Saikkonen, Pentti; Lütkepohl, Helmut

Working Paper

Testing for the cointegrating rank of a VAR process with structural shifts

SFB 373 Discussion Paper, No. 1998,82

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Saikkonen, Pentti; Lütkepohl, Helmut (2001) : Testing for the cointegrating rank of a VAR process with structural shifts, SFB 373 Discussion Paper, No. 1998,82, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,
https://nbn-resolving.de/urn:nbn:de:kobv:11-10060635

This Version is available at:
https://hdl.handle.net/10419/61232

July 23, 2001

Testing for the Cointegrating Rank of a VAR Process with Structural Shifts*

by

Pentti Saikkonen

Department of Statistics
University of Helsinki
P.O. Box 54
SF-00014 University of Helsinki
FINLAND

Tel.: +358-9-1918867
Fax: +358-9-1918872

and

Helmut Lütkepohl

Institut für Statistik und Ökonometrie
Wirtschaftswissenschaftliche Fakultät
Humboldt–University
Spandauer Str. 1
10178 Berlin
GERMANY

Tel.: +49-30-2093-5718
Fax: +49-30-2093-5712

Abstract

Tests for the cointegrating rank of a vector autoregressive process are considered which allow for possible exogenous shifts in the mean of the data generation process. The break points are assumed to be known a priori. It is proposed to estimate and remove the deterministic terms such as mean, linear trend term and a shift in a first step. Then systems cointegration tests are applied to the adjusted series. The resulting tests are shown to have known limiting null distributions which are free of nuisance parameters and do not depend on the break point. The tests are applied for analyzing the number of cointegrating relations in a German money demand system.

*We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213), for financial support. We thank Kirstin Hubrich for helping with the computations. Part of this research was done while the first author was visiting the Humboldt University in Berlin.

1 Introduction

Many economic time series exhibit breaks or shifts in their levels which are not consistent with standard types of data generation processes (DGPs). Such breaks are often caused by exogenous events that have occurred during the observation period. For example, the German unification has caused shifts in a number of macroeconomic time series such as Gross National Product (GNP) and measures of the money stock. In this example the timing and the reasons for the shifts are known. In other situations neither the timing nor the fact whether a shift actually has occurred are known at the outset of an analysis.

Corresponding to the various different types of structural changes and problems related to them there is an extensive literature dealing with the consequences of structural shifts for estimation and testing procedures in univariate and multivariate time series models as well as in regression models for time series variables (see, e.g., Hackl & Westlund (1989) for a large number of references to the earlier literature). In particular, in many studies testing for unit roots and breaks in univariate time series is considered. Examples are Perron (1989, 1990), Perron & Vogelsang (1992), Rappoport & Reichlin (1989), Zivot & Andrews (1992), Banerjee, Lumsdaine & Stock (1992), Amsler & Lee (1995), Ghysels & Perron (1996) to name just a few. In these papers different assumptions regarding the DGP are made. For instance, the break point may be known or unknown, it may be a shift in the level of a series or it may be a break in the deterministic trend component. Structural shifts in the context of cointegration analysis are considered, e.g., by Hansen (1992), Gregory & Hansen (1996), Campos, Ericsson & Hendry (1996) and Johansen & Nielsen (1993) among others. In all these papers, except Johansen & Nielsen (1993), tests for cointegration are discussed in a single equation framework. In contrast, Johansen & Nielsen (1993) study the consequences of structural breaks in a systems context and derive likelihood ratio (LR) tests for the number of cointegrating relations in a system of variables.

The overall message from these studies is that structural breaks can distort standard inference procedures substantially and, hence, it is necessary to make appropriate adjustments if structural shifts are known to have occurred or are suspected. If there is just one break point in the observation period it may be tempting to analyze the two regimes before and after the shift separately. This may, however, result in a substantial loss in efficiency and/or power. Therefore procedures that take structural changes into account by adjusting

the inference methods are often preferable to eliminating parts of the sample.

In this study we will consider the problem of how to test for the number of cointegrating relations in a system of variables if some of them have a shift in the mean at a known time point. This situation is relevant, for instance, in the aforementioned case of German data if the sampling period covers the German unification. Assuming that the DGP is a finite order vector autoregressive (VAR) process with a shift in the mean, tests for the number of cointegrating relations will be proposed with an asymptotic null distribution which is free of nuisance parameters and does not depend on where the break point has occurred. In this respect our tests contrast with the tests analyzed by Johansen & Nielsen (1993) whose asymptotic null distribution depends on when the shift has occurred. If the break point is assumed to be known this does not necessarily mean that the tests depend on unknown nuisance parameters. It means, however, that new critical values have to be obtained for each specific situation. Even if just a few new observations become available generating new critical values will generally be necessary in order to perform the tests. In contrast, the tests proposed in the following have a known asymptotic distribution which does not depend on the break date so that no new simulations are required. Tables with critical values are available elsewhere in the literature. Moreover, our test can be adopted for time series with more than one shift in the mean or to series with individual outlying observations.

The idea underlying our new tests is to estimate and remove the deterministic parts including the shifts in a first step and then perform a test for the cointegrating rank on the adjusted series. The deterministic part of the DGP may, in fact, include a linear time trend in addition to shifts in the mean term. Similar ideas were used by Amsler & Lee (1995) in unit root tests in the presence of structural breaks and by Lütkepohl & Saikkonen (1997) and Saikkonen & Lütkepohl (1997) to construct tests for VAR processes with deterministic linear trends and without structural shifts. Since we will refer to the latter two articles several times in the following we will abbreviate them as L&S and S&L, respectively.

The structure of the paper is as follows. In the next section the DGP is precisely specified and the assumptions underlying our analysis are laid out. The estimation of the parameters of the deterministic parts is discussed in Section 3 and the cointegration tests are presented in Section 4. An example based on a German money demand system is given in Section 5 and conclusions follow in Section 6. The proofs of our theorems are provided in Appendix

A and the sources of the data used in the empirical example are given in Appendix B.

The following general notation is used. The lag and differencing operators are denoted by L and Δ, respectively, that is, for a time series or stochastic process y_t we have $Ly_t = y_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ denotes an integrated process of order d, that is, the process is stationary or asymptotically stationary after differencing d times while it is still nonstationary after differencing just $d-1$ times. Convergence in distribution or weak convergence is signified by $\xrightarrow{d}$. The maximal eigenvalue, the trace and the rank of the matrix A are denoted by $\lambda_{max}(A)$, $\mathrm{tr}(A)$ and $\mathrm{rk}(A)$, respectively. Moreover, $\|\cdot\|$ denotes the Euclidean norm. If A is an $(n \times m)$ matrix of full column rank $(n > m)$ we denote an orthogonal complement by $A_\perp$ so that $A_\perp$ is an $(n \times (n-m))$ matrix of full column rank and such that $A'A_\perp = 0$. The orthogonal complement of a nonsingular square matrix is zero and the orthogonal complement of a zero matrix is an identity matrix of suitable dimension. An $(n \times n)$ identity matrix is denoted by I_n. LS, GLS and RR are used to abbreviate least squares, generalized least squares and reduced rank, respectively. LR and LM test are short for likelihood ratio and Lagrange multiplier test. DGP stands for data generation process and ECM abbreviates error correction model. A sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 The Model

Suppose an observed n-dimensional time series $y_t = (y_{1t}, \ldots, y_{nt})'$ $(t = 1, \ldots, T)$ is generated by the following mechanism:

$$y_t = \mu_0 + \mu_1 t + \delta_0 d_{0t} + \delta_1 d_{1t} + x_t, \quad t = 1, 2, \ldots, \tag{2.1}$$

where μ_i and δ_i $(i = 0, 1)$ are unknown $(n \times 1)$ parameter vectors, d_{0t} and d_{1t} are dummy variables defined as

$$d_{0t} = \begin{cases} 1, & t = T_0 \\ 0, & t \neq T_0 \end{cases} \tag{2.2}$$

and

$$d_{1t} = \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases}, \tag{2.3}$$

that is, d_{0t} is an impulse dummy and d_{1t} is a step dummy variable. These terms allow to take into account sudden changes in the mean of the process which occur, for instance, in German macroeconomic time series at the time of the reunification.

The term x_t is an unobservable error process which is assumed to have a VAR(p) representation,

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t. \tag{2.4}$$

Here the A_j are $(n \times n)$ coefficient matrices and, for simplicity, we assume that $x_t = 0$ for $t \leq 0$ and $\varepsilon_t \sim NID(0, \Omega)$, that is, the ε_t are independent, identically distributed (i.i.d.) Gaussian vectors with zero mean and covariance matrix Ω. The normality assumption is made here for convenience although it could be replaced by an i.i.d. assumption and sufficient moment conditions. The assumption regarding the initial values is also a convenient simplification. The asymptotic results derived later remain valid if the initial values are assumed to be from a fixed probability distribution which does not depend on the sample size. The usual error correction form of (2.4), obtained by subtracting x_{t-1} from both sides of (2.4) and rearranging terms, is given by

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \quad t = p+1, p+2, \ldots, \tag{2.5}$$

where $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ $(j = 1, \ldots, p-1)$ are $(n \times n)$ matrices.

