)} + o_p(1) \\ T^{-1}\zeta\bar{\beta}_{\perp}^{(1)\prime}S_{11}^{*}\beta^{(1)} + o_p(1) & T^{-1}\zeta\bar{\beta}_{\perp}^{(1)\prime}S_{11}^{*}\bar{\beta}_{\perp}^{(1)} + o_p(1) \end{bmatrix} - \begin{bmatrix} o_p(1) & o_p(1) \\ o_p(1) & o_p(1) \end{bmatrix} \right| \quad (6.30)$$

$$= \left| T^{-1}\zeta B^{*\prime}S_{11}^{*}B^{*} + o_p(1) \right|$$

where $B^* = \begin{bmatrix} \beta^{(1)} & \bar{\beta}_{\perp}^{(1)} \end{bmatrix}$, $p^* \times p$. The second matrix in (6.30) is $o_p(1)$ because its blocks are products of averages of products of either two $I(0)$ processes (S_{00}^*) or an $I(0)$ and an $I(1)$ process ($B^{*\prime}S_{10}^*$), which are $O_p(1)$ (see (A.10) in Appendix A), thus after scaling by Υ_T they all become $o_p(1)$.

Then we have

$$|\Upsilon_T' B_T' S^*(\zeta) B_T \Upsilon_T| = \left| T^{-1}B^{*\prime}S_{11}^{*}B^{*} + o_p(1) \right|$$

$$\xrightarrow{d} \left| \zeta B^{*\prime}C^{*}\int_{0}^{1}\tilde{W}\tilde{W}'C^{*\prime}B^{*}du \right| = 0 \quad (6.31)$$

by (6.29).

From (6.31) we find that in the limit there are p roots at zero k of which exist by construction. This suggests that performing the LR tests for cointegration using the under-specified model will lead to the rejection of the hypothesis of cointegration (i.e. acceptance of $r = 0$) as the sample size becomes larger.

Case (ii)

In what follows we will use the row equivalent form of β that appears in (6.23). Consequently in a 2×2 block-partition of β the lower left block of β or equivalently the upper right block of β' is zero. Thus,

$$\beta = \begin{bmatrix} \underset{p^* \times q}{\beta_{11}} & \underset{p^* \times (r-q)}{\beta_{12}} \\ \underset{k \times q}{\beta_{21}} & \underset{k \times (r-q)}{\beta_{22}} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ 0 & \beta_{22} \end{bmatrix}$$

or

$$\beta' = \begin{bmatrix} \underset{q \times p^*}{\beta'_{11}} & \underset{q \times k}{0} \\ \underset{(r-q) \times p^*}{\beta'_{12}} & \underset{(r-q) \times k}{\beta'_{22}} \end{bmatrix}.$$

We then have the following partitions: $\beta^{(1)} = \begin{bmatrix} \beta_{11} & \beta_{12} \end{bmatrix}$ defined above and $\beta^{(2)} = \begin{bmatrix} \underset{k \times q}{\beta_{21}} & \underset{k \times (r-q)}{\beta_{22}} \end{bmatrix} = \begin{bmatrix} 0 & \beta_{22} \end{bmatrix}$. Note that β_{11} must satisfy the condition $\beta'_{11} C^* = 0$ so that $\beta'_{11} X_t^* = \beta'_{11} C_1^*(L) \varepsilon_t \sim I(0)$, by (6.28).

Then (6.26) becomes

$$|B'_T S^*(\zeta) B_T| =$$

$$\left|\left[\begin{array}{ccc} \zeta\beta_{11}'S_{11}^*\beta_{11} & \zeta\beta_{11}'S_{11}^*\beta_{12} & T^{-1/2}\zeta\beta_{11}'S_{11}^*\bar{\beta}_\perp^{(1)} \\ \zeta\beta_{12}'S_{11}^*\beta_{11} & \zeta(\beta_{12}'S_{11}^*\beta_{12}+\beta_{22}'\beta_{22}) & T^{-1/2}\zeta(\beta_{12}'S_{11}^*\bar{\beta}_\perp^{(1)}+\beta_{22}'\bar{\beta}_\perp^{(2)}) \\ T^{-1/2}\zeta\bar{\beta}_\perp^{(1)'}S_{11}^*\beta_{11} & T^{-1/2}\zeta(\bar{\beta}_\perp^{(1)'}S_{11}^*\beta_{12}+\bar{\beta}_\perp^{(2)'}\beta_{22}) & T^{-1}\zeta(\bar{\beta}_\perp^{(1)'}S_{11}^*\bar{\beta}_\perp^{(1)}+\bar{\beta}_\perp^{(2)'}\bar{\beta}_\perp^{(2)}) \end{array}\right]\right.$$

$$\left.-\left[\begin{array}{ccc} \beta_{11}'S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{11} & \beta_{11}'S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{12} & \beta_{11}'S_{10}^*S_{00}^{*-1}S_{01}^*\bar{\beta}_\perp^{(1)} \\ \beta_{12}'S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{11} & \beta_{12}'S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{12} & \beta_{12}'S_{10}^*S_{00}^{*-1}S_{01}^*\bar{\beta}_\perp^{(1)} \\ \bar{\beta}_\perp^{(1)'}S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{11} & \bar{\beta}_\perp^{(1)'}S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{12} & \bar{\beta}_\perp^{(1)'}S_{10}^*S_{00}^{*-1}S_{01}^*\bar{\beta}_\perp^{(1)} \end{array}\right]\right|. \tag{6.32}$$

Since $\beta_{12}'X_t^*$ is assumed to be $I(1)$ the first term of (6.32) needs to be rescaled. Let now $\Upsilon_T = \begin{bmatrix} I_q & 0 & 0 \\ 0 & T^{-1/2}I_{r-q} & 0 \\ 0 & 0 & I_{p-r} \end{bmatrix}$ then

$$|\Upsilon_T'B_T'S^*(\zeta)B_T\Upsilon_T| =$$

$$\left|\left[\begin{array}{ccc} \zeta\beta_{11}'S_{11}^*\beta_{11} & o_p(1) & o_p(1) \\ o_p(1) & \zeta T^{-1}\beta_{12}'S_{11}^*\beta_{12}+o_p(1) & \zeta T^{-1}\beta_{12}'S_{11}^*\bar{\beta}_\perp^{(1)}+o_p(1) \\ o_p(1) & \zeta T^{-1}\bar{\beta}_\perp^{(1)'}S_{11}^*\beta_{12}+o_p(1) & \zeta T^{-1}\bar{\beta}_\perp^{(1)'}S_{11}^*\bar{\beta}_\perp^{(1)}+o_p(1) \end{array}\right]\right.$$

$$\left.-\left[\begin{array}{ccc} \beta_{11}'S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{11} & o_p(1) & o_p(1) \\ o_p(1) & o_p(1) & o_p(1) \\ o_p(1) & o_p(1) & o_p(1) \end{array}\right]\right|$$

$$= \left|\begin{array}{cc} \zeta\beta_{11}'S_{11}^*\beta_{11}-\beta_{11}'S_{10}^*S_{00}^{*-1}S_{01}^*\beta_{11} & o_p(1) \\ o_p(1) & \zeta T^{-1}B^{*'}S_{11}^*B^*+o_p(1) \end{array}\right| \tag{6.33}$$

where now $B^* = \begin{bmatrix} \beta_{12} & \bar{\beta}_\perp^{(1)} \end{bmatrix}$, $p^* \times (p-q)$.

The $o_p(1)$ blocks are blocks that were $O_p(1)$ before scaling by Υ_T because they were products of averages of products of either two $I(0)$ processes ($\beta_{11}'S_{10}^*$, S_{00}^*) or an $I(0)$ and an $I(1)$ process ($B^{*'}S_{10}^*$, $B^{*'}S_{11}^*\beta_{11}$), see (A.9) and (A.10) in Appendix A.

In order to find the limit of (6.33) we need the following:

$$S_{00}^* \xrightarrow{p} \Sigma_{00}^* = H'\Sigma_{00}H \tag{6.34}$$

$$\beta_{11}'S_{10}^* \xrightarrow{p} \Sigma_{\beta_{11}0}^* = H'\Sigma_{\beta 0}H \tag{6.35}$$

$$\beta_{11}'S_{11}^*\beta_{11} \xrightarrow{p} \Sigma_{\beta_{11}\beta_{11}}^* = H'\Sigma_{\beta\beta}H \tag{6.36}$$

and $S_{00} \xrightarrow{p} \Sigma_{00}$, $\beta' S_{10} \xrightarrow{p} \Sigma_{\beta 0}$ and $\beta' S_{11}\beta \xrightarrow{p} \Sigma_{\beta\beta}$ by the WLLN (see also Johansen (1996, Lemma 10.3)). Furthermore, we define

$$Var\begin{bmatrix} \Delta X_t^* \\ \beta_{11}' X_t^* \end{bmatrix} = \begin{bmatrix} \Sigma_{00}^* & \Sigma_{0\beta_{11}}^* \\ \Sigma_{\beta_{11}0}^* & \Sigma_{\beta_{11}\beta_{11}}^* \end{bmatrix}$$

for the SM and

$$Var\begin{bmatrix} \Delta X_t \\ \beta' X_t \end{bmatrix} = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}$$

for the DGP.

Thus,

$$|\Upsilon_T' B_T' S^*(\zeta) B_T \Upsilon_T| =$$

$$\begin{vmatrix} \zeta\beta_{11}' S_{11}^* \beta_{11} - \beta_{11}' S_{10}^* S_{00}^{*-1} S_{01}^* \beta_{11} & o_p(1) \\ o_p(1) & \zeta T^{-1} B^{*'} S_{11}^* B^* + o_p(1) \end{vmatrix} \xrightarrow{d}$$

$$= \begin{vmatrix} \zeta\Sigma_{\beta_{11}\beta_{11}}^* - \Sigma_{\beta_{11}0}^* \Sigma_{00}^{*-1} \Sigma_{0\beta_{11}}^* & 0 \\ 0 & \zeta B^{*'} C^* \int_0^1 \tilde{W}\tilde{W}' du C^{*'} B^* \end{vmatrix}$$

$$= |\zeta\Sigma_{\beta_{11}\beta_{11}}^* - \Sigma_{\beta_{11}0}^* \Sigma_{00}^{*-1} \Sigma_{0\beta_{11}}^*||\zeta B^{*'} C^* \int_0^1 \tilde{W}\tilde{W}' du C^{*'} B^*| = 0 \qquad (6.37)$$

by (6.34)-(6.36) for the first factor and by (6.29) for the second.

Thus (6.37) indicates that there are q non-zero and $(p-q)$ zero roots in the limit, which suggests that q cointegrating vectors can be detected in the underspecified model as the sample size becomes large. The stochastic matrix $B^{*'} C^* \int_0^1 \tilde{W}\tilde{W}' du C^{*'} B^*$ with dimensions $(p-q) \times (p-q)$ has rank $(p^* - q)$ almost surely (see also section 6.1.1) and the $k \equiv (p - p^*)$ zero roots appear in the second factor of (6.37) by construction.

6.2.2 Consistency

The analysis of consistency is carried out only for case (ii) where some cointegrating vectors can be detected.

For the analysis of consistency we use the partition of β that appears in subsection 6.2.1,

$$\beta = \begin{bmatrix} \underset{p^*\times q}{\beta_{11}} & \underset{p^*\times(r-q)}{\beta_{12}} \\ \underset{k\times q}{\beta_{21}} & \underset{k\times(r-q)}{\beta_{22}} \end{bmatrix}$$

where $\beta_{21} = 0$. We define $B = \begin{bmatrix} \underset{p^*\times q}{\beta_{11}} & \underset{p^*\times(p^*-q)}{\bar{\beta}_{11\perp}} \end{bmatrix}$ and $B^{-1} = \begin{bmatrix} \underset{q\times p^*}{\bar{\beta}'_{11}} \\ \underset{(p^*-q)\times p^*}{\beta'_{11\perp}} \end{bmatrix}$ where $\bar{\beta}_{11\perp} = \beta_{11\perp}(\beta'_{11\perp}\beta_{11\perp})^{-1}$, $\bar{\beta}_{11} = \beta_{11}(\beta'_{11}\beta_{11})^{-1}$ and $\beta'_{11}\beta_{11\perp} = 0$. B and B^{-1} are such that the following relationship holds

$$B^{-1}B = BB^{-1} = \beta_{11}\bar{\beta}'_{11} + \bar{\beta}_{11\perp}\beta'_{11\perp} = I_{p^*}. \tag{6.38}$$

We have shown in subsection 6.2.1 that the tests detect q cointegrating vectors, hence under the assumption of cointegration Π^* in (6.22) has rank q. Thus, Π^* can be expressed as $\Pi^* = \alpha_{11}\beta'_{11}$, where α_{11}[9] and β_{11} are $p^* \times q$ matrices of rank q. The SM then takes the form

$$\Delta X^*_t = \alpha_{11}\beta'_{11}X^*_{t-1} + e^*_t \tag{6.39}$$

with $Var(e^*_t) \equiv \Lambda^*$.

---

[9] Partitioning α similarly to β we obtain $\alpha = \begin{bmatrix} \underset{p^*\times q}{\alpha_{11}} & \underset{p^*\times(r-q)}{\alpha_{12}} \\ \underset{k\times q}{\alpha_{21}} & \underset{k\times(r-q)}{\alpha_{22}} \end{bmatrix}$ where $H'\alpha = \alpha^{(1)} = \begin{bmatrix} \underset{p^*\times q}{\alpha_{11}} & \underset{p^*\times(r-q)}{\alpha_{12}} \end{bmatrix}$ and α_{11} are the adjustment coefficients that correspond to the cointegrating vectors detectable in the under-specified model.

Let $\hat{\beta}_{11}$, $\hat{\alpha}_{11}$ and $\hat{\Lambda}^*$ be the maximum likelihood estimators of β_{11}, α_{11} and Λ^* calculated from the SM (6.22) (using (6.24)). The parameters β_{11} and α_{11} correspond to the $p^* \times q$ submatrices of β, α in the DGP.

For the analysis of consistency we use a linear transformation of the columns of $\hat{\beta}_{11}$, which also maximises the likelihood function (see subsection 6.1.2, footnote 8), given by

$$\begin{aligned} \tilde{\beta}_{11} &= \hat{\beta}_{11}(\bar{\beta}_{11}'\hat{\beta}_{11})^{-1} \qquad (6.40) \\ &= \beta_{11} + \bar{\beta}_{11\perp}\beta_{11\perp}'\hat{\beta}_{11}(\bar{\beta}_{11}'\hat{\beta}_{11})^{-1} \\ &= \beta_{11} + \bar{\beta}_{11\perp}b_1 \end{aligned}$$

where the second equality follows by using (6.38) and $b_1 = \beta_{11\perp}'\tilde{\beta}_{11}$.

We also define $\tilde{\alpha}_{11} = \hat{\alpha}_{11}\hat{\beta}_{11}'\bar{\beta}_{11}$ such that $\tilde{\alpha}_{11}\tilde{\beta}_{11}' = \hat{\alpha}_{11}\hat{\beta}_{11}'$ and

$$\begin{aligned} \tilde{\alpha}_{11} &= S_{01}^*\hat{\beta}_{11}(\hat{\beta}_{11}'S_{11}^*\hat{\beta}_{11})^{-1}\hat{\beta}_{11}'\bar{\beta}_{11} \\ &= S_{01}^*\tilde{\beta}_{11}(\tilde{\beta}_{11}'S_{11}^*\tilde{\beta}_{11})^{-1} \end{aligned}$$

where the first equality follows from the fact that $\hat{\alpha}_{11} = S_{01}^*\hat{\beta}_{11}(\hat{\beta}_{11}'S_{11}^*\hat{\beta}_{11})^{-1}$ (see equation (2.13) in section 2.3) given that we can estimate β_{11} by solving (6.24).

In addition,

$$\begin{aligned} \hat{\Lambda}^* &= S_{00}^* - S_{01}^*\hat{\beta}_{11}(\hat{\beta}_{11}'S_{11}^*\hat{\beta}_{11})^{-1}\hat{\beta}_{11}'S_{10}^* \\ &= S_{00}^* - S_{01}^*\tilde{\beta}_{11}(\tilde{\beta}_{11}'S_{11}^*\tilde{\beta}_{11})^{-1}\tilde{\beta}_{11}'S_{10}^* \end{aligned}$$

where the first equality follows from the expression for the estimator of the variance-covariance matrix of the errors in the SM (see equation (2.14) in section 2.3) and the second equality follows from the definition of $\tilde{\beta}_{11}$.

The proposition below establishes the consistency of the maximum likelihood estimator for the cointegrating vectors in the sense that the estimator from the SM converges in probability to a submatrix of the parameter, β, in the DGP, which is associated with the included variables.

Proposition 6.2. The estimator of the cointegrating vectors, $\tilde{\beta}_{11}$, associated with the underspecified model (6.22) *converges to vectors in $sp(\beta)$, i.e. $\tilde{\beta}_{11} \xrightarrow{p} \beta_{11}$.*

Proof. The equations (6.25) and (6.26) have the same eigenvalues but (6.26) has eigenvectors $B_T^{-1}\hat{V}$ where $\hat{V} = \left[\underset{p\times q}{\hat{\beta}_q} \quad \underset{p\times(p-q)}{\hat{V}_2} \right]$ is the matrix whose columns are the eigenvectors of (6.25) and $\hat{\beta}_q = H\hat{\beta}_{11} = \begin{bmatrix} \hat{\beta}_{11} \\ 0 \end{bmatrix}$. The eigenvalues of (6.26) converge to the eigenvalues of (6.37). Thus, the space spanned by the q first eigenvectors of (6.26) which correspond to the q largest eigenvalues converges to the space spanned by vectors with zeros in the last $(p-q)$ positions. The space spanned by the first q eigenvectors of (6.26) is $sp(B_T^{-1}\hat{\beta}_q) = sp(B_T^{-1}\tilde{\beta}_q)$ where $\tilde{\beta}_q = H\tilde{\beta}_{11}$ and

$$B_T^{-1}\tilde{\beta}_q = \begin{bmatrix} \bar{\beta}' \\ T^{1/2}\bar{\beta}_{\perp}' \end{bmatrix} \tilde{\beta}_q = \begin{bmatrix} (\beta'\beta)^{-1}\beta'\tilde{\beta}_q \\ T^{1/2}\bar{\beta}_{\perp}'\tilde{\beta}_q \end{bmatrix}.$$

First we analyse block (1,1). Using the formula for the partitioned inverse we have,

$$\begin{aligned}
(\beta'\beta)^{-1} &= \begin{bmatrix} \beta_{11}'\beta_{11} & \beta_{11}'\beta_{12} \\ \beta_{12}'\beta_{11} & \beta_{12}'\beta_{12} + \beta_{22}'\beta_{22} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} (\beta_{11}'\beta_{11})^{-1}[I_q + \beta_{11}'\beta_{12}F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1}] & -(\beta_{11}'\beta_{11})^{-1}\beta_{11}'\beta_{12}F \\ -F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1} & F \end{bmatrix}
\end{aligned}$$

where

$$
\begin{aligned}
F &= [\beta_{12}'\beta_{12} + \beta_{22}'\beta_{22} - \beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1}\beta_{11}'\beta_{12}]^{-1} \\
&= [\beta_{22}'\beta_{22} + \beta_{12}'(I_{p^*} - \beta_{11}(\beta_{11}'\beta_{11})^{-1}\beta_{11}')\beta_{12}]^{-1} \\
&= [\beta_{22}'\beta_{22} + \beta_{12}'\bar{\beta}_{11\perp}\beta_{11\perp}'\beta_{12}]^{-1}
\end{aligned}
$$

and the last equality follows from the relationship in (6.38).

Thus,

$$
\begin{aligned}
&(\beta'\beta)^{-1}\beta'\tilde{\beta}_q = \\
&\begin{bmatrix} (\beta_{11}'\beta_{11})^{-1}[I_q + \beta_{11}'\beta_{12}F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1}] & -(\beta_{11}'\beta_{11})^{-1}\beta_{11}'\beta_{12}F \\ -F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1} & F \end{bmatrix} \begin{bmatrix} \beta_{11}'\tilde{\beta}_{11} \\ \beta_{12}'\tilde{\beta}_{11} \end{bmatrix} \\
&= \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}
\end{aligned}
$$

where

$$
\begin{aligned}
A_1 &= (\beta_{11}'\beta_{11})^{-1}[I_q+\beta_{11}'\beta_{12}F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1}]\beta_{11}'\tilde{\beta}_{11}-(\beta_{11}'\beta_{11})^{-1}\beta_{11}'\beta_{12}F\beta_{12}'\tilde{\beta}_{11} \\
&= (\beta_{11}'\beta_{11})^{-1}\beta_{11}'[I_{p^*} + \beta_{12}F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1}\beta_{11}' - \beta_{12}F\beta_{12}']\tilde{\beta}_{11} \\
&= \bar{\beta}_{11}'[I_{p^*} - \beta_{12}F\beta_{12}'(I_{p^*} - \beta_{11}(\beta_{11}'\beta_{11})^{-1}\beta_{11}')]\tilde{\beta}_{11} \\
&= \bar{\beta}_{11}'(I_{p^*} - \beta_{12}F\beta_{12}'\bar{\beta}_{11\perp}\beta_{11\perp}')(\beta_{11} + \bar{\beta}_{11\perp}b_1) \\
&= I_q - \bar{\beta}_{11}'\beta_{12}F\beta_{12}'\bar{\beta}_{11\perp}b_1
\end{aligned}
$$

and

$$
\begin{aligned}
A_2 &= -F\beta_{12}'\beta_{11}(\beta_{11}'\beta_{11})^{-1}\beta_{11}'\tilde{\beta}_{11} + F\beta_{12}'\tilde{\beta}_{11} \\
&= F\beta_{12}'[I_{p^*} - \beta_{11}(\beta_{11}'\beta_{11})^{-1}\beta_{11}']\tilde{\beta}_{11} \\
&= F\beta_{12}'\bar{\beta}_{11\perp}\beta_{11\perp}'(\beta_{11} + \bar{\beta}_{11\perp}b_1) \\
&= F\beta_{12}'\bar{\beta}_{11\perp}b_1 \,.
\end{aligned}
$$

Then we analyse $\beta_{\perp}'\tilde{\beta}_q$ which appears in block (2,1). Partitioning $\beta_{\perp}'$ as in $\beta_{\perp} = \left[\begin{array}{cc} \underset{(p-r)\times p^*}{\beta_{\perp}^{(1)'}} & \underset{(p-r)\times k}{\beta_{\perp}^{(2)'}} \end{array}\right]$ we obtain

$$\beta_{\perp}'\tilde{\beta}_q = \left[\begin{array}{cc} \beta_{\perp}^{(1)'} & \beta_{\perp}^{(2)'} \end{array}\right] \left[\begin{array}{c} \tilde{\beta}_{11} \\ 0 \end{array}\right] = \beta_{\perp}^{(1)'}\tilde{\beta}_{11} = \beta_{\perp}^{(1)'}(\beta_{11} + \bar{\beta}_{11\perp}b_1) = \beta_{\perp}^{(1)'}\bar{\beta}_{11\perp}b_1$$

by the assumption $\beta'\beta_{\perp} = 0$ (or $\beta_{\perp}'\beta = 0$) which gives

$$\beta_{\perp}'\beta = \left[\begin{array}{cc} \beta_{\perp}^{(1)'} & \beta_{\perp}^{(2)'} \end{array}\right] \left[\begin{array}{cc} \beta_{11} & \beta_{12} \\ 0 & \beta_{22} \end{array}\right] = \left[\begin{array}{cc} \beta_{\perp}^{(1)'}\beta_{11} & \beta_{\perp}^{(2)'}\beta_{12} + \beta_{\perp}^{(2)'}\beta_{22} \end{array}\right] = 0$$

and therefore $\beta_{\perp}^{(1)'}\beta_{11} = 0$.

Thus,

$$B_T^{-1}\tilde{\beta}_q = \left[\begin{array}{c} \underset{q\times q}{I_q - \bar{\beta}_{11}'\beta_{12}F\beta_{12}'\bar{\beta}_{11\perp}b_1} \\ \underset{(r-q)\times q}{F\beta_{12}'\bar{\beta}_{11\perp}b_1} \\ \underset{(p-r)\times q}{T^{1/2}\beta_{\perp}^{(1)'}\bar{\beta}_{11\perp}b_1} \end{array}\right]. \tag{6.41}$$

By the form of (6.37) the last two blocks of (6.41) should converge to zero (in other words $sp(B_T^{-1}\tilde{\beta}_q)$ should converge to the space spanned by vectors with zeros in the last $(p-q)$ coordinates. A necessary condition for this is $T^{1/2}b_1 \xrightarrow{p} 0$. Then $sp(B_T^{-1}\tilde{\beta}_q) \xrightarrow{p} sp(\left[\begin{array}{c} I_q \\ 0 \end{array}\right])$.

From (6.40) we obtain $T^{1/2}(\tilde{\beta}_{11} - \beta_{11}) = \bar{\beta}_{11\perp}(T^{1/2}b_1) \xrightarrow{p} 0$ and that $(\tilde{\beta}_{11} - \beta_{11}) = o_p(T^{-1/2})$. ■

We then consider the probability limits of $\tilde{\alpha}_{11}$ and $\hat{\Lambda}^*$ obtained from the underspecified model. We first partition α and β conformably with $X_t = \left[\begin{array}{c} X_t^* \\ X_t^{(k)} \end{array}\right]$ (see also subsection 6.2.1) and we use the transformed, row equivalent form of β. Then, the DGP (6.21) becomes,

$$\left[\begin{array}{c} \Delta X_t^* \\ \Delta X_t^{(k)} \end{array}\right] = \left[\begin{array}{cc} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{array}\right] \left[\begin{array}{cc} \beta_{11}' & 0 \\ \beta_{12}' & \beta_{22}' \end{array}\right] \left[\begin{array}{c} X_{t-1}^* \\ X_{t-1}^{(k)} \end{array}\right] + \left[\begin{array}{c} \varepsilon_t^* \\ \varepsilon_t^{(k)} \end{array}\right]^{10}.$$

The part of the DGP that corresponds to the included variables is

$$\Delta X_t^* = \alpha_{11}\beta_{11}^{'}X_{t-1}^* + \alpha_{12}(\beta_{12}^{'}X_{t-1}^* + \beta_{22}^{'}X_{t-1}^{(k)}) + \varepsilon_t^*$$

or

$$\Delta X_t^* = \alpha_{11}\beta_{11}^{'}X_{t-1}^* + \alpha_{12}Z_{t-1} + \varepsilon_t^* \tag{6.42}$$

where $\varepsilon_t^* = H^{'}\varepsilon_t \sim i.i.d.(0, \Omega^*)$, $\Omega^* = H^{'}\Omega H$ and $Z_{t-1} = \beta_{12}^{'}X_{t-1}^* + \beta_{22}^{'}X_{t-1}^{(k)} \sim I(0)$, is the part of the DGP that cannot be estimated due to the omission of $X_t^{(k)}$. Using the full sample, (6.42) can be written as

$$\Delta X^* = \alpha_{11}\beta_{11}^{'}X_{-1}^* + \alpha_{12}Z_{-1} + \varepsilon^* \tag{6.43}$$

where ΔX^*, X_{-1}^*, ε^* are $p^* \times T$, Z_{-1} is $(r-q)\times T$ and they are the full sample counterparts of ΔX_t^*, X_{t-1}^*, ε_t^* and Z_{t-1} respectively.

Using the partitioned form of X_t and β,

$$\begin{aligned}
\Sigma_{\beta\beta} &= Var(\beta^{'}X_{t-1}) = E(\beta^{'}X_{t-1}X_{t-1}^{'}\beta) \qquad (6.44)\\
&= \begin{bmatrix} E(\beta_{11}^{'}X_{t-1}^*X_{t-1}^{*'}\beta_{11}) & E(\beta_{11}^{'}X_{t-1}^*Z_{t-1}^{'}) \\ E(Z_{t-1}X_{t-1}^{*'}\beta_{11}) & E(Z_{t-1}Z_{t-1}^{'}) \end{bmatrix} \\
&\equiv \begin{bmatrix} \Sigma_{\beta_{11}\beta_{11}}^* & \Sigma_{\beta_{11}Z}^* \\ \Sigma_{Z\beta_{11}}^* & \Sigma_{ZZ}^* \end{bmatrix}
\end{aligned}$$

and the second equality follows from the fact that there are no deterministic terms in the DGP.

The proposition below relates to the 'inconsistency' of $\tilde{\alpha}_{11}$ and $\hat{\Lambda}^*$ in the sense that their probability limits are different from the parameters, in the underspecified model, that they aim to estimate.

Proposition 6.3. The estimators $\tilde{\alpha}_{11}$ *and* $\hat{\Lambda}^*$ *are 'inconsistent' for the parameters* α_{11} and Ω^* *in* (6.42) *in the sense that they do not converge to the submatrices of* α *and* Ω *(parameters of the DGP) that correspond to the included variables i.e.* plim $\tilde{\alpha}_{11} \neq \alpha_{11}$ *and* plim $\hat{\Lambda}^* \neq \Omega^*$.

Proof. Since β_{11} can be estimated consistently (see Proposition 6.2)

$$plim\ \tilde{\alpha}_{11} = plim\ S^*_{01}\beta_{11}(\beta^{'}_{11}S^*_{11}\beta_{11})^{-1} = plim\ [T^{-1}\Delta X^* X^{*'}_{-1}\beta_{11}(T^{-1}\beta^{'}_{11}X^*_{-1}X^{*'}_{-1}\beta_{11})^{-1}]$$

where the second equality is due to the absence of deterministic terms in the SM. Substituting for ΔX^* as it is given in (6.43) and using Slutsky's Theorem,

$$\begin{aligned} & plim\tilde{\alpha}_{11} \qquad (6.45) \\ = \ & \alpha_{11} + \alpha_{12} plim[(T^{-1}Z_{-1}X^{*'}_{-1}\beta_{11})][plim(T^{-1}\beta^{'}_{11}X^*_{-1}X^{*'}_{-1}\beta_{11})]^{-1} \\ = \ & \alpha_{11} + \alpha_{12}\Sigma^*_{Z\beta_{11}}\Sigma^{*-1}_{\beta_{11}\beta_{11}} \end{aligned}$$

and the probability limits equal the corresponding population moments since the process $\beta^{'}X_{t-1}$ (and therefore $\beta_{11}X^*_{t-1}$ and Z_{t-1}) is stationary and ergodic (see subsection 6.1.1). (6.45) shows that $\tilde{\alpha}_{11}$ is 'inconsistent' (or asymptotically biased) unless $\alpha_{12} = 0$ or $plim$ $(T^{-1}Z_{-1}X^{*'}_{-1}\beta_{11}) = 0$. A stronger condition to achieve consistency is $Z_{-1}X^{*'}_{-1}\beta_{11} = 0$ i.e. Z_{-1} is orthogonal to $X^{*'}_{-1}\beta_{11}$.

For the estimator of the variance-covariance matrix of the errors (again using the consistency of $\tilde{\beta}_{11}$) we have

$$\begin{aligned} plim\ \hat{\Lambda}^* &= plim\ [S^*_{00} - S^*_{01}\beta_{11}(\beta^{'}_{11}S^*_{11}\beta_{11})^{-1}\beta^{'}_{11}S^*_{10}] \\ &= plim\ (T^{-1}\Delta X^*\Delta X^{*'}) \\ &\quad -plim\ [T^{-1}\Delta X^* X^{*'}_{-1}\beta_{11}(T^{-1}\beta^{'}_{11}X^*_{-1}X^{*'}_{-1}\beta_{11})^{-1}T^{-1}\beta^{'}_{11}X^*_{-1}\Delta X^{*'}] \\ &= plim\ T^{-1}\Delta X^* M^* \Delta X^{*'} \end{aligned}$$

where $M^* = I_T - X^{*'}_{-1}\beta_{11}(\beta^{'}_{11}X^*_{-1}X^{*'}_{-1}\beta_{11})^{-1}\beta^{'}_{11}X^*_{-1}$. Substituting for ΔX^* using (6.43),

$$plim\ \hat{\Lambda}^* = plim\ T^{-1}[\alpha_{12}Z_{-1}M^*Z^{'}_{-1}\alpha^{'}_{12} + \alpha_{12}Z_{-1}M^*\varepsilon^{*'} + \varepsilon^* M^* Z^{'}_{-1}\alpha^{'}_{12} + \varepsilon^* M^* \varepsilon^{*'}]$$

and $M^*Z^{'}_{-1}$ can be viewed as the residuals from the regression of $Z^{'}_{-1}$ on $\beta^{'}_{11}X^*_{-1}$. By the WLLN we have

$$plim\ T^{-1}Z_{-1}M^*\varepsilon^{*'} = E(Z_{-1}M^*\varepsilon^{*'}) = 0$$

since $E(Z_{-1}M^*\varepsilon^{*'}) = E[E(Z_{-1}M^*\varepsilon^{*'}|\mathcal{X}_{t-1})] = E[Z_{-1}M^*E(\varepsilon^{*'}|\mathcal{X}_{t-1})] = 0$, where $\mathcal{X}_{t-1}$ is the minimal σ-field generated by the random vector X_{t-1}. Furthermore,

$$plim\ T^{-1}\beta^{'}_{11}X^*_{-1}\varepsilon^{*'} = E(\beta^{'}_{11}X^*_{-1}\varepsilon^{*'}) = 0$$

since $E(\beta^{'}_{11}X^*_{-1}\varepsilon^{*'}) = E[E(\beta^{'}_{11}X^*_{-1}\varepsilon^{*'}|\mathcal{X}_{t-1})] = E[\beta^{'}_{11}X^*_{-1}E(\varepsilon^{*'}|\mathcal{X}_{t-1})] = 0$ (see also footnote 10). Hence,

$$\begin{aligned} plim\ \hat{\Lambda}^* &= plim\ (T^{-1}\varepsilon^*\varepsilon^{*'}) + plim\ (T^{-1}\alpha_{12}Z_{-1}M^*Z^{'}_{-1}\alpha^{'}_{12}) \qquad (6.46) \\ &= \Omega^* + \alpha_{12}(\Sigma^*_{ZZ} - \Sigma^*_{Z\beta_{11}}\Sigma^{*-1}_{\beta_{11}\beta_{11}}\Sigma^*_{\beta_{11}Z})\alpha^{'}_{12} \end{aligned}$$

since ε^* and Z_{-1} are stationary random variables and by the WLLN the probability limits in (6.46) equal their corresponding population moments. Therefore, $\hat{\Lambda}^*$ is 'inconsistent' unless $\alpha_{12} = 0$. ■

From (6.45) and (6.46) we observe that in order to gauge the magnitude of the inconsistency (or the asymptotic bias), ($p$lim $\tilde{\alpha}_{11} - \alpha_{11}$) and ($p$lim $\hat{\Lambda}^* - \Omega^*$) we need to estimate α_{12}, $\Sigma^*_{\beta_{11}\beta_{11}}$, $\Sigma^*_{Z\beta_{11}}$ and Σ^*_{ZZ} which is infeasible.