We assume that x_t is at most $I(1)$ and cointegrated with cointegrating rank r. This implies in particular that the Granger representation theorem of Johansen (1995, Chapter 4) is assumed to hold. Hence, the matrix Π can be written as

$$\Pi = \alpha \beta', \tag{2.6}$$

where α and β are $(n \times r)$ matrices of full column rank. As is well-known, $\beta' x_t$ and Δx_t are then zero mean (asymptotically) stationary processes. Moreover, defining

$$\Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} j A_{j+1}$$

and $C = \beta_\perp (\alpha_\perp' \Psi \beta_\perp)^{-1} \alpha_\perp'$, we have

$$x_t = C \sum_{j=1}^{t} \varepsilon_j + \xi_t, \quad t = 1, 2, \ldots, \tag{2.7}$$

where ξ_t is a zero mean (asymptotically) stationary process.

Now we consider the dummy variables d_{0t} and d_{1t}. We assume the values of the integers T_0 and T_1 are known a priori and, if y_t is observed for $t = 1, \ldots, T$, then $T_0 \leq T$ and $T_1 \leq T$. The case $T_0 = T_1$ is possible unless $T_0 = T_1 = T$. We exclude this latter possibility by assuming $T_1 < T - p$. It is also convenient to assume that $T_0 > p$ and $T_1 > p$ and, furthermore, that

$$\lim_{T\to\infty} \frac{T_1}{T} = a_1 \quad \text{with} \quad 0 < a_1 \leq 1. \tag{2.8}$$

In other words, the break point T_1 may be thought of as occurring at a fixed proportion of the full sample size even if an asymptotic analysis is performed where $T \to \infty$. Alternatively, the break may be viewed as having occurred a fixed number of periods before the end of the sample period. In the latter case $a_1 = 1$. These assumptions for the integers T_0 and T_1 are not restrictive from a practical point of view. This is obvious for the aforementioned inequalities. Condition (2.8) may look somewhat restrictive because it implies that the jump in the dummy d_{1t} is not allowed to take place at the beginning of the sample so that, for example, T_1 would be only slightly larger than p. This would mean that the parameters μ_0 and δ_0 would became asymptotically indistinguishable. However, this problem can readily be handled by redefining d_{1t} so that it takes the value 1 for $t < T_1$ and 0 for $t \geq T_1$. Then the inequalities in (2.8) should be changed to $0 \leq a < 1$ and it is easy to see that our subsequent derivations apply with only minor and fairly obvious modifications. In summary, the above discussion shows that the jump in the dummy variable d_{1t} can also take place at the beginning or at the end of the sample but, for ease of exposition, we exclude one of these two possibilities. Thus, our assumptions about the integers T_0 and T_1 are quite general and weaker than in some previous studies, where condition (2.8) is required with $0 < a_1 < 1$. Finally, note that we assume for convenience and for expository purposes that there is just one impulse dummy d_{0t} and one step dummy d_{1t}. It is not difficult to see that our results can be adopted to models with any number of (linearly independent) impulse dummies and step dummies. Moreover, our results can easily be modified to the case where the model contains only step dummies or impulse dummies. It is also possible to exclude the trend term from the model, that is, $\mu_1 = 0$ may be assumed a priori. This case will be briefly discussed later on.

For our analysis it is convenient to define the lag polynomial

$$\begin{aligned} A(L) &= I_n - A_1 L - \cdots - A_p L^p \\ &= I_n \Delta - \Pi L - \Gamma_1 \Delta L - \cdots - \Gamma_{p-1} \Delta L^{p-1} \end{aligned} \tag{2.9}$$

and notice that the relation between the two representations is given by

$$\begin{aligned} A_1 &= I_n + \alpha\beta' + \Gamma_1 \\ A_j &= \Gamma_j - \Gamma_{j-1}, \qquad j = 2, \ldots, p-1, \\ A_p &= -\Gamma_{p-1} \end{aligned} \tag{2.10}$$

where Π is expressed as in (2.6). Multiplying (2.1) by $A(L)$ yields

$$\begin{aligned} \Delta y_t &= \nu_0 + \nu_1 t + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \delta_0 d_{0t} - \sum_{j=1}^{p} A_j \delta_0 d_{0,t-j} \\ &\quad + \delta_1 \Delta d_{1t} - \sum_{j=1}^{p-1} \Gamma_j \delta_1 \Delta d_{1,t-j} - \Pi \delta_1 d_{1,t-1} + \varepsilon_t, \quad t = p+1, p+2, \ldots, \end{aligned}$$

where $\nu_0 = -\Pi\mu_0 + (\Psi + \Pi)\mu_1$ and $\nu_1 = -\Pi\mu_1$. To be able to write this model in a simpler form, define

$$\gamma_{0j} = \begin{cases} \delta_0, & j = 0 \\ -A_j \delta_0, & j = 1, \ldots, p \end{cases}$$

and

$$\gamma_{1j} = \begin{cases} \delta_1, & j = 0 \\ -\Gamma_j \delta_1, & j = 1, \ldots, p-1 \end{cases}.$$

Using (2.6), we can then write

$$\begin{aligned} \Delta y_t &= \nu + \alpha(\beta' y_{t-1} - \tau(t-1) - \theta d_{1,t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} \\ &\quad + \sum_{j=0}^{p} \gamma_{0j} d_{0,t-j} + \sum_{j=0}^{p-1} \gamma_{1j} \Delta d_{1,t-j} + \varepsilon_t, \qquad t = p+1, p+2, \ldots, \end{aligned} \tag{2.11}$$

where $\nu = -\Pi\mu_0 + \Psi\mu_1$, $\tau = \beta'\mu_1$ and $\theta = \beta'\delta_1$. Notice that here $\Delta d_{1,t-j}$ is an impulse dummy which takes the value one at $t = T_1 + j$ and zero elsewhere.

Equation (2.11) specifies an ECM for the observed series y_t. We shall use this form of the model to obtain first stage estimators for the parameters of the error process x_t, that is, for α, β, Γ_j $(j = 1, \ldots, p-1)$ and Ω. Some remarks on the ECM (2.11) and the estimation of its parameters are therefore in order. Using equation (2.10) and the definitions it can first be seen that a conventional RR regression cannot be used to obtain the ML estimators because there are nonlinear restrictions between the parameters in (2.11). To obtain the

above mentioned first stage estimators we shall simply ignore these restrictions. This should not cause any great loss of efficiency because the restrictions occur in coefficient vectors of impulse dummies only. Before computing the estimators one should check, however, that the dummy variables on the right-hand side of (2.11) are linearly independent. Since both d_{0t} and Δd_{1t} are impulse dummies it is possible that some impulse dummies appear twice in (2.11) and can be combined. This, of course, has no effect on the estimation of the parameters α, β, Γ_j and Ω which are of interest at this point. For simplicity, we assume that all dummy variables in (2.11) are linearly independent. The assumption $p < T_1 < T - p$ guarantees that there cannot be linear dependencies between the step dummy $d_{1,t-1}$ and the impulse dummies. Such linear dependencies could be eliminated by excluding impulse dummies so that the assumption $T_1 < T-p$ is actually not needed here. However, it appears convenient later (see the proof of Theorem 3.1) and is therefore imposed because assuming $T_1 < T - p$ seems harmless from a practical point of view.

In the framework of our model we are interested in testing whether the assumption made for the rank of the matrix Π is correct. In other words, for some prespecified rank r_0, we wish to consider testing the null hypothesis

$$H_0(r_0) : \mathrm{rk}(\Pi) = r_0 \quad vs. \quad H_1(r_0) : \mathrm{rk}(\Pi) > r_0. \tag{2.12}$$

In this context it turns out to be very useful to make explicit use of the assumption that the DGP is of the form (2.1). As will be seen later, it is then possible to obtain tests with convenient asymptotic properties. As mentioned in the introduction, our formulation of the model allows to estimate the deterministic part of the DGP first and then apply cointegration tests to the process adjusted for deterministic terms. In the next section estimators of the parameters of the deterministic part will be presented and the cointegration tests will be considered in Section 4.

3 Estimating the Parameters of the Deterministic Part of the Model

We shall estimate the parameters μ_i and δ_i $(i = 0, 1)$ in (2.1) by using a feasible GLS approach similar to that in S&L. To derive the estimator we define

$$a_{0t} = \begin{cases} 1 & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases} \qquad \text{and} \qquad a_{1t} = \begin{cases} t & \text{for} \quad t \geq 1 \\ 0 & \text{for} \quad t \leq 0 \end{cases} .$$

Multiplying (2.1) from the left by $A(L)$ gives

$$A(L)y_t = H_{0t}\mu_0 + H_{1t}\mu_1 + K_{0t}\delta_0 + K_{1t}\delta_1 + \varepsilon_t, \qquad t = 1, 2, \ldots, \tag{3.1}$$

where the y_t are set to zero for $t \leq 0$, $H_{it} = A(L)a_{it}$ and $K_{it} = A(L)d_{it}$, $i = 0, 1$. As in S&L we also define the matrix

$$Q = [\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1/2} : \alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1/2}] \tag{3.2}$$

with the property $QQ' = \Omega^{-1}$. Premultiplying (3.1) by Q' transforms the covariance matrix of the error term to an identity matrix so that, as required in GLS estimation, we have a transformation which results in a (multivariate) regression model with standard properties of the error term. As will be discussed later, the above transformation is not the only possibility but it appears convenient when asymptotic properties of the estimators are derived.

To make the above transformation feasible, suitable estimators of the parameters α, β, Γ_j $(j = 1, \ldots, p-1)$ and Ω are needed. Such estimators can be obtained by a RR regression of (2.11) in the way discussed in the previous section. These estimators are denoted by $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\Gamma}_j$ and $\tilde{\Omega}$. Substituting them for the corresponding theoretical parameters in (2.10) gives estimators for the coefficient matrices A_j. Denoting these estimators by $\tilde{A}_j$ we can define $\tilde{A}(L) = I_n - \tilde{A}_1 L - \cdots - \tilde{A}_p L^p$ and, furthermore, $\tilde{H}_{it} = \tilde{A}(L)a_{it}$ and $\tilde{K}_{it} = \tilde{A}(L)d_{it}$ $(i = 0, 1)$. Note that $\tilde{A}(L)$ satisfies the cointegrating restrictions. Thus we are able to construct a feasible analog of (3.1). Moreover, a suitable estimator of the matrix Q can be readily obtained by forming $\tilde{\alpha}_\perp$ from $\tilde{\alpha}$ and replacing Ω, α and $\alpha_\perp$ in the definition of Q by their estimators. If $\tilde{Q}$ is used to denote the resulting estimator of Q we can finally introduce the multivariate auxiliary regression model

$$\tilde{Q}'\tilde{A}(L)y_t = \tilde{Q}'\tilde{H}_{0t}\mu_0 + \tilde{Q}'\tilde{H}_{1t}\mu_1 + \tilde{Q}'\tilde{K}_{0t}\delta_0 + \tilde{Q}'\tilde{K}_{1t}\delta_1 + \eta_t, \qquad t = 1, \ldots, T. \tag{3.3}$$

The LS estimators of the parameters μ_i and δ_i will be denoted by $\hat{\mu}_i$ and $\hat{\delta}_i$ $(i = 0, 1)$, respectively. They are used in the next section to obtain tests for the cointegrating rank. It is, of course, apparent that the estimator $\hat{\delta}_0$ which estimates the coefficient vector of the impulse dummy d_{0t} cannot be consistent. From S&L it is also clear that the estimator $\hat{\mu}_0$ is generally not consistent although it is consistent in the direction of β, that is, $\beta'\hat{\mu}_0$ is a consistent estimator of $\beta'\mu_0$. On the basis of this result it is to be expected that the consistency property of the estimators $\hat{\delta}_1$ are similar to those of $\hat{\mu}_0$ provided $T - T_1 \to \infty$. The asymptotic properties of the estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ are given in the following theorem whose proof requires suitable consistency results of the estimators $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\Gamma}_j$ and $\tilde{\Omega}$. These results are first stated in the following lemma.

Lemma 3.1. Suppose the assumptions made in Section 2 hold and the null hypothesis $H_0(r_0)$ is true. Suppose further that if (2.8) holds with $a_1 = 1$ then $T - T_1$ either tends to infinity or converges to a finite constant. Define the (infeasible) estimators $\tilde{\beta}_\xi = \tilde{\beta}(\xi'\tilde{\beta})^{-1}$ and $\tilde{\alpha}_\xi = \tilde{\alpha}\tilde{\beta}'\xi$, where $\xi' = (\beta'\beta)^{-1}\beta'$. Then $\tilde{\beta}_\xi = \beta + O_p(T^{-1})$, $\tilde{\alpha}_\xi = \alpha + O_p(T^{-1/2})$, $\tilde{\Gamma}_j = \Gamma_j + O_p(T^{-1/2})$ $(j = 1, \ldots, p-1)$ and $\tilde{\Omega} = \Omega + O_p(T^{-1/2})$. □

The lemma is proven in the Appendix. The results are similar to those one obtains from a model without dummy variables. They can be used to show that the same consistency properties apply with any normalization of the estimators (cf. Johansen (1995, p. 184)). In the present context the normalization does not matter, however, since we use the estimators $\tilde{\alpha}$ and $\tilde{\beta}$ in situations which are invariant to a particular normalization. Now we can state the properties of the estimators of the parameters of the deterministic part of (2.1). Again the proof is given in the Appendix.

Theorem 3.1. Under the conditions of Lemma 3.1,

$$\beta'(\hat{\mu}_0 - \mu_0) = O_p(T^{-1/2}) \tag{3.4}$$

$$\beta'_\perp(\hat{\mu}_0 - \mu_0) = O_p(1) \tag{3.5}$$

$$\hat{\delta}_0 - \delta_0 = O_p(1) \tag{3.6}$$

$$\beta'(\hat{\delta}_1 - \delta_1) = O_p((T - T_1)^{-1/2}) \tag{3.7}$$

$$\beta_\perp'(\hat{\delta}_1 - \delta_1) = O_p(1) \tag{3.8}$$

$$\beta'(\hat{\mu}_1 - \mu_1) = O_p(T^{-3/2}) \tag{3.9}$$

$$T^{1/2}\beta_\perp'(\hat{\mu}_1 - \mu_1) \xrightarrow{d} N(0, \beta_\perp' C\Omega C'\beta_\perp) \tag{3.10}$$

Here $C = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$ as before and all quantities converge jointly in distribution upon appropriate standardization. □

The properties of the estimators $\hat{\mu}_0$ and $\hat{\mu}_1$ are entirely similar to those obtained by S&L, Theorem 1. Whereas $\hat{\mu}_1$ is consistent, the same is not true for $\hat{\mu}_0$. The latter is consistent only in the direction of β and not in the direction of $\beta_\perp$. However, even in the direction of $\beta_\perp$ the estimator $\hat{\mu}_0$ is bounded in probability and this property turns out to be sufficient for our purposes. As to the estimators $\hat{\delta}_0$ and $\hat{\delta}_1$, the obvious inconsistency of the former is implied by (3.6) while (3.7) and (3.8) show that the asymptotic behavior of the latter is similar to that of $\hat{\mu}_0$ when $T - T_1 \to \infty$ or, in other words, when information on the parameter δ_1 in the direction of β increases with the sample size. A general conclusion of (3.7) and (3.8) is that the estimator $\hat{\delta}_1$ is never consistent in the direction of $\beta_\perp$ and it can also be inconsistent in the direction of β if $T - T_1$ does not go to infinity. As in the case of $\hat{\mu}_0$, from our point of view it is important that the estimator $\hat{\delta}_1$ is bounded in probability, particularly in the direction of $\beta_\perp$. We will now discuss how these estimators can be used in constructing tests for the pair of hypotheses in (2.12).

4 Test Procedures

When the estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ $(i = 0, 1)$ are available one can form a sample analog of the series x_t as

$$\hat{x}_t = y_t - \hat{\mu}_0 - \hat{\mu}_1 t - \hat{\delta}_0 d_{0t} - \hat{\delta}_1 d_{1t}$$

and use it to obtain LM type or LR type test statistics for the hypothesis $H_0(r_0)$ in the same way as in L&S and S&L. The LM type test statistic requires estimators of the parameters α, β and Ω. The RR regression estimators based on (2.11) and discussed in the previous section can be used for this purpose. Alternatively, the LR type test statistic may be obtained in the same way as the usual LR test statistic from the feasible counterpart of the ECM (2.5),

that is, it is determined from

$$\Delta\hat{x}_t = \Pi\hat{x}_{t-1} + \sum_{j=1}^{p-1}\Gamma_j\Delta\hat{x}_{t-j} + e_t, \quad t = p+1,\ldots,T, \tag{4.1}$$

where e_t is an error term defined explicitly in the appendix. The following general formulation discusses LM and LR type test statistics obtained in both of the aforementioned ways.

Let $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\Omega}$ be any estimators of the parameters α, β and Ω, respectively, satisfying the consistency results in Lemma 2.1. Consider the auxiliary regression model

$$\tilde{\alpha}_\perp'\Delta\hat{x}_t = \varphi^*\hat{u}_{t-1} + \rho^*\hat{v}_{t-1} + \sum_{j=1}^{p-1}\Gamma_j^*\Delta\hat{x}_{t-j} + \tilde{\alpha}_\perp' e_t^*, \quad t = p+1,\ldots,T, \tag{4.2}$$

where $\hat{u}_t = \tilde{\beta}'\hat{x}_t$, $\hat{v}_t = \tilde{\beta}_\perp'\hat{x}_t$ and $e_t^* = e_t - \alpha(\tilde{\beta}-\beta)'\hat{x}_{t-1}$. The LM type test statistic of S&L is obtained by forming the usual LM test statistic of the multivariate linear model for the null hypothesis $\rho^* = 0$ in (4.2). If the usual LR test statistic of the multivariate linear model is used instead an asymptotically equivalent test statistic is obtained. In general this test statistic is not the LR test statistic based on (4.1). However, as shown in Saikkonen & Lütkepohl (1998), this is actually the case if $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\Omega}$ are RR regression estimators based on (4.1). Thus, to be able to use the above general framework to construct the LR type test based on (4.1) we have to show that the consistency results of Lemma 3.1 hold for the RR regression estimators of α, β and Ω obtained from (4.1). This will be done in the Appendix. Thus, testing the null hypothesis $\rho^* = 0$ in (4.2) by conventional methods of the multivariate linear model gives both the LM type and the LR type tests discussed in S&L. For convenience the following test statistic assumes a LM type (or Wald type) form but the LR type form is of course asymptotically equivalent under the null hypothesis.

Following S&L we now introduce the test statistic

$$LM(r_0) = \mathrm{tr}\{\hat{\rho}^*\hat{M}_{vv\cdot z}\hat{\rho}^{*\prime}(\tilde{\alpha}_\perp'\tilde{\Omega}\tilde{\alpha}_\perp)^{-1}\}, \tag{4.3}$$

where $\hat{\rho}^*$ is the LS estimator of ρ^* from (4.2) and

$$\hat{M}_{vv\cdot z} = \sum_{t=p+1}^{T}\hat{v}_{t-1}\hat{v}_{t-1}' - \sum_{t=p+1}^{T}\hat{v}_{t-1}\hat{z}_t'\left(\sum_{t=p+1}^{T}\hat{z}_t\hat{z}_t'\right)^{-1}\sum_{t=p+1}^{T}\hat{z}_t\hat{v}_{t-1}'$$

with $\hat{z}_t = [\hat{u}_{t-1}' : \Delta\hat{x}_{t-1}' : \cdots : \Delta\hat{x}_{t-p+1}']'$. S&L also discuss an asymptotically equivalent variant of the above test statistic obtained by deleting the regressor $\hat{u}_{t-1}$ from (4.2). We

shall not consider this modification because it was found to give very similar results in small samples.