6.3 Monte Carlo simulations

In this section we present the results of some Monte Carlo experiments in order to illustrate the asymptotic results presented in sections 6.1 and 6.2 and to give some idea about the consequences of possible misspecifications of the SM, in finite samples, in the case of irrelevant or omitted variables.

We use experimental designs similar to those in Podivinsky (1998, p. 6)[11], which allow for up to two cointegrating vectors among up to three variables. All calculations were done using Ox 3.00 (see Doornik (1999)). The number of replications is 10,000 for all experiments. We use the 95% tabulated asymptotic critical values from Osterwald-Lenum (1992, Case 0), thus the tests are carried out at 5% significance level.

[11] For DGP1 we use different cointegrating vectors, β, and for DGP3 and DGP4 we use a different adjustment coefficient matrix, α.

6.3.1 Inference about the cointegrating rank (r)

Irrelevant variables

The first DGP (DGP1) consists of two variables and one cointegrating vector and the second (DGP2) of three variables and two cointegrating vectors. These are given below in error correction forms,

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} -0.4 \\ 0.1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \qquad \text{(DGP1)}$$

and

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \end{bmatrix} = \begin{bmatrix} -0.4 & 0.1 \\ 0.1 & 0.2 \\ 0.1 & 0.3 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ 1 & -0.5 & -0.5 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \qquad \text{(DGP2)}$$

where $t = 1, 2, \ldots, T$, $\varepsilon_t \sim i.i.d.N_j(0, I)$, $\varepsilon_t^{'} = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix}$ with $j = 2$ for DGP1 and $\varepsilon_t^{'} = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} & \varepsilon_{3t} \end{bmatrix}$ with $j = 3$ for DGP2.

The SMs used for performing the cointegration tests consist of three variables for DGP1 and four variables for DGP2. Thus, we augment the DGPs with an independent random walk which has innovations with zero mean and unit variance.

Tables 6.1 and 6.2 show the rejection frequencies using the trace and the maximal eigenvalue statistics for different rank hypotheses and different sample sizes. The simulation results agree with the asymptotic analysis of section 6.1, according to which the LR tests for cointegration should detect the true number of cointegrating vectors, r (i.e. the cointegrating rank in the DGP) as the sample size becomes large, when an overspecified SM is used for cointegration testing.

Table 6.1. Rejection frequencies using the trace and the maximal eigenvalue statistics (DGP1).

Sample size / Rank hypothesis	50	100	150	500	800
			Trace statistic		
$r = 0$	0.8024	0.9999	1	1	1
$r \leq 1$	0.0524	0.0498	0.0499	0.04550	0.04670
			Maximal eigenvalue statistic		
$r = 0$	0.8515	1	1	1	1
$r \leq 1$	0.0516	0.0471	0.0474	0.0449	0.0440

Table 6.2. Rejection frequencies using the trace and the maximal eigenvalue statistics (DGP2).

Sample size / Rank hypothesis	50	100	150	500	800
			Trace statistic		
$r = 0$	1	1	1	1	1
$r \leq 1$	0.9945	1	1	1	1
$r \leq 2$	0.0989	0.0762	0.0618	0.0556	0.0446
			Maximal eigenvalue statistic		
$r = 0$	1	1	1	1	1
$r \leq 1$	0.9956	1	1	1	1
$r \leq 2$	0.0960	0.0732	0.0612	0.0540	0.0460

From Tables 6.1 and 6.2 we can see that we tend to accept the hypothesis of $r = 1$ and $r = 2$ for DGP1 and DGP2 respectively, since the corresponding rejection frequencies for these hypotheses are quite close to the nominal size of the tests.

Next we conduct another experiment in which we control the local power of the test. We use the same DGP as in DGP1 but we let the adjustment coefficients vary so that the single cointegrating vector can be detected with high, medium or low asymptotic local power. Thus, the setup of the DGP is such that the cointegrating vector has adjustment coefficients that tend to zero as the sample size becomes large. In other words there is no cointegration ($r = 0$) under the null hypothesis and under the local alternative there is one cointegrating vector ($r = 1$). The DGP then takes the form,

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \alpha_{(1)}\beta'_{(1)} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \qquad \text{(DGP1}^*\text{)}$$

where $t = 1, 2, \ldots, T$, $\alpha'_{(1)} = \begin{bmatrix} \alpha_1 & \alpha_2 \end{bmatrix}$, $\beta'_{(1)} = \begin{bmatrix} 1 & -1 \end{bmatrix}$ and $\varepsilon'_t = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix} \sim$ $i.i.d.N_2(0, I)$.

Under the local alternative of one cointegrating vector the asymptotic local power depends only on two parameters f and g given by[12]

$$T^{-1}f = \beta'_{(1)}\alpha_{(1)} = \alpha_1 - \alpha_2$$

and

$$(T^{-1}g)^2 = \alpha'_{(1)}\alpha_{(1)}\beta'_{(1)}\beta_{(1)} - (\beta'_{(1)}\alpha_{(1)})^2 = (\alpha_1 + \alpha_2)^2$$

see Johansen (1996, p. 209)[13]. Therefore we can express the adjustment coefficients in terms of the parameters that affect the local power as

$$\alpha_1 = (f + g)/2T$$

and

$$\alpha_2 = (g - f)/2T.$$

We can then control local power by choosing combinations of f and g that correspond to a particular level of local power and use them in the DGP. We use six pairs of (f, g), two pairs for each power level, high, medium and low as shown in Table 6.3. The values in Table 6.3

[12] Since under the null we assume $r = 0$, the matrices $C = I_2$, $\beta_\perp$ and $\alpha_\perp$ that appear in the definitions of f and g in Johansen (1996, p. 209) have full rank and from the properties of errors in DGP1* we have $\Omega = I_2$ (see also sections 5.1 and 5.2).

[13] In Johansen's (1996, equation 14.2) notation, the deviation from the null is $T^{-1}\alpha_1\beta'_1$, corresponding to $\alpha_{(1)}\beta'_{(1)}$ used here. Thus, $f = \beta'_1\alpha_1$ in Johansen (1996), after simplification, corresponds to $T\beta'_{(1)}\alpha_{(1)}$. A similar adjustment is required for g. Hence, f and g change with T across the simulations.

were taken from Johansen (1996, Table 15.6) and given DGP1* we have $(p - r) = 2$. The values that appear in Table 6.3 were computed using $T = 400$ and 2,000 replications.

Table 6.3. f, g and asymptotic local power

$(f,\ g)$	power
(-3, 12)	0.850 (high)
(-18, 12)	0.830 (high)
(-18, 0)	0.565 (medium)
(-15, 6)	0.513 (medium)
(-6, 6)	0.272 (low)
(-12, 0)	0.269 (low)

For the particular pairs and sample sizes used we also calculated the local power (using 10,000 replications) as the rejection frequencies of $r = 0$ under DGP1* to verify the distinction among high, medium and low power levels. The results of this experiment show that the distinction made to the power levels applies, since the rejection frequencies for all sample sizes are approximately 0.8 for (-3, 12) and (-18, 12), 0.5 for (-18, 0) and (-15, 6) and 0.25 for (-6, 6) and (-12, 0). However, the rejection frequencies do not always approach the limit expected monotonically. The detailed tables for this experiment appear in Appendix F, Tables F.1 and F.2.

Tables 6.4 and 6.5 show the rejection frequencies of the hypothesis of $r = 0$ against the alternative of $r = 1$ computed from the overspecified model, using the trace and the maximal eigenvalue statistics respectively. We observe that for both statistics the rejection frequencies are systematically below the prespecified level, which suggests that including an irrelevant variable in the SM reduces the power of the LR tests for cointegration.

Table 6.4. Rejection frequencies of the hypothesis $r = 0$ using the trace statistic.

Sample size / Power level	50	100	150	500	800
0.850	0.6601	0.6715	0.6801	0.6787	0.6845
0.830	0.5314	0.4973	0.5024	0.4816	0.4787
0.565	0.2986	0.2756	0.2532	0.2512	0.2472
0.513	0.2823	0.2582	0.2515	0.2359	0.2392
0.272	0.1503	0.1381	0.1386	0.1267	0.1315
0.269	0.1690	0.1623	0.1584	0.1525	0.1491

Table 6.5. Rejection frequencies of the hypothesis $r = 0$ using the maximal eigenvalue statistic.

Sample size / Power level	50	100	150	500	800
0.850	0.6457	0.6684	0.6765	0.6764	0.6811
0.830	0.5547	0.5268	0.5197	0.5041	0.5000
0.565	0.2986	0.2610	0.2557	0.2494	0.2424
0.513	0.2699	0.2492	0.2400	0.2256	0.2270
0.272	0.1381	0.1222	0.1236	0.1146	0.1216
0.269	0.1443	0.1395	0.1399	0.1308	0.1214

Omitted variables

Again we use two DGPs which are chosen on the basis of the asymptotic analysis to reflect the cases $(r-k) = 0$ and $(r-k) > 0$, treated in section 6.2. Both DGPs consist of three variables, but the first one (DGP3) has one cointegrating vector involving all three variables whereas the second one (DGP4) has two cointegrating vectors both involving all three variables. Thus,

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0.1 \\ -0.7 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \tag{DGP3}$$

and

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \end{bmatrix} = \begin{bmatrix} 0.433 & 0.233 \\ 0.5 & 0.3 \\ 0.366 & 0.366 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ 1 & -0.5 & -0.5 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \tag{DGP4}$$

where $t = 1, 2, \ldots, T$, $\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} & \varepsilon_{3t} \end{bmatrix}' \sim i.i.d.N_3(0, I)$ for DGP3 and DGP4.

The SMs used for the calculation of the trace and maximal eigenvalue statistics include only X_{1t} and X_{2t}.

Tables 6.6 and 6.7 show the rejection frequencies for various rank hypotheses using the trace and the maximal eigenvalue statistics, for different sample sizes.

Table 6.6. Rejection frequencies using the trace and the maximal eigenvalue statistics (DGP3).

Sample size / Rank hypothesis	50	100	150	500	800
	Trace statistic				
$r = 0$	0.1363	0.1474	0.1517	0.1571	0.1606
$r \leq 1$[14]	0.0166	0.0168	0.0178	0.0162	0.0164
	Maximal eigenvalue statistic				
$r = 0$	0.1379	0.1503	0.1563	0.1583	0.1627
$r \leq 1$	0.0166	0.0168	0.0178	0.0162	0.0164

Table 6.7. Rejection frequencies using the trace and the maximal eigenvalue statistics (DGP4).

Sample size / Rank hypothesis	50	100	150	500	800
	Trace statistic				
$r = 0$	1	1	1	1	1
$r \leq 1$	0.0747	0.0686	0.0669	0.0722	0.0686
	Maximal eigenvalue statistic				
$r = 0$	1	1	1	1	1
$r \leq 1$	0.0747	0.0686	0.0669	0.0722	0.0686

From Table 6.6 we can see that the tests might not detect any cointegrating vectors (low rejection frequencies of $r = 0$, especially for small sample sizes) which is what we expected since $(r - k) = 0$ (see section 6.2). From Table 6.7 we conclude that with DGP4 the LR tests are very likely to detect one cointegrating vector and this is in accordance with the theoretical finding which suggests that if $(r - k) > 0$ the tests detect at least $(r - k)$ (2-1=1, in this case) cointegrating vectors.

[14] Since there are only two variables in the SM the trace and the maximal eigenvalue statistics for $r \leq 1$ coincide, similarly for the corresponding hypothesis in Table 6.7.

6.3.2 Consistency

Irrelevant variables

We use DGP1 and DGP2 to check the consistency of the cointegrating vectors suggested in Proposition 6.1(i). In DGP1, there is one cointegrating vector and one irrelevant variable, and in DGP2 there are two cointegrating vectors and one irrelevant variable. We use $T = 5{,}000$ and compute the 1%, 5%, 10%, 25%, 50%, 75%, 90% and 99% quantiles of the elements of the estimated cointegrating vector(s) (i.e. the elements of the eigenvectors that correspond to the largest eigenvalue(s)) for each DGP in 10,000 replications. In particular we use the normalised form of the estimated cointegrating vectors ($\tilde{\beta}^{+}$ instead of $\hat{\beta}^{+}$) that is given in (6.15). The reason for using this normalisation is (as shown in subsection 6.1.2) that we can achieve convergence to the true (known) cointegrating vectors and not just the space spanned by them. The simulation results appear in Tables 6.8 and 6.9.

Table 6.8. Quantiles of the elements of the estimated cointegrating vector (DGP1).

$\tilde{\beta}^{+\ddagger}$ / Quantiles	$\tilde{\beta}^{+}_{11}$	$\tilde{\beta}^{+}_{21}$	$\tilde{\beta}^{+}_{31(ir)}$
1%	0.9975	-1.0024	-0.0039
5%	0.9985	-1.0015	-0.0025
10%	0.9989	-1.0011	-0.0017
25%	0.9995	-1.0005	-0.0008
50%	1.0000	-1.0000	-0.0000
75%	1.0005	-0.9995	0.0007
90%	1.0011	-0.9989	0.0017
95%	1.0015	-0.9985	0.0024
99%	1.0024	-0.9975	0.0041

‡Note. $\tilde{\beta}^{+\prime} = \begin{bmatrix} \tilde{\beta}^{+}_{11} & \tilde{\beta}^{+}_{21} & \tilde{\beta}^{+}_{31(ir)} \end{bmatrix}$ is the normalised estimated cointegrating vector. $\tilde{\beta}^{+}_{31(ir)}$ is the element of $\tilde{\beta}^{+}$ that corresponds to the irrelevant variable.

Table 6.9. Quantiles of the elements of
the estimated cointegrating vectors (DGP2).

$\tilde{\beta}^{+\ddagger}$ Quantiles	$\tilde{\beta}^{+}_{11}$	$\tilde{\beta}^{+}_{21}$	$\tilde{\beta}^{+}_{31}$	$\tilde{\beta}^{+}_{41(iv)}$	$\tilde{\beta}^{+}_{12}$	$\tilde{\beta}^{+}_{22}$	$\tilde{\beta}^{+}_{32}$	$\tilde{\beta}^{+}_{42(iv)}$
1%	0.9996	-2.0004	0.9996	-0.0040	0.9995	-0.5004	-0.5004	-0.0045
5%	0.9997	-2.0002	0.9997	-0.0023	0.9997	-0.5002	-0.5002	-0.0027
10%	0.9998	-2.0002	0.9998	-0.0017	0.9998	-0.5002	-0.5002	-0.0020
25%	0.9999	-2.0001	0.9999	0.0007	0.9999	-0.5000	-0.5000	-0.0009
50%	1.0000	-2.0000	1.0000	0.0000	1.0000	-0.5000	-0.5000	0.0000
75%	1.0001	-1.9999	1.0001	0.0007	1.0001	-0.4999	-0.4999	0.0008
90%	1.0002	-1.9998	1.0002	0.0017	1.0002	-0.4998	-0.4998	0.0019
95%	1.0003	-1.9997	1.0003	0.0023	1.0003	-0.4997	-0.4997	0.0026
99%	1.0004	1.9996	1.0004	0.0038	1.0005	-0.4995	0.4995	0.0044

‡Note. The rows of $\tilde{\beta}^{+'} = \begin{bmatrix} \tilde{\beta}^{+}_{11} & \tilde{\beta}^{+}_{21} & \tilde{\beta}^{+}_{31} & \tilde{\beta}^{+}_{41(iv)} \\ \tilde{\beta}^{+}_{12} & \tilde{\beta}^{+}_{22} & \tilde{\beta}^{+}_{32} & \tilde{\beta}^{+}_{42(iv)} \end{bmatrix}$ are the estimated cointegrating vectors. $\tilde{\beta}^{+}_{41(iv)}$ and $\tilde{\beta}^{+}_{42(iv)}$ are the elements of $\tilde{\beta}^{+}$ that correspond to the irrelevant variable.

From Tables 6.8 and 6.9, we observe that the normalised estimated cointegrating vectors converge to the true cointegrating vectors and the elements of $\tilde{\beta}^{+}$ that correspond to the irrelevant variable are zero to four decimal places for the 50% quantile.

Omitted variables

According to the asymptotic analysis in section 6.2 and the simulation results in Table 6.6, using DGP3 along with a two-variable SM, the eigenvalue equation should yield eigenvectors sufficiently close to zero. Table 6.10 shows the quantiles of the elements of the estimated eigenvectors in 10,000 replications using $T = 5{,}000$.

Table 6.10. Quantiles of the elements of the estimated eigenvectors (DGP3).

$\hat{v}^{\ddagger}$ / Quantiles	$\hat{v}_{11}$	$\hat{v}_{21}$	$\hat{v}_{12}$	$\hat{v}_{22}$
1%	-0.0563	-0.1026	-0.0457	-0.0428
5%	-0.0398	-0.0820	-0.0341	-0.0310
10%	-0.0310	-0.0696	-0.0283	-0.0252
25%	-0.0133	-0.0442	-0.0198	-0.0159
50%	0.0099	0.0079	-0.0097	-0.0028
75%	0.0327	0.0454	0.0109	0.0113
90%	0.0496	0.0685	0.0215	0.0219
95%	0.0587	0.0814	0.0269	0.0282
99%	0.0739	0.1051	0.0367	0.0400

‡Note. $\hat{v} = \begin{bmatrix} \hat{v}_{11} & \hat{v}_{12} \\ \hat{v}_{21} & \hat{v}_{22} \end{bmatrix}$ is the matrix whose columns hold the estimated eigenvectors and $\hat{v}' \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \sim I(1)$.

Next we use a modified form of DGP4, particularly, we use a matrix whose rows are linear transformations of the rows of β' found by adding to the first row twice the second row i.e.

$$\begin{bmatrix} 1 & -2 & 1 \\ 1 & -0.5 & -0.5 \end{bmatrix} \approx \begin{bmatrix} 3 & -3 & 0 \\ 1 & -0.5 & -0.5 \end{bmatrix},$$

where $\approx$ denotes a row equivalent matrix. Based on the asymptotic analysis of section 6.2, if we omit variable X_{3t} we should expect one cointegrating vector whose estimator converges to the space spanned by β_{11} in the notation of section 6.2, and in this case $\beta_{11}' = \begin{bmatrix} 3 & -3 \end{bmatrix}$. Table 6.11 shows the quantiles of the elements of the estimated cointegrating vector, $\tilde{\beta}_{11} = \begin{bmatrix} \tilde{\beta}_{11}^{(1)} \\ \tilde{\beta}_{11}^{(2)} \end{bmatrix}$ (associated with the largest eigenvalue) and the elements of the eigenvector corresponding to the smallest eigenvalue. In fact we use the normalised form of the estimated cointegrating vectors, $\tilde{\beta}_{11}$ given in (6.40), in order to achieve convergence to the true (known) submatrix of the true β, β_{11}, instead of a linear combination of it. Again the estimation is carried out using $T = 5{,}000$ and 10,000 replications.

Table 6.11. Quantiles of the elements of the estimated eigenvectors (DGP4).

$\hat{v}$‡ / Quantiles	$\tilde{\beta}_{11}^{(1)}$	$\tilde{\beta}_{11}^{(2)}$	$\hat{v}_{12}$	$\hat{v}_{22}$
1%	2.9999	-3.0001	-0.0303	-0.0315
5%	3.0000	-3.0000	-0.0186	-0.0219
10%	3.0000	-3.0000	-0.0127	-0.0157
25%	3.0001	-2.9999	-0.0058	-0.0057
50%	3.0001	-2.9999	0.0001	-0.0000
75%	3.0003	-2.9997	0.0061	0.0053
90%	3.0005	-2.9995	0.0133	0.0151
95%	3.0007	-2.9993	0.0194	0.0209
99%	3.0011	-2.9989	0.0296	0.0321

‡Note. The first column of $\hat{v} = \begin{bmatrix} \tilde{\beta}_{11}^{(1)} & \hat{v}_{12} \\ \tilde{\beta}_{11}^{(2)} & \hat{v}_{22} \end{bmatrix}$ holds the eigenvector which corresponds to the largest eigenvalue, i.e. the normalised estimated cointegrating vector, $\tilde{\beta}_{11}$ whereas $(\hat{v}_{12}X_{1t} + \hat{v}_{22}X_{2t}) \sim I(1)$.

In Table 6.10 the elements of the estimated eigenvectors are sufficiently close to zero which is in accordance with the absence of any cointegrating vectors. In Table 6.11 we can see that the elements of the estimated cointegrating vector, after normalisation converge to the appropriate elements of the submatrix of β in the DGP namely $\beta_{11}' = \begin{bmatrix} 3 & -3 \end{bmatrix}$. The elements of the other estimated eigenvector, which is associated with the smallest eigenvalue seem to be sufficiently small.

Next we use DGP4 and a SM with only X_{1t} and X_{2t} to compute the quantiles of the elements of the estimated adjustment coefficient matrix. The estimator of α_{11} used in the simulations is given by $\tilde{\alpha}_{11} = \hat{\alpha}_{11}\hat{\beta}_{11}'\bar{\beta}_{11}$ (see subsection 6.2.2) which is a transformation of $\hat{\alpha}_{11}$ such that $\tilde{\alpha}_{11}\tilde{\beta}_{11}' = \hat{\alpha}_{11}\hat{\beta}_{11}'$. For $T = 5{,}000$ and 10,000 replications the estimated adjustment coefficients seem to converge to the sum of the true adjustment coefficient matrix (i.e. the part of α, α_{11} say, in the DGP that corresponds to the single cointegrating vector that can be detected using the misspecified SM) and the asymptotic bias, which is computed using $T = 5{,}000$ and 10,000 replications (this sum is given by the right-hand side of (6.45)). For this case we have $\alpha_{11} = \begin{bmatrix} \alpha_{11}^{(1)} \\ \alpha_{11}^{(2)} \end{bmatrix} = \begin{bmatrix} 0.433 \\ 0.5 \end{bmatrix}$, and $\tilde{\alpha}_{11} = \begin{bmatrix} \tilde{\alpha}_{11}^{(1)} \\ \tilde{\alpha}_{11}^{(2)} \end{bmatrix}$ is the transformed estimator of α_{11}. The results appear in Tables 6.12 and 6.13.

Table 6.12. Quantiles of the estimated adjustment coefficients.

$\tilde{\alpha}_{11}$ / Quantiles	$\tilde{\alpha}_{11}^{(1)}$	$\tilde{\alpha}_{11}^{(2)}$
1%	0.4879	0.5730
5%	0.4901	0.5752
10%	0.4914	0.5763
25%	0.4935	0.5783
50%	0.4957	0.5804
75%	0.4980	0.5826
90%	0.5002	0.5847
95%	0.5014	0.5859
99%	0.5036	0.5880

Tables 6.12 and 6.13 provide an illustration of Proposition 6.3 namely that the estimator of the adjustment coefficients in an underspecified SM is inconsistent or asymptotically biased. From Table 6.12 we can see that the normalised estimated adjustment coefficients are biased upwards.

Table 6.13. Quantiles of α_{11} plus estimated asymptotic bias.

α_{11} + est. as. bias / Quantiles	$\alpha_{11}^{(1)}$ + est. as. bias	$\alpha_{11}^{(2)}$ + est. as. bias
1%	0.49102	0.5741
5%	0.49251	0.5760
10%	0.4932	0.5770
25%	0.4944	0.5785
50%	0.4958	0.5803
75%	0.4971	0.5820
90%	0.4984	0.5837
95%	0.4992	0.5847
99%	0.5005	0.5864

6.4 Concluding remarks

This chapter has considered the effects of overspecifying (inclusion of irrelevant variables) or underspecifying (omission of relevant variables) the SM on the LR tests for cointegration proposed by Johansen (1988, 1996). We showed that including irrelevant variables in the SM will affect neither the inference about the cointegrating rank nor the consistency of the estimated cointegrating vectors and adjustment coefficients as the sample size becomes large. However, simulations showed that overspecifying the SM reduces the power of cointegration tests for both small/medium ($T = 50$, 100) and large sample sizes ($T = 500$, 800). We also showed that omitting relevant variables from the SM will lead to either no detection of cointegrating relationships, if the true cointegrating rank is smaller than or equal to the number of omitted variables ($r \leq k$) or the detection of $q < r$ cointegrating relationships, if the true cointegrating rank is greater than the number of omitted variables ($r > k$). In addition, the use of an underspecified SM does not affect the consistency of the estimated cointegrating vectors since they still converge to a subspace of $sp(\beta)$ but it does affect the consistency of the estimators of the adjustment coefficient matrix and variance of the errors.

Although the analytical results are asymptotic, small sample simulations show that the theoretical findings also arise in sample sizes used in empirical work.

The omitted variables can also be $I(0)$. Since the inclusion of a stationary variable increases the dimensions of the cointegrating space by one, the omission of only $I(0)$ variables will lead to the underestimation of the cointegrating rank by the number of omitted $I(0)$ variables.

Overall we conclude that the omission of relevant variables from the SM has more serious consequences (especially when followed by tests for linear restrictions on α and β conditional on the wrong cointegrating rank) on cointegration analysis than the inclusion of irrelevant variables, which is in accordance with the simulation results of Podivinsky (1998) as well as with the known 'verdict' in the standard regression analysis.

Chapter 7
Conclusions

This chapter provides an overview of the aims and findings of the thesis along with some limitations and possible extensions of the results herein.

7.1 Aim

The thesis aimed to study the effects of two types of misspecifications on the LR tests of cointegration proposed by Johansen (1988, 1996), implemented using the trace or the maximal eigenvalue statistic. In other words we assume that the SM used for cointegration testing differs from the DGP (since the DGP is unknown to the modeller) and we examine the sensitivity of the tests to misspecifications of the SM.

The misspecifications under consideration are: (i) intercept shifts present in the DGP but ignored in modelling (absence from the SM of step dummy variables accounting for the intercept shifts), (ii) presence of irrelevant variables in the SM or omission of relevant variables (present in the DGP) from the SM.

In investigating the effects of the above misspecifications we use (i) asymptotic analysis i.e. we examine the asymptotic behaviour of the eigenvalue equation and therefore the behaviour of the eigenvalues used in the trace and maximal eigenvalue statistics and (ii) Monte Carlo simulations to check the asymptotic findings and evaluate the impact of the misspecifications in finite samples.

Since we analyse the effects of misspecifications, it is desirable to have some degree of control over the power of the tests, in carrying out the Monte Carlo experiments. Thus, where appropriate, the experiments were designed in a way that we can control the level of asymptotic local power.

For the first type of misspecification (intercept shifts) both methods of analysis were employed to examine the effects of shifts at different or common dates along with alternative specifications of the deterministic term.

For the second type of misspecification (irrelevant or omitted variables) both methods of analysis were used to study the consistency of the estimators of the parameters in the ECM.

7.2 Findings

In Chapter 4 we show that under the first type of misspecification the tests reject the true null hypothesis of cointegrating rank r $(0 < r < p)$ with probability one as the sample size tends to infinity. Thus, we tend to accept spurious cointegrating relations/vectors not present in the DGP. An upper bound is found for the number (b) of spurious cointegrating vectors that arise asymptotically and it is given by the number of variables with intercept shifts (p_1). In the case of shifts at a common date the upper bound is given by the number distinct/different shifts in the DGP. The attainment of the upper bound depends on the weak exogeneity status of the variables. It is found that (i) when none of the variables with intercept shifts are weakly exogenous or (ii) when all the variables free of shifts are weakly exogenous no spurious cointegration occurs $(b = 0)$.

In Chapter 5, using Monte Carlo simulations we find that for a given level of asymptotic local power, the frequency of rejecting the true null hypothesis $r = 1$ increases as the sample size becomes larger. This finding is in accordance with the asymptotic analysis of Chapter 4. Furthermore, the frequency with which the true null hypothesis is rejected rises as the magnitude of the shift increases. These patterns arise under the assumption of both distinct shifts and shifts at a common date as well as under all constant term specifications (no constant, restricted and unrestricted constant) considered. For sample sizes and constant term specifications commonly used in applied works, together with a magnitude of shift (e.g. $\delta = 0.5$) that is difficult to detect, the rejection frequencies of the true null hypothesis are far-off (sometimes they exceed 30%) the asymptotic size of the tests (5%). O'Brien (1999) argues that shifts of this size may be difficult to detect visually, and when their location is not known, difficult to detect by testing.

In Chapter 6 we show that inclusion of irrelevant variables does not affect the inference about the cointegrating rank but it does affect the magnitude of the probability limit of the positive eigenvalues. In addition, the consistency of the estimators of the parameters in the ECM is not affected. However, simulations show reduction in the power of the tests when irrelevant variables are included in the SM. Moreover, we find that omitting relevant variables from the SM affects the inference about the cointegrating rank. The tests either fail to detect cointegration when $r \leq k$, or they detect $q < r$ cointegrating vectors when $r > k$. In the case that $q < r$ cointegrating vectors can be detected their estimators are 'consistent' in the sense that they converge to a subspace of $sp(\beta)$. Nevertheless, the estimators of the adjustment coefficients and variance of the errors of the SM are inconsistent. The results of the Monte Carlo simulations conform with the asymptotic results and the effects of this misspecification are apparent for sample sizes used in empirical works ($T = 50{,}100$).

7.3 Limitations and extensions

No asymptotic distributions were derived under the types of misspecifications considered.

In the case of intercept shifts the asymptotic distribution was derived for the null case ($r = 0$) and was tabulated (for $p_1 = 1$) by O'Brien (1999). For $r > 0$, that is the case analysed in Chapter 4, the limit of the eigenvalue equation is not the determinant of a block diagonal matrix, which indicates the need for adopting a different scaling than that used in Johansen (1996). However, it was shown (see section 4.4) that the asymptotic distribution for $r = 0$ can be derived as a sub-case using the asymptotic results derived for $0 < r < p$. An extension of Chapter 4 would be the derivation of the asymptotic distribution by re-defining the directions in $\mathbb{R}^p$ appropriately.

Even though the Monte Carlo investigation in Chapter 5 is quite extensive, a response surface analysis would provide useful insights into the dependency of the rejection frequencies on the sample size, design and deterministic term specification.

For the irrelevant variables case presented in Chapter 6 it would be useful to derive the variances of the estimators and compare them with those from the correctly specified model in order to make more specific efficiency statements. For the omitted variables case we could gain more understanding if the asymptotic distribution under misspecification was derived. Moreover, for both cases analysed in Chapter 6 an investigation of the behaviour of the estimators in small samples would be informative. Finally an extension of Chapter 6 would be to analyse the case that the SM includes irrelevant variables and at the same time does not take into account relevant variables.

7.4 Contribution

Although the LR tests for cointegration proposed by Johansen (1988, 1996) are routinely used in applied works, the literature concerning the effects of misspecifications on these tests is limited to some Monte Carlo studies with the exception of O'Brien (1996, 1997, 1999). The contribution of this thesis is to provide analytical (asymptotic results) and numerical (Monte Carlo results) evidence about the robustness of these tests under misspecifications.

The asymptotic analysis (Chapters 4 and 6) provides knowledge as to which parameters of the model play key roles under misspecification and this knowledge is utilised to design informative Monte Carlo experiments. The asymptotic analysis proved to be useful since the parameter space, especially in the case of ECMs is impossible to be fully explored.