The LR type statistic based on (4.1) is obtained in the usual way by solving the generalized eigenvalue problem $\det(\bar{\Pi}\hat{M}_T\bar{\Pi}' - \lambda\bar{\Omega}) = 0$, where $\bar{\Pi}$ is the LS estimator of Π obtained from (4.1), $\bar{\Omega}$ is the corresponding residual covariance matrix and

$$\hat{M}_T = \sum_{t=p+1}^{T} \hat{x}_{t-1}\hat{x}'_{t-1} - \sum_{t=p+1}^{T} \hat{x}_{t-1}\Delta\hat{X}'_{t-1} \left(\sum_{t=p+1}^{T} \Delta\hat{X}_{t-1}\Delta\hat{X}'_{t-1} \right)^{-1} \sum_{t=p+1}^{T} \Delta\hat{X}_{t-1}\hat{x}'_{t-1}$$

with $\Delta\hat{X}_t = [\Delta\hat{x}'_{t-1} : \cdots : \Delta\hat{x}'_{t-p+1}]'$. Denoting the resulting eigenvalues by $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$, the LR type statistic becomes

$$LR(r_0) = \sum_{j=r_0+1}^{n} \log(1 + \hat{\lambda}_j). \tag{4.4}$$

Now we can state the following theorem where $\mathbf{B}(s)$ is an $(n - r_0)$-dimensional standard Brownian motion.

Theorem 4.1. If $H_0(r_0)$ in (2.12) is true,

$$LM(r_0), LR(r_0) \xrightarrow{d} \operatorname{tr}\left\{ \left(\int_0^1 \mathbf{B}_*(s)d\mathbf{B}_*(s)' \right)' \left(\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'ds \right)^{-1} \left(\int_0^1 \mathbf{B}_*(s)d\mathbf{B}_*(s)' \right) \right\},$$

where $\mathbf{B}_*(s) = \mathbf{B}(s) - s\mathbf{B}(1)$ is an $(n - r_0)$-dimensional Brownian bridge and $d\mathbf{B}_*(s) = d\mathbf{B}(s) - ds\mathbf{B}(1)$ is as in Theorem 2 of S&L and Theorem 5.1 of L&S. Note that $\int_0^1 \mathbf{B}_*(s)d\mathbf{B}_*(s)'$ abbreviates $\int_0^1 \mathbf{B}(s)d\mathbf{B}(s)' - \mathbf{B}(1)\int_0^1 sd\mathbf{B}(s)' - \int_0^1 \mathbf{B}(s)ds\mathbf{B}(1)' + \frac{1}{2}\mathbf{B}(1)\mathbf{B}(1)'$. □

The limiting distribution obtained in Theorem 4.1 is free of unknown nuisance parameters and actually the same as the one obtained by S&L and L&S in a model without any dummy variables. Critical values are given in Table 1 of L&S. Thus, in our framework, including step dummies and impulse dummies in the model and estimating their coefficients has no effect on the limiting distribution of the cointegration tests. This is very convenient and contrasts with the LR tests proposed by Johansen & Nielsen (1993). They include dummy variables in the error correction model for y_t and show that in this case the asymptotic null distribution depends on the break point. Hence, a new set of critical values is required for each break point. Our theorem extends previous results of S&L who noticed that the limiting distribution of the corresponding tests for models without dummy variables is not affected by the limiting properties of the (GLS) estimator of the mean parameter μ_0.

The preceding discussion also suggests that if the a priori restriction $\mu_1 = 0$ is employed in (2.1) and the GLS estimation of Section 3 as well as the above test procedure are modified accordingly, the limiting distribution of the resulting test statistic is the same as in a model without any deterministic terms, that is, the limiting distribution obtained by replacing the Brownian bridge $\mathbf{B}_*(s)$ in Theorem 1 by the Brownian motion $\mathbf{B}(s)$. This situation was also studied by S&L and, from the proofs given in the appendix and in that paper, it can be seen that the above conclusion actually holds. The same result can also be obtained by replacing the GLS estimator considered in this paper by one proposed by Saikkonen & Luukkonen (1997). This GLS estimation is similar to the one developed in Section 3 except that the estimator $\tilde{A}(L)$ is obtained by ignoring the cointegration structure of (2.11) and applying LS. It should be noted, however, that, to the best of our knowledge, this GLS estimation has not been studied in a model with a time trend. Further, the modified GLS estimator discussed at the end of Section 3 explicitly assumes that the estimator $\tilde{A}(L)$ satisfies the cointegrating restrictions implied by the null hypothesis. Thus, this modification cannot be applied if an unrestricted LS estimator of $A(L)$ is employed.

5 Illustration: A German Money Demand System

Based on a single equation analysis of Wolters, Teräsvirta & Lütkepohl (1998), Lütkepohl & Wolters (1998) construct a small macroeconomic model to investigate the channels of German monetary policy. They build a vector error correction model for M3, GNP, an inflation rate, an interest rate spread variable and import price inflation. In a demand relation for M3, GNP is a proxi for the transactions volume, the inflation rate and the interest rate spread are opportunity cost variables and the import price inflation is included as a measure for the real exchange rate to account for the openness of the German economy. The variable M3 is used to measure the money stock because it is the target variable of the Deutsche Bundesbank in executing its monetary policy. The interest rate spread and the import price inflation variables turned out to be stationary whereas the other three variables were found to be $I(1)$ in the aforementioned studies. Therefore we shall focus on M3, GNP and the inflation rate in the following and investigate the number of cointegration relations among these variables.

We use quarterly, seasonally unadjusted data for the period 1975(1) to 1996(4) as in Lütkepohl & Wolters (1998). The initial period was chosen because the Bundesbank started its policy of monetary targeting in 1975. Specifically the following variables are used: m_t represents the logarithm of real M3, gnp_t is the logarithm of real GNP and p_t is the logarithm of the GNP deflator, hence, Δp_t is the inflation rate which will be used here. In other words, the vector of variables considered in the following cointegration analysis is $y_t = (m_t, gnp_t, \Delta p_t)'$. The data sources are given in Appendix B. The variables are plotted in Fig. 1. Obviously, m_t and gnp_t have clear shifts in the third quarter of 1990. These shifts are due to the German reunification. Note that, although the political reunification took place in October 1990, the monetary unification occurred already on July 1, 1990. Since then all variables refer to the unified Germany and, hence, the shift in the third quarter of 1990 is quite natural.

If the full sample is to be used in the cointegration tests, the structural shifts in 1990 have to be taken into account. Therefore Lütkepohl & Wolters (1998) and Wolters, Teräsvirta & Lütkepohl (1998) include a shift dummy $d_{1t} = 0$ until the second quarter of 1990 and $d_{1t} = 1$ afterwards. They also include an impulse dummy $d_{0t} = 1$ in 1990(3) and 0 elsewhere. We will use these dummy variables in our analysis as well. Because m_t and gnp_t potentially have a deterministic trend term we also include a linear trend as in our original model (2.1). In addition to the deterministic terms included in that model we also need seasonal dummy variables here because our data are not seasonally adjusted and have quite pronounced seasonal components (see Fig. 1). As noted in Johansen (1991), seasonal dummy variables do not have an impact on the asymptotic behaviour of LR cointegration tests and it is easy to see that the same is true for the LM type tests considered in the previous section. In the following we will estimate the parameters associated with the seasonal dummy variables by the GLS procedure described in Section 3 for the other parameters of the deterministic part of the model. Then y_t is adjusted for all deterministic terms including the seasonal terms to get $\hat{x}_t$.

In Table 1 the results of various cointegration tests are provided. They are based on models of order $p = 2$ which was also used by Wolters, Teräsvirta & Lütkepohl (1998). Therefore the first two values of our time series are used as presample values. In addition to tests for the full period, we also give results for the pre-unification period using data up

Table 1. Cointegration Tests for German Macroeconomic Data $(m_t, y_t, \Delta p_t)$.

	critical values[a]		1975(3) - 1990(2) ($T = 60$)		1975(3) - 1996(4) ($T = 86$)	
H_0	90%	95%	LR	LM	LR	LM
$r_0 = 0$	25.90	28.47	50.25	96.16	73.14	133.69
$r_0 = 1$	13.89	15.92	2.61	1.11	10.15	3.46
$r_0 = 2$	5.43	6.83	0.24	0.72	2.79	2.89

[a]) from Table 1 of L&S

to 1990(2) only. In these tests the dummies d_{it} $(i = 0, 1)$ are not included. Obviously, one cointegration relation is found regardless of the observation period. Moreover, for common significance levels both versions of the tests clearly reach the same conclusion regarding the cointegrating rank. Still it is pleasing that the new tests enable us to use the full sample information.

6 Conclusions

In this study we have proposed and applied tests for the cointegrating rank of a system of variables in the presence of structural shifts. Under the assumption that the break point is known a priori it is suggested to estimate the deterministic part of the DGP first, subtract the estimated deterministic part from the original series and then perform standard systems cointegration tests for the adjusted series. We have considered LR and LM type tests in this context and find that they have asymptotic null distributions which are tabulated elsewhere in the literature and do not depend on the break point. Hence, the tests are conveniently applied without the need for simulating new critical values. For illustrative purposes, the tests are applied to a system of German macroeconomic variables which may be thought of as a money demand system.

Appendix A. Proofs

A.1 Proof of Lemma 3.1

We shall first demonstrate that assuming the true values of μ_i and δ_i, denoted by $\mu_i^{(0)}$ and $\delta_i^{(0)}$, respectively, to be zero ($i = 0, 1$) does not imply a loss of generality. To this end, write for $t \geq p + 1$,

$$\begin{aligned}
&A(L)(y_t - \mu_0 - \mu_1 t - \delta_0 d_{0t} - \delta_1 d_{1t}) \\
&\quad = A(L)[x_t - (\mu_0 - \mu_0^{(0)}) - (\mu_1 - \mu_1^{(0)})t - (\delta_0 - \delta_0^{(0)})d_{0t} - (\delta_1 - \delta_1^{(0)})d_{1t}] \\
&\quad = \Delta x_t - \alpha\beta' x_{t-1} - \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} \\
&\qquad + \alpha\beta'(\mu_0 - \mu_0^{(0)}) + \alpha\beta'(\mu_1 - \mu_1^{(0)})(t-1) - \Psi(\mu_1 - \mu_1^{(0)}) \\
&\qquad - (\delta_0 - \delta_0^{(0)})d_{0t} + \sum_{j=1}^{p} A_j(\delta_0 - \delta_0^{(0)})d_{0,t-j} \\
&\qquad - (\delta_1 - \delta_1^{(0)})\Delta d_{1t} + \alpha\beta'(\delta_1 - \delta_1^{(0)})d_{1,t-1} + \sum_{j=1}^{p-1} \Gamma_j(\delta_1 - \delta_1^{(0)})\Delta d_{1,t-j}.
\end{aligned}$$

This shows that by a linear reparameterization we can transform (2.11) to a form in which y_t is replaced by x_t and the true values of all parameters related to deterministic variables are zero. This reparameterization is of course infeasible but it can be used in theoretical derivations and, in particular, to study asymptotic properties of RR regression estimators of α, β, Γ_j and Ω. Thus, we can proceed by assuming that the true values of μ_i and δ_i are zero so that in (2.11) the true values of ν, τ, θ and γ_{ij} are all zero.

Next we introduce some notation. Define

$$\begin{aligned}
X_t &= \left[y'_{t-1} : t-1 : d_{1,t-1}\right]', \\
Z_{1t} &= \left[1 : \Delta y'_{t-1} : \cdots : \Delta y'_{t-p+1}\right]', \\
Z_{2t} &= [d_{0t} : \cdots : d_{0,t-p} : \Delta d_{1t} : \cdots : \Delta d_{1,t-p+1}]',
\end{aligned}$$

and set $Z_t = [Z'_{1t} : Z'_{2t}]'$. Then we can write equation (2.11) as

$$\Delta y_t = \alpha\psi' X_t + \Phi Z_t + \varepsilon_t, \quad t = p+1, p+2, ..., \tag{A.1}$$

where $\psi' = [\beta' : -\tau : -\theta]$ and $\Phi = [\Phi_1 : \Phi_2]$ with $\Phi_1 = [\nu : \Gamma_1 : \cdots : \Gamma_{p-1}]$ and $\Phi_2 = [\gamma_{00} : ... : \gamma_{0p} : \gamma_{10} : ... : \gamma_{1,p-1}]$. The RR regression estimators of α, ψ and Ω can be obtained as follows. Define

$$S_{ij} = N^{-1} \sum_{t=p+1}^{T} R_{it} R'_{jt}, \quad i, j = 0, 1,$$

where $N = T - p$ and R_{0t} and R_{1t} are the LS residuals obtained by regressing Δy_t and X_t on Z_t, respectively. As is well known, the RR regression estimator of ψ is based on the eigenvectors corresponding to the r largest eigenvalues of the determinantal equation

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0 \tag{A.2}$$

(e.g., Johansen (1995, Chapter 6) or Lütkepohl (1991, Appendix A.14)). When the RR regression estimator of ψ is available those of α and Ω are obtained by replacing ψ by its estimator in the formulas $S_{01}\psi(\psi' S_{11}\psi)^{-1}$ and $S_{00} - S_{01}\psi(\psi' S_{11}\psi)^{-1}\psi' S_{10}$, respectively. Recall that we have assumed that all impulse dummies in (2.11) are linearly independent so that S_{00} is nonsingular and the above estimators are well defined.

In the same way as in the proof of Lemma 13.1 of Johansen (1995) we now transform the equation (A.2). To this end, we define the matrix

$$A_T = \begin{bmatrix} \beta & T^{-1/2}\beta_\perp(\beta_\perp'\beta_\perp)^{-1} & 0 & 0 \\ 0 & 0 & T^{-1} & 0 \\ 0 & 0 & 0 & T^{1/2}/(T-T_1)^{1/2} \end{bmatrix}.$$

Pre- and postmultiplying (A.2) by A_T' and A_T, respectively, gives

$$\det(\lambda A_T' S_{11} A_T - A_T' S_{10} S_{00}^{-1} S_{01} A_T) = 0. \tag{A.3}$$

The eigenvalues of (A.2) and (A.3) are identical and the eigenvectors of (A.3) are obtained from those of (A.2) by premultiplying by A_T^{-1}. The next step is to study the weak limit of (A.3). For ease of exposition and without loss of generality, we shall assume that the initial values of x_t are such that $\beta' x_t$ and Δx_t are stationary.

We shall demonstrate next that

$$S_{00} = \tilde{S}_{00} + o_p(1) \quad \text{and} \quad S_{01}A_T = \tilde{S}_{01}A_T + o_p(1), \tag{A.4}$$

where $\tilde{S}_{00}$ and $\tilde{S}_{01}$ are analogs of S_{00} and S_{01}, respectively, obtained by omitting Z_{2t} from (A.1) (i.e., $Z_t = Z_{1t}$). To show (A.4), denote $D_T^{-1} = \text{diag}[N^{-1/2}I : I]$ where the partition is conformable to that of $Z_t = [Z_{1t}', Z_{2t}']'$. Then note that

$$D_T^{-1} \sum_{t=p+1}^{T} Z_t Z_t' D_T^{-1} \xrightarrow{p} \text{diag}[E(Z_{1t}Z_{1t}') : I], \tag{A.5}$$

where the expectation is well defined because $y_t = x_t$ by assumption. This result is a straightforward consequence of the definitions and well-known properties of stationary processes. Notice that $(A.5)$ holds even if the time points T_0 and T_1 are not fixed because $T^{-1/2} \max_{1 \leq t \leq T} \Delta x_t = o_p(1)$. The same line of reasoning also gives

$$N^{-1/2} \sum_{t=p+1}^{T} \Delta y_t Z_t' D_T^{-1} \xrightarrow{p} [E(\Delta x_t Z_{1t}') : 0]. \tag{A.6}$$

Because

$$S_{00} = N^{-1} \sum_{t=p+1}^{T} \Delta y_t \Delta y_t' - N^{-1} \sum_{t=p+1}^{T} \Delta y_t Z_t' \left(\sum_{t=p+1}^{T} Z_t Z_t' \right)^{-1} \sum_{t=p+1}^{T} Z_t \Delta y_t', \tag{A.7}$$

the first equality of $(A.4)$ follows from $(A.5)$ and $(A.6)$.

To justify the second equality in $(A.4)$, we consider the matrix $N^{-1/2} D_T^{-1} \sum_{t=p+1}^{T} Z_t X_t' A_T$, where $X_t' A_T$ consists of four components analyzed separately below. First note that

$$N^{-1/2} D_T^{-1} \sum_{t=p+1}^{T} Z_t y_{t-1}' \beta \xrightarrow{p} \begin{bmatrix} E(Z_{1t} x_t' \beta) \\ 0 \end{bmatrix}, \tag{A.8}$$

where the convergence follows in the same way as in $(A.6)$. Next we have

$$N^{-1} D_T^{-1} \sum_{t=p+1}^{T} Z_t y_{t-1}' \beta_\perp = \begin{bmatrix} N^{-3/2} \sum_{t=p+1}^{T} Z_{1t} x_{t-1}' \beta_\perp \\ o_p(1) \end{bmatrix}, \tag{A.9}$$

where the order result follows from the definition of Z_{2t} and the well-known fact that $T^{-1} \max_{1 \leq t \leq T} \beta_\perp' x_t = o_p(1)$. Similarly,

$$N^{-3/2} D_T^{-1} \sum_{t=p+1}^{T} (t-1) Z_t = \begin{bmatrix} N^{-2} \sum_{t=p+1}^{T} (t-1) Z_{1t} \\ o(1) \end{bmatrix}. \tag{A.10}$$

Finally, we have

$$\begin{aligned} (T - T_1)^{-1/2} D_T^{-1} \sum_{t=p+1}^{T} Z_t d_{1,t-1} &= \begin{bmatrix} (T - T_1)^{-1/2} N^{-1/2} \sum_{t=T_1+1}^{T} Z_{1t} \\ (T - T_1)^{-1/2} \sum_{t=T_1+1}^{T} Z_{2t} \end{bmatrix} \\ &= \begin{bmatrix} (T - T_1)^{-1/2} N^{-1/2} \sum_{t=T_1+1}^{T} Z_{1t} \\ O(1) \end{bmatrix}, \end{aligned} \tag{A.11}$$

where the term $O(1)$ converges either to zero (if $T - T_1 \longrightarrow \infty$) or to a positive constant (if $T - T_1 \longrightarrow c < \infty$). Now, since $S_{01}A_T$ is obtained by replacing $\Delta y_t'$ with $X_t'A_T$ in $(A.7)$, the second equality in $(A.4)$ can be obtained from $(A.5)$ - $(A.11)$.