In addition, the thesis can act as a caveat for the applied worker since the cointegration analysis is shown to be distorted under the misspecifications mentioned above. For the first type of misspecification to be avoided, the modellers should perform tests for shifts on the univariate processes included in the SM (see e.g. Perron (1989), Perron and Vogelsang (1992), Zivot and Andrews (1992)) and/or cointegration tests that allow for shifts in the mean of the vector processes (see e.g. Johansen et al. (2000), Inoue (1999), Saikkonen and Lütkepohl (1998)).

For the second type of misspecification considered, the impact of omitted variables on cointegration analysis seems more serious than that of irrelevant variables, since in the case of the former only part of the model can be recovered and the modeller might use the 'inadequate' model to test structural hypotheses on β and α and reach misleading conclusions. Since the inclusion of irrelevant variables does not appear to distort cointegration analysis (except for a reduction in the power of the tests), this finding can be used as an advocate of general-to-specific approach to modelling (see Hendry (1995)).

Overall we can conclude that the LR tests for cointegration are sensitive to the misspecifications considered. The use of a misspecified model affects the analysis in various ways such as the inference about the cointegrating rank, the consistency of the estimators or the power of the tests. Thus, application of pre-tests on the univariate processes, diagnostic tests or modified tests for cointegration are necessary to avoid misspecifications or limit their effects.

Appendix A: Preliminary results

In the proofs we repeatedly use some asymptotic properties of linear processes which are stated in the following theorems. In addition we provide some results about the order of magnitude of linear functions of step dummy variables needed in the proofs of Lemma 4.1 and 4.2. Note that any terms in parentheses written as subscripts or superscripts indicate indices.

Theorem A.1. Let $\{\varepsilon_t\}$ *be a sequence of p-dimensional i.i.d.*[15] *random vectors with mean zero and variance matrix* Ω. *Let* $W(u)$ *be the p-dimensional Brownian motion, with variance* Ω, *on* $C[0,1]$. *Define* $\xi_t = \sum_{i=1}^{t} \varepsilon_i$. *Let* $\{f_T(t)\}_{t=1}^{T}$ *be a sequence of deterministic functions such that* $f_T([Tu]) \to f(u)$, *with* $f(\cdot)$ *defined on* $[0,1]$. *Then,*

$$T^{-1/2} \sum_{i=1}^{[Tu]} \varepsilon_i = T^{-1/2} \xi_{[Tu]} \xrightarrow{d} W(u) \tag{A.1}$$

$$T^{-2} \sum_{t=1}^{T} \xi_t \xi_t' \xrightarrow{d} \int_0^1 W(u) W(u)' du \tag{A.2}$$

$$T^{-1} \sum_{t=1}^{T} \xi_{t-1} \varepsilon_t' \xrightarrow{d} \int_0^1 W(dW)' \tag{A.3}$$

$$T^{-1/2} \sum_{t=1}^{T} \varepsilon_t f_T(t)' \xrightarrow{d} \int_0^1 (dW) f' \tag{A.4}$$

$$T^{-3/2} \sum_{t=1}^{T} \xi_t f_T(t)' \xrightarrow{d} \int_0^1 W(u) f(u)' du. \tag{A.5}$$

[15] Functional central limit theorems can be derived by imposing weaker conditions on the process $\{\varepsilon_t\}$ and also by considering convergence in the space of cadlag functions, see e.g. Billingsley (1968, Chapters 3 and 4), Phillips and Durlauf (1986), Phillips (1987), Phillips and Solo (1992).

Theorem A.2. Let ε_t and $f_T(\cdot)$ be defined as in Theorem A.1 and in addition let ε_t have finite fourth moments. Let $u_t = \sum_{i=0}^{\infty} e_i \varepsilon_{t-i}$ and $v_t = \sum_{i=0}^{\infty} h_i \varepsilon_{t-i}$ with coefficients e_i and h_i that decrease exponentially, such that $e(y) = \sum_{i=0}^{\infty} e_i y^i$ and $h(y) = \sum_{i=0}^{\infty} h_i y^i$ are convergent for $|y| < 1 + \omega$, $\omega > 0$. Then,

$$T^{-1/2} \max_{1 \leq t \leq T} |u_t| \xrightarrow{p} 0 \tag{A.6}$$

$$T^{-1/2} \sum_{t=1}^{[Tu]} \begin{bmatrix} u_t \\ v_t \end{bmatrix} \xrightarrow{d} \begin{bmatrix} e(1) \\ h(1) \end{bmatrix} W(u) \tag{A.7}$$

$$T^{-1/2} \sum_{t=1}^{T} u_t f_T(t)' \xrightarrow{d} e(1) \int_0^1 (dW) f' \tag{A.8}$$

$$T^{-1} \sum_{t=1}^{T} u_t v_{t+s}' \xrightarrow{p} E(u_t v_{t+s}') = \sum_{i=0}^{\infty} e_i \Omega h_{i+s}' = \Gamma_s, \ s = 0, 1, 2, \ldots \tag{A.9}$$

$$T^{-1} \sum_{t=1}^{T} (\sum_{i=1}^{t-1} u_i) v_t' \xrightarrow{d} e(1) \int_0^1 W(dW)' h(1)' + \sum_{s=1}^{\infty} \Gamma_s. \tag{A.10}$$

Another useful result that will be used in the proofs is given by the following theorem.

Theorem A.3. Continuous Mapping Theorem (CMT). Let $g(\cdot)$ be a continuous functional on $C[0,1]$ such that $g(\cdot) : C[0,1] \longmapsto \mathbb{R}^p$ or $g(\cdot) : C[0,1] \longmapsto C[0,1]$. If $X_T \xrightarrow{d} X$, with $X \in C[0,1]$, then $g(X_T) \xrightarrow{d} g(X)$.

Theorems A.1, A.2 and A.3 are given in Johansen (1996, Theorems B.12, B.13 and B.5) and in Hamilton (1994, Proposition 18.1).

In the proofs of Lemma 4.1 and Lemma 4.2 we encounter products of lag polynomials, step (or shift) dummy variables (z_t's) and error terms. In Lemma A.2 we establish some results concerning the order of magnitude of these terms. First we explain the notation to be adopted.

The $p \times p$ matrix lag polynomial $C_1(L)$ is partitioned conformably with $\Phi D_t = \begin{bmatrix} z_t \\ \varphi \end{bmatrix}$, (with z_t, $p_1 \times 1$ subvector of step dummies and φ, $p_2 \times 1$ subvector of constants, $p_1 + p_2 = p$) into $C_1(L) = \begin{bmatrix} C_1^1(L) & C_1^2(L) \end{bmatrix}$, where $C_1^1(L)$ is $p \times p_1$ and $C_1^2(L)$ is $p \times p_2$. Moreover, $C_1^1(L) = \begin{bmatrix} e_{11}(L) & \cdots & e_{1p_1}(L) \\ \vdots & \ddots & \vdots \\ e_{p1}(L) & \cdots & e_{pp_1}(L) \end{bmatrix}$.

Note that the submatrices $C_1^1(L)$ and $C_1^2(L)$ as well as the elements $e_{ij}(L)$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p_1$ inherit the properties of $C_1(L)$ (see Hamilton (1994, pp. 258, 545)) stated in Theorem 2.1.

Below we give Lemma A.1 which is used for the proofs of some of the results given in Lemma A.2.

Lemma A.1. Let $S = \sum\limits_{i=0}^{p} S_i$ *and* $S_i = \sum\limits_{k=a_i}^{b_i} e_{mi}^{(k)} \Lambda_i$. *If* $a_i < b_i$, $a_0 = 0$, $b_p = \infty$, $a_n = b_{n-1} + 1$, $n = 1, 2, \ldots, p$ *and* $a_i \to \infty$, $b_i \to \infty$ *as* $T \to \infty$; *then* $S \to \sum\limits_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_0$.

Proof.

$$S = S_0 + S_1 + \ldots + S_p =$$

$$\sum_{k=0}^{b_0} e_{mi}^{(k)} \Lambda_0 + \sum_{k=a_1}^{b_1} e_{mi}^{(k)} \Lambda_1 + \cdots + \sum_{k=a_p}^{\infty} e_{mi}^{(k)} \Lambda_p =$$

$$\sum_{k=0}^{b_0} e_{mi}^{(k)} \Lambda_0 + [\sum_{k=0}^{b_1} e_{mi}^{(k)} \Lambda_1 - \sum_{k=0}^{b_0} e_{mi}^{(k)} \Lambda_1] + [\sum_{k=0}^{b_2} e_{mi}^{(k)} \Lambda_2 - \sum_{k=0}^{b_1} e_{mi}^{(k)} \Lambda_2] +$$

$$\cdots + [\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_p - \sum_{k=0}^{b_{p-1}} e_{mi}^{(k)} \Lambda_p] =$$

$$\sum_{k=0}^{b_0} e_{mi}^{(k)}(\Lambda_0 - \Lambda_1) + \sum_{k=0}^{b_1} e_{mi}^{(k)}(\Lambda_1 - \Lambda_2) + \cdots + \sum_{k=0}^{b_{p-1}} e_{mi}^{(k)}(\Lambda_{p-1} - \Lambda_p) + \sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_p$$
$$\rightarrow \sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_0, \; as\ T \rightarrow \infty. \quad \blacksquare$$

Lemma A.1 requires Λ_i to be $O(1)$. Later we use the symbol Λ_k for expressions that are not $O(1)$ but separate demonstrations are given therein.

For the proofs of A(v), A(vi), B(i), B(ii), C(i), C(ii), E(ii), E(iii) and F in Lemma A.2 we use an algebraic decomposition of $C_1(L)$,which is referred to as the Beveridge-Nelson decomposition (see Beveridge and Nelson (1981)) in econometric literature. Note that this decomposition is initially applied on $C(L)$ that appears in Theorem 2.1 (Granger Representation Theorem) and has the form $C(L) = C(1) + (1-L)C_1(L)$. Then $C_1(L) = \sum_{i=0}^{\infty} C_{1i}L^i$ which is also convergent (see Theorem 2.1) can be expressed as

$$C_1(L) = C_1(1) + (1-L)C_1^*(L)$$

where $C_1^*(L) = \sum_{k=0}^{\infty} C_{1k}^* L^k$, $C_{1k}^* = -\sum_{i=k+1}^{\infty} C_{1i} = \sum_{i=k+1}^{\infty}\sum_{j=i+1}^{\infty} C_j = \sum_{j=2}^{\infty}(j-1)C_{k+j}$ and $C_1^*(1) = \sum_{k=0}^{\infty} C_{1k}^* = \sum_{j=1}^{\infty} jC_{1j} = \frac{1}{2}\sum_{j=2}^{\infty} j(j-1)C_j < \infty$. The validity of the above decomposition hinges on the fact that $C_1^*(L)$ is convergent, which follows from Lemma 4.1 in Johansen (1996).

Moreover, a typical element of $C_1^1(y)$ would be $e_{ij}(y) = \sum_{k=0}^{\infty} y^k e_{ij}^{(k)}$, which is convergent for $|y| < 1+\omega$, $\omega > 0$, by definition (see Theorem 2.1). It then follows that

$|e_{ij}^{(k)}| < a^k$, where $a = (1+\omega)^{-1}$ and $0 < a < 1$. We then have

$$\sum_{k=0}^{\infty} k e_{ij}^{(k)} \leq \sum_{k=0}^{\infty} k|e_{ij}^{(k)}| < \sum_{k=0}^{\infty} k a^k = a(1-a)^{-2} < \infty.$$

The expressions of Lemma A.2 involve only the submatrix, $C_1^1(L)$, of $C_1(L)$, therefore the decomposition of the lag polynomial is $C_1^1(L) = C_1^1(1) + C_1^{1*}(L)(1-L)$, which also applies to each element of $C_1^1(L)$. $C_1^{1*}(L)$ is $p \times p_1$ and consists of the first p_1 columns of $C_1^*(L)$ (that appears in the Beveridge-Nelson decomposition of $C_1(L)$, shown above).

Finally, note that for an arbitrary breakpoint $t_0 = [T\lambda]$, $\lambda \in (0,1)$, $[T\lambda] = T\lambda$ for $T\lambda$ integer, which implies $\lambda = t_0/T$, and $[T\lambda] = (T-1)\lambda$, otherwise, which implies $\lambda = t_0/(T-1)$. In deriving the asymptotic results we write $\lambda = t_0/T$ to avoid complicating the demonstration since $t_0 = T\lambda = O(T)$ for $T\lambda$ integer and $t_0 = T\lambda - \lambda = O(T) + O(1) = O(T)$, otherwise.

Lemma A.2. Let $v_t = C_1(L)\varepsilon_t$. Under the assumptions of Theorem 2.1 about $C_1(L)$, the assumptions of section 4.1 about z_t, Z_t and the assumptions of Theorems A.1 and A.2 about ε_t,

A. (i) $C_1^1(L)z_t$

(ii) $\sum_{t=1}^{T} C_1^1(L)(1-L)z_t$

(iii) $\sum_{t=1}^{T} z_t[z_t'(1-L)C_1^1(L)']$

(iv) $\sum_{t=1}^{T}[C_1^1(L)(1-L)z_t][z_t'(1-L)C_1^1(L)']$

(v) $\sum_{t=1}^{T} C_1^1(L)z_t$

(vi) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][z_{t-1}'(1-L)C_1^1(L)']$ *are* $O(1)$,

B. (i) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][z_{t-1}'C_1^1(L)']$

(ii) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}]z_{t-1}^{'}$

(iii) $\sum_{t=1}^{T} Z_{t-1}[z_t^{'}(1-L)C_1^1(L)^{'}]$

(iv) $\sum_{t=1}^{T}(t-1)[z_t^{'}(1-L)C_1^1(L)^{'}]$ *are* $O(T)$,

C. (i) $\sum_{t=1}^{T} Z_{t-1}[z_{t-1}^{'}C_1^1(L)^{'}]$

(ii) $\sum_{t=1}^{T}(t-1)[z_{t-1}^{'}C_1^1(L)^{'}]$ *are* $O(T^2)$,

D. (i) $\sum_{t=1}^{T} \varepsilon_t[z_t^{'}(1-L)C_1^1(L)^{'}]$

(ii) $\sum_{t=1}^{T}[(1-L)v_t]z_t^{'}$

(iii) $\sum_{t=1}^{T}[(1-L)v_t][z_t^{'}(1-L)C_1^1(L)^{'}]$

(iv) $\sum_{t=1}^{T} v_{t-1}[z_{t-1}^{'}(1-L)C_1^1(L)^{'}]$

(v) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][v_t^{'}(1-L)]$ *are* $O_p(1)$,

E. (i) $\sum_{t=1}^{T} v_{t-1}z_{t-1}^{'}$

(ii) $\sum_{t=1}^{T} v_{t-1}[z_{t-1}^{'}C_1^1(L)^{'}]$

(iii) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}]\varepsilon_t^{'}$

(iv) $\sum_{t=1}^{T} \xi_{t-1}[z_t^{'}(1-L)C_1^1(L)^{'}]$ *are* $O_p(T^{1/2})$,

F. $\sum_{t=1}^{T} \xi_{t-1}[z_{t-1}^{'}C_1^1(L)^{'}]$ *is* $O_p(T^{3/2})$.

Proof. A. (i) We analyse $C_1^1(L)z_t$.

$$C_1^1(L)z_t = \begin{bmatrix} e_{11}(L) & \cdots & e_{1p_1}(L) \\ \vdots & \ddots & \vdots \\ e_{p1}(L) & \cdots & e_{pp_1}(L) \end{bmatrix} \begin{bmatrix} z_{1t} \\ \vdots \\ z_{p_1t} \end{bmatrix}$$

$$= \sum_{i=1}^{p_1} \begin{bmatrix} e_{1i}(L)z_{it} \\ \vdots \\ e_{pi}(L)z_{it} \end{bmatrix}.$$

The (i,j)-th element of $C_1^1(y)$, $e_{ij}(y) = \sum_{k=0}^{\infty} y^k e_{ij}^{(k)}$ is convergent for $|y| < 1+\omega$, $\omega > 0$ and the coefficients $e_{ij}^{(k)}$'s are exponentially decreasing. From $|y| < 1+\omega$, it follows that $|e_{ij}^{(k)}| < (1+\omega)^{-k}$ and $\sum_{k=0}^{\infty} |e_{ij}^{(k)}| < \sum_{k=0}^{\infty}(1+\omega)^{-k} = 1+\omega^{-1}$, which shows that $e_{ij}^{(k)}$'s are absolutely summable. From the definition of the typical j-th step dummy, $z_{j(t-k)}$, we have $-\delta_j < z_{j(t-k)} < \delta_j$ which implies $|z_{j(t-k)}| < |\delta_j|$ which implies $|e_{ij}^{(k)} z_{j(t-k)}| < |e_{ij}^{(k)}\delta_j| = |\delta_j||e_{ij}^{(k)}|$. Taking the infinite sum both sides, $|\sum_{k=0}^{\infty} e_{ij}^{(k)} z_{j(t-k)}| < |\delta_j||\sum_{k=0}^{\infty} e_{ij}^{(k)}| \le |\delta_j|\sum_{k=0}^{\infty}|e_{ij}^{(k)}| < \infty$ as $e_{ij}^{(k)}$'s are absolutely summable. So, there exists $0 < m < \infty$ such that $|e_{ij}(L)z_{jt}| < m$, therefore $e_{ij}(L)z_{jt}$ is bounded, hence $O(1)$. It then follows that $C_1^1(L)z_t$ is $O(1)$.

A. (ii) $\sum_{t=1}^{T} C_1^1(L)(1-L)z_t =$

$$\sum_{i=1}^{p_1}\begin{bmatrix} \sum_{t=1}^{T} e_{1i}(L)(1-L)z_{it} \\ \vdots \\ \sum_{t=1}^{T} e_{pi}(L)(1-L)z_{it} \end{bmatrix}.$$

As $(z_{it} - z_{i(t-1)}) = \begin{cases} \delta_i(\lambda_i - 1),\ t = 1 \\ \delta_i,\ t = t_{0i}+1 \\ 0, \text{ otherwise} \end{cases}$,
analysing the typical, j-th, element of this vector we have

$$\sum_{t=1}^{T} e_{ji}(L)(z_{it} - z_{i(t-1)}) = \sum_{t=1}^{T}\sum_{k=0}^{\infty} e_{ji}^{(k)}(z_{i(t-k)} - z_{i(t-1-k)})$$

$$= \sum_{k=0}^{\infty} e_{ji}^{(k)} \sum_{t=k+1}^{T} (z_{i(t-k)} - z_{i(t-1-k)})$$

$$= \sum_{k=0}^{\infty} e_{ji}^{(k)}[(z_{i(1)} - z_{i(0)}) + (z_{i(2)} - z_{i(1)}) + \cdots + (z_{i(T-k)} - z_{i(T-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{ji}^{(k)} \Lambda_k$$

where,

$$\Lambda_k = \begin{cases} \delta_i(\lambda_i - 1), t_{0i} \geq T - k \Rightarrow k \geq T(1-\lambda_i) \\ \delta_i \lambda_i, \text{otherwise} \end{cases}.$$

So,

$$\sum_{k=0}^{\infty} e_{ji}^{(k)} \Lambda_k = \delta_i \lambda_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{ji}^{(k)} + \delta_i(\lambda_i - 1) \sum_{k=T(1-\lambda_i)}^{\infty} e_{ji}^{(k)} \rightarrow$$

$$\delta_i \lambda_i \sum_{k=0}^{\infty} e_{ji}^{(k)} < \infty,$$

by Lemma A.1. Thus, $\sum_{t=1}^{T} C_1^1(L)(1-L)z_t$ is $O(1)$.

A. (iii) $\sum_{t=1}^{T} z_t[z_t'(1-L)C_1^1(L)'] =$

$$\sum_{i=1}^{p_1} \begin{bmatrix} \sum_{t=1}^{T} z_{1t}[(z_{it} - z_{i(t-1)})e_{1i}(L)] & \cdots & \sum_{t=1}^{T} z_{1t}[(z_{it} - z_{i(t-1)})e_{pi}(L)] \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^{T} z_{p_1 t}[(z_{it} - z_{i(t-1)})e_{1i}(L)] & \cdots & \sum_{t=1}^{T} z_{p_1 t}[(z_{it} - z_{i(t-1)})e_{pi}(L)] \end{bmatrix}.$$

Analysing the typical (l, m)-th element of the above matrix we have,

$$\begin{aligned} \sum_{t=1}^{T} z_{lt}[e_{mi}(L)(z_{it} - z_{i(t-1)})] &= \sum_{t=1}^{T} z_{lt} \sum_{k=0}^{\infty} e_{mi}^{(k)}((z_{i(t-k)} - z_{i(t-k-1)})) \\ &= \sum_{k=0}^{\infty} e_{mi}^{(k)} \sum_{t=k+1}^{T} z_{lt}(z_{i(t-k)} - z_{i(t-k-1)}) \end{aligned}$$

$$= \sum_{k=0}^{\infty} e_{mi}^{(k)}[z_{l(1+k)}(z_{i(1)} - z_{i(0)}) + z_{l(k+2)}(z_{i(2)} - z_{i(1)}) + \cdots + z_{lT}(z_{i(T-k)} - z_{i(T-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k^{(j)}, j = 1, 2, 3, 4, 5,$$

with

$$\Lambda_k^{(j)} = \begin{cases} \delta_l\delta_i\lambda_l(\lambda_i - 1),\, t_{0l} < k+1, t_{0i} \geq T-k,\ j=1 \\ \delta_l\delta_i\lambda_l\lambda_i,\, t_{0l} < k+1, t_{0i} < T-k,\ j=2 \\ \delta_l\delta_i(\lambda_l - 1)(\lambda_i - 1),\, t_{0l} \geq k+1, t_{0i} \geq T-k,\ j=3 \\ \delta_l\delta_i(\lambda_l - 1)\lambda_i,\, t_{0l} \geq k+1, t_{0i} < T-k,\ t_{0l} \geq t_{0i}+k,\ j=4 \\ \delta_l\delta_i(\lambda_l\lambda_i - \lambda_i + 1),\, t_{0l} \geq k+1, t_{0i} < T-k,\ t_{0l} < t_{0i}+k,\ j=5 \end{cases}.$$

Rewriting the restrictions with respect to k,

$$\Lambda_k^{(j)} = \begin{cases} \delta_l\delta_i\lambda_l(\lambda_i - 1),\, k \geq t_{0l}, k \geq T-t_{0i},\ j=1 \\ \delta_l\delta_i\lambda_l\lambda_i,\, k \geq t_{0l}, k < T-t_{0i},\ j=2 \\ \delta_l\delta_i(\lambda_l - 1)(\lambda_i - 1),\, k < t_{0l}, k \geq T-t_{0i},\ j=3 \\ \delta_l\delta_i(\lambda_l - 1)\lambda_i,\, k < t_{0l}, k < T-t_{0i},\ k \leq t_{0l}-t_{0i},\ j=4 \\ \delta_l\delta_i(\lambda_l\lambda_i - \lambda_i + 1),\, k < t_{0l}, k < T-t_{0i},\ k > t_{0l}-t_{0i},\ j=5 \end{cases}.$$

The ranges for k depend on whether $t_{0l} \gtreqless T - t_{0i}$ and $t_{0l} \gtrless t_{0i}$[16]. Thus, we distinguish four cases as follows, for $j = 1, 2, 3, 4, 5$

(a) $t_{0l} \geq T - t_{0i}, t_{0l} > t_{0i}$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_k^{(j)} =$$

$$\sum_{k=0}^{t_{0l}-t_{0i}} e_{mi}^{(k)}\Lambda_k^{(4)} + \sum_{k=t_{0l}-t_{0i}+1}^{T-t_{0i}-1} e_{mi}^{(k)}\Lambda_k^{(5)} + \sum_{k=T-t_{0i}}^{t_{0l}-1} e_{mi}^{(k)}\Lambda_k^{(3)} + \sum_{k=t_{0l}}^{\infty} e_{mi}^{(k)}\Lambda_k^{(1)} \rightarrow$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_k^{(4)} = \delta_l\delta_i(\lambda_l - 1)\lambda_i \sum_{k=0}^{\infty} e_{mi}^{(k)} < \infty.$$

(b) $t_{0l} \geq T - t_{0i}, t_{0l} < t_{0i}$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_k^{(j)} =$$

$$\sum_{k=0}^{T-t_{0i}-1} e_{mi}^{(k)}\Lambda_k^{(5)} + \sum_{k=T-t_{0i}}^{t_{0l}-1} e_{mi}^{(k)}\Lambda_k^{(3)} + \sum_{k=t_{0l}}^{\infty} e_{mi}^{(k)}\Lambda_k^{(1)} \rightarrow$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_k^{(5)} = \delta_l\delta_i(\lambda_l\lambda_i - \lambda_i + 1) \sum_{k=0}^{\infty} e_{mi}^{(k)} < \infty.$$

---

16 The case where $t_{0l} = t_{0i}$ is not considered since we assume distinct shifts.

(c) $t_{0l} < T - t_{0i}$, $t_{0l} > t_{0i}$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k^{(j)} =$$

$$\sum_{k=0}^{t_{0l}-t_{0i}} e_{mi}^{(k)} \Lambda_k^{(4)} + \sum_{k=t_{0l}-t_{0i}+1}^{t_{0l}-1} e_{mi}^{(k)} \Lambda_k^{(5)} + \sum_{k=t_{0l}}^{T-t_{0i}-1} e_{mi}^{(k)} \Lambda_k^{(2)} + \sum_{k=T-t_{0i}}^{\infty} e_{mi}^{(k)} \Lambda_k^{(1)} \to$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k^{(4)} = \delta_l \delta_i (\lambda_l - 1)\lambda_i \sum_{k=0}^{\infty} e_{mi}^{(k)} < \infty.$$

(d) $t_{0l} < T - t_{0i}$, $t_{0l} < t_{0i}$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k^{(j)} =$$

$$\sum_{k=0}^{t_{0l}-1} e_{mi}^{(k)} \Lambda_k^{(5)} + \sum_{k=t_{0l}}^{T-t_{0i}-1} e_{mi}^{(k)} \Lambda_k^{(2)} + \sum_{k=T-t_{0i}}^{\infty} e_{mi}^{(k)} \Lambda_k^{(1)} \to$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k^{(5)} = \delta_l \delta_i (\lambda_l \lambda_i - \lambda_i + 1) \sum_{k=0}^{\infty} e_{mi}^{(k)} < \infty.$$

In order to find the limits in cases (a)-(d), we apply Lemma A.1. Since the deterministic process in all cases converges to a bounded sum, $\sum_{t=1}^{T} z_t[z_t'(1-L)C_1^1(L)']$ is $O(1)$.

A. (iv) $\sum_{t=1}^{T}[C_1^1(L)(1-L)z_t][z_t'(1-L)C_1^1(L)'] =$

$$\sum_{i=1}^{p_1}\sum_{j=1}^{p_1}\sum_{t=1}^{T}\begin{bmatrix} e_{1i}(L)(1-L)z_{it}z_{jt}(1-L)e_{1j}(L) & \cdots & e_{1i}(L)(1-L)z_{it}z_{jt}(1-L)e_{pj}(L) \\ \vdots & \ddots & \vdots \\ e_{pi}(L)(1-L)z_{it}z_{jt}(1-L)e_{1j}(L)] & \cdots & e_{pi}(L)(1-L)z_{it}z_{jt}(1-L)e_{pj}(L) \end{bmatrix}.$$

The (l, m)-th element of the above matrix has the form

$$\sum_{t=1}^{T}[e_{li}(L)(z_{it} - z_{i(t-1)})][(z_{jt} - z_{j(t-1)})e_{mj}(L)] =$$

$$\sum_{t=1}^{T}[\sum_{k=0}^{\infty} e_{li}^{(k)}((z_{i(t-k)} - z_{i(t-k-1)})][\sum_{s=0}^{\infty} e_{mj}^{(s)}(z_{j(t-s)} - z_{j(t-s-1)})] =$$

$$\sum_{k=0}^{\infty}\sum_{s=0}^{\infty} e_{li}^{(k)} e_{mj}^{(s)} [\sum_{t=n+1}^{T} (z_{i(t-k)} - z_{i(t-k-1)})(z_{j(t-s)} - z_{j(t-s-1)})], \text{ if } n = \max(k, s).$$

For A(iv) the finite sum and consequently the infinite sum are not non-zero for the whole range of k and s. So we have the following non-trivial cases,

(a) $\Lambda^{(1)} = \delta_i\delta_j(\lambda_i - 1)(\lambda_j - 1) \sum_{k=0}^{\infty} e_{li}^{(k)} \sum_{s=0}^{\infty} e_{mj}^{(s)}$, for $k = s$, $t_{0i} \neq t_{0j}$,
(b) $\Lambda^{(2)} = \delta_i\delta_j(\lambda_j - 1) \sum_{k=0}^{\infty} e_{li}^{(k)} \sum_{s=0}^{\infty} e_{mj}^{(s)}$, for $s = k + t_{0i}$,
(c) $\Lambda^{(3)} = \delta_i\delta_j(\lambda_i - 1) \sum_{k=0}^{\infty} e_{li}^{(k)} \sum_{s=0}^{\infty} e_{mj}^{(s)}$, for $k = s + t_{0j}$,
(d) $\Lambda^{(4)} = \delta_i\delta_j \sum_{k=0}^{\infty} e_{li}^{(k)} \sum_{s=0}^{\infty} e_{mj}^{(s)}$, for $s = k + t_{0i} - t_{0j}$ and $t_{0i} > t_{0j}$,
(e) $\Lambda^{(5)} = \delta_i\delta_j \sum_{k=0}^{\infty} e_{li}^{(k)} \sum_{s=0}^{\infty} e_{mj}^{(s)}$, for $k = s + t_{0j} - t_{0i}$ and $t_{0j} > t_{0i}$.

Given the above cases we have,

$$\sum_{k=0}^{\infty}\sum_{s=0}^{\infty} e_{li}^{(k)} e_{mj}^{(s)} [\sum_{t=n+1}^{T} (z_{i(t-k)} - z_{i(t-k-1)})(z_{j(t-s)} - z_{j(t-s-1)})] =$$

$$\Lambda^{(h)} < \infty, \ h = 1, 2, 3, 4, 5.$$

Thus, $\sum_{t=1}^{T}[C_1^1(L)(1 - L)z_t][z_t'(1 - L)C_1^1(L)']$ is $O(1)$.

A. (v) $\sum_{t=1}^{T} C_1^1(L)z_t =$

$$\sum_{t=1}^{T}[C_1^1(1) + C_1^{1*}(L)(1 - L)]z_t =$$

$$C_1^1(1)\sum_{t=1}^{T} z_t + \sum_{t=1}^{T} C_1^{1*}(L)(1 - L)z_t.$$

The first term sums to zero by definition (see (4.5)) and the second term is $O(1)$ because it has the same form as A(ii). So, $\sum_{t=1}^{T} C_1^1(L)z_t$ is $O(1)$.

A. (vi) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][z_{t-1}'(1-L)C_1^1(L)'] =$

$$\sum_{t=1}^{T}\{[C_1^1(1)+C_1^{1*}(L)(1-L)]z_{t-1}\}[z_{t-1}'(1-L)C_1^1(L)'] =$$

$$C_1^1(1)\sum_{t=1}^{T} z_{t-1}z_{t-1}'(1-L)C_1^1(L)' + \sum_{t=1}^{T} C_1^{1*}(L)(1-L)z_{t-1}z_{t-1}'(1-L)C_1^1(L)'$$

which by A(iii) and A(iv) is the sum of two $O(1)$ terms, therefore $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][z_{t-1}'(1-L)C_1^1(L)']$ is $O(1)$.

B. (i) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][z_{t-1}'C_1^1(L)'] =$

$$\sum_{t=1}^{T}[C_1^1(1)z_{t-1}+C_1^{1*}(L)(1-L)z_{t-1}][z_{t-1}'C_1^1(1)'+z_{t-1}'(1-L)C_1^{1*}(L)'] =$$

$$C_1^1(1)\sum_{t=1}^{T} z_{t-1}z_{t-1}'C_1^1(1)' + C_1^1(1)\sum_{t=1}^{T} z_{t-1}z_{t-1}'(1-L)C_1^{1*}(L)'$$

$$+\sum_{t=1}^{T} C_1^{1*}(L)(1-L)z_{t-1}z_{t-1}'C_1^1(1)' + \sum_{t=1}^{T} C_1^{1*}(L)(1-L)z_{t-1}z_{t-1}'(1-L)C_1^{1*}(L)'.$$

The first term is $O(T)$ (see O'Brien (1997, p. 25)) and the remaining terms are $O(1)$ by A(iii) and A(iv).So,

$T^{-1}\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][z_{t-1}'C_1^1(L)'] \rightarrow T^{-1}C_1^1(1)\sum_{t=1}^{T} z_{t-1}z_{t-1}'C_1^1(1)' \rightarrow C_1^1(1)gC_1^1(1)'$, with the (i,j)-th element of g being $\delta_i\delta_j\lambda_l(1-\lambda_m)$ and $\lambda_l = \min(\lambda_i,\lambda_j)$ and $\lambda_m = \max(\lambda_i,\lambda_j)$, see O'Brien (1997, p. 25).