From $(A.4)$ we can conclude that the asymptotic behaviour of the latter matrix in $(A.3)$ is similar to that in the conventional model without any dummies. Thus, in the same way as in Johansen (1995, p. 158, 180) we find that

$$A_T' S_{10} S_{00}^{-1} S_{01} A_T \xrightarrow{p} \begin{bmatrix} \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix}, \qquad (A.12)$$

where Σ_{00} and $\Sigma_{0\beta}$ are conditional covariance matrices defined by

$$\mathrm{Cov}\left(\begin{bmatrix} \Delta x_t \\ \beta' x_t \end{bmatrix} | \Delta x_{t-1}, ..., \Delta x_{t-p+1} \right) = \begin{bmatrix} \Sigma_{00} & \Sigma_{0\beta} \\ \Sigma_{\beta 0} & \Sigma_{\beta\beta} \end{bmatrix}.$$

Next we have to consider the asymptotic behaviour of the matrix $A_T' S_{11} A_T$. For this purpose we note that

$$A_T' X_t = \begin{bmatrix} \beta' y_{t-1} \\ T^{-1/2}(\beta_\perp'\beta_\perp)^{-1}\beta_\perp' y_{t-1} \\ T^{-1}(t-1) \\ \left(\frac{T}{T-T_1}\right)^{1/2} d_{1,t-1} \end{bmatrix}, \qquad (A.13)$$

where $y_t = x_t$ can be assumed. It appears convenient to analyze separately the cases where $a_1 < 1$ and $a_1 = 1$ in (2.8).

<u>$a_1 < 1$</u>

If $a_1 < 1$ then the term $O(1)$ in $(A.11)$ can be changed to $o(1)$ and we have $A_T' S_{11} A_T = A_T' \tilde{S}_{11} A_T + o_p(1)$ with $\tilde{S}_{11}$ defined by omitting Z_{2t} in the same way as in $\tilde{S}_{00}$ and $\tilde{S}_{01}$. Because we may assume that $y_t = x_t$, the weak limit of $A_T' \tilde{S}_{11} A_T$ can be obtained by arguments similar to those used in the conventional case (see Johansen (1995, p. 158, 180)). It is first easy to check that the matrix $A_T' \tilde{S}_{11} A_T$ is asymptotically block diagonal with the $(r \times r)$ block in the upper left hand corner converging in probability to $\Sigma_{\beta\beta}$. To obtain the weak limit of the lower right hand corner, let $\mathbf{W}(s)$ be an n-dimensional Brownian motion with covariance matrix Ω, denote by $I(\cdot)$ the indicator function and let $[c]$ be the integer part of c. From the

multivariate invariance principle we then find that

$$\begin{bmatrix} T^{-1/2}\beta_\perp' x_{[Ts]-1} \\ T^{-1}([Ts]-1) \\ (\frac{T}{T-T_1})^{1/2} d_{1,[Ts]-1} \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \beta_\perp' C\mathbf{W}(s) \\ s \\ (1-a_1)^{-1/2} I(s \geq a_1) \end{bmatrix} \stackrel{def}{=} \mathbf{G}(s), \tag{A.14}$$

where convergence is in the Skorohod topology of $D[0,1]$. Hence, we can conclude that

$$\lambda A_T' S_{11} A_T \xrightarrow{d} \begin{bmatrix} \lambda\Sigma_{\beta\beta} & 0 \\ 0 & \lambda \int_0^1 \bar{\mathbf{G}}(s)\bar{\mathbf{G}}(s)' ds \end{bmatrix}, \tag{A.15}$$

where $\bar{\mathbf{G}}(s) = \mathbf{G}(s) - \int_0^1 \mathbf{G}(u)du$ (cf. Johansen (1995, p. 158 and 180)).

Now, let $\tilde{\psi} = [\tilde{\beta}_\xi', -\tilde{\tau}, -\tilde{\theta}]'$ be the (normalized) RR regression estimator of ψ described above $(A.2)$. In the same way as in Johansen (1995, p. 180) we can then use $(A.12)$ and $(A.15)$ to conclude that

$$A_T^{-1}\tilde{\psi} = \begin{bmatrix} I_r \\ T^{1/2}\beta_\perp'\tilde{\beta}_\xi \\ -T\tilde{\tau}' \\ -(\frac{T-T_1}{T})^{1/2}\tilde{\theta}' \end{bmatrix} = \begin{bmatrix} I_r \\ o_p(1) \\ o_p(1) \\ o_p(1) \end{bmatrix}, \tag{A.16}$$

where the first equality follows from the definitions of A_T and $\tilde{\psi}$. This shows that the estimators $\tilde{\beta}_\xi, \tilde{\tau}$ and $\tilde{\theta}$ are consistent of orders $o_p(T^{-1/2}), o_p(T^{-1})$ and $o_p(1)$, respectively. (Recall that the true values of τ and θ are zero.) From these results it is further straightforward to show that $\tilde{\alpha}_\xi, \tilde{\Gamma}_j$ and $\tilde{\Omega}$ are also consistent (cf. Johansen (1995, p. 181)). Before strengthening the above consistency results to the required form we shall obtain similar intermediate results in the case $a_1 = 1$.

$\underline{a_1 = 1}$

Recall that $(A.12)$ holds even if $a_1 = 1$ but the previous derivation of $(A.15)$ fails because the limit in $(A.14)$ is not defined. To find the weak limit of $A_T' S_{11} A_T$ in this case, we consider again $(A.13)$ and note that the second sample moments between the last component of $A_T' X_t$ and the remaining ones converge to zero in probability. To demonstrate this, note for

example that

$$\begin{aligned} E\left\| N^{-1}(T-T_1)^{-1/2}\sum_{t=p+1}^{T}\beta_\perp' y_{t-1}d_{1,t-1}\right\| &\leq N^{-1}(T-T_1)^{-1/2}\sum_{t=T_1+1}^{T}E\|\beta_\perp' x_{t-1}\| \\ &\leq \text{const}\cdot N^{-1}(T-T_1)^{-1/2}\sum_{t=T_1+1}^{T}t^{1/2} \\ &\leq \text{const}\cdot N^{-1}(T-T_1)^{-1/2}T^{1/2}(T-T_1) \\ &\leq \text{const}\cdot(\tfrac{T-T_1}{T})^{1/2}=o(1) \end{aligned}$$

when $T_1/T \to 1$. This shows that the second sample moments between the second and last components of $A_T'X_t$ are of order $o_p(1)$. The other cases can also be handled in a straightforward manner so that the arguments used to obtain $(A.15)$ now yield

$$N^{-1}\sum_{t=p+1}^{T}A_T'X_tX_t'A_T \xrightarrow{d} \begin{bmatrix} \Sigma_{\beta\beta} & 0 & 0 \\ 0 & \int_0^1 \mathbf{G}_1(s)\mathbf{G}_1(s)'ds & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{A.17}$$

where $\mathbf{G}_1(s)$ $((n-r+1)\times 1)$ is defined by the first $n-r+1$ components of $\mathbf{G}(s)$. Using $(A.5)$ and $(A.8)$ - $(A.11)$ we next note that

$$\begin{aligned} &N^{-1}\sum_{t=p+1}^{T}A_T'X_tZ_t'\left(\sum_{t=p+1}^{T}Z_tZ_t'\right)^{-1}\sum_{t=p+1}^{T}Z_tX_t'A_T \\ &\quad= N^{-1}\sum_{t=p+1}^{T}A_T'X_tZ_{1t}'\left(\sum_{t=p+1}^{T}Z_{1t}Z_{1t}'\right)^{-1}\sum_{t=p+1}^{T}Z_{1t}X_t'A_T \\ &\qquad+N^{-1}\sum_{t=p+1}^{T}A_T'X_tZ_{2t}'\sum_{t=p+1}^{T}Z_{2t}X_t'A_T+o_p(1). \end{aligned} \tag{A.18}$$