B. (ii) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}]z_{t-1}' =$

$$\sum_{t=1}^{T}[C_1^1(1)+C_1^{1*}(L)(1-L)]z_{t-1}z_{t-1}' =$$

$$C_1^1(1)\sum_{t=1}^{T} z_{t-1}z_{t-1}' + \sum_{t=1}^{T} C_1^{1*}(L)(1-L)z_{t-1}z_{t-1}'.$$

The first term is $O(T)$ (see O'Brien (1997, p. 25)) and the second term is $O(1)$ by A(iii).

So, $T^{-1}\sum_{t=1}^{T}[C_1^1(L)z_{t-1}]z_{t-1}' \rightarrow T^{-1}C_1^1(1)\sum_{t=1}^{T} z_{t-1}z_{t-1}' \rightarrow C_1^1(1)g$.

B. (iii) $\sum\limits_{t=1}^{T} Z_{t-1}[z_t^{'}(1-L)C_1^1(L)^{'}] =$

$$\sum_{i=1}^{p_1}\begin{bmatrix} \sum\limits_{t=1}^{T} Z_{1(t-1)}[z_{it}(1-L)e_{1i}(L)] & \cdots & \sum\limits_{t=1}^{T} Z_{1(t-1)}[z_{it}(1-L)e_{pi}(L)] \\ \vdots & \ddots & \vdots \\ \sum\limits_{t=1}^{T} Z_{p_1(t-1)}[z_{it}(1-L)e_{1i}(L)] & \cdots & \sum\limits_{t=1}^{T} Z_{1(t-1)}[z_{it}(1-L)e_{pi}(L)] \end{bmatrix}.$$

The (l,m)-th element of the matrix is

$$\sum_{t=1}^{T} Z_{l(t-1)}[z_{it}(1-L)e_{mi}(L)] = \sum_{t=1}^{T}[\sum_{k=0}^{\infty} e_{mi}^{(k)}(z_{i(t-k)} - z_{i(t-k-1)})Z_{l(t-1)}]$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}[\sum_{t=k+1}^{T}(z_{i(t-k)} - z_{i(t-k-1)})Z_{l(t-1)}] =$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}[(z_{i(1)} - z_{i(0)})Z_{l(k)} + \cdots + (z_{i(T-k)} - z_{i(T-k-1)})Z_{l(T-1)}] =$$

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_k$$

where

$$\Lambda_k = \begin{cases} \delta_i(\lambda_i - 1)Z_{l(k)},\ t_{0i} \geq T-k \Rightarrow k \geq T(1-\lambda_i) \\ \delta_i(\lambda_i - 1)Z_{l(k)} + \delta_i Z_{l(t_{0i}+k)},\ \text{otherwise} \end{cases}.$$

So,

$$\sum_{k=0}^{\infty} e_{mi}^{(k)}\Lambda_k =$$

$$\delta_i(\lambda_i - 1)\sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)}Z_{l(k)} + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)}Z_{l(t_{0i}+k)} + \delta_i(\lambda_i - 1)\sum_{k=T(1-\lambda_i)}^{\infty} e_{mi}^{(k)}Z_{l(k)}$$

$$= \delta_i(\lambda_i - 1)\sum_{k=0}^{\infty} e_{mi}^{(k)}Z_{l(k)} + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)}Z_{l(t_{0i}+k)}.$$

For the first term above we have,

$$Z_{l(k)} = \begin{cases} k\delta_l(\lambda_l - 1),\ k \le t_{0l} = T\lambda_l \\ (k-T)\delta_l\lambda_l, \text{ otherwise} \end{cases},$$

then

$$\delta_i(\lambda_i - 1)\sum_{k=0}^{\infty} e_{mi}^{(k)} Z_{l(k)} =$$

$$\delta_i\delta_l(\lambda_i - 1)(\lambda_l - 1)\sum_{k=0}^{T\lambda_l} k e_{mi}^{(k)} + \delta_i\delta_l(\lambda_i - 1)\lambda_l \sum_{k=T\lambda_l+1}^{\infty} k e_{mi}^{(k)} - T\delta_i\delta_l(\lambda_i - 1)\lambda_l \sum_{k=T\lambda_l+1}^{\infty} e_{mi}^{(k)} =$$

$O(1)$

because $\sum_{k=0}^{T\lambda_l} k e_{mi}^{(k)} < \sum_{k=0}^{\infty} k e_{mi}^{(k)} < \infty$ (see properties of $e_{ij}(y)$ above); $\sum_{k=T\lambda_l+1}^{\infty} k e_{mi}^{(k)} < \sum_{k=T\lambda_l+1}^{\infty} k a^k = (T\lambda_l + 1)a^{T\lambda_l+1}/(1-a) + a^{T\lambda_l+2}/(1-a)^2 \to 0$ as $T \to \infty$ and $T \sum_{k=T\lambda_l+1}^{\infty} e_{mi}^{(k)} < T \sum_{k=T\lambda_l+1}^{\infty} a^k = Ta^{T\lambda_l+1}/(1-a) \to 0$ as $T \to \infty$ and $|e_{mi}^{(k)}| < a^k$, $0 < a < 1$. For the second term we have,

$$Z_{l(t_{0i}+k)} = \begin{cases} (t_{0i}+k)\delta_l(\lambda_l - 1),\ t_{0l} \ge t_{0i} + k \\ t_{0l}\delta_l(\lambda_l - 1) + (t_{0i} + k - t_{0l})\delta_l\lambda_l,\ t_{0l} < t_{0i} + k \end{cases}$$

and

$$\delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)} Z_{l(t_{0i}+k)} =$$

$$\delta_i\delta_l(\lambda_l - 1)\sum_{k=0}^{t_{0l}-t_{0i}} (t_{0i}+k)e_{mi}^{(k)} + \sum_{k=t_{0l}-t_{0i}+1}^{T(1-\lambda_i)-1} [t_{0l}\delta_l(\lambda_l - 1) + (t_{0i} + k - t_{0l})\delta_l\lambda_l]e_{mi}^{(k)}$$

$$= T\delta_i\delta_l\lambda_l(\lambda_l - 1)\sum_{k=0}^{T(\lambda_l-\lambda_i)} e_{mi}^{(k)} + \delta_i\delta_l(\lambda_l - 1)\sum_{k=0}^{T(\lambda_l-\lambda_i)} k e_{mi}^{(k)}$$

$$+T\delta_l\lambda_l(\lambda_l - 1)\sum_{k=T(\lambda_l-\lambda_i)+1}^{T(1-\lambda_i)-1} e_{mi}^{(k)} + \delta_l\lambda_l \sum_{k=T(\lambda_l-\lambda_i)+1}^{T(1-\lambda_i)-1} k e_{mi}^{(k)} + T(\lambda_i - \lambda_l)\delta_l\lambda_l \sum_{k=T(\lambda_l-\lambda_i)+1}^{T(1-\lambda_i)-1} e_{mi}^{(k)}$$

which is $O(T)$ because of the first term. The second term is convergent and the remaining terms tend to zero as $T \to \infty$ (using similar arguments to those used for the components of the first term above). Therefore, $\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k$ and hence $\sum_{t=1}^{T} Z_{t-1}[z_t'(1-L)C_1^1(L)']$ are $O(T)$.

B. (iv) $\sum_{t=1}^{T}(t-1)[z_t'(1-L)C_1^1(L)'] =$

$$\sum_{i=1}^{p_1}\sum_{t=1}^{T}(t-1)\left[\begin{array}{ccc} (z_{it}-z_{i(t-1)})e_{1i}(L) & \cdots & (z_{it}-z_{i(t-1)})e_{pi}(L) \end{array}\right]$$

Extracting a typical element e.g. the j-th we have

$$\sum_{t=1}^{T}(t-1)(z_{it}-z_{i(t-1)})e_{ji}(L) = \sum_{t=1}^{T}(t-1)[\sum_{k=0}^{\infty} e_{ji}^{(k)}(z_{i(t-k)}-z_{i(t-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{ji}^{(k)}[\sum_{t=k+1}^{T}(t-1)(z_{i(t-k)}-z_{i(t-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{ji}^{(k)}[k(z_{i(1)}-z_{i(0)}) + (k+1)(z_{i(2)}-z_{i(1)}) + \cdots + (T-1)(z_{i(T-k)}-z_{i(T-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{ji}^{(k)}\Lambda_k$$

where,

$$\Lambda_k = \begin{cases} \delta_i(\lambda_i - 1)k,\ t_{0i} \geq T-k \Rightarrow k \geq T(1-\lambda_i) \\ \delta_i\lambda_i k + T\delta_i\lambda_i, \text{ otherwise} \end{cases}.$$

Then,

$$\sum_{k=0}^{\infty} e_{ji}^{(k)}\Lambda_k =$$

$$\delta_i\lambda_i \sum_{k=0}^{T(1-\lambda_i)-1} k e_{ji}^{(k)} + T\delta_i\lambda_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{ji}^{(k)} + \delta_i(\lambda_i-1)\sum_{k=T(1-\lambda_i)}^{\infty} k e_{ji}^{(k)} =$$

$$\delta_i\lambda_i\sum_{k=0}^{\infty}ke_{ji}^{(k)}-\delta_i\sum_{k=T(1-\lambda_i)}^{\infty}ke_{ji}^{(k)}+T\delta_i\lambda_i\sum_{k=0}^{T(1-\lambda_i)-1}e_{ji}^{(k)}=O(T)$$

because $T\delta_i\lambda_i\sum\limits_{k=0}^{T(1-\lambda_i)-1}e_{ji}^{(k)}<T\delta_i\lambda_i\sum\limits_{k=0}^{T(1-\lambda_i)-1}a^k=T\delta_i\lambda_i(\frac{1-a^{T(1-\lambda_i)}}{1-a})=O(T)$. Thus, the sum $\sum\limits_{t=1}^{T}(t-1)[(z_{it}-z_{i(t-1)})e_{ji}(L)]$ has to be scaled by T^{-1} to be convergent. It then follows that $\sum\limits_{t=1}^{T}(t-1)[z_t'(1-L)C_1^1(L)']$ is $O(T)$.

C. (i) $\sum\limits_{t=1}^{T}Z_{t-1}[z_{t-1}'C_1^1(L)']=$

$$\sum_{t=1}^{T}Z_{t-1}\{z_{t-1}'[C_1^1(1)'+(1-L)C_1^{1*}(L)']\}=$$

$$\sum_{t=1}^{T}Z_{t-1}z_{t-1}'C_1^1(1)'+\sum_{t=1}^{T}Z_{t-1}z_{t-1}'(1-L)C_1^{1*}(L)'.$$

The first term is $O(T^2)$ (see O'Brien (1997, p. 23)) and the second term is $O(T)$ by B(iii). Scaling by T^{-2} we get

$$T^{-2}\sum_{t=1}^{T}Z_{t-1}[z_{t-1}'C_1^1(L)'] \rightarrow$$

$$T^{-2}\sum_{t=1}^{T}Z_{t-1}z_{t-1}'C_1^1(1)' \rightarrow \int_0^1 Z(u)z(u)'duC_1^1(1)'.$$

For the (i,j)-th element of the limit above we have

$$T^{-2}\sum_{t=1}^{T}Z_{i(t-1)}z_{j(t-1)}' \rightarrow [\delta_i\delta_j\lambda_l(\lambda_m-1)h(\lambda_i,\lambda_j)]/2,$$

with $h(\lambda_i,\lambda_j)=\left\{\begin{array}{c}\lambda_i-\lambda_j,\ \lambda_i\neq\lambda_j\\ T^{-1},\ \lambda_i=\lambda_j\end{array}\right.$ and λ_l, λ_m defined as above. A detailed derivation of this result can be found in O'Brien (1997, pp. 23-24).

C. (ii) $\sum\limits_{t=1}^{T}(t-1)[z_{t-1}'C_1^1(L)']=$

$$\sum_{t=1}^{T}(t-1)\{z_{t-1}'[C_1^1(1)'+(1-L)C_1^{1*}(L)']\}=$$

$$\sum_{t=1}^{T}(t-1)z_{t-1}'C_1^1(1)' + \sum_{t=1}^{T}(t-1)z_{t-1}'(1-L)C_1^{1*}(L)'.$$

The first term is $O(T^2)$ (see O'Brien (1996, p. 35) and the second term is $O(T)$ by B(iv). Therefore, $\sum_{t=1}^{T}(t-1)[z_{t-1}'C_1^1(L)'] = O(T^2)$ since it is the sum of an $O(T^2)$ and an $O(T)$ term.

D. (i) $\sum_{t=1}^{T}\varepsilon_t[z_t'(1-L)C_1^1(L)'] =$

$$\sum_{i=1}^{p_1}\begin{bmatrix} \sum_{t=1}^{T}\varepsilon_{1t}[(z_{it}-z_{i(t-1)})e_{1i}(L)] & \cdots & \sum_{t=1}^{T}\varepsilon_{1t}[(z_{it}-z_{i(t-1)})e_{pi}(L)] \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^{T}\varepsilon_{pt}[(z_{it}-z_{i(t-1)})e_{1i}(L)] & \cdots & \sum_{t=1}^{T}\varepsilon_{pt}[(z_{it}-z_{i(t-1)})e_{pi}(L)] \end{bmatrix}.$$

Analysing the typical (l,m)-th element gives

$$\sum_{t=1}^{T}\varepsilon_{lt}[(z_{it}-z_{i(t-1)})e_{mi}(L)] = \sum_{t=1}^{T}\varepsilon_{lt}[\sum_{k=0}^{\infty}e_{li}^{(k)}(z_{i(t-k)}-z_{i(t-k-1)})]$$

$$= \sum_{k=0}^{\infty}e_{li}^{(k)}[\sum_{t=k+1}^{T}\varepsilon_{lt}(z_{i(t-k)}-z_{i(t-k-1)})]$$

$$= \sum_{k=0}^{\infty}e_{li}^{(k)}[\varepsilon_{l(k+1)}(z_{i(1)}-z_{i(0)}) + \varepsilon_{l(k+2)}(z_{i(2)}-z_{i(1)}) + \cdots + \varepsilon_{l(T)}(z_{i(T-k)}-z_{i(T-k-1)})]$$

$$= \sum_{k=0}^{\infty}e_{li}^{(k)}\Lambda_k$$

where,

$$\Lambda_k = \begin{cases} \delta_i(\lambda_i-1)\varepsilon_{l(k+1)}, & t_{0i} \geq T-k \Rightarrow k \geq T(1-\lambda_i) \\ \delta_i(\lambda_i-1)\varepsilon_{l(k+1)} + \delta_i\varepsilon_{l(t_{0i}+k+1)}, & \text{otherwise} \end{cases}.$$

Then,

$$\sum_{k=0}^{\infty}e_{ji}^{(k)}\Lambda_k =$$

$$\delta_i(\lambda_i - 1)\sum_{k=0}^{T(1-\lambda_i)-1} e_{li}^{(k)}\varepsilon_{l(k+1)} + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{li}^{(k)}\varepsilon_{l(t_{0i}+k+1)} + \delta_i(\lambda_i - 1)\sum_{k=T(1-\lambda_i)}^{\infty} e_{li}^{(k)}\varepsilon_{l(k+1)}$$

$$= \delta_i(\lambda_i - 1)\sum_{k=0}^{\infty} e_{li}^{(k)}\varepsilon_{l(k+1)} + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{li}^{(k)}\varepsilon_{l(t_{0i}+k+1)} = O_p(1)$$

because it is the sum of an infinite and a finite weighted sum of $i.i.d.$ random variables (with exponentially decreasing weights) with zero means and finite variances $\delta_i^2(\lambda_i - 1)^2\omega_{ll}\sum_{k=0}^{\infty}(e_{li}^{(k)})^2$ and $\delta_i^2\omega_{ll}\sum_{k=0}^{T(1-\lambda_i)-1}(e_{li}^{(k)})^2 < \delta_i^2\omega_{ll}(1-a^{2T(1-\lambda_i)})/(1-a^2)$ for the first and second term respectively, where ω_{ll} is the (l,l)-th element of $\Omega = Var(\varepsilon_t)$ and $|e_{li}^{(k)}| < a^k$, $0 < a < 1$. Therefore, $\sum_{t=1}^{T}\varepsilon_t[z_t'(1-L)C_1^1(L)']$ is $O_p(1)$.

D. (ii) $\sum_{t=1}^{T}[(1-L)v_t]z_t' =$

$$\sum_{t=1}^{T}(1-L)\begin{bmatrix} v_{1t}z_{1t} & \cdots & v_{1t}z_{p_1t} \\ \vdots & \ddots & \vdots \\ v_{pt}z_{1t} & \cdots & v_{pt}z_{p_1t} \end{bmatrix}$$

and the (l,m)-th element has the form

$$\sum_{t=1}^{T}(v_{lt} - v_{l(t-1)})z_{mt} =$$

$$[(v_{l(1)} - v_{l(0)})z_{m1} + (v_{l(2)} - v_{l(1)})z_{m2} + \cdots + (v_{l(T)} - v_{l(T-1)})z_{mT}] =$$

$$(v_{l(t_{0m})} - v_{l(0)})\delta_m(\lambda_m - 1) + (v_{l(T)} - v_{l(t_{0m})})\delta_m\lambda_m =$$

$$\delta_m[\lambda_m(v_{l(T)} - v_{l(0)}) - (v_{l(t_{0m})} - v_{l(0)})]$$

which is $O_p(1)$ because it has zero mean and finite variance since $v_t = C_1(L)\varepsilon_t$ with $E(v_t) = 0$, $Var(v_t) = \sum_{i=0}^{\infty}C_{1i}\Omega C_{1i}'$, $Cov(v_t, v_{t+h}) = \sum_{i=0}^{\infty}C_{1i}\Omega C_{1(i+h)}'$ and ε_t is $i.i.d.$ with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \Omega$. Hence, $\sum_{t=1}^{T}[(1-L)v_t]z_t'$ is $O_p(1)$.

D. (iii) $\sum_{t=1}^{T}[(1-L)v_t][z_t'(1-L)C_1^1(L)'] =$

$$\sum_{i=1}^{p_1}\begin{bmatrix} \sum_{t=1}^{T}[(1-L)v_{1t}][z_{it}(1-L)e_{1i}(L)] & \cdots & \sum_{t=1}^{T}[(1-L)v_{1t}][z_{it}(1-L)e_{pi}(L)] \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^{T}[(1-L)v_{pt}][z_{it}(1-L)e_{1i}(L)] & \cdots & \sum_{t=1}^{T}[(1-L)v_{pt}][z_{it}(1-L)e_{pi}(L)] \end{bmatrix}$$

and the (l,m)-th element is given by

$$\sum_{t=1}^{T}(v_{lt}-v_{l(t-1)})[(z_{it}-z_{i(t-1)})e_{mi}(L)] =$$

$$\sum_{t=1}^{T}(v_{lt}-v_{l(t-1)})[\sum_{k=0}^{\infty}e_{mi}^{(k)}(z_{i(t-k)}-z_{i(t-k-1)})] =$$

$$\sum_{k=0}^{\infty}e_{mi}^{(k)}[\sum_{t=k+1}^{T}(v_{lt}-v_{l(t-1)})(z_{i(t-k)}-z_{i(t-k-1)})] =$$

$$\sum_{k=0}^{\infty}e_{mi}^{(k)}[(v_{l(k+1)}-v_{l(k)})(z_{i(1)}-z_{i(0)})+\cdots+(v_{l(T)}-v_{l(T-1)})(z_{i(T-k)}-z_{i(T-k-1)})]$$

$$=\sum_{k=0}^{\infty}e_{mi}^{(k)}\Lambda_k$$

where,

$$\Lambda_k=\begin{cases}\delta_i(\lambda_i-1)(v_{l(k+1)}-v_{l(k)}), & t_{0i}\geq T-k\Rightarrow k\geq T(1-\lambda_i)\\ \delta_i(\lambda_i-1)(v_{l(k+1)}-v_{l(k)})+\delta_i(v_{l(t_{0i}+k+1)}-v_{l(t_{0i}+k)}), & \text{otherwise}\end{cases}.$$

Then,

$$\sum_{k=0}^{\infty}e_{mi}^{(k)}\Lambda_k =$$

$$\delta_i(\lambda_i - 1) \sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)}(v_{l(k+1)} - v_{l(k)}) + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)}(v_{l(t_{0i}+k+1)} - v_{l(t_{0i}+k)})$$
$$+\delta_i(\lambda_i - 1) \sum_{k=T(1-\lambda_i)}^{\infty} e_{mi}^{(k)}(v_{l(k+1)} - v_{l(k)}) =$$

$$\delta_i(\lambda_i - 1) \sum_{k=0}^{\infty} e_{mi}^{(k)}(v_{l(k+1)} - v_{l(k)}) + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} e_{mi}^{(k)}(v_{l(t_{0i}+k+1)} - v_{l(t_{0i}+k)}) =$$

$$\delta_i(\lambda_i - 1) \sum_{k=0}^{\infty}(e_{mi}^{(k-1)} - e_{mi}^{(k)})v_{l(k)} + \delta_i \sum_{k=0}^{T(1-\lambda_i)-1} (e_{mi}^{(k-1)} - e_{mi}^{(k)})v_{l(t_{0i}+k)} + e^{T(1-\lambda_i)-1} v_{l(t_{0i}+T(1-\lambda_i))}$$

with $e_{mi}^{(-1)} = 0$. The above expression is $O_p(1)$ because it is the sum of an infinite and a finite weighted sum (with exponentially decreasing weights) of random variables, which can be expressed as moving averages and therefore they have zero means and finite variances.

D. (iv) $\sum_{t=1}^{T} v_{t-1}[z_{t-1}'(1-L)C_1^1(L)']$ is $O_p(1)$. The proof parallels that of D(i) since ε_t and v_t are of the same stochastic order of magnitude, i.e. $O_p(1)$.

D. (v) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}][v_t'(1-L)] =$

$$\sum_{t=1}^{T}\{[C_1^1(1) + C_1^{1*}(L)(1-L)]z_{t-1}\}v_t'(1-L) =$$

$$C_1^1(1)\sum_{t=1}^{T} z_{t-1}v_t'(1-L) + \sum_{t=1}^{T} C_1^{1*}(L)(1-L)z_{t-1}v_t'(1-L),$$

which is $O_p(1)$ because it is the sum of two $O_p(1)$ terms, by D(ii) and D(iii).

E. (i) From (A.8) it follows that the sum $\sum_{t=1}^{T} v_{t-1}z_{t-1}'$ is $O_p(T^{1/2})$.

E. (ii) $\sum_{t=1}^{T} v_{t-1}[z_{t-1}^{'}C_1^1(L)^{'}] =$

$$\sum_{t=1}^{T} v_{t-1}\{z_{t-1}^{'}[C_1^1(1)^{'} + (1-L)C_1^{1*}(L)^{'}]\} =$$

$$\sum_{t=1}^{T} v_{t-1}z_{t-1}^{'}C_1^1(1)^{'} + \sum_{t=1}^{T} v_{t-1}z_{t-1}^{'}(1-L)C_1^{1*}(L)^{'},$$

which is $O_p(T^{1/2})$ because the first term is $O_p(T^{1/2})$, by E(i), and the second is $O_p(1)$, by D(iv).

E. (iii) $\sum_{t=1}^{T}[C_1^1(L)z_{t-1}]\varepsilon_t^{'} =$

$$\sum_{t=1}^{T}\{[C_1^1(1) + C_1^{1*}(L)(1-L)]z_{t-1}\}\varepsilon_t^{'} =$$

$$C_1^1(1)\sum_{t=1}^{T} z_{t-1}\varepsilon_t^{'} + \sum_{t=1}^{T} C_1^{1*}(L)(1-L)z_{t-1}\varepsilon_t^{'},$$

which is $O_p(T^{1/2})$ because the first term is $O_p(T^{1/2})$, by (A.4), and the second is $O_p(1)$, by D(i).

E. (iv) $\sum_{t=1}^{T} \xi_{t-1}[z_t^{'}(1-L)C_1^1(L)^{'}] =$

$$\sum_{i=1}^{p_1}\begin{bmatrix} \sum_{t=1}^{T} \xi_{1(t-1)}[(z_{it} - z_{i(t-1)})e_{1i}(L)] & \cdots & \sum_{t=1}^{T} \xi_{1(t-1)}[(z_{it} - z_{i(t-1)})e_{pi}(L)] \\ \vdots & \ddots & \vdots \\ \sum_{t=1}^{T} \xi_{p(t-1)}[(z_{it} - z_{i(t-1)})e_{1i}(L)] & \cdots & \sum_{t=1}^{T} \xi_{p(t-1)}[(z_{it} - z_{i(t-1)})e_{pi}(L)] \end{bmatrix}$$

with (l,m)-th element

$$\sum_{t=1}^{T} \xi_{l(t-1)}[(z_{it} - z_{i(t-1)})e_{mi}(L)] = \sum_{t=1}^{T} \xi_{l(t-1)}[\sum_{k=0}^{\infty} e_{mi}^{(k)}(z_{i(t-k)} - z_{i(t-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{mi}^{(k)}[\sum_{t=k+1}^{T} \xi_{l(t-1)}(z_{i(t-k)} - z_{i(t-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{mi}^{(k)} [\xi_{l(k)}(z_{i(1)} - z_{i(0)}) + \xi_{l(k+1)}(z_{i(2)} - z_{i(1)}) + \cdots + \xi_{l(T-1)}(z_{i(T-k)} - z_{i(T-k-1)})]$$

$$= \sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k$$

where,

$$\Lambda_k = \begin{cases} \delta_i(\lambda_i - 1)\xi_{l(k)}, \ t_{0i} \geq T - k \Rightarrow k \geq T(1 - \lambda_i) \\ \delta_i(\lambda_i - 1)\xi_{l(k)} + \delta_i \xi_{l(t_{0i}+k)}, \text{ otherwise} \end{cases} .$$

Then,

$$\sum_{k=0}^{\infty} e_{mi}^{(k)} \Lambda_k =$$

$$\delta_i(\lambda_i - 1) \sum_{k=0}^{T(1-\lambda i)-1} e_{mi}^{(k)} \xi_{l(k)} + \delta_i \sum_{k=0}^{T(1-\lambda i)-1} e_{mi}^{(k)} \xi_{l(t_{0i}+k)} + \delta_i(\lambda_i - 1) \sum_{k=T(1-\lambda i)}^{\infty} e_{mi}^{(k)} \xi_{l(k)} =$$

$$\delta_i(\lambda_i - 1) \sum_{k=0}^{\infty} e_{mi}^{(k)} \xi_{l(k)} + \delta_i \sum_{k=0}^{T(1-\lambda i)-1} e_{mi}^{(k)} \xi_{l(t_{0i}+k)} =$$

$$\delta_i(\lambda_i - 1) \sum_{k=0}^{\infty} (\sum_{j=k}^{\infty} e_{mi}^{(j)}) \varepsilon_{l(k)} + \delta_i(\sum_{k=0}^{T(1-\lambda i)-1} e_{mi}^{(k)}) \xi_{l(t_{0i})} + \delta_i \sum_{k=1}^{T(1-\lambda i)-1} (\sum_{j=k}^{T(1-\lambda i)-1} e_{mi}^{(j)}) \varepsilon_{l(t_{0i}+k)}$$

which is $O_p(T^{1/2})$ because of $\xi_{l(t_{0i})}$ (see (A.1)). Thus, $\sum_{t=1}^{T} \xi_{t-1}[z_t'(1-L)C_1^1(L)']$ is $O_p(T^{1/2})$.

F. $\sum_{t=1}^{T} \xi_{t-1}[z_{t-1}' C_1^1(L)'] =$

$$\sum_{t=1}^{T} \xi_{t-1}\{z_{t-1}'[C_1^1(1)' + (1 - L)C_1^{1*}(L)']\} =$$

$$\sum_{t=1}^{T} \xi_{t-1} z_{t-1}' C_1^1(1)' + \sum_{t=1}^{T} \xi_{t-1} z_{t-1}' (1 - L) C_1^{1*}(L)'.$$

The first term is $O_p(T^{3/2})$ by (A.5) and the second term is $O_p(T^{1/2})$ by E(iv). So, the first term dominates asymptotically and $\sum_{t=1}^{T} \xi_{t-1}[z_{t-1}' C_1^1(L)']$ is $O_p(T^{3/2})$. Scaling by $T^{-3/2}$ and using (A.5) we have

$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}[z_{t-1}^{'}C_1^1(L)^{'}] \xrightarrow{d}$$
$$T^{-3/2}\sum_{t=1}^{T}\xi_{t-1}z_{t-1}^{'}C_1^1(1)^{'} \xrightarrow{d} \int_0^1 W(u)z(u)^{'}duC_1^1(1)^{'}.$$

For the (i,j)-th element of the above limit we have

$$T^{-3/2}\sum_{t=1}^{T}\xi_{i(t-1)}z_{j(t-1)}^{'} \xrightarrow{d}$$
$$\int_0^1 W_i(u)z_j(u)du = \delta_j(\lambda_j-1)\int_0^{\lambda_j} W_i(u)du + \delta_j\lambda_j\int_{\lambda_j}^1 W_i(u)du$$

see O'Brien (1997, p. 29). ■

Next, let

$$Var\begin{bmatrix} \Delta X_t \\ \beta^{'}X_{t-1} \end{bmatrix} = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}$$

which is the covariance matrix of the stationary components when X_t is $I(1)$ and cointegrated. Because of the absence of short-run dynamics in the model (VAR(1)) the conditional means and variances coincide with the unconditional. From the representation (2.21) we have

$$E(\Delta X_t) = C(L)\Phi D_t$$

and

$$Var(\Delta X_t) = E\{[C(L)\varepsilon_t][\varepsilon_t^{'}C(L)^{'}]\} = \sum_{i=0}^{\infty} C_i\Omega C_i^{'} = \Sigma_{00}. \qquad \text{(A.11)}$$

From the representation (2.22) we have

$$E(\beta^{'}X_{t-1}) = \beta^{'}C_1(L)\Phi D_{t-1}$$

and

$$Var(\beta' X_{t-1}) = E\{\beta'[C_1(L)\varepsilon_{t-1}][\varepsilon'_{t-1}C_1(L)']\beta\} = \beta' \sum_{i=0}^{\infty} C_{1i}\Omega C'_{1i}\beta = \Sigma_{\beta\beta}. \quad \text{(A.12)}$$

We also have

$$Cov(\Delta X_t, \beta' X_{t-1}) = E\{[C(L)\varepsilon_t][\varepsilon'_{t-1}C_1(L)']\beta\} = \sum_{i=0}^{\infty} C_{i+1}\Omega C'_{1i}\beta = \Sigma_{0\beta}. \quad \text{(A.13)}$$

The mean of the stationary functions of X_t is affected by a smoothed version of the intercept shift since the step dummies enter in an infinite lag form. To give an example, we can analyse

$$C^1(L)z_t =$$

$$\begin{bmatrix} h_{11}(L) & h_{12}(L) & \cdots & h_{1p_1}(L) \\ \vdots & & & \vdots \\ h_{p1}(L) & h_{p2}(L) & \cdots & h_{pp_1}(L) \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \\ \vdots \\ z_{p_1t} \end{bmatrix} =$$

$$\sum_{j=1}^{p_1} \begin{bmatrix} h_{1j}(L)z_{jt} \\ h_{2j}(L)z_{jt} \\ \vdots \\ h_{pj}(L)z_{jt} \end{bmatrix} \quad \text{(A.14)}$$

where $C^1(L)$ is the $p \times p_1$ submatrix of $C(L)$ in the partition $C(L) = \left[\underset{p\times p_1}{C^1(L)} \quad \underset{p\times p_2}{C^2(L)} \right]$ and $h_{ij}(L) = \sum_{k=0}^{\infty} h_{ij}^{(k)} L^k$. Then the i-th element of (A.14) takes the form

$$\sum_{j=1}^{p_1} h_{ij}(L)z_{jt} = \begin{cases} \sum_{j=1}^{p_1} \delta_i(\lambda_i - 1) \sum_{k=0}^{t-t_{0i}-1} h_{ij}^{(k)},\ t \leq t_{0i} + k \\ \sum_{j=1}^{p_1} \delta_i\lambda_i \sum_{k=t-t_{0i}}^{\infty} h_{ij}^{(k)},\ t > t_{0i} + k \end{cases} ,i = 1, \ldots, p. \quad \text{(A.15)}$$

Thus $E(\Delta X_t) = C(L)\Phi D_t = C^1(L)z_t + C^2(1)\varphi$ (where the second equality follows by introducing the partition of $C(L)$ mentioned above) is affected by the 'smoothed' intercept shifts (step dummy variables), by (A.14) and (A.15).

Appendix B: Proofs of lemma 4.1 and 4.2

Proof of Lemma 4.1.

Using the representation (2.22) we analyse the limiting behaviour of $X_{[Tu]}$ in the directions of γ and τ. First we need to show that $T^{-1/2}\bar{\gamma}' X_{[Tu]} \xrightarrow{d} \bar{\gamma}' CW(u)$. This is the direction that annihilates the deterministic components.

$$\begin{aligned} T^{-1/2}\bar{\gamma}' X_{[Tu]} &= T^{-1/2}[\bar{\gamma}' C \sum_{i=1}^{[Tu]} \varepsilon_i + \bar{\gamma}' C_1(L)(\varepsilon_{[Tu]} + \Phi D_{[Tu]}) + \bar{\gamma}' A] \\ &= T^{-1/2}\bar{\gamma}'[C\xi_{[Tu]} + v_{[Tu]} + C_1^1(L)z_{[Tu]} + C_1^2(1)\varphi + A] \\ &\xrightarrow{d} \bar{\gamma}' CW(u) \end{aligned}$$

by (A.1) for the first term, (A.6) and A(i) which show that in the second equality above the second and third terms are $o_p(T^{1/2})$ and $O(1)$ respectively.

Then, we need to show that $T^{-1}\bar{\tau}' X_{[Tu]} \xrightarrow{d} \begin{bmatrix} Z(u) \\ u \end{bmatrix}$. This is the direction in which deterministic terms dominate stochastic trends (random walk process).

$$\begin{aligned} T^{-1}\bar{\tau}' X_{[Tu]} &= T^{-1}[\bar{\tau}' C \sum_{t=1}^{[Tu]} \varepsilon_i + \bar{\tau}'\tau \begin{bmatrix} Z_{[Tu]} \\ [Tu] \end{bmatrix} + \bar{\tau}'(C_1(L)\varepsilon_{[Tu]} + \Phi D_{[Tu]}) + \bar{\tau}' A] \\ &= T^{-1}[\bar{\tau}' C\xi_{[Tu]} + \begin{bmatrix} Z_{[Tu]} \\ [Tu] \end{bmatrix} + \bar{\tau}'(v_{[Tu]} + C_1^1(L)z_{[Tu]} + C_1^2(1)\varphi + A)] \\ &\xrightarrow{d} \begin{bmatrix} Z(u) \\ u \end{bmatrix} \end{aligned}$$

by (4.7) for the second term, (A.1), (A.6) and A(i) which show that in the second equality above the first, third and fourth terms are $O_p(T^{1/2})$, $o_p(T^{1/2})$ and $O(1)$ respectively. Combining the two limiting results above we get $G_0(u)$ given in (4.15).

Finally we need to show that

$$T^{-1/2}B_T\bar{X} = \begin{bmatrix} T^{-1/2}\bar{\gamma}'\bar{X} \\ T^{-1}\bar{\tau}'\bar{X} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \bar{\gamma}'C\int_0^1 W(u)du \\ \int_0^1 Z(u)du \\ 1/2 \end{bmatrix}.$$

This result follows directly from the CMT (Theorem A.3) because the mapping $\mathcal{I}_1 : x \mapsto \int_0^1 x(u)du$ is continuous, for $x(\cdot)$ a continuous function on $[0,1]$. Therefore, combining the CMT with (4.15) we get $\bar{G}_0$. Putting together $G_0(u)$ and $\bar{G}_0$ we get (4.16). ■

Proof of Lemma 4.2.

Proof of (4.17).

We need to show that $S_{00} \xrightarrow{p} \Sigma_{00} + C\Phi R\Phi'C'$.

$S_{00} = T^{-1}\sum_{t=1}^{T}\Delta X_t\Delta X_t' - \bar{\Delta X}\bar{\Delta X}'$. We analyse the first term by substituting the representation (2.21) for ΔX_t. So,

$$T^{-1}\sum_{t=1}^{T}\Delta X_t\Delta X_t' = T^{-1}\sum_{t=1}^{T}C(L)(\varepsilon_t+\Phi D_t)(\varepsilon_t+\Phi D_t)'C(L)' =$$

$$T^{-1}\sum_{t=1}^{T}[C(L)\varepsilon_t\varepsilon_t'C(L)' + C(L)\varepsilon_t D_t'\Phi'C(L)' + C(L)\Phi D_t\varepsilon_t'C(L)' + C(L)\Phi D_t D_t'\Phi'C(L)'].$$

The terms of the above expression are numbered (1)-(4) and we analyse each of these terms separately.

(1) $T^{-1}\sum_{t=1}^{T}C(L)\varepsilon_t\varepsilon_t'C(L)' \xrightarrow{p} \Sigma_{00}$ by (A.9) and (A.11).

(2) $T^{-1}\sum_{t=1}^{T}C(L)\varepsilon_t D_t'\Phi'C(L)' =$

$$T^{-1}\sum_{t=1}^{T}[C\varepsilon_t D_t'\Phi'C' + C\varepsilon_t D_t'\Phi'(1-L)C_1(L)' + C_1(L)(1-L)\varepsilon_t D_t'\Phi'C'$$
$$+C_1(L)(1-L)\varepsilon_t D_t'\Phi'(1-L)C_1(L)']$$

and the four terms in this expression are numbered 2(i)-2(iv).

2(i) $T^{-1}\sum_{t=1}^{T} C\varepsilon_t D_t'\Phi' C' = T^{-1}\sum_{t=1}^{T} C\left[\begin{array}{cc} \varepsilon_t z_t' & \varepsilon_t \varphi' \end{array}\right] C' \xrightarrow{p} 0$ by (A.4) for block (1,1) and (A.1) for block (1,2).

2(ii) $T^{-1}\sum_{t=1}^{T} C\varepsilon_t D_t'\Phi'(1-L)C_1(L)' = T^{-1}\sum_{t=1}^{T} C\varepsilon_t(z_t - z_{t-1})'C_1^1(L)' \xrightarrow{p} 0$ by D(i).

2(iii) $T^{-1}\sum_{t=1}^{T} C_1(L)(1-L)\varepsilon_t D_t'\Phi' C' = T^{-1}\sum_{t=1}^{T}\left[\begin{array}{cc} (1-L)v_t z_t' & (1-L)v_t\varphi' \end{array}\right] C' \xrightarrow{p} 0$ by D(ii) for block (1,1) and (A.6) for block (1,2).

2(iv) $T^{-1}\sum_{t=1}^{T} C_1(L)(1-L)\varepsilon_t D_t'\Phi'(1-L)C_1(L)' = T^{-1}\sum_{t=1}^{T}(v_t - v_{t-1})(z_t - z_{t-1})'C_1^1(L)' \xrightarrow{p}$ 0 by D(iii).

So, $T^{-1}\sum_{t=1}^{T} C(L)\varepsilon_t D_t'\Phi' C(L)' \xrightarrow{p} 0$ since all of its components converge in probability to zero.

(3) $T^{-1}\sum_{t=1}^{T} C(L)\Phi D_t\varepsilon_t' C(L)' = T^{-1}\sum_{t=1}^{T}[C(L)\varepsilon_t D_t'\Phi' C(L)']' \xrightarrow{p} 0$, by (2) above.

(4) $T^{-1}\sum_{t=1}^{T} C(L)\Phi D_t D_t'\Phi' C(L)' =$

$$T^{-1}\sum_{t=1}^{T}[C\Phi D_t D_t'\Phi' C' + C\Phi D_t D_t'\Phi'(1-L)C_1(L)' + C_1(L)(1-L)\Phi D_t D_t'\Phi' C'$$
$$+C_1(L)(1-L)\Phi D_t D_t'\Phi'(1-L)C_1(L)'].$$

The terms in this expression are numbered 4(i)-4(iv) and we analyse each one below.

4(i) $T^{-1}\sum_{t=1}^{T} C\Phi D_t D_t'\Phi' C' = T^{-1}\sum_{t=1}^{T} C\left[\begin{array}{cc} z_t z_t' & z_t\varphi' \\ \varphi z_t' & \varphi\varphi' \end{array}\right] C' \to C\left[\begin{array}{cc} g & 0 \\ 0 & \varphi\varphi' \end{array}\right] C'$ $= C\Phi\left[\begin{array}{cc} g & 0 \\ 0 & 1 \end{array}\right]\Phi' C'$, where $T^{-1}\sum_{t=1}^{T} z_t z_t' \to g$, a $p_1 \times p_1$ matrix with (i,j)-th element $\delta_i\delta_j\lambda_l(1-\lambda_m)$ and $\lambda_l = \min(\lambda_i, \lambda_j)$, $\lambda_m = \max(\lambda_i, \lambda_j)$ (see O'Brien (1997, 1999)).

4(ii) $T^{-1}\sum_{t=1}^{T} C\Phi D_t D_t'\Phi'(1-L)C_1(L)' = T^{-1}\sum_{t=1}^{T} C\left[\begin{array}{c} z_t(z_t - z_{t-1})'C_1^1(L)' \\ \varphi(z_t - z_{t-1})'C_1^1(L)' \end{array}\right] \to 0$ by A(iii) for block (1,1) and A(ii) for block (2,1).

4(iii) $T^{-1}\sum_{t=1}^{T} C_1(L)(1-L)\Phi D_t D_t'\Phi' C' = T^{-1}\sum_{t=1}^{T}[C\Phi D_t D_t'\Phi'(1-L)C_1(L)']' \to 0$ because it is the transpose of 4(ii) above.

4(iv) $T^{-1}\sum_{t=1}^{T} C_1(L)(1-L)\Phi D_t D_t'\Phi'(1-L)C_1(L)' =$

$T^{-1}\sum_{t=1}^{T} C_1^1(L)(z_t - z_{t-1})(z_t - z_{t-1})'C_1^1(L)' \to 0$ by A(iv).

Thus, the first term of S_{00}, $T^{-1}\sum_{t=1}^{T}\Delta X_t \Delta X_t' \xrightarrow{p} \Sigma_{00}+C\Phi\begin{bmatrix} g & 0 \\ 0 & 1 \end{bmatrix}\Phi' C'$. Next, we analyse the average, $\bar{\Delta}X$ in order to find a limiting expression for $\bar{\Delta}X\bar{\Delta}X'$.

$$\begin{aligned}
\bar{\Delta}X &= T^{-1}\sum_{t=1}^{T}\Delta X_t = T^{-1}\sum_{t=1}^{T}C(L)(\varepsilon_t + \Phi D_t) \\
&= T^{-1}\sum_{t=1}^{T}[C(L)\varepsilon_t + C\Phi D_t + (1-L)C_1(L)\Phi D_t] \\
&= T^{-1}\sum_{t=1}^{T}C(L)\varepsilon_t + T^{-1}\sum_{t=1}^{T}C\begin{bmatrix} z_t \\ \varphi \end{bmatrix} + T^{-1}\sum_{t=1}^{T}C_1^1(L)(z_t - z_{t-1}) \\
\xrightarrow{p} C\begin{bmatrix} 0 \\ \varphi \end{bmatrix} &= C\Phi\begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{aligned} \tag{B.1}$$

by (4.5) for the second term in the fourth equality, (A.7) and A(ii) which show that in the fourth equality above the first and third terms are $O_p(T^{1/2})$ and $O(1)$ respectively. It then follows from Slutsky's Theorem (see Davidson (2000, pp. 39, 46)), that the second term of S_{00},

$$\bar{\Delta}X\bar{\Delta}X' \xrightarrow{p} C\Phi\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\Phi' C'.$$

Combining the limits of the two terms of S_{00} we have $S_{00} \xrightarrow{p} \Sigma_{00} + C\Phi\begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix}\Phi' C' =$ $\Sigma_{00} + C\Phi R\Phi' C'$, where $R = \begin{bmatrix} g & 0 \\ 0 & 0 \end{bmatrix}$.

Proof of (4.18).

We need to show $\beta' S_{11}\beta \xrightarrow{p} \Sigma_{\beta\beta} + \beta' C_1^1(1) g C_1^1(1)'\beta$.

$\beta' S_{11}\beta = T^{-1}\sum_{t=1}^{T}\beta' X_{t-1}X_{t-1}'\beta - \beta'\bar{X}\bar{X}'\beta$. Since X_t is cointegrated we use representation (2.22) in the first term of $\beta' S_{11}\beta$ and we have

$$T^{-1}\sum_{t=1}^{T}\beta' X_{t-1}X_{t-1}'\beta = T^{-1}\sum_{t=1}^{T}\beta' C_1(L)(\varepsilon_{t-1}+\Phi D_{t-1})(\varepsilon_{t-1}+\Phi D_{t-1})' C_1(L)'\beta =$$

$$T^{-1}\sum_{t=1}^{T}[\beta' C_1(L)\varepsilon_{t-1}\varepsilon_{t-1}' C_1(L)'\beta + \beta' C_1(L)\varepsilon_{t-1}D_{t-1}'\Phi' C_1(L)'\beta$$
$$+\beta' C_1(L)\Phi D_{t-1}\varepsilon_{t-1}' C_1(L)'\beta + \beta' C_1(L)\Phi D_{t-1}D_{t-1}'\Phi' C_1(L)'\beta].$$

The terms in the above expression are numbered (1)-(4).

(1) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\varepsilon_{t-1}\varepsilon_{t-1}' C_1(L)'\beta \xrightarrow{p} \Sigma_{\beta\beta}$, by (A.9) and (A.12).

(2) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\varepsilon_{t-1}D_{t-1}'\Phi' C_1(L)'\beta = T^{-1}\sum_{t=1}^{T}\beta'(v_{t-1}z_{t-1}' C_1^1(L)' + v_{t-1}\varphi' C_1^2(1)')\beta \xrightarrow{p} 0$, by E(ii) and (A.7).

(3) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\Phi D_{t-1}\varepsilon_{t-1}' C_1(L)'\beta = T^{-1}\sum_{t=1}^{T}(\beta' C_1(L)\varepsilon_{t-1}D_{t-1}'\Phi' C_1(L)'\beta)' \xrightarrow{p} 0$, because it is the transpose of (2) above.

(4) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\Phi D_{t-1}D_{t-1}'\Phi' C_1(L)'\beta =$

$$= T^{-1}\sum_{t=1}^{T}\beta'[C_1^1(L)z_{t-1}z_{t-1}' C_1^1(L)' + C_1^2(1)\varphi z_{t-1}' C_1^1(L)' + C_1^1(L)z_{t-1}\varphi' C_1^2(1)'$$
$$+C_1^2(1)\varphi\varphi' C_1^2(1)']\beta \rightarrow \beta'[C_1^1(1)gC_1^1(1)' + C_1^2(1)\varphi\varphi' C_1^2(1)']\beta,$$

by B(i) and A(v). $C_1^1(L)$ and $C_1^2(L)$ are $p\times p_1$ and $p\times p_2$ respectively and they constitute a partition of $C_1(L)$ i.e. $C_1(L) = \left[\begin{array}{cc} \underset{p\times p_1}{C_1^1(L)} & \underset{p\times p_2}{C_1^2(L)} \end{array}\right]$. So for the first term of $\beta' S_{11}\beta$ we have

$T^{-1}\sum_{t=1}^{T}\beta' X_{t-1}X_{t-1}'\beta' \xrightarrow{p} \Sigma_{\beta\beta} + \beta'[C_1^1(1)gC_1^1(1)' + C_1^2(1)\varphi\varphi' C_1^2(1)']\beta$. Then we analyse the limiting behaviour of $\beta'\bar{X}$,

$$\beta'\bar{X} = T^{-1}\sum_{t=1}^{T}\beta' X_{t-1} = T^{-1}\sum_{t=1}^{T}\beta' C_1(L)(\varepsilon_{t-1} + \Phi D_{t-1})$$

$$= T^{-1}\sum_{t=1}^{T}\beta' v_{t-1} + T^{-1}\sum_{t=1}^{T}\beta' C_1^1(L)z_{t-1} + T^{-1}\sum_{t=1}^{T}\beta' C_1^2(1)\varphi$$

$$\xrightarrow{p} \beta' C_1^2(1)\varphi \tag{B.2}$$

since by (A.7) and A(v) the first and second terms of the third equality above are $O_p(T^{1/2})$ and $O(1)$ respectively. For the product of the averages we have

$$\beta'\bar{X}\bar{X}'\beta \xrightarrow{p} \beta' C_1^2(1)\varphi\varphi' C_1^2(1)'\beta.$$

By combining the asymptotic results for the sum of the products and the product of the averages we get the limiting result for $\beta' S_{11}\beta$, which is $\beta' S_{11}\beta \xrightarrow{p} \Sigma_{\beta\beta}+\beta' C_1^1(1)gC_1^1(1)'\beta$.

Proof of (4.19).

We need to show $\beta' S_{10} \xrightarrow{p} \Sigma_{\beta 0} + \beta' C_1^1(1)gC_1'$.

$\beta' S_{10} = T^{-1}\sum_{t=1}^{T}\beta' X_{t-1}\Delta X_t' - \beta'\bar{X}\bar{\Delta X}'$. Using the representations (2.21) and (2.22) we can write the first term of $\beta' S_{10}$ as follows

$$T^{-1}\sum_{t=1}^{T}\beta' X_{t-1}\Delta X_t'$$

$$= T^{-1}\sum_{t=1}^{T}\beta' C_1(L)(\varepsilon_{t-1} + \Phi D_{t-1})(\varepsilon_t + \Phi D_{t-1})' C(L)'$$

$$= T^{-1}\sum_{t=1}^{T}\beta'(C_1(L)\varepsilon_{t-1}\varepsilon_t' C(L)' + C_1(L)\varepsilon_{t-1}D_t'\Phi' C(L)' + C_1(L)\Phi D_{t-1}\varepsilon_t' C(L)'$$
$$+C_1(L)\Phi D_{t-1}D_t'\Phi' C(L)')$$

and the terms are numbered (1)-(4).

(1) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\varepsilon_{t-1}\varepsilon_t' C(L)' \xrightarrow{p} \Sigma_{\beta 0}$ by (A.9) and (A.13).

(2) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\varepsilon_{t-1}D_t'\Phi' C(L)'$

$= T^{-1}\sum_{t=1}^{T}\beta' v_{t-1}D_t'\Phi' C' + T^{-1}\sum_{t=1}^{T}\beta' v_{t-1}D_t'\Phi'(1-L)C_1(L)'$

$= T^{-1}\sum_{t=1}^{T}\beta' \left[\begin{array}{cc} v_{t-1}z_t' & v_{t-1}\varphi' \end{array}\right] C + T^{-1}\sum_{t=1}^{T}\beta' \left[\begin{array}{cc} v_{t-1}(z_t - z_{t-1})' C_1^1(L)' & 0 \end{array}\right] \xrightarrow{p} 0,$

by E(i), (A.7) and D(iv).

(3) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\Phi D_{t-1}\varepsilon_t' C(L)'$

$= T^{-1}\sum_{t=1}^{T}\beta'(C_1^1(L)z_{t-1}\varepsilon_t' C' + C_1^2(1)\varphi\varepsilon_t' C') + T^{-1}\sum_{t=1}^{T}\beta'[C_1^1(L)z_{t-1}(v_t - v_{t-1})'$

$+C_1^2(1)\varphi(v_t - v_{t-1})'] \xrightarrow{p} 0$, by E(iii), (A.1), D(v) and (A.6).

(4) $T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\Phi D_{t-1}D_t'\Phi' C(L)' =$

$T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\Phi D_{t-1}D_t'\Phi' C' + T^{-1}\sum_{t=1}^{T}\beta' C_1(L)\Phi D_{t-1}D_t'\Phi'(1-L)C_1(L)'$

$= T^{-1}\sum_{t=1}^{T}\beta'(C_1^1(L)z_{t-1}z_t' C_1' + C_1^2(1)\varphi z_t' C_1' + C_1^1(L)z_{t-1}\varphi' C_2' + C_1^2(1)\varphi\varphi' C_2')$

$+T^{-1}\sum_{t=1}^{T}\beta'[C_1^1(L)z_{t-1}(z_t - z_{t-1})' C_1^1(L)' + C_1^2(1)\varphi(z_t - z_{t-1})' C_1^1(L)']$

$\xrightarrow{p} \beta'(C_1^1(1)gC_1' + C_1^2(1)\varphi\varphi' C_2')$, by B(ii), (4.5), A(v), A(vi) and A(ii). C_1 and C_2 are $p\times p_1$ and $p \times p_2$ respectively and constitute the partition of C i.e. $C = \left[\begin{array}{cc} \underset{p\times p_1}{C_1} & \underset{p\times p_2}{C_2} \end{array}\right]$.

So the first term of $\beta' S_{10}$ asymptotically takes the form

$$T^{-1}\sum_{t=1}^{T}\beta' X_{t-1}\Delta X_t' \xrightarrow{p} \Sigma_{\beta 0} + \beta'(C_1^1(1)gC_1' + C_1^2(1)\varphi\varphi' C_2').$$

From (B.1) and (B.2) the second term of $\beta' S_{10}$ has the following limiting form, $\beta'\bar{X}\bar{\Delta}X' \xrightarrow{p} \beta' C_1^2(1)\varphi\varphi' C_2'$. Therefore, $\beta' S_{10} \xrightarrow{p} \Sigma_{\beta 0} + \beta' C_1^1(1)gC_1'$.

Proof of (4.20).

We need to show that $T^{-1}B_T' S_{11}B_T \xrightarrow{d} \int_0^1 GG' du$.

$T^{-1}B_T'S_{11}B_T = \begin{bmatrix} \underset{(p-r-q)\times(p-r-q)}{T^{-1}\bar{\gamma}'S_{11}\bar{\gamma}} & \underset{(p-r-q)\times q}{T^{-3/2}\bar{\gamma}'S_{11}\bar{\tau}} \\ \underset{q\times(p-r-q)}{T^{-3/2}\bar{\tau}'S_{11}\bar{\gamma}} & \underset{q\times q}{T^{-2}\bar{\tau}'S_{11}\bar{\tau}} \end{bmatrix}$. We analyse separately each block.

For block (1,1) we have (see also (A.2))

$$\begin{aligned} T^{-1}\bar{\gamma}'S_{11}\bar{\gamma} &= T^{-1}\sum_{t=1}^{T} T^{-1/2}\bar{\gamma}'(X_{t-1}-\bar{X})(X_{t-1}-\bar{X})'\bar{\gamma}T^{-1/2} \\ &\xrightarrow{d} \int_0^1 \bar{\gamma}'C(W(u)-\bar{W})(W(u)-\bar{W})'C'\bar{\gamma}du \end{aligned}$$

by block (1,1) of (4.16) and the CMT since the mapping $\mathcal{I}_2 : x \longmapsto \int_0^1 x(u)x(u)'du$ is continuous, and $x(\cdot)$ is a continuous function on $[0,1]$. For block (1,2) we have

$$\begin{aligned} T^{-3/2}\bar{\gamma}'S_{11}\bar{\tau} &= T^{-1}\sum_{t=1}^{T} T^{-1/2}\bar{\gamma}'(X_{t-1}-\bar{X})(X_{t-1}-\bar{X})'\bar{\tau}T^{-1} \\ &\xrightarrow{d} \int_0^1 \bar{\gamma}'C(W(u)-\bar{W})\left[\ (Z(u)-\bar{Z})' \quad u-1/2\ \right]du \end{aligned}$$

by (4.16) and the CMT, since the mappings $\mathcal{I}_3 : (x,y) \longmapsto \int_0^1 x(u)y(u)'du$ and $\mathcal{I}_4 : x \longmapsto \int_0^1 x(u)udu$ are continuous for $x(\cdot)$ and $y(\cdot)$ continuous functions on $[0,1]$. For block (2,2) we have

$$\begin{aligned} T^{-2}\bar{\tau}'S_{11}\bar{\tau} &= T^{-1}\sum_{t=1}^{T} T^{-1}\bar{\tau}'(X_{t-1}-\bar{X})(X_{t-1}-\bar{X})'\bar{\tau}T^{-1} \\ &\xrightarrow{d} \begin{bmatrix} \int_0^1(Z(u)-\bar{Z})(Z(u)-\bar{Z})'du & \int_0^1(Z(u)-\bar{Z})(u-1/2)du \\ \int_0^1(u-1/2)(Z(u)-\bar{Z})'du & 1/12 \end{bmatrix} \end{aligned}$$

by blocks (2,1) and (3,1) of (4.16) and the CMT. Assembling the results for all the blocks we get

$T^{-1}B_T'S_{11}B_T \xrightarrow{d} \int_0^1 GG'du$, where $G = \begin{bmatrix} \bar{\gamma}'C(W(u)-\bar{W}) \\ Z(u)-\bar{Z} \\ u-1/2 \end{bmatrix}$, defined in (4.16).

Proof of (4.21).

We need to show that $T^{-1/2}B_T'S_{11}\beta \xrightarrow{d} VC_1^1(1)'\beta$.

$T^{-1/2}B_T'S_{11}\beta = \begin{bmatrix} T^{-1/2}\bar{\gamma}'S_{11}\beta \\ T^{-1}\bar{\tau}'S_{11}\beta \end{bmatrix}$. Block (1,1) can be written as

$$T^{-1/2}\bar{\gamma}'S_{11}\beta = T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'X_{t-1}X_{t-1}'\beta - T^{-1/2}\bar{\gamma}'\bar{X}\bar{X}'\beta.$$

Using the representation (2.22) the first term can be expressed as

$$T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'X_{t-1}X_{t-1}'\beta$$

$$= T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'(C\sum_{i=1}^{t-1}\varepsilon_i + C_1(L)\varepsilon_{t-1} + C_1(L)\Phi D_{t-1} + A)(C_1(L)\varepsilon_{t-1} + C_1(L)\Phi D_{t-1})'\beta$$

$$\begin{aligned} = \; & T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'(C\xi_{t-1}v_{t-1}' + C\xi_{t-1}D_{t-1}'\Phi'C_1(L)' + v_{t-1}v_{t-1}' \\ & + v_{t-1}D_{t-1}'\Phi'C_1(L)' + C_1(L)\Phi D_{t-1}v_{t-1}' \\ & + C_1(L)\Phi D_{t-1}D_{t-1}'\Phi'C_1(L)' + Av_{t-1}' + AD_{t-1}'\Phi'C_1(L)')\beta. \end{aligned}$$

The corresponding orders of magnitude for each of the terms in the expression above are as follows:

$T^{-3/2}(O_p(T)+O_p(T^{3/2})+O_p(T)+O_p(T^{1/2})+O_p(T^{1/2})+O_p(T)+O_p(T^{1/2})+O_p(T))$.

Except for the second term, all the others converge in probability to zero by (A.10) (which is the generalisation of (A.3) when the error process is autocorrelated), with $e(1) = 1$, $h(1) = C_1(1)$ and $\Gamma_s = 0$, $s = 1, 2, \ldots$ (first term); (A.9) (third term), E(ii) and (A.7) (fourth and fifth term); B(i) and A(v) (sixth term); (A.7) (seventh term) and A(v) (eighth term). The asymptotically non-degenerate term (second) has the following limit,

$$T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C\xi_{t-1}D_{t-1}'\Phi'C_1(L)'\beta = T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C(\xi_{t-1}z_{t-1}'C_1^1(L)'+\xi_{t-1}\varphi'C_1^2(1)')\beta$$

$\xrightarrow{d} \bar{\gamma}' C(\int_0^1 W(u)z(u)'duC_1^1(1)' + \int_0^1 W(u)du\varphi' C_1^2(1)')\beta$, by F and (A.1) and the CMT.

From block (1,1) of (4.16) and the CMT, and (B.2) the second term of $T^{-1/2}\bar{\gamma}' S_{11}\beta$ converges in distribution,

$$T^{-1/2}\bar{\gamma}'\bar{X}\bar{X}'\beta \xrightarrow{d} \bar{\gamma}' C \int_0^1 W(u)du\varphi' C_1^2(1)'\beta.$$

So for block (1,1) $T^{-1/2}\bar{\gamma}' S_{11}\beta \xrightarrow{d} \bar{\gamma}' C \int_0^1 W(u)z(u)'duC_1^1(1)'\beta$.

The element of block (1,2) of $T^{-1/2}B_T' S_{11}\beta$ is given by

$$T^{-1}\bar{\tau}' S_{11}\beta = T^{-2}\sum_{t=1}^{T}\bar{\tau}' X_{t-1}X_{t-1}'\beta - T^{-1}\bar{\tau}'\bar{X}\bar{X}'\beta.$$

Analysing the first term we get

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\bar{\tau}' X_{t-1}X_{t-1}'\beta &= T^{-2}\sum_{t=1}^{T}(\bar{\tau}' C\sum_{i=1}^{t-1}\varepsilon_i + \sum_{i=1}^{t-1} D_i + \bar{\tau}' C_1(L)\varepsilon_{t-1} \\ &\quad + \bar{\tau}' C_1(L)\Phi D_{t-1} + \bar{\tau}' A)(\varepsilon_{t-1} + \Phi D_{t-1})' C_1(L)'\beta \end{aligned}$$

$$\begin{aligned} = \; & T^{-2}\sum_{t=1}^{T}(\bar{\tau}' C\xi_{t-1}v_{t-1}' + \bar{\tau}' C\xi_{t-1}D_{t-1}'\Phi' C_1(L)' + \sum_{i=1}^{t-1} D_i v_{t-1}'\beta + \sum_{i=1}^{t-1} D_i D_{t-1}'\Phi' C_1(L)' \\ & + \bar{\tau}' v_{t-1}v_{t-1}' + \bar{\tau}' v_{t-1}D_{t-1}'\Phi' C_1(L)' + \bar{\tau}' C_1(L)\Phi D_{t-1}v_{t-1}' \\ & + \bar{\tau}' C_1(L)\Phi D_{t-1}D_{t-1}'\Phi' C_1(L)' + \bar{\tau}' A v_{t-1}' + \bar{\tau}' A D_{t-1}'\Phi' C_1(L)')\beta \end{aligned}$$

with the following orders of magnitude,

$T^{-2}(O_p(T) + O_p(T^{3/2}) + O_p(T^{3/2}) + O_p(T^2) + O_p(T) + O_p(T^{1/2}) + O_p(T^{1/2}) + O_p(T) + O_p(T^{1/2}) + O_p(T))$.

All the terms except for the fourth converge in probability to zero by (A.10) with $e(1) = 0$, $h(1) = C_1(1)$ and $\Gamma_s = 0$, $s = 1, 2, \ldots$ (first term); F, (A.1) and the CMT (second term); the proof of (10) in O'Brien (1999, p. 29) (third term); (A.9) (fifth term); E(ii) and (A.7) (sixth and seventh term); B(i) and A(v) (eighth term); (A.7) (ninth term) and A(v) (tenth term). Next we analyse the non-degenerate (fourth) term

$$\begin{aligned} & T^{-2}\sum_{t=1}^{T}(\sum_{i=1}^{t-1} D_i)D_{t-1}'\Phi' C_1(L)'\beta \\ = & T^{-2}\sum_{t=1}^{T}\begin{bmatrix} Z_{t-1}z_{t-1}'C_1^1(L)' + Z_{t-1}\varphi' C_1^2(1)' \\ (t-1)z_{t-1}'C_1^1(L)' + (t-1)\varphi' C_1^2(1)' \end{bmatrix}\beta \\ \xrightarrow{d} & \begin{bmatrix} \int_0^1 Z(u)z(u)duC_1^1(1)' + \int_0^1 Z(u)du\varphi' C_1^2(1)' \\ \int_0^1 uz(u)'duC_1^1(1)' + 1/2\varphi' C_1^2(1)' \end{bmatrix}\beta, \end{aligned}$$

by C(i) and equation (1) in O'Brien (1997, p. 23) for block (1,1) and C(ii) for the first term in block (1,2). Therefore, for the first term of $T^{-1}\bar{\tau}' S_{11}\beta$ we have

$$T^{-2}\sum_{t=1}^{T}\bar{\tau}' X_{t-1}X_{t-1}'\beta \xrightarrow{d} \begin{bmatrix} \int_0^1 Z(u)z(u)'duC_1^1(1)' + \int_0^1 Z(u)du\varphi' C_1^2(1)' \\ \int_0^1 uz(u)'duC_1^1(1)' + 1/2\varphi' C_1^2(1)' \end{bmatrix}\beta.$$

The limit of the product of the averages (second term of $T^{-1}\bar{\tau}' S_{11}\beta$) is

$$T^{-1}\bar{\tau}'\bar{X}\bar{X}'\beta \xrightarrow{d} \begin{bmatrix} \int_0^1 Z(u)du\varphi' C_1^2(1)'\beta \\ 1/2\varphi' C_1^2(1)'\beta \end{bmatrix},$$

by using blocks (2,1) and (3,1) of (4.16) and the CMT.

Thus, $T^{-1}\bar{\tau}' S_{11}\beta \xrightarrow{d} \begin{bmatrix} \int_0^1 Z(u)z(u)'duC_1^1(1)' \\ \int_0^1 uz(u)'duC_1^1(1)' \end{bmatrix}\beta.$

Combining the asymptotic results for $T^{-1/2}\bar{\gamma}' S_{11}\beta$ and $T^{-1}\bar{\tau}' S_{11}\beta$ we get (4.21) i.e.

$$T^{-1/2}B_T' S_{11}\beta \xrightarrow{d} VC_1^1(1)'\beta$$

where $V = \begin{bmatrix} \bar{\gamma}' C\int_0^1 W(u)z(u)'du \\ \int_0^1 Z(u)z(u)'du \\ \int_0^1 uz(u)'du \end{bmatrix} = \int_0^1 G_0(u)z(u)'du.$

Proof of (4.22).