From $(A.8)$ - $(A.11)$ it can be seen that the elements of the second matrix on the r.h.s. converge to zero in probability except the one in the lower right hand corner which converges to c_1^2, say, the limit of the squared norm of the term $O(1)$ in the last expression of $(A.11)$. On the other hand, arguments similar to those used to show the asymptotic block diagonality of the term on the l.h.s. of $(A.17)$ can be used to show that the elements in the last row and last column of the first term on the r.h.s. of $(A.18)$ converge to zero in probability while the remaining elements converge in probability to the same limits as in a model without any dummy variables. Because $A_T'S_{11}A_T$ is the difference between the l.h.s. of $(A.17)$ and the l.h.s. of $(A.18)$ the above discussion as a whole in conjunction with $(A.17)$ shows that

$$\lambda A_T'S_{11}A_T \xrightarrow{d} \begin{bmatrix} \lambda\Sigma_{\beta\beta} & 0 & 0 \\ 0 & \lambda\int_0^1 \bar{\mathbf{G}}_1(s)\bar{\mathbf{G}}_1(s)'ds & 0 \\ 0 & 0 & \lambda(1-c_1^2) \end{bmatrix}, \tag{A.19}$$

where $\bar{\mathbf{G}}_1(s) = \mathbf{G}_1(s) - \int_0^1 \mathbf{G}_1(u)du$. (Notice that $1 - c_1^2 > 0$ because it is the limit of the residual variance in a regression of $[T/(T-T_1)]^{1/2} d_{1,t-1}$ on Z_{2t} and because these variables are linearly independent by assumption.) Thus, since we still have $(A.12)$ the only difference between the present case ($a_1 = 1$) and the previous one ($a_1 < 1$) is that the weak limit of $\lambda A_T' S_{11} A_T$ differs from its previous counterpart obtained in $(A.15)$. This, however, does not affect the arguments used to derive $(A.16)$ which therefore hold in the present context as well. This implies that $\tilde{\beta}_\xi$ and $\tilde{\tau}$ are consistent of the same orders as previously but, since $T_1/T \to 1$ nothing can be concluded about the consistency of $\tilde{\theta}$. Fortunately, however, the consistency of $\tilde{\alpha}_\xi, \tilde{\Gamma}_j$ and $\tilde{\Omega}$ can still be proved in the same way as before. To see this, note that $\tilde{\alpha}_\xi$ and $\tilde{\Gamma}_j$ can be obtained by LS from the auxiliary regression model

$$\Delta y_t = \alpha \tilde{\psi}_T' A_T' X_t + \nu + \sum_{j=1}^{p-1} \Gamma_i \Delta y_{t-j} + \Phi_2 Z_{2t} + \text{error}, \qquad t = p+1, \ldots, T, \tag{A.20}$$

where $\tilde{\psi}_T = A_T^{-1}\tilde{\psi}$. Since $A_T^{-1}\psi = [I_r : 0]'$, $\tilde{\psi}_T$ is a consistent estimator of $A_T^{-1}\psi$ [see $(A.16)$]. This consistency result and our previous results about the second sample moments of the variables in the auxiliary regression model $(A.20)$ can be used to prove the consistency of the estimators $\tilde{\alpha}_\xi$ and $\tilde{\Gamma}_j$ (and also $\tilde{\nu}$). After this the consistency of $\tilde{\Omega}$ can be proved in a standard fashion. Details of these derivations are straightforward and therefore omitted.

To complete the proof we have to establish the stated orders of consistency. Define

$$U_T = \begin{bmatrix} T\beta_\perp' \tilde{\beta}_\xi \\ T^{2/3}\tilde{\tau}' \\ (T-T_1)^{1/2}\tilde{\theta}' \end{bmatrix} \qquad ((n-r+2) \times r).$$

We shall first show that $U_T = O_p(1)$ which implies that $T(\tilde{\beta}_\xi - \beta) = O_p(1)$. Define the $((n+2) \times (n-r+2))$ matrix

$$B_T = \begin{bmatrix} \beta_\perp(\beta_\perp'\beta_\perp)^{-1} & 0 & 0 \\ 0 & T^{-1/2} & 0 \\ 0 & 0 & \frac{T}{(T-T_1)^{1/2}} \end{bmatrix}$$

and notice that the matrix A_T can be written as $A_T = [b : T^{-1/2} B_T]$ with $b = [\beta' \quad 0 \quad 0]'$. Since U_T is formed by the last $n-r+2$ rows of $T^{1/2} A_T^{-1}(\tilde{\psi} - \psi)$ it follows from the above definitions that (cf. Johansen, 1995, p. 179)

$$\tilde{\psi} - \psi = T^{-1} B_T U_T. \tag{A.21}$$

As in Johansen (1995, p. 182) one obtains the first order conditions for $\tilde{\psi}$ as

$$\tilde{\alpha}_\xi' \tilde{\Omega}^{-1}(S_{01} - \tilde{\alpha}_\xi \tilde{\psi}' S_{11}) = 0.$$

Define next

$$S_{\varepsilon 1} = T^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_t R_{1t}' = S_{01} - \alpha \psi' S_{11}$$

where $\hat{\varepsilon}_t$ is the LS residual in the regression of ε_t on Z_t. Thus, the first order conditions can be expressed as (cf. Johansen (1995, p. 182))

$$0 = \tilde{\alpha}_\xi' \tilde{\Omega}^{-1}(S_{\varepsilon 1} - \tilde{\alpha}_\xi(\tilde{\psi} - \psi)' S_{11} - (\tilde{\alpha}_\xi - \alpha)\psi' S_{11})$$

or, postmultiplying by B_T and using $(A.21)$,

$$0 = \tilde{\alpha}_\xi' \tilde{\Omega}^{-1}(S_{\varepsilon 1} B_T - \tilde{\alpha}_\xi U_T' [T^{-1} B_T' S_{11} B_T] - (\tilde{\alpha}_\xi - \alpha)\psi' S_{11} B_T). \qquad (A.22)$$

Recall that we can assume that $\psi' X_t$ is zero mean stationary and the matrix $T^{-1/2} B_T$ is formed by the last $n - r + 2$ columns of A_T. Using these two facts it is straightforward to check that $\psi' S_{11} B_T = O_p(1)$ both when $a_1 < 1$ and $a_1 = 1$ in (2.8). Thus by the consistency of $\tilde{\alpha}_\xi$ the last term on the right-hand side of $(A.22)$ is of order $O_p(1)$. By similar arguments one can show that $S_{\varepsilon 1} B_T = O_p(1)$ (cf. the derivation of $(A.12)$), so that $U_T = O_p(1)$ follows if the matrix in brackets in $(A.22)$ converges weakly to a positive definite limit. This, however, follows from $(A.15)$ (when $a_1 < 1$) and $(A.17)$ (when $a_1 = 1$) and the definition of the matrix B_T. Thus we have proved that $\tilde{\beta}_\xi = \beta + O_p(T^{-1})$.

Since $U_T = O_p(1)$ implies that $\tilde{\psi}_T$ in $(A.20)$ is a consistent estimator of $A_T^{-1}\psi$ of order $O_p(T^{-1/2})$ the remaining orders of consistency are straightforward to establish by considering the LS estimators obtained from $(A.20)$. First note that here the vector Z_{2t} can be ignored and that using arguments similar to those previously used it can be seen that the matrix of second sample moments of the other regressors converges in probability to a positive definite limit. Thus, it suffices to check that the second sample moments between the regressors and the error term are of order $O_p(N^{-1/2})$. Since the error term equals $\varepsilon_t - \alpha(\tilde{\psi}_T - A_T^{-1}\psi)' A_T X_t$, where $\tilde{\psi}_T - A_T \psi = O_p(T^{-1/2})$ this can be readily seen. This completes the proof of Lemma 3.1. □

Remark A.1 (i) From the proof it is also possible to derive the limiting distributions of $\tilde{\alpha}_\xi$ and $\tilde{\beta}_\xi$ and thereby similar results for other normalizations. The limiting distribution of

$\tilde{\beta}_\xi$ is mixed normal but its explicit form differs for the cases $a_1 < 1$ and $a_1 = 1$ (cf. $(A.15)$ and $(A.17)$).

(ii) From the proof it can also be seen that $\tilde{\tau}-\tau = O_p(T^{-3/2})$ and $\tilde{\theta}-\theta = O_p((T-T_1)^{-1/2})$. Since $\tau = \beta'\mu_1$ and $\theta = \beta'\delta_1$ these results may be viewed as RR analogs of (3.9) and (3.7).

A.2 Proof of Theorem 3.1

Since all relevant quantities will be invariant to normalizations of $\tilde{\alpha}$ and $\tilde{\beta}$ we can assume that some kind of normalization has been imposed so that, by Lemma 3.1, we can also assume that $\tilde{\alpha} = \alpha + O_p(T^{-1/2})$ and $\tilde{\beta} = \beta + O_p(T^{-1})$.

Using the definitions we first note that explicit forms of the variables $\tilde{H}_{it}$ and $\tilde{K}_{it}$ $(i = 0, 1)$ in (3.3) are given by

$$\tilde{H}_{0t} = \begin{cases} I_n, & t = 1 \\ I_n - \sum_{j=1}^{t-1} \tilde{A}_j, & t = 2, \ldots, p \\ -\tilde{\alpha}\tilde{\beta}', & t = p+1, \ldots, T \end{cases},$$

$$\tilde{H}_{1t} = \begin{cases} I_n, & t = 1 \\ tI_n - \sum_{j=1}^{t-1}(t-j)\tilde{A}_j, & t = 2, \ldots, p \\ \tilde{\Psi} - (t-1)\tilde{\alpha}\tilde{\beta}', & t = p+1, \ldots, T \end{cases},$$

$$\tilde{K}_{0t} = \begin{cases} 0, & t < T_0 \\ I_n, & t = T_0 \\ -\tilde{A}_j, & t = T_0 + j, \quad j = 1, \ldots, p \\ 0, & t = T_0 + p + 1, \ldots, T \end{cases},$$

$$\tilde{K}_{1t} = \begin{cases} 0, & t < T_1 \\ I_n, & t = T_1 \\ I_n - \sum_{j=1}^{t-T_1} \tilde{A}_j, & t = T_1 + 1, \ldots, T_1 + p - 1 \\ -\tilde{\alpha}\tilde{\beta}', & t = T_1 + p, \ldots, T \end{cases}.$$

The proof is based on ideas similar to those used in the proof of Theorem 1 of S&L.