We need to show $T^{-1/2}B_T' S_{10} \xrightarrow{d} \begin{bmatrix} V & 0 \end{bmatrix} C'$.

$$T^{-1/2}B_T'S_{10} = \begin{bmatrix} T^{-1/2}\bar{\gamma}'S_{10} \\ T^{-1}\bar{\tau}'S_{10} \end{bmatrix}$$. Block (1,1) has the form

$$T^{-1/2}\bar{\gamma}'S_{10} = T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'X_{t-1}\Delta X_t' - T^{-1/2}\bar{\gamma}'\bar{X}\bar{\Delta}X'.$$

The first term can be written as

$$\begin{aligned}
& T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'X_{t-1}\Delta X_t' \\
= \; & T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'(C\sum_{i=1}^{t-1}\varepsilon_i + C_1(L)\varepsilon_{t-1} + C_1(L)\Phi D_{t-1} + A)(\varepsilon_t + \Phi D_t)'C(L)' \\
= \; & T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'(C\xi_{t-1}\varepsilon_t'C(L)' + C\xi_{t-1}D_t'\Phi'C(L)' + v_{t-1}\varepsilon_t'C(L)' \\
& + v_{t-1}D_t'\Phi'C(L)' + C_1(L)\Phi D_{t-1}\varepsilon_t'C(L)' \\
& + C_1(L)\Phi D_{t-1}D_t'\Phi'C(L)' + A\varepsilon_t'C(L)' + AD_t'\Phi'C(L)')
\end{aligned}$$

and the terms above have the following orders of magnitude

$T^{-3/2}(O_p(T) + O_p(T^{3/2}) + O_p(T) + O_p(T^{1/2}) + O_p(T^{1/2}) + O_p(T) + O_p(T^{1/2}) + O_p(T))$.

So, only the second term does not vanish asymptotically, the remaining terms converge in probability to zero by (A.10) (first term); (A.9) (third term); (A.8), (A.7) and D(iv) (fourth term); E(iii), (A.1), D(v) and (A.6) (fifth term); B(ii), A(v), A(vi) and A(ii) (sixth term); (A.7) (seventh term); (4.5) and A(ii) (eighth term). For the non-degenerate second term we have

$$\begin{aligned}
T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C\xi_{t-1}D_t'\Phi'C(L)' &= T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C\xi_{t-1}D_t'\Phi'C' \\
& +T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C\xi_{t-1}D_t'\Phi'C_1(L)'(1-L)
\end{aligned}$$

$$
\begin{aligned}
&= T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C\left[\begin{array}{cc}\xi_{t-1}z_t' & \xi_{t-1}\varphi'\end{array}\right]C' + T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'C\xi_{t-1}(z_t - z_{t-1})'C_1^1(L)' \\
&\xrightarrow{d} \bar{\gamma}'C\left[\begin{array}{cc}\int_0^1 W(u)z(u)'du & \int_0^1 W(u)\varphi'du\end{array}\right]C'
\end{aligned}
$$

by (A.5) (see also O'Brien (1999, pp. 30-31)), (A.1) and the CMT and E(iv). For the first term of $T^{-1/2}\bar{\gamma}'S_{10}$ we have

$$
T^{-3/2}\sum_{t=1}^{T}\bar{\gamma}'X_{t-1}\Delta X_t' \xrightarrow{d} \bar{\gamma}'C\left[\begin{array}{cc}\int_0^1 W(u)z(u)'du & \int_0^1 W(u)\varphi'du\end{array}\right]C'.
$$

From block (1,1) of (4.16) and the CMT, and (B.1) it follows that the second term of $T^{-1/2}\bar{\gamma}'S_{10}$ has the following limit

$$
T^{-1/2}\bar{\gamma}'\bar{X}\bar{\Delta}X' \xrightarrow{d} \bar{\gamma}'C\left[\begin{array}{cc}0 & \int_0^1 W(u)\varphi'du\end{array}\right]C'.
$$

Thus, $T^{-1/2}\bar{\gamma}'S_{10} \xrightarrow{d} \bar{\gamma}'C\left[\begin{array}{cc}\int_0^1 W(u)z(u)'du & 0\end{array}\right]C'$.

Next we analyse block (1,2) of $T^{-1/2}B_TS_{10}$ which is given by

$$
T^{-1}\bar{\tau}'S_{10} = T^{-2}\sum_{t=1}^{T}\bar{\tau}'X_{t-1}\Delta X_t' - T^{-1}\bar{\tau}'\bar{X}\bar{\Delta}X'.
$$

We analyse the first term

$$
T^{-2}\sum_{t=1}^{T}\bar{\tau}'X_{t-1}\Delta X_t' = T^{-2}\sum_{t=1}^{T}(\bar{\tau}'C\sum_{i=1}^{t-1}\varepsilon_i + \sum_{i=1}^{t-1}D_i + \bar{\tau}'C_1(L)\varepsilon_{t-1} + \bar{\tau}'C_1(L)\Phi D_{t-1} +
$$

$$
\bar{\tau}'A)(\varepsilon_t + \Phi D_t)'C(L)'
$$

$$
\begin{aligned}
= \; & T^{-2}\sum_{t=1}^{T}(\bar{\tau}'C\xi_{t-1}\varepsilon_t'C(L)' + \bar{\tau}'C\xi_{t-1}D_t'\Phi'C(L)' + \sum_{i=1}^{t-1}D_i\varepsilon_t'C(L)' + \\
& \sum_{i=1}^{t-1}D_iD_t'\Phi'C(L)' + \bar{\tau}'v_{t-1}\varepsilon_t'C(L)' + \bar{\tau}'v_{t-1}D_t'\Phi'C(L)' \\
& + \bar{\tau}'C_1(L)\Phi D_{t-1}\varepsilon_t'C(L)' + \bar{\tau}'C_1(L)\Phi D_{t-1}D_t'\Phi'C(L)' \\
& + \bar{\tau}'A\varepsilon_t'C(L)' + \bar{\tau}'AD_t'\Phi'C(L)',
\end{aligned}
$$

which gives the following orders of magnitude

$T^{-2}(O_p(T)+O_p(T^{3/2})+O_p(T^{3/2})+O_p(T^2)+O_p(T)+O_p(T^{1/2})+O_p(T^{1/2})+O_p(T)+$
$O_p(T^{1/2})+O_p(T))$.

All terms but the fourth converge in probability to zero by (A.10) with $e(1) = 1$, $h(1) = C_1(1)$ and $\Gamma_s = 0$, $s = 1, 2, \ldots$ (first term); (A.5), (A.1) and the CMT and E(iv) (second term); (A.7) (third term); (A.9) (fifth term); (A.8), (A.7) and D(iv) (sixth term); E(iii), (A.1), D(v) and (A.6) (seventh term); B(ii), A(v), A(vi) and A(ii) (eighth term); (A.7) (ninth term); (4.5) and A(ii) (tenth term).

The limit of the fourth term is

$$T^{-2}\sum_{t=1}^{T}(\sum_{i=1}^{t-1} D_i)D_t'\Phi' C(L)' =$$
$$T^{-2}\sum_{t=1}^{T}\begin{bmatrix} Z_{t-1}z_t' & Z_{t-1}\varphi' \\ (t-1)z_t' & (t-1)\varphi' \end{bmatrix} C' + T^{-2}\sum_{t=1}^{T}\begin{bmatrix} Z_{t-1}z_t'(1-L)C_1^1(L)' \\ (t-1)z_t'(1-L)C_1^1(L)' \end{bmatrix}$$
$$\xrightarrow{d} \begin{bmatrix} \int_0^1 Z(u)z(u)'du & \int_0^1 Z(u)du\varphi' \\ \int_0^1 uz(u)'du & 1/2\varphi' \end{bmatrix} C'$$

by O'Brien (1997, pp. 23-24) for block (1,1) (see also the proof of C(i)), result (2) in O'Brien (1997, p. 23) for block (1,2), (4.6) and the CMT for block (2,1), for the first term. The second term converges to zero by B(iii) for block (1,1) and B(iv) for block (2,1).

Therefore the first term of $T^{-1}\bar{\tau}' S_{10}$ in the limit is

$$T^{-2}\sum_{t=1}^{T}\bar{\tau}' X_{t-1}\Delta X_t' \xrightarrow{d} \begin{bmatrix} \int_0^1 Z(u)z(u)'du & \int_0^1 Z(u)du\varphi' \\ \int_0^1 uz(u)'du & 1/2\varphi' \end{bmatrix} C'.$$

The limit of the second term of $T^{-1}\bar{\tau}' S_{10}$ is found from blocks (2,1) and (3,1) of (4.16) and the CMT, and (B.1) as

$$T^{-1}\bar{\tau}'\bar{X}\bar{\Delta}X' \xrightarrow{d} \begin{bmatrix} 0 & \int_0^1 Z(u)du\varphi' \\ 0 & 1/2\varphi' \end{bmatrix} C'.$$

Combining the last two results we have $T^{-1}\bar{\tau}' S_{10} \xrightarrow{d} \begin{bmatrix} \int_0^1 Z(u)z(u)'du & 0 \\ \int_0^1 uz(u)'du & 0 \end{bmatrix} C'$ and then assembling the results for $T^{-1/2}\bar{\gamma}' S_{10}$ and $T^{-1}\bar{\tau}' S_{10}$ we get (4.22),

$$T^{-1/2} B_T' S_{10} \xrightarrow{d} \begin{bmatrix} V & 0 \end{bmatrix} C'$$

where $\begin{bmatrix} V & 0 \end{bmatrix} = \begin{bmatrix} \bar{\gamma}' C \int_0^1 W(u)z(u)'du & 0 \\ \int_0^1 Z(u)z(u)'du & 0 \\ \int_0^1 uz(u)'du & 0 \end{bmatrix} = \begin{bmatrix} \int_0^1 G_0(u)z(u)'du & 0 \end{bmatrix}$. ■

Appendix C: DGPs

The DGPs used for the simulation experiments in Chapter 4 are of the form:

$$\Delta X_t = \alpha\beta' X_{t-1} + \Phi D_t + \varepsilon_t, \;\; t = 1, 2, \ldots, T$$

where all the components are defined as in Chapter 4.

The DGP used for Figures 4.1 and 4.2 is

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \\ \Delta X_{4t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -0.75 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \\ X_{4(t-1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_t \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}$$

where $\varepsilon_{jt} \sim i.i.d.N(0,1)$ for $j = 1, \ldots 4$ and $z_t = \begin{cases} -0.25, & 1 \le t \le [T/2] \\ 0.25, & [T/2]+1 \le t \le T \end{cases}$.

For Figures 4.3 and 4.4 the DGP is

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \\ \Delta X_{4t} \\ \Delta X_{5t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -0.75 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \\ X_{4(t-1)} \\ X_{5(t-1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \end{bmatrix}$$

where $\varepsilon_{jt} \sim i.i.d.N(0,1)$ for $j = 1, \ldots 5$, $z_{1t} = \begin{cases} -0.333, & 1 \le t \le [T/3] \\ 0.166, & [T/3]+1 \le t \le T \end{cases}$ and $z_{2t} = \begin{cases} -0.166, & 1 \le t \le [2T/3] \\ 0.333, & [2T/3]+1 \le t \le T \end{cases}$.

For Figure 4.5 the DGP takes the form

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \\ \Delta X_{4t} \end{bmatrix} = \begin{bmatrix} -0.75 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \\ X_{4(t-1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_t \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}$$

where $\varepsilon_{jt} \sim i.i.d.N(0,1)$ for $j = 1, \ldots 4$ and $z_t = \begin{cases} -0.25, & 1 \le t \le [T/2] \\ 0.25, & [T/2]+1 \le t \le T \end{cases}$.

For Figures 4.6 and 4.7 the DGP is

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \\ \Delta X_{4t} \\ \Delta X_{5t} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -0.75 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} X_{1(t-1)} \\ X_{2(t-1)} \\ X_{3(t-1)} \\ X_{4(t-1)} \\ X_{5(t-1)} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z_t \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \end{bmatrix}$$

where $\varepsilon_{jt} \sim i.i.d.N(0,1)$ for $j = 1, \ldots 5$ and $z_t = \begin{cases} -0.25, & 1 \le t \le [T/2] \\ 0.25, & [T/2]+1 \le t \le T \end{cases}$.

For all of the above DGPs the corresponding SM is estimated with unrestricted constant term and for $T = 100, 200, 300, 400, 500, 600, 700, 800, 900, 1{,}000$. The critical values used can be found in Osterwald-Lenum (1992, Table 1).

Appendix D: The power function for $(p-r)=4$

Table D.1. The power function for the trace statistic, case (i).

$\frac{g}{f}$	0	6	12	18	24	30	36	42	48	54	60
0	0.048	0.443	0.820	0.959	0.992	0.999	0.999	1	1	1	1
-3	0.035	0.162	0.579	0.872	0.971	0.996	0.999	1	1	1	1
-6	0.049	0.099	0.395	0.752	0.931	0.989	0.998	1	1	1	1
-9	0.066	0.107	0.295	0.649	0.885	0.977	0.996	0.999	1	1	1
-12	0.085	0.115	0.288	0.572	0.841	0.965	0.993	0.998	0.999	1	1
-15	0.115	0.145	0.279	0.547	0.802	0.937	0.988	0.998	1	1	1
-18	0.152	0.188	0.312	0.527	0.783	0.929	0.985	0.998	1	1	1
-21	0.192	0.226	0.347	0.547	0.775	0.925	0.979	0.997	0.999	1	1
-24	0.247	0.285	0.383	0.576	0.774	0.913	0.974	0.995	0.999	1	1
-27	0.318	0.356	0.437	0.615	0.786	0.912	0.977	0.994	0.998	1	1
-30	0.395	0.429	0.514	0.658	0.803	0.921	0.981	0.996	0.999	1	1
-36	0.554	0.581	0.641	0.752	0.862	0.936	0.981	0.995	0.999	1	1
-42	0.705	0.716	0.774	0.842	0.912	0.957	0.984	0.998	0.999	1	1
-48	0.815	0.834	0.864	0.907	0.949	0.975	0.990	0.997	0.999	1	1
-54	0.914	0.921	0.938	0.958	0.974	0.988	0.995	0.999	0.999	1	1
-60	0.964	0.963	0.968	0.981	0.989	0.995	0.999	0.999	1	1	1

Table D.2. The power function for the maximal eigenvalue statistic, case (i).

$\frac{g}{f}$	0	6	12	18	24	30	36	42	48	54	60
0	0.054	0.415	0.819	0.963	0.994	0.999	1	1	1	1	1
-3	0.044	0.135	0.579	0.885	0.977	0.998	1	1	1	1	1
-6	0.053	0.096	0.387	0.782	0.952	0.995	0.999	1	1	1	1
-9	0.069	0.101	0.304	0.685	0.915	0.988	0.999	1	1	1	1
-12	0.085	0.109	0.287	0.608	0.893	0.983	0.998	0.999	1	1	1
-15	0.118	0.151	0.291	0.607	0.869	0.969	0.996	0.999	1	1	1
-18	0.158	0.197	0.338	0.598	0.850	0.968	0.996	1	1	1	1
-21	0.209	0.256	0.389	0.627	0.854	0.964	0.994	0.999	1	1	1
-24	0.281	0.326	0.441	0.660	0.860	0.964	0.994	0.999	1	1	1
-27	0.369	0.405	0.529	0.709	0.877	0.962	0.993	0.998	1	1	1
-30	0.460	0.516	0.605	0.765	0.899	0.969	0.995	1	1	1	1
-36	0.666	0.686	0.771	0.862	0.943	0.981	0.995	1	1	1	1
-42	0.827	0.841	0.887	0.937	0.970	0.989	0.996	1	1	1	1
-48	0.917	0.933	0.952	0.971	0.986	0.996	1	1	1	1	1
-54	0.978	0.979	0.987	0.993	0.996	0.999	1	1	1	1	1
-60	0.994	0.993	0.998	0.998	0.999	1	1	1	1	1	1

Table D.3. The power function for the trace statistic, case (ii).

$\frac{g}{f}$	0	6	12	18	24	30	36	42	48	54	60
0	0.049	0.411	0.789	0.945	0.985	0.997	0.999	1	1	1	1
-3	0.044	0.143	0.525	0.825	0.955	0.988	0.998	1	1	1	1
-6	0.050	0.091	0.327	0.675	0.894	0.973	0.996	0.999	1	1	1
-9	0.056	0.087	0.241	0.553	0.826	0.956	0.991	0.999	1	1	1
-12	0.070	0.096	0.232	0.464	0.764	0.926	0.981	0.997	1	1	1
-15	0.088	0.113	0.204	0.445	0.695	0.886	0.970	0.994	1	1	1
-18	0.114	0.132	0.235	0.410	0.673	0.865	0.963	0.992	0.999	1	1
-21	0.147	0.178	0.252	0.431	0.663	0.842	0.951	0.989	0.997	1	1
-24	0.182	0.211	0.295	0.442	0.654	0.826	0.941	0.983	0.997	1	1
-27	0.227	0.268	0.322	0.483	0.656	0.820	0.936	0.981	0.995	0.999	1
-30	0.289	0.322	0.392	0.519	0.678	0.839	0.934	0.983	0.996	0.999	1
-36	0.411	0.433	0.493	0.605	0.730	0.859	0.937	0.977	0.996	0.999	1
-42	0.547	0.565	0.631	0.717	0.809	0.885	0.951	0.986	0.999	0.999	1
-48	0.678	0.696	0.735	0.797	0.861	0.926	0.965	0.988	0.997	0.999	1
-54	0.790	0.813	0.847	0.879	0.912	0.959	0.981	0.994	0.997	0.999	1
-60	0.891	0.889	0.907	0.933	0.957	0.975	0.992	0.993	0.998	0.999	1

Table D.4. The power function for the maximal eigenvalue statistic, case (ii).

$\frac{g}{f}$	0	6	12	18	24	30	36	42	48	54	60
0	0.053	0.370	0.778	0.347	0.989	0.999	1	1	1	1	1
-3	0.052	0.110	0.495	0.841	0.961	0.994	0.999	1	1	1	1
-6	0.051	0.076	0.303	0.702	0.920	0.988	0.998	1	1	1	1
-9	0.059	0.076	0.219	0.581	0.861	0.975	0.997	1	1	1	1
-12	0.062	0.081	0.205	0.482	0.815	0.956	0.994	0.999	1	1	1
-15	0.089	0.103	0.202	0.470	0.770	0.935	0.989	0.998	1	1	1
-18	0.107	0.128	0.237	0.450	0.743	0.923	0.988	0.997	1	1	1
-21	0.144	0.175	0.267	0.474	0.743	0.914	0.979	0.997	1	1	1
-24	0.185	0.214	0.312	0.500	0.745	0.907	0.980	0.996	1	1	1
-27	0.253	0.279	0.374	0.563	0.763	0.906	0.977	0.997	0.999	1	1
-30	0.310	0.362	0.456	0.612	0.788	0.916	0.980	0.998	0.999	1	1
-36	0.483	0.525	0.612	0.731	0.856	0.937	0.984	0.997	1	1	1
-42	0.676	0.688	0.762	0.838	0.917	0.965	0.988	0.998	1	1	1
-48	0.815	0.825	0.870	0.914	0.957	0.980	0.993	0.997	1	1	1
-54	0.914	0.925	0.943	0.963	0.980	0.992	0.998	1	1	1	1
-60	0.970	0.972	0.981	0.987	0.993	0.995	0.999	1	1	1	1

Table D.5. The power function for the trace statistic, case (iii).

$\frac{g}{f}$	0	6	12	18	24	30	36	42	48	54	60
0	0.055	0.176	0.573	0.858	0.960	0.991	0.998	1	1	1	1
-3	0.057	0.113	0.366	0.701	0.901	0.977	0.996	0.999	1	1	1
-6	0.069	0.095	0.281	0.594	0.838	0.953	0.993	0.999	1	1	1
-9	0.077	0.104	0.239	0.510	0.787	0.932	0.987	0.998	0.999	1	1
-12	0.098	0.118	0.247	0.465	0.745	0.912	0.975	0.997	1	1	1
-15	0.112	0.141	0.238	0.459	0.692	0.879	0.969	0.992	0.999	1	1
-18	0.148	0.171	0.264	0.446	0.685	0.868	0.962	0.991	0.998	1	1
-21	0.189	0.215	0.304	0.470	0.693	0.858	0.951	0.988	0.997	1	1
-24	0.227	0.268	0.343	0.496	0.681	0.846	0.947	0.987	0.996	1	1
-27	0.286	0.326	0.384	0.537	0.697	0.850	0.945	0.983	0.997	0.999	1
-30	0.356	0.389	0.452	0.569	0.741	0.860	0.949	0.984	0.996	0.999	1
-36	0.485	0.508	0.568	0.676	0.786	0.885	0.952	0.984	0.997	0.999	1
-42	0.627	0.641	0.700	0.776	0.858	0.923	0.965	0.992	0.999	1	1
-48	0.752	0.766	0.795	0.850	0.903	0.950	0.978	0.992	0.998	0.999	1
-54	0.851	0.860	0.894	0.917	0.942	0.973	0.988	0.996	0.998	1	1
-60	0.928	0.929	0.937	0.957	0.972	0.985	0.995	0.997	0.999	1	1

Table D.6. The power function for the maximal eigenvalue statistic, case (iii).

$\frac{g}{f}$	0	6	12	18	24	30	36	42	48	54	60
0	0.051	0.148	0.556	0.859	0.970	0.995	0.999	1	1	1	1
-3	0.056	0.081	0.305	0.701	0.918	0.985	0.999	1	1	1	1
-6	0.054	0.069	0.218	0.593	0.862	0.971	0.995	0.999	1	1	1
-9	0.062	0.078	0.190	0.497	0.809	0.957	0.997	0.999	1	1	1
-12	0.068	0.087	0.192	0.440	0.769	0.938	0.991	0.998	1	1	1
-15	0.096	0.112	0.204	0.446	0.739	0.917	0.986	0.998	1	1	1
-18	0.118	0.141	0.245	0.444	0.731	0.913	0.983	0.996	1	1	1
-21	0.163	0.187	0.281	0.475	0.737	0.909	0.974	0.996	1	1	1
-24	0.207	0.238	0.328	0.511	0.747	0.902	0.976	0.996	1	1	1
-27	0.277	0.304	0.395	0.558	0.772	0.910	0.975	0.996	1	1	1
-30	0.343	0.391	0.476	0.636	0.798	0.919	0.982	0.998	0.999	1	1
-36	0.517	0.561	0.643	0.754	0.872	0.945	0.987	0.997	1	1	1
-42	0.714	0.722	0.788	0.860	0.930	0.970	0.990	0.999	1	1	1
-48	0.839	0.856	0.888	0.928	0.963	0.982	0.996	0.998	1	1	1
-54	0.933	0.941	0.953	0.972	0.985	0.993	0.998	1	1	1	1
-60	0.979	0.980	0.986	0.990	0.995	0.996	1	1	1	1	1

Appendix E: A graphical representation of the results of the experiments in 5.2

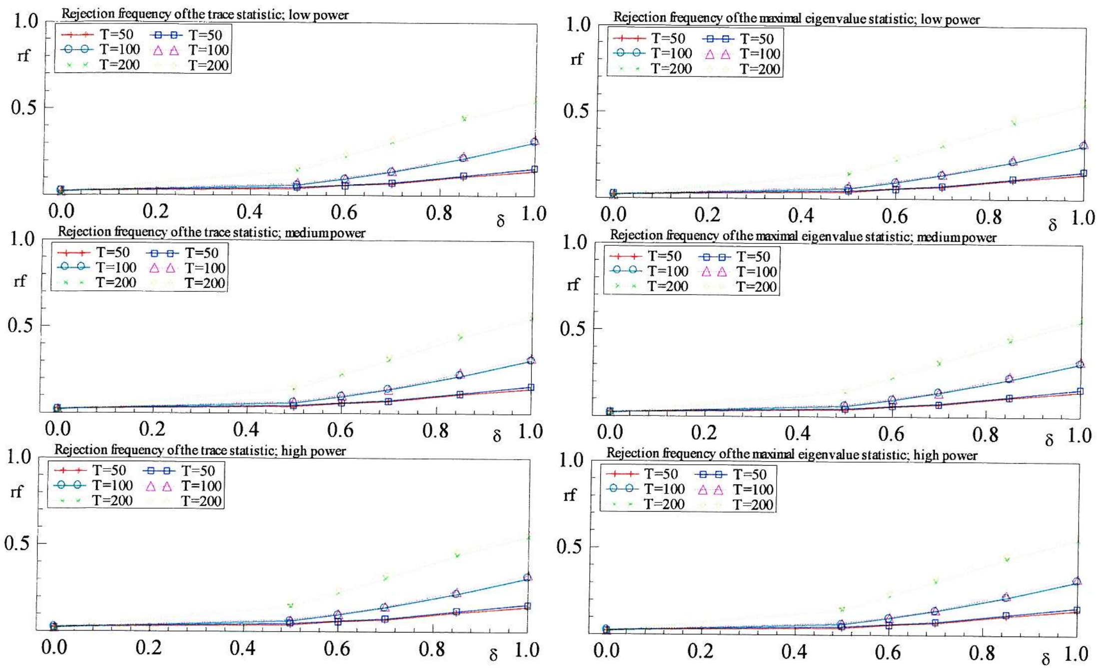

Figure E.1. Frequency of rejecting the true null hypothesis $r \leq 1$ for D1(i) (two different shifts).

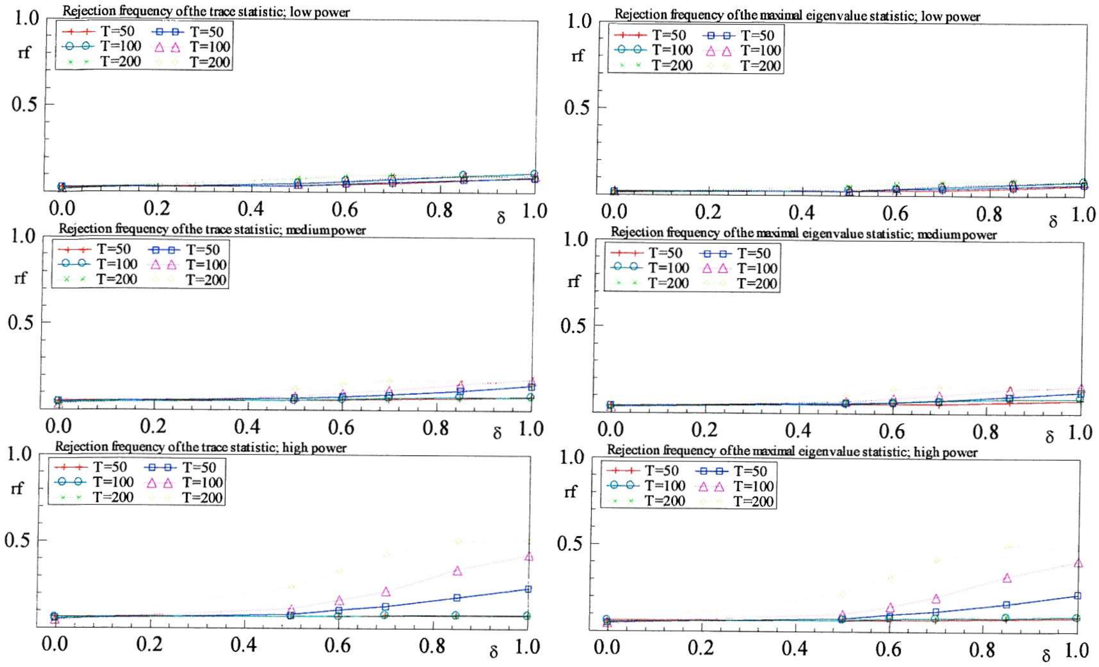

Figure E.2. Frequency of rejecting the true null hypothesis $r \leq 1$ for D2(i) (two different shifts).

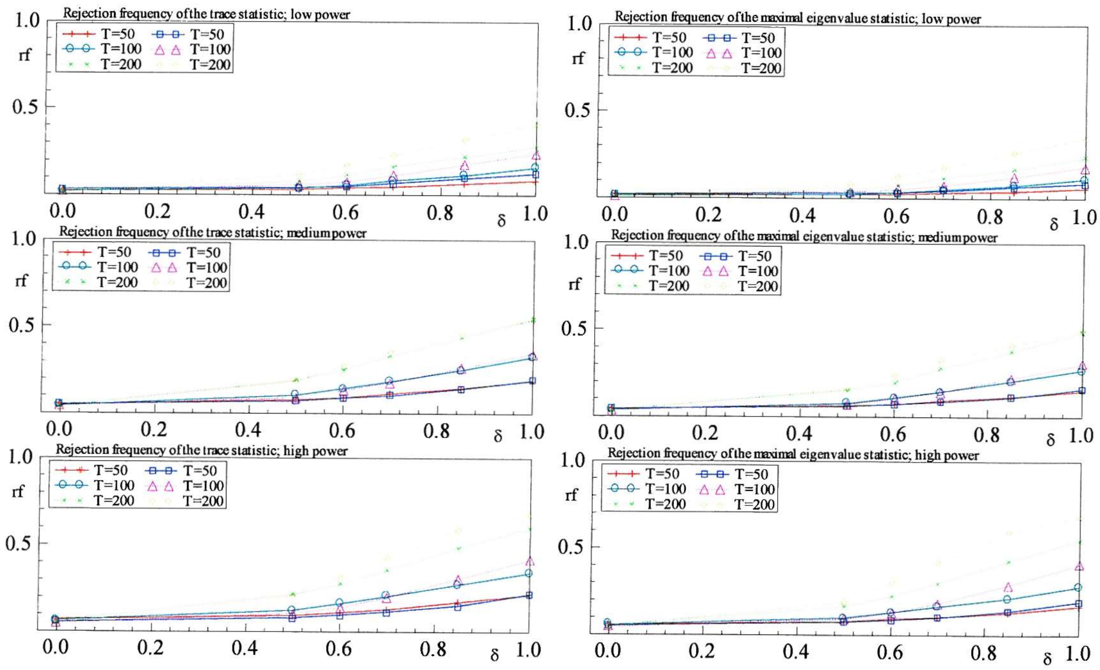

Figure E.3. Frequency of rejecting the true null hypothesis $r \leq 1$ for D3(i) (two different shifts).

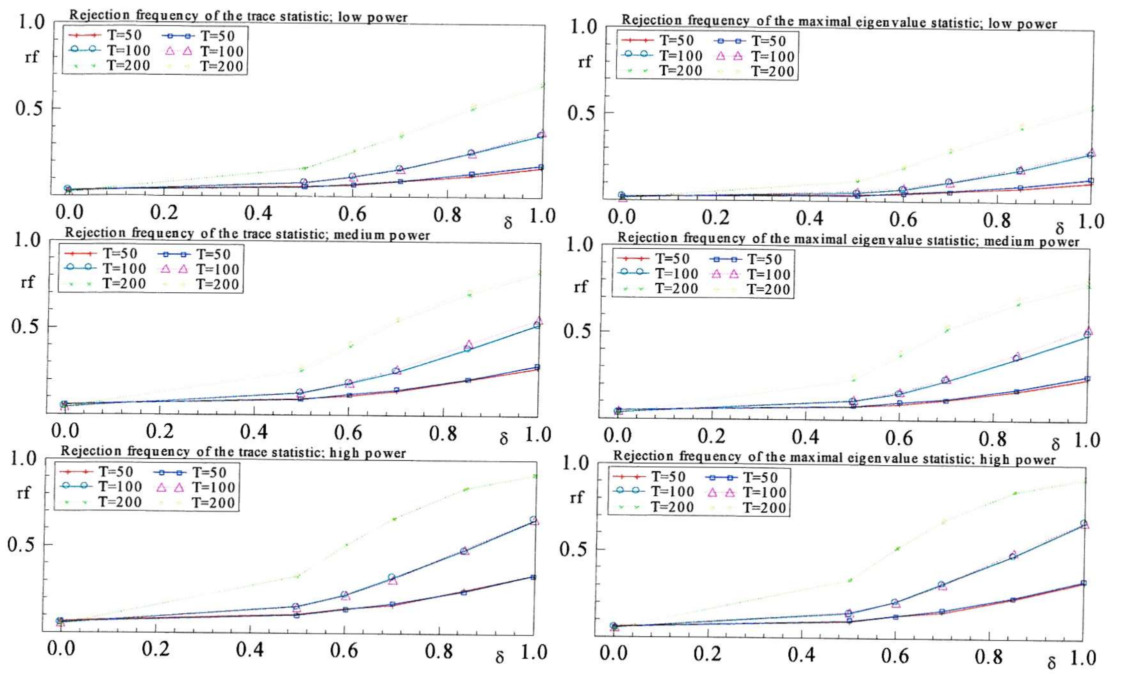

Figure E.4. Frequency of rejecting the true null hypothesis $r \leq 1$ for D1(ii) (two different shifts).

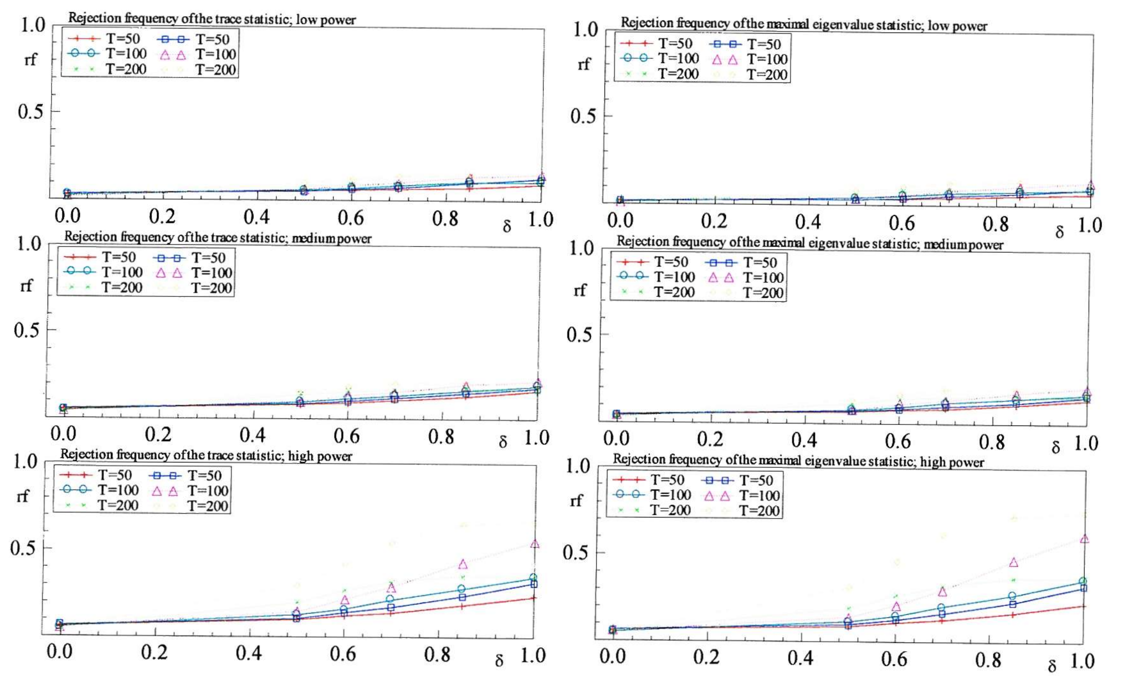

Figure E.5. Frequency of rejecting the true null hypothesis $r \leq 1$ for D2(ii) (two different shifts).

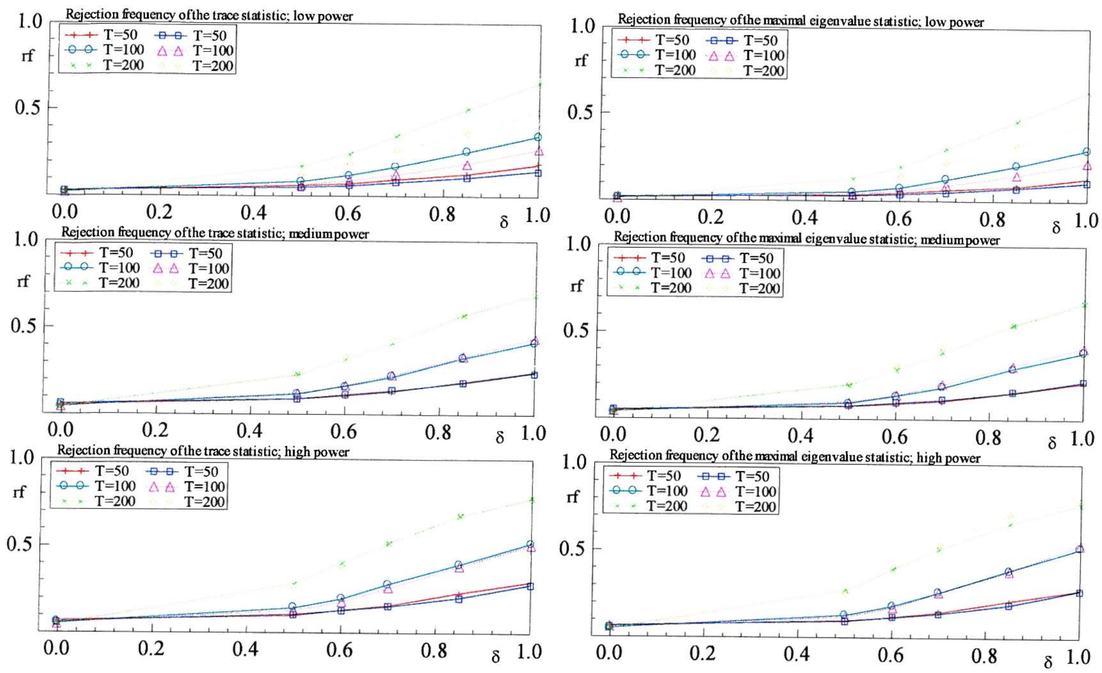

Figure E.6. Frequency of rejecting the true null hypothesis $r \leq 1$ for D3(ii) (two different shifts).

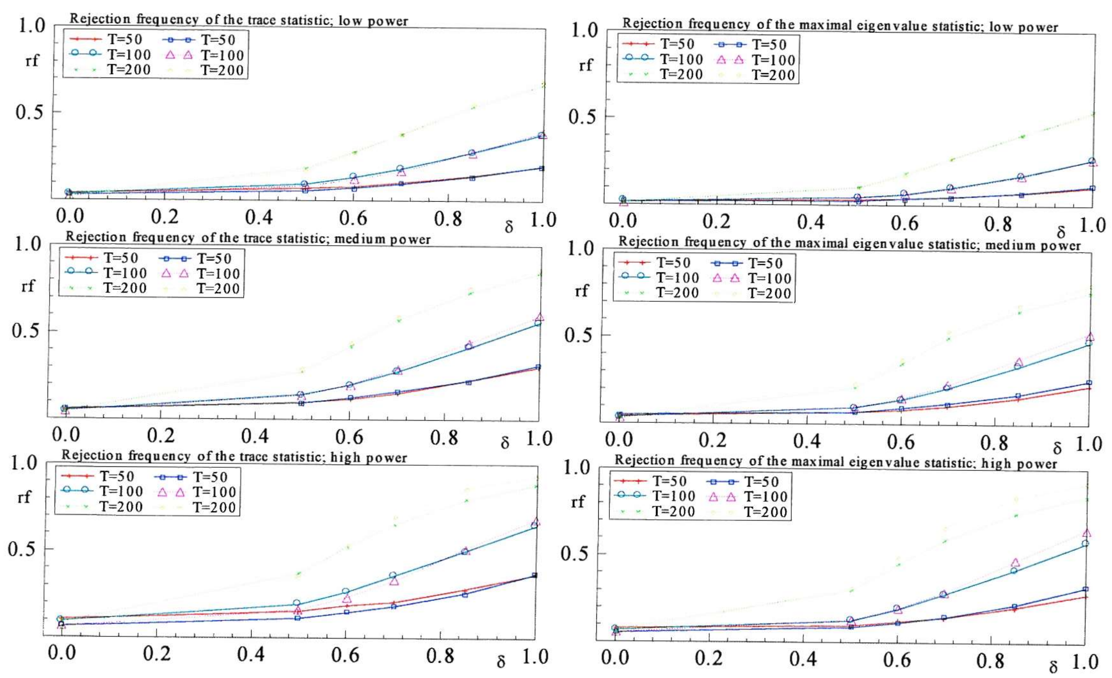

Figure E.7. Frequency of rejecting the true null hypothesis $r \leq 1$ for D1(iii) (two different shifts).

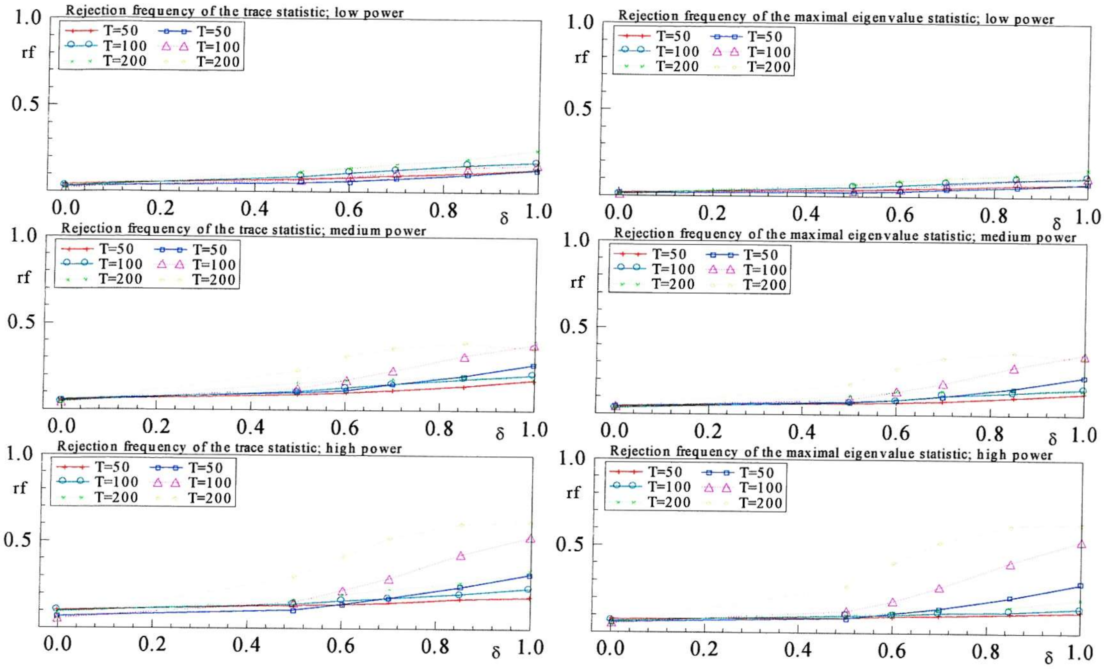

Figure E.8. Frequency of rejecting the true null hypothesis $r \leq 1$ for D2(iii) (two different shifts).

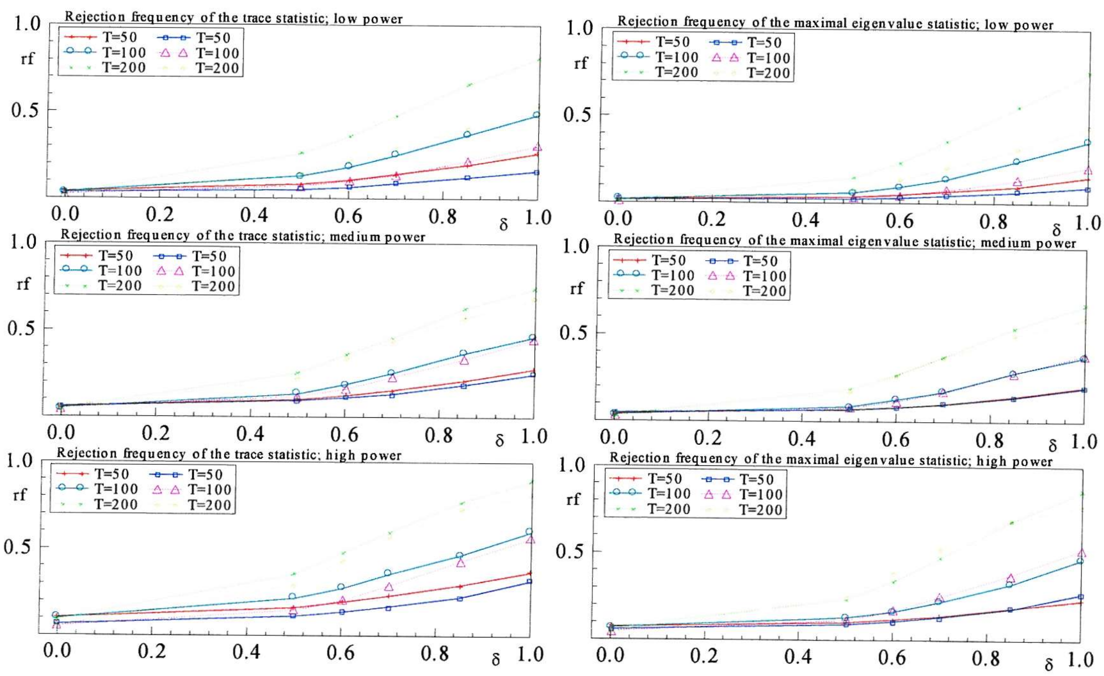

Figure E.9. Frequency of rejecting the true null hypothesis $r \leq 1$ for D3(iii) (two different shifts).

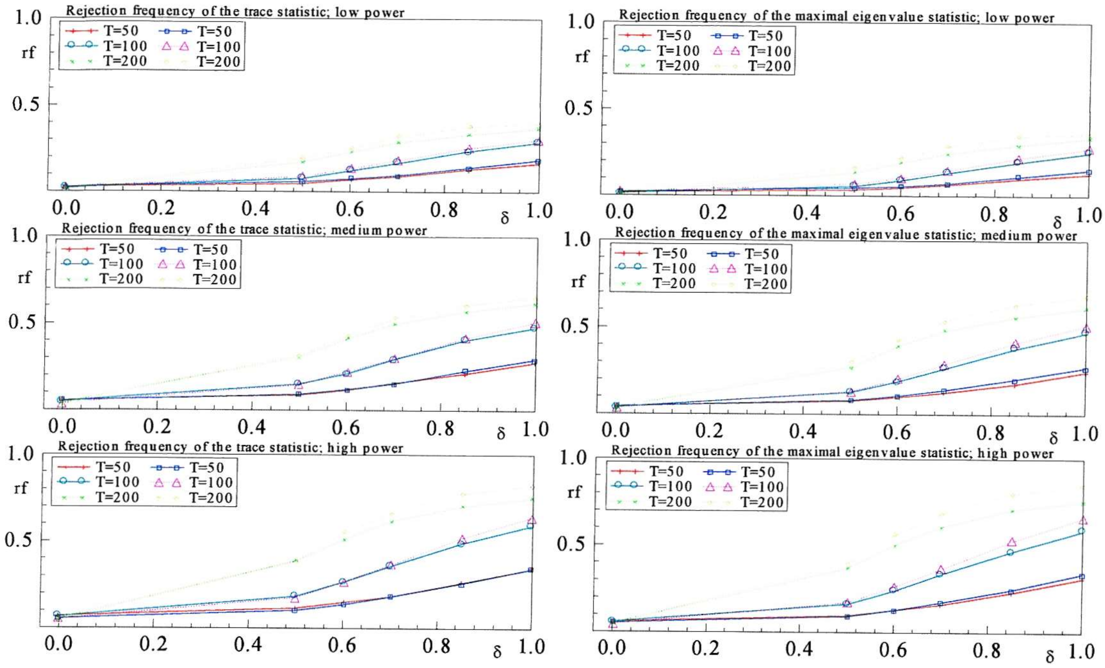

Figure E.10. Frequency of rejecting the true null hypothesis $r \leq 1$ for D1(i) (a common shift).

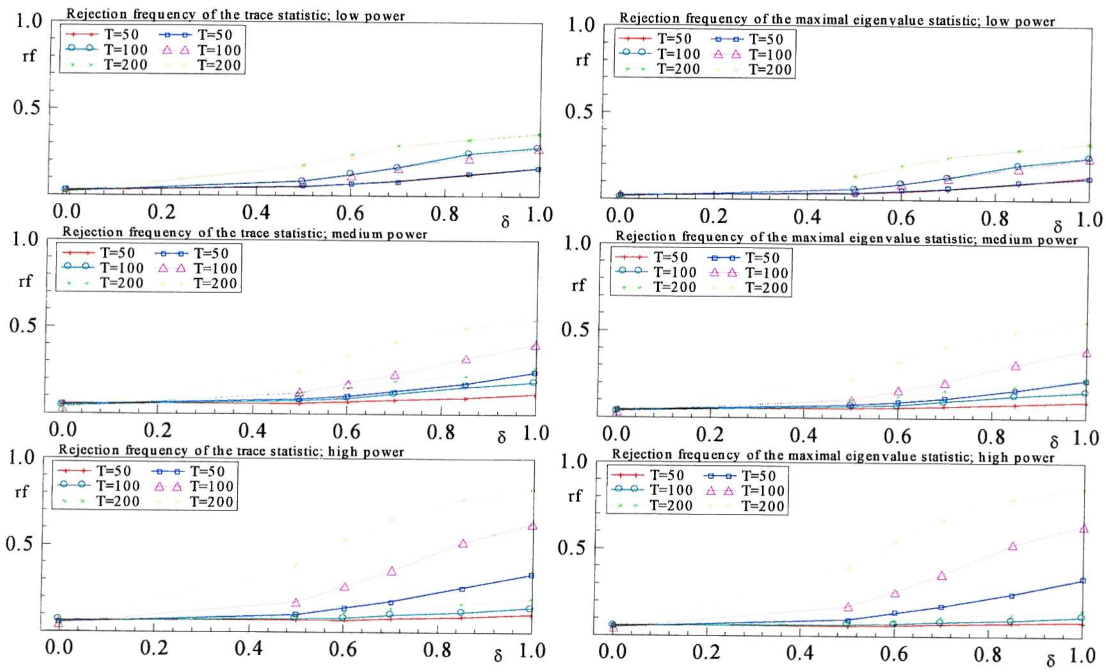

Figure E.11. Frequency of rejecting the true null hypothesis $r \leq 1$ for D2(i) (a common shift).

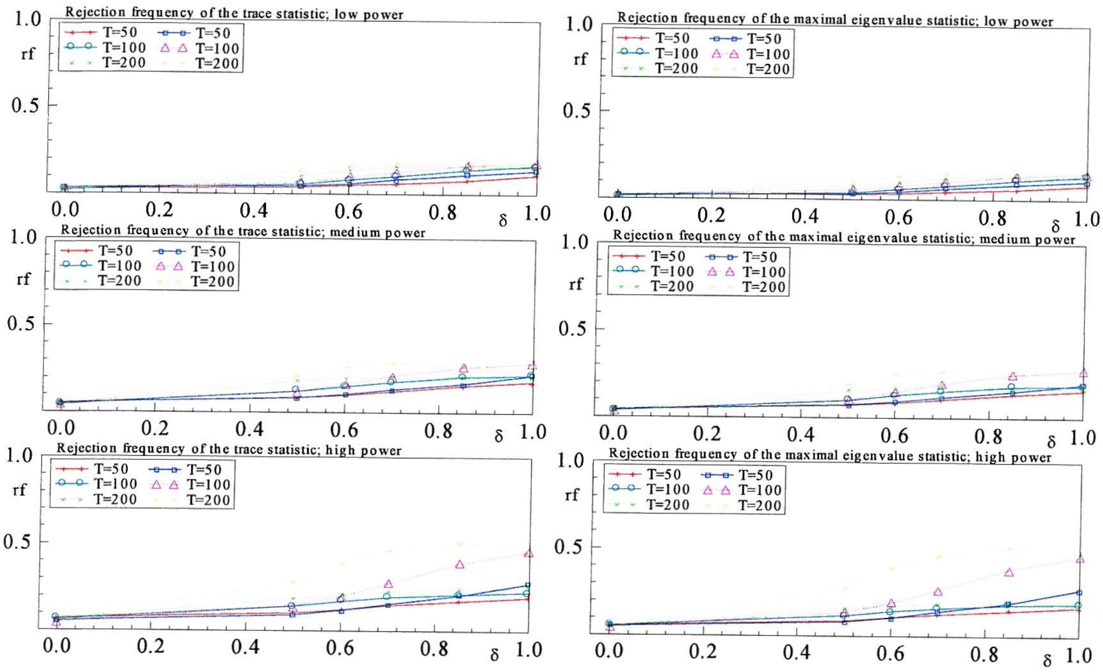

Figure E.12. Frequency of rejecting the true null hypothesis $r \leq 1$ for D3(i) (a common shift).

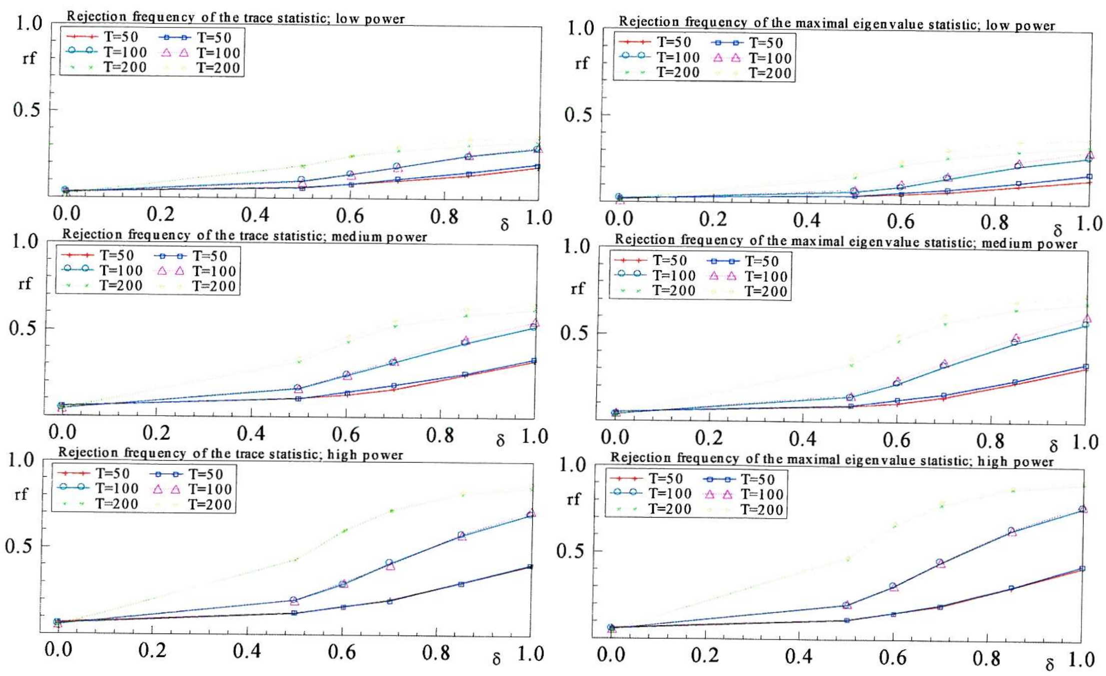

Figure E.13. Frequency of rejecting the true null hypothesis $r \leq 1$ for D1(ii) (a common shift).

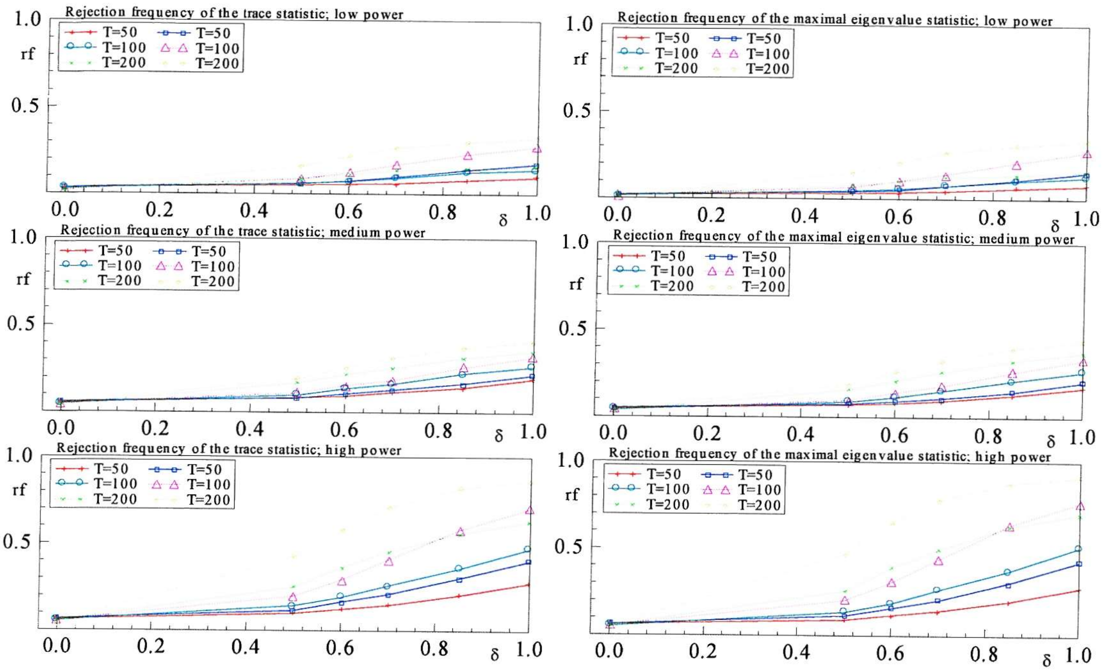

Figure E.14. Frequency of rejecting the true null hypothesis $r \leq 1$ for D2(ii) (a common shift).

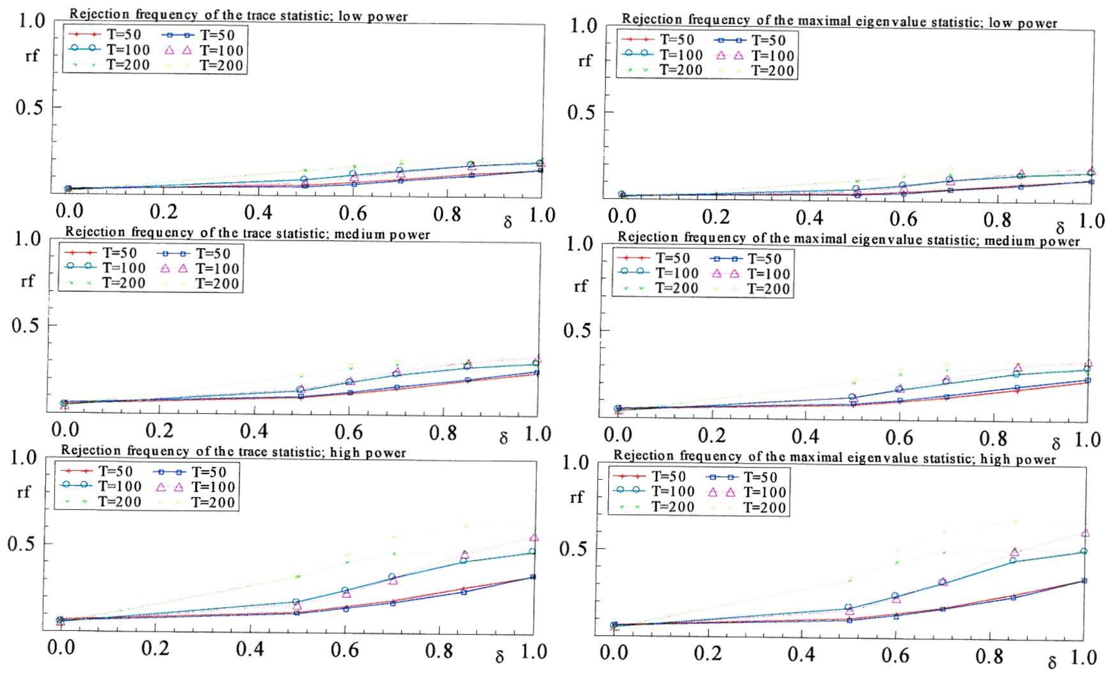

Figure E.15. Frequency of rejecting the true null hypothesis $r \leq 1$ for D3(ii) (a common shift).

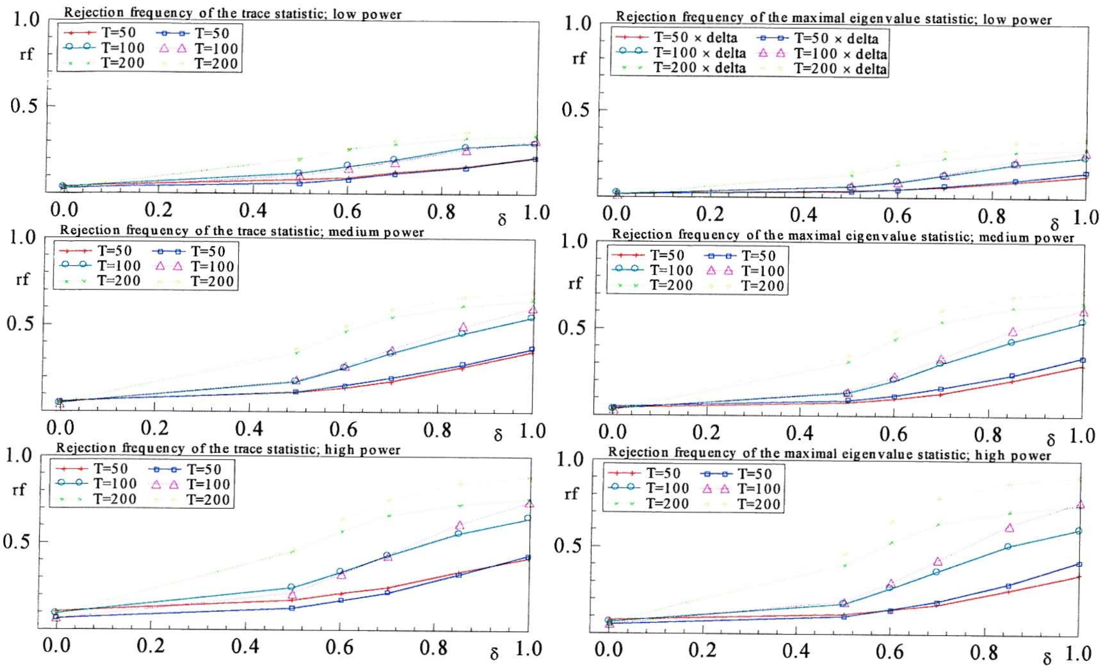

Figure E.16. Frequency of rejecting the true null hypothesis $r \leq 1$ for D1(iii) (a common shift).

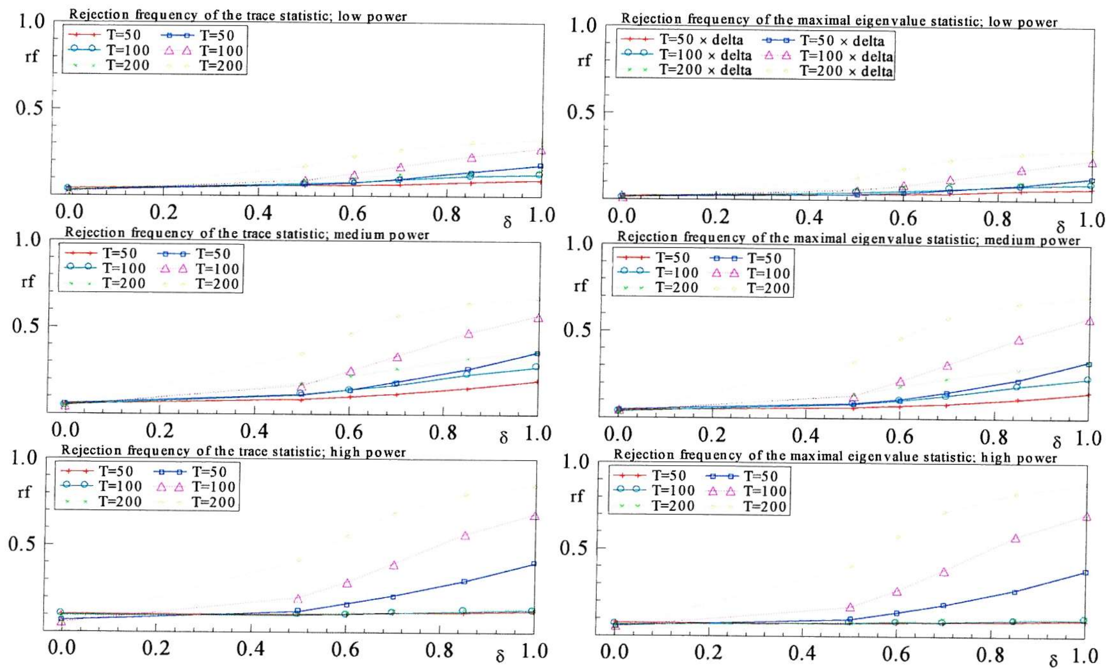

Figure E.17. Frequency of rejecting the true null hypothesis $r \leq 1$ for D2(iii) (a common shift).