Define

$$\underset{\sim}{\gamma}_1 = \begin{bmatrix} \underset{\sim}{\gamma}_{11} \\ \underset{\sim}{\gamma}_{12} \\ \underset{\sim}{\gamma}_{13} \end{bmatrix} = \begin{bmatrix} \delta_0 \\ \tilde{\beta}_\perp'\delta_1 \\ \tilde{\beta}_\perp'\mu_0 \end{bmatrix}$$

and

$$\underline{\gamma}_2 = \begin{bmatrix} \underline{\gamma}_{21} \\ \underline{\gamma}_{22} \\ \underline{\gamma}_{23} \\ \underline{\gamma}_{24} \end{bmatrix} = \begin{bmatrix} \tilde{\beta}'\delta_1 \\ \tilde{\beta}'\mu_0 \\ \tilde{\beta}'\mu_1 \\ \tilde{\beta}'_\perp\mu_1 \end{bmatrix}.$$

We shall first obtain asymptotic properties of the LS estimator of the "parameters" $\underline{\gamma}_1$ and $\underline{\gamma}_2$ and then use Theorem 1 of S&L to derive the stated results. To express (3.3) in terms of $\underline{\gamma}_1$ and $\underline{\gamma}_2$ we transform the variables $\tilde{H}_{it}$ and $\tilde{K}_{it}$ accordingly and define

$$\begin{aligned}
\tilde{F}_{11t} &= \tilde{Q}'\tilde{K}_{0t}, \\
\tilde{F}_{12t} &= \tilde{Q}'\tilde{K}_{1t}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}, \\
\tilde{F}_{13t} &= \tilde{Q}'\tilde{H}_{0t}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{F}_{21t} &= \tilde{Q}'\tilde{K}_{1t}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}, \\
\tilde{F}_{22t} &= \tilde{Q}'\tilde{H}_{0t}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}, \\
\tilde{F}_{23t} &= \tilde{Q}'\tilde{H}_{1t}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}, \\
\tilde{F}_{24t} &= \tilde{Q}'\tilde{H}_{1t}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}.
\end{aligned}$$

Then setting $\tilde{F}_{1t} = [\tilde{F}_{11t} : \tilde{F}_{12t} : \tilde{F}_{13t}]$ and $\tilde{F}_{2t} = [\tilde{F}_{21t} : \tilde{F}_{22t} : \tilde{F}_{23t} : \tilde{F}_{24t}]$ we can write (3.3) as

$$\tilde{d}_t := \tilde{Q}'\tilde{A}(L)y_t = \tilde{F}_{1t}\underline{\gamma}_1 + \tilde{F}_{2t}\underline{\gamma}_2 + \eta_t, \quad t = 1, \ldots, T. \tag{A.23}$$

From the definitions it follows that $\tilde{F}_{1t}$ takes nonzero values only for a fixed number of time indices t.

We shall next study the sums of cross products between $\tilde{F}_{1t}$ and the error term η_t which is identical to its counterpart in S&L. Thus,

$$\eta_t = \tilde{Q}'\varepsilon_t - \tilde{Q}'\tilde{\alpha}(\tilde{\beta} - \beta)'x_{t-1} - \tilde{Q}'(\tilde{\alpha} - \alpha)\tilde{\beta}'x_{t-1} - \tilde{Q}'\sum_{j=1}^{p-1}(\tilde{\Gamma}_j - \Gamma_j)\Delta x_{t-j}. \tag{A.24}$$

Using this expression, Lemma 3.1 and the above mentioned property of $\tilde{F}_{1t}$ it can be seen that

$$\sum_{t=1}^{T}\tilde{F}'_{1t}\eta_t = O_p(1). \tag{A.25}$$

For $\tilde{F}_{2t}$ the corresponding result is

$$\Upsilon_T^{-1}\sum_{t=1}^{T}\tilde{F}'_{2t}\eta_t = O_p(1), \tag{A.26}$$

where $\Upsilon_T^{-1} = \text{diag}[(T-T_1)^{-1/2}I : T^{-1/2}I : T^{-3/2}I : T^{-1/2}I]$ and the partition is conformable to that of $\tilde{F}_{2t}$. To justify $(A.26)$, note first that for $\tilde{F}_{22t}$, $\tilde{F}_{23t}$, $\tilde{F}_{24t}$ the result is obtained directly from $(A.5)$, $(A.6)$ and $(A.7)$ of S&L, respectively. When (2.8) holds with $a_1 < 1$, $(A.26)$ can be justified for $\tilde{F}_{21t}$ in the same way as for $\tilde{F}_{22t}$ because $\tilde{F}_{21t} = \tilde{F}_{22,t-T_1+1}$ for $t \geq T_1$ and zero for $t = 1, ..., T_1 - 1$. If (2.8) holds with $a_1 = 1$, the desired result can be obtained by using arguments similar to those used to obtain $(A.17)$.

The next step is to study the moment matrix related to the auxiliary regression model $(A.23)$. By the definitions and Lemma 3.1 it is clear that $\sum_{t=1}^{T} \tilde{F}_{1t}'\tilde{F}_{1t}$ converges in probability. The limit is block diagonal because $\tilde{F}_{13t}$ is orthogonal to $\tilde{F}_{11t}$ and $\tilde{F}_{12t}$. The lower right hand block, related to $\tilde{F}_{13t}$, is positive definite, as discussed in S&L. That the upper left hand block is positive definite can also be seen easily. If $T_0 \neq T_1$ this is obvious by the definitions of $\tilde{F}_{11t}$ and $\tilde{F}_{12t}$. If $T_0 = T_1$ and $A_p \neq 0$ the same conclusion is again obvious. If $A_p = 0$ but $A_{p-1} \neq 0$ the situation is reduced to the preceding one so that repeating this argument it can be seen that the upper left hand block of $\sum_{t=1}^{T} \tilde{F}_{1t}'\tilde{F}_{1t}$ and hence the whole matrix converges in probability to a positive definite limit.

To analyze the other parts of the above mentioned moment matrix, we first assume that $a_1 < 1$ in (2.8). Using the definitions and Lemma 3.1 it is then straightforward to check that $\Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{F}_{2t}'\tilde{F}_{1t} = o_p(1)$ so that, upon appropriate standardization, the moment matrix is asymptotically block diagonal. In order to study the matrix $\Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{F}_{2t}'\tilde{F}_{2t}\Upsilon_T^{-1}$ we first conclude from the definitions that, for $t \geq p+1$, $\tilde{F}_{22t} = -\tilde{Q}'\tilde{\alpha}$, $\tilde{F}_{23t} = \tilde{Q}'\tilde{\Psi}\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1} - (t-1)\tilde{Q}'\tilde{\alpha}$ and $\tilde{F}_{24t} = \tilde{Q}'\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}$. As in S&L we can proceed by assuming that the above expressions of $\tilde{F}_{22t}$ and $\tilde{F}_{24t}$ hold for all t and also that $\tilde{F}_{23t}$ is replaced by $-(t-1)\tilde{Q}'\tilde{\alpha}$. Furthermore, since $\tilde{F}_{21t} = 0$ for $t < T_1$ and $\tilde{F}_{21t} = -\tilde{Q}'\tilde{\alpha}d_{1t}$ for $t \geq T_1 + p$, we can act as if $\tilde{F}_{21t} = -\tilde{Q}'\tilde{\alpha}d_{1t}$ would hold for all t. Thus, we have

$$\Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{F}_{2t}'\tilde{F}_{2t}\Upsilon_T^{-1} = \sum_{t=1}^{T} \begin{bmatrix} -(T-T_1)^{-1/2}\tilde{\alpha}'\tilde{Q}d_{1t} \\ -T^{-1/2}\tilde{\alpha}'\tilde{Q} \\ -T^{-3/2}(t-1)\tilde{\alpha}'\tilde{Q} \\ -T^{-1/2}(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{Q} \end{bmatrix} \begin{bmatrix} -(T-T_1)^{-1/2}\tilde{\alpha}'\tilde{Q}d_{1t} \\ -T^{-1/2}\tilde{\alpha}'\tilde{Q} \\ -T^{-3/2}(t-1)\tilde{\alpha}'\tilde{Q} \\ -T^{-1/2}(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\beta}_\perp'\tilde{\Psi}'\tilde{Q} \end{bmatrix}' + o_p(1)$$

$$= \begin{bmatrix} E_T \otimes \tilde{A}_{11} & e_T \otimes \tilde{A}_{12} \\ e_T' \otimes \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} + o_p(1). \qquad (A.27)$$

Here the notation is defined as follows. First, as in S&L,

$$\begin{aligned}\tilde{A}_{11} &= \tilde{\alpha}'\tilde{\Omega}\tilde{\alpha} \\ \tilde{A}_{12} &= \tilde{A}'_{21} = \tilde{\alpha}'\tilde{\Omega}^{-1}\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}\end{aligned}$$

and

$$\begin{aligned}\tilde{A}_{22} &= (\tilde{\beta}'_\perp\tilde{\beta})^{-1}\tilde{\beta}'_\perp\tilde{\Psi}'\tilde{\Omega}^{-1}\tilde{\Psi}\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1} \\ &= \tilde{A}_{21}\tilde{A}_{11}^{-1}\tilde{A}_{12} + \tilde{B}^{-1},\end{aligned}$$

where

$$\begin{aligned}\tilde{B} &= \tilde{\beta}'_\perp\tilde{\beta}_\perp(\tilde{\alpha}'_\perp\tilde{\Psi}\tilde{\beta}_\perp)^{-1}\tilde{\alpha}'_\perp\tilde{\Omega}\tilde{\alpha}_\perp(\tilde{\beta}'_\perp\tilde{\Psi}'\tilde{\alpha}_\perp)^{-1}\tilde{\beta}'_\perp\tilde{\beta}_\perp \\ &= \tilde{\beta}'_\perp\tilde{C}\tilde{\Omega}\tilde{C}'\tilde{\beta}_\perp\end{aligned}$$

with $\tilde{C} = \tilde{\beta