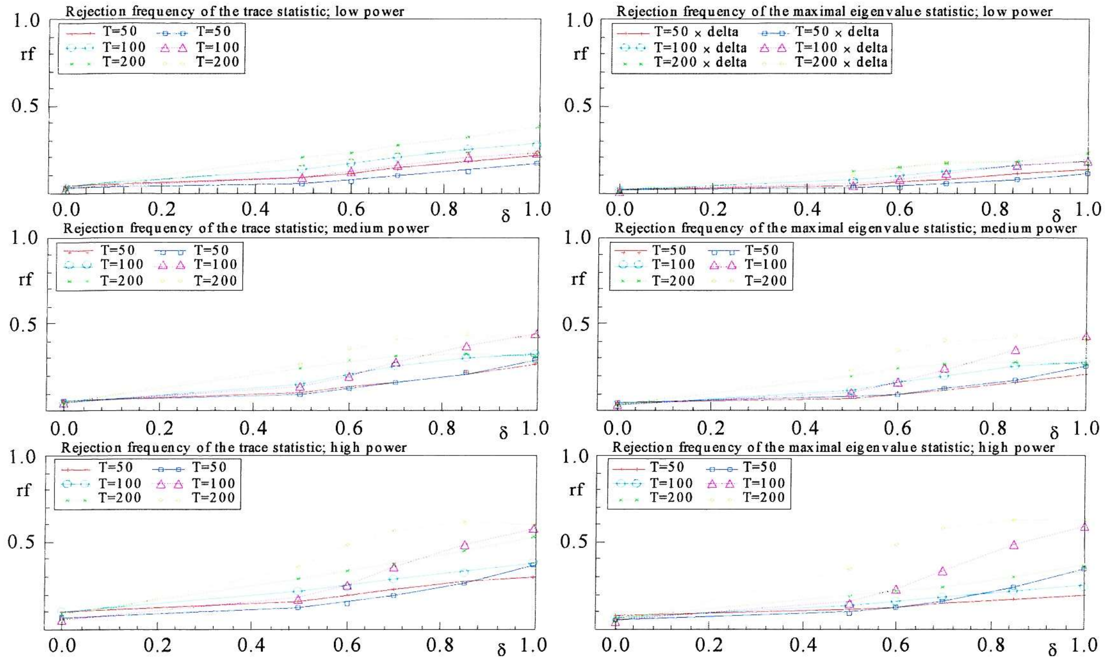

Figure E.18. Frequency of rejecting the true null hypothesis $r \leq 1$ for D3(iii) (a common shift).

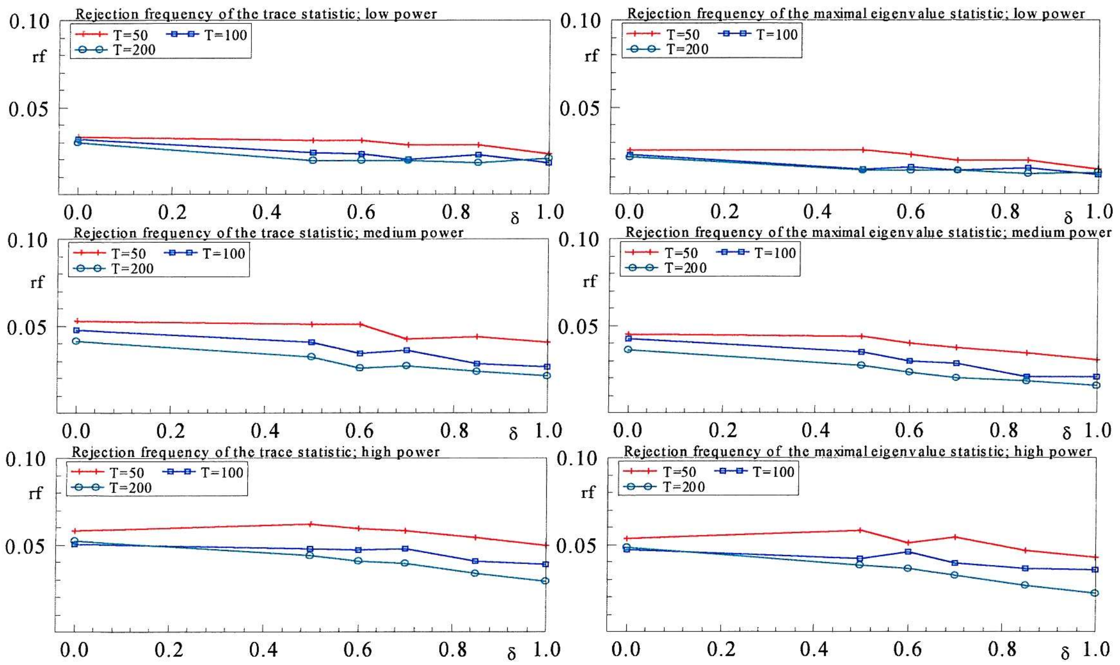

Figure E.19. Frequency of rejecting the true null hypothesis $r \leq 1$ with $\alpha_{\perp}^{(1)} = 0$; case (i) (a single shift).

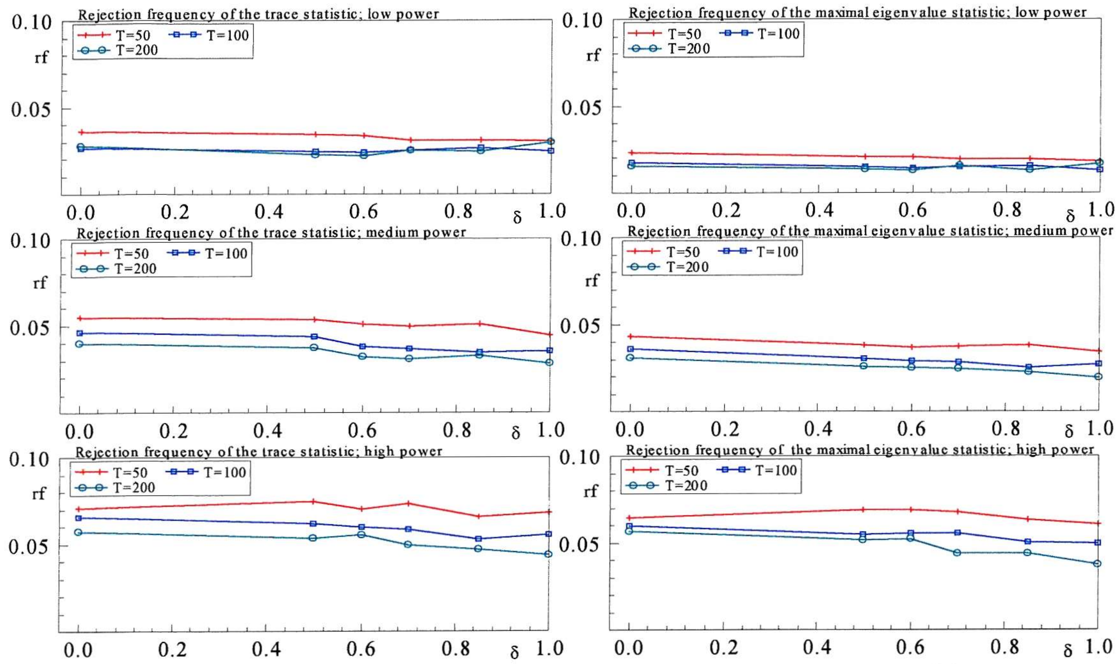

Figure E.20. Frequency of rejecting the true null hypothesis $r \leq 1$ with $\alpha_{\perp}^{(1)} = 0$; case (ii) (a single shift).

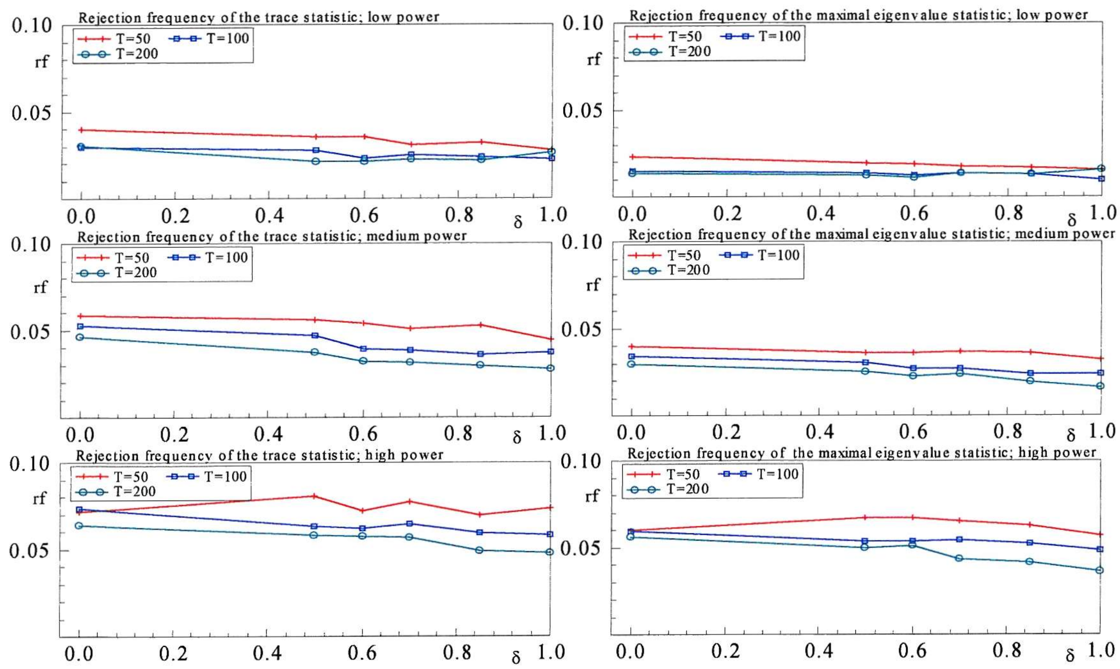

Figure E.21. Frequency of rejecting the true null hypothesis $r \leq 1$ with $\alpha_{\perp}^{(1)} = 0$; case (iii) (a single shift).

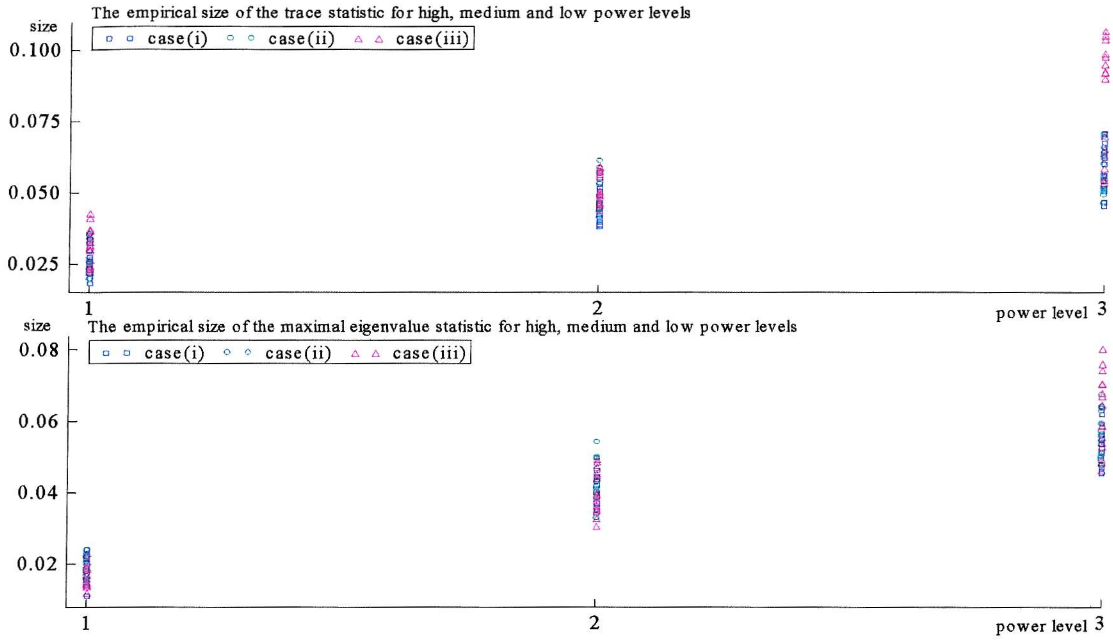

Figure E.22. The empirical size for low, medium and high power levels.

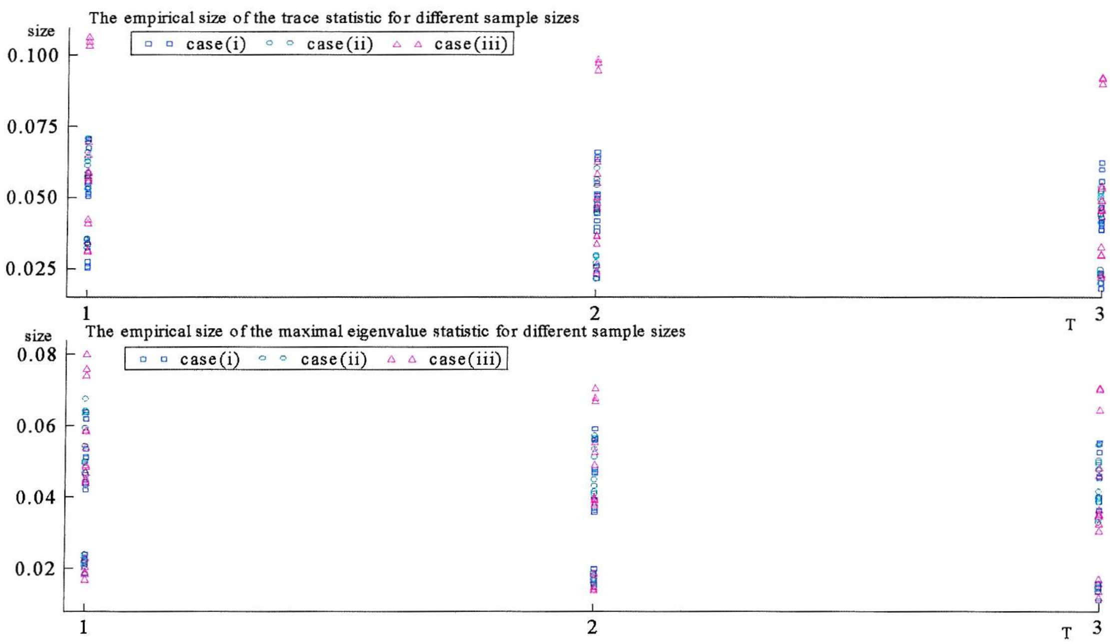

Figure E.23. The empirical size for different sample sizes.

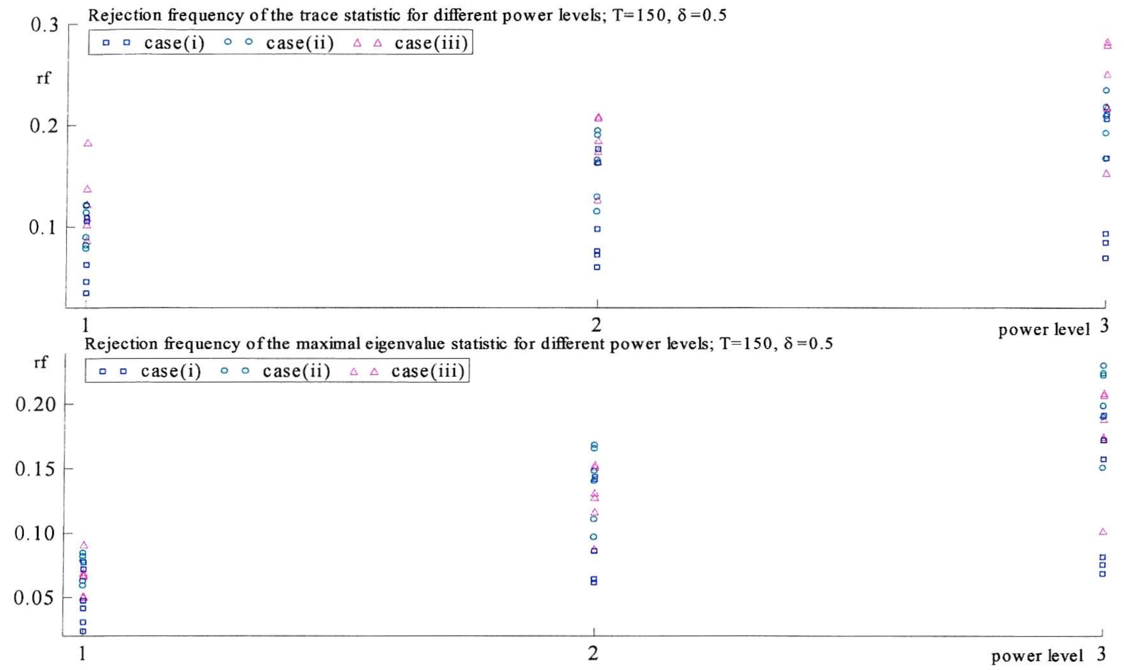

Figure E.24. Rejection frequency for different power levels (two different shifts; $T = 150$, $\delta = 0.5$).

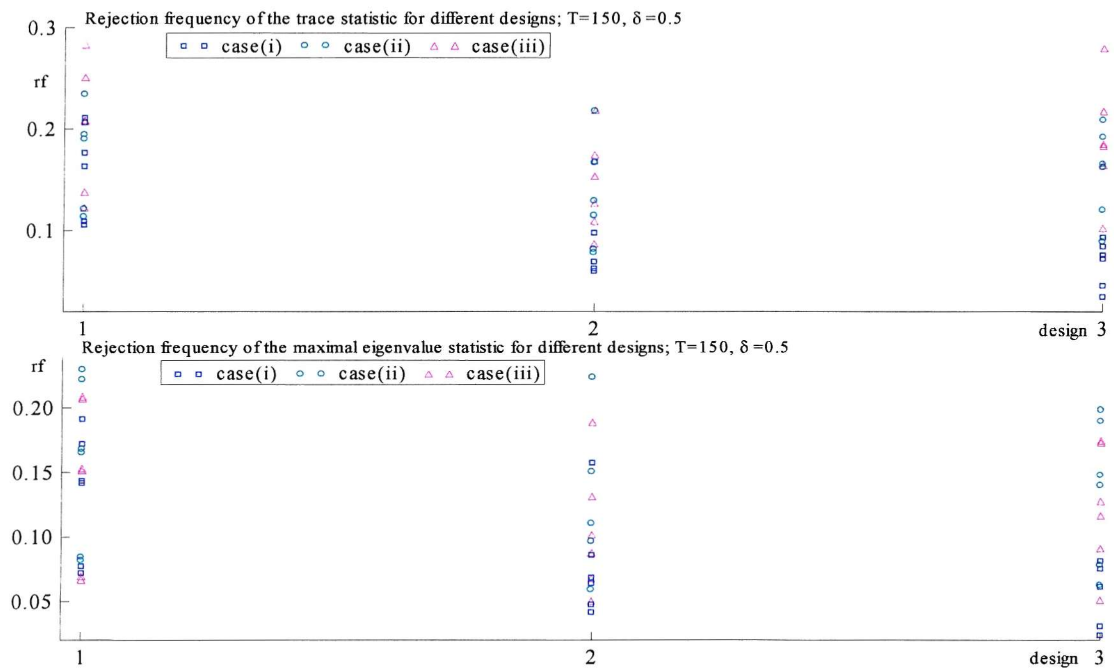

Figure E.25. Rejection frequency for different designs (two different shifts; $T = 150$, $\delta = 0.5$).

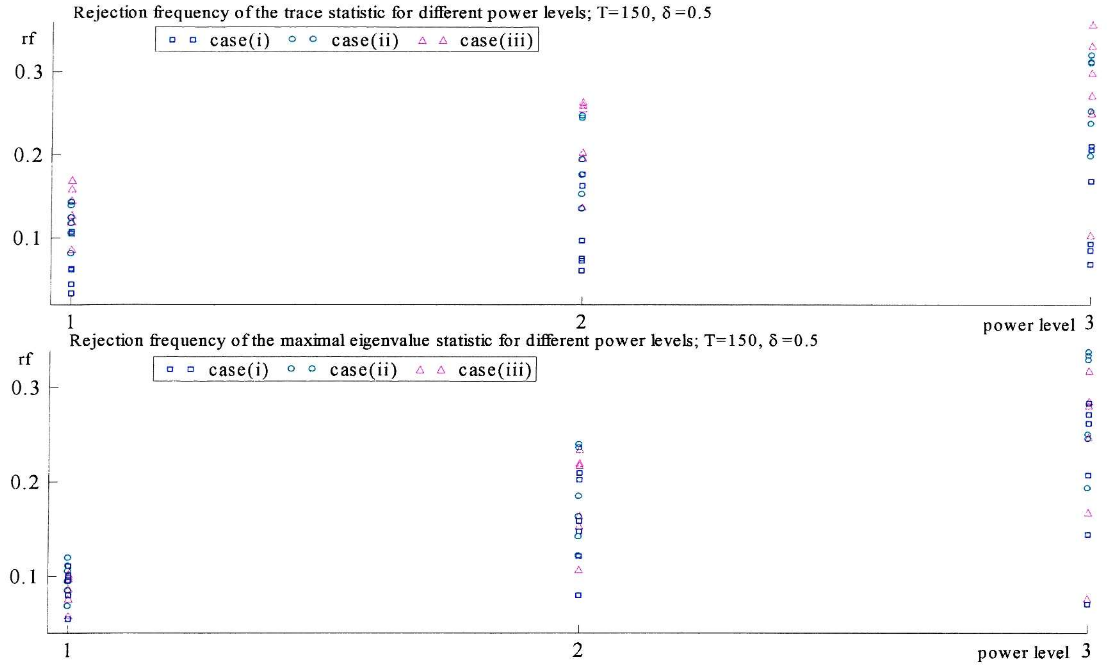

Figure E.26. Rejection frequency for different power levels (a common shift; $T = 150$, $\delta = 0.5$).

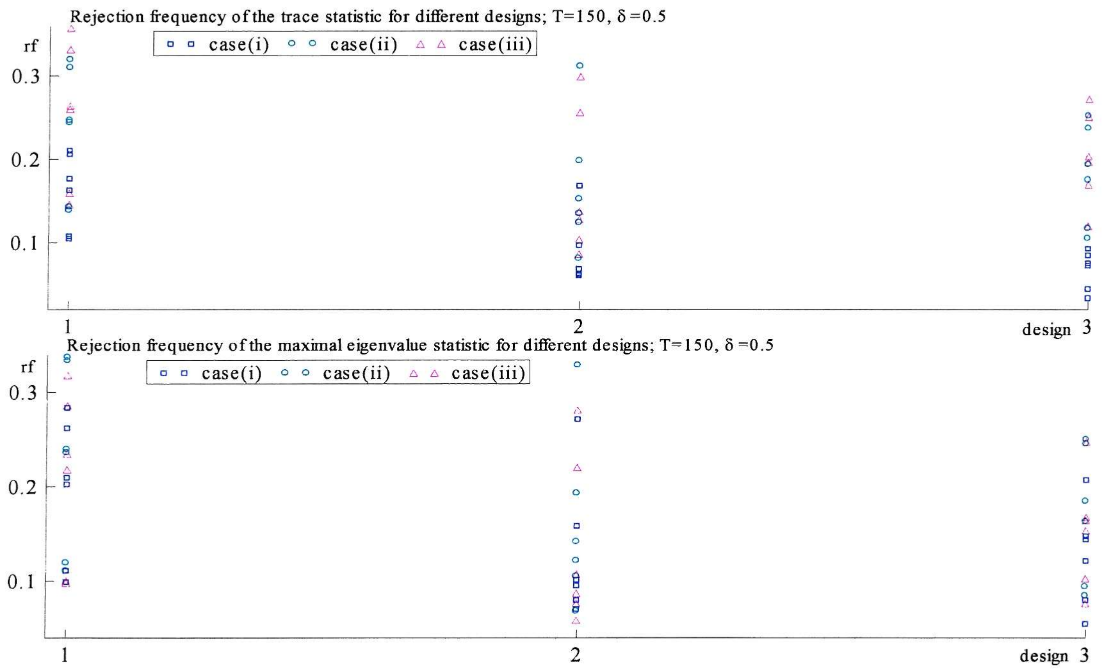

Figure E.27. Rejection frequency for different designs (a common shift; $T = 150$, $\delta = 0.5$).

Appendix F: Estimates of the local power

Table F.1. Rejection frequencies of the hypothesis $r = 0$
using the trace statistic when DGP1* ≡SM.

Sample size / (f, g,power)	50	100	150	500	800
(-3, 12, 0.850)	0.8136	0.8270	0.8295	0.8274	0.8331
(-18, 12, 0.830)	0.8359	0.8284	0.8215	0.8078	0.8090
(-15, 6, 0.565)	0.5354	0.5095	0.4991	0.4855	0.4801
(-18, 0, 0.513)	0.5889	0.5509	0.5364	0.5194	0.5156
(-6, 6, 0.272)	0.2657	0.2635	0.2539	0.2582	0.2604
(-12, 0, 0.269)	0.2789	0.2714	0.2702	0.2588	0.2548

Table F.2. Rejection frequencies of the hypothesis $r = 0$
using the maximal eigenvalue statistic when DGP1* ≡SM.

Sample size / (f, g,power)	50	100	150	500	800
(-3, 12, 0.850)	0.8073	0.8210	0.8258	0.8253	0.8333
(-18, 12, 0.830)	0.8457	0.8345	0.8318	0.8208	0.8187
(-15, 6, 0.565)	0.5272	0.5036	0.4899	0.4800	0.4758
(-18, 0, 0.513)	0.5920	0.5550	0.5398	0.5227	0.5110
(-6, 6, 0.272)	0.2463	0.2432	0.2412	0.2447	0.2446
(-12, 0, 0.269)	0.2653	0.2571	0.2473	0.2473	0.2384

References

Anderson, T. W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distribution. Annals of Mathematical Statistics 22, 327-351.

Anderson, T. W., 1984. An Introduction to Multivariate Statistical Analysis. 2nd ed., John Wiley, New York.

Andersson, S. A., Brøns, H. K., Jensen, S. T., 1983. Distribution of eigenvalues in multivariate statistical analysis. Annals of Statistics 11, 392-415.

Andrade, I. C., O'Brien, R. J., Podivinsky, J. M., 1994. Cointegration tests and mean shifts. University of Southampton Department of Economics Discussion Paper, USDP 9405.

Banerjee, A., Lumsdaine, R. L., Stock, J. H., 1992. Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence. Journal of Business and Economic Statistics 10, 271-287.

Beveridge, S., Nelson, C. R., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. Journal of Monetary Economics 7, 151-174.

Billingsley, P., 1968. Convergence of Probability Measures. John Wiley, New York.

Boswijk, P., Franses, P. H., 1992. Dynamic specification and cointegration. Oxford Bulletin of Economics and Statistics 53, 369-382.

Campos, J., Ericsson, N. R., Hendry, D. F., 1996. Cointegration tests in the presence of structural breaks. Journal of Econometrics 70, 187-220.

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

Clements, M. P., Hendry, D. F., 1999. Forecasting Non-stationary Economic Time Series. MIT Press, Cambridge MA.

Davidson, J., 2000. Econometric Theory. Blackwell, Oxford.

DeLoach, S. B., 2001. More evidence in favor of the Balassa-Samuelson hypothesis. Review of International Economics 9, 336-342.

Doornik, J. A., 1998. Approximations to the asymptotic distributions of cointegration tests. Journal of Economic Surveys 12, 573-593.

Doornik, J. A., 1999. Object-Oriented Matrix Programming using Ox version 2.1. Timberlake Consultants Ltd, London and www.nuff.ox.ac.uk/Users/Doornik.

Doornik, J. A., O'Brien, R. J., 2001. Numerically stable cointegration analysis. Forthcoming in Computational Statistics and Data Analysis.

Engle, R. F., Granger, C. W. J., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55, 251-276.

Gregory, A. W., Hansen, B. E., 1996a. Residual-based tests for cointegration in models with regime shifts. Journal of Econometrics 70, 99-126.

Gregory, A. W., Hansen, B. E., 1996b. Tests for cointegration in models with regime and trend shifts. Oxford Bulletin of Economics and Statistics 58, 555-560.

Gregory, A. W., Nason, J. M., Watt, D. G., 1996. Testing for structural breaks in cointegrated relationships. Journal of Econometrics 71, 321-341.

Hamilton, J. D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Hansen, B. E., 1992. Tests for parameter instability in regressions with $I(1)$ processes. Journal of Business and Economic Statistics 10, 321-335.

Hansen, H., Johansen, S., 1999. Some tests for parameter constancy in cointegrated VAR-models. Econometrics Journal 2, 306-333.

Hao, K., 1996. Testing for structural change in cointegrated regression models: some comparisons and generalizations. Econometric Reviews 15, 401-429.

Hendry, D. F., 1995. Dynamic Econometrics. Oxford University Press, Oxford.

Hendry, D. F., Neale, A. J., 1991. A monte carlo study of the effects of structural breaks on tests for unit roots. In P. Hackl and A. H. Westlund (eds.), Economic Structural Change: Analysis and Forecasting, pp. 95-119, Springer Verlag, Berlin.

Inoue, A., 1999. Tests of cointegrating rank with trend-break. Journal of Econometrics 90, 215-237.

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

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

Johansen, S., 1991b. The power function of the likelihood ratio test for cointegration. In J. Gruber (ed.), Econometric Decision Models: New Methods of Modelling and Applications, pp. 323-335, Springer Verlag, New York.

Johansen, S., 1994. The role of the constant and linear terms in cointegration analysis of nonstationary variables. Econometric Reviews 13, 205-229.

Johansen, S., 1996. Likelihood-based Inference in Cointegrated Vector Autoregressive Models. 2nd printing. Oxford University Press, Oxford.

Johansen, S., 1999. The Bartlett correction factor for the test for cointegrating rank. Discussion Paper. European University Institute.

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

Johansen, S., Mosconi, R., Nielsen, B., 2000. Cointegration analysis in the presence of structural breaks in the deterministic trend. Econometrics Journal 3, 216-249.

Kuo, B., 1998. Test for partial parameter instability in regressions with $I(1)$ processes. Journal of Econometrics 86, 337-368.

Larsson, R., 1999. Approximation of the asymptotic distribution of the log likelihood ratio test for cointegration. Econometric Theory 15, 789-813.

MacKinnon, J. G., Haug, A. A., Michelis, L., 1999. Numerical distribution functions of likelihood ratio tests for cointegration. Journal of Applied Econometrics 14, 563-577.

Nelson, R., Plosser, C. I., 1982. Trends and random walks in macroeconomic time series: some evidence and implications. Journal of Monetary Economics 10, 139-162.

Nielsen B., 1997. On the distribution of tests for cointegration rank. Working Paper W10. Nuffield College.

Nielsen B., Rahbek, A., 2000. Similarity issues in cointegration analysis. Oxford Bulletin of Economics and Statistics 62, 5-22.

O'Brien, R. J., 1996. The curious case of 'spurious' cointegration. University of Southampton Department of Economics Discussion Paper, USDP 9638.

O'Brien, R. J., 1997. Further curiosa in 'spurious' cointegration. University of Southampton Department of Economics Discussion Paper, USDP 9704.

O'Brien, R. J.,1999. The curious case of 'spurious' cointegration. Paper presented at the Far Eastern Meeting of the Econometric Society, Singapore, July 1999.

O'Brien, R. J., 2001. Is my asymptosis terminal?. Mimeo. University of Southampton Department of Economics.

Ohanian, L. E., 1988. The spurious effects of unit roots on vector autoregressions: A Monte Carlo study. Journal of Econometrics 39, 251-266.

Osterwald-Lenum, M., 1992. A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. Oxford Bulletin of Economics and Statistics 54, 461-471.

Paruolo, P., 2001. The power of lambda max. Oxford Bulletin of Economics and Statistics 63, 395-403.

Perron, P., 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361-1401.

Perron, P., Campbell, J. Y., 1993. A note on Johansen's cointegration procedure when trends are present. Empirical Economics 18, 777-789.

Perron, P., Vogelsang, T. J., 1992. Nonstationarity and level shifts with an application to purchasing power parity. Journal of Business and Economic Statistics 10, 301-320.

Phillips, P. C. B., 1987. Time series regression with a unit root. Econometrica 55, 277-301.

Phillips, P. C. B., Durlauf, S. N., 1986. Multiple time series regression with integrated processes. Review of Economic Studies 53, 473-495.

Phillips, P. C. B., Solo, V., 1992. Asymptotics for linear processes. Annals of Statistics 20, 971-1001.

Podivinsky, J. M., 1998. Testing misspecified cointegrating relationships. Economics Letters 60, 1-9.

Quintos, C. E., 1997. Stability tests in error correction models. Journal of Econometrics 82, 289-315.

Quintos, C. E., Phillips, P. C. B., 1993. Parameter constancy in cointegrating regressions. Empirical Economics 18, 675-706.

Rahbek, A. C., 1994. The power of some multivariate cointegration tests. Discussion Paper, Institute of Mathematical Statistics, University of Copenhagen.

Reinsel, G. C., Ahn, S. K., 1989. Vector autoregressive models with unit roots and reduced rank structure: estimation, likelihood ratio test, and forecasting. Journal of Time Series Analysis 13, 353-375.

Saikkonen, P., Lütkepohl, H., 1998. Testing for the cointegrating rank of a VAR process with structural shifts. Discussion Paper 84/98, SFB 373, Humboldt - Universität zu Berlin.

Saikkonen, P., Lütkepohl, H., 1999. Local power of likelihood ratio tests for the cointegrating rank of a VAR process. Econometric Theory 15, 50-78.

Seo, B., 1998. Tests for structural change in cointegrated systems. Econometric Theory 14, 222-259.

Stock, J. H., Watson, M. W., 1988. Testing for common trends. Journal of the American Statistical Association 83, 1097-1107.

Toda, H. Y., 1994. Finite sample properties of the likelihood ratio tests for cointegrating ranks when linear trends are present. Review of Economics and Statistics 76, 66-79.

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

Toda, H. Y., Phillips, P. C. B., 1993. The spurious effects of unit roots on vector autoregressions: An analytical study. Journal of Econometrics 59, 229-255.

Zivot, E., Andrews, D. W. K., 1992. Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. Journal of Business and Economic Statistics 10, 251-270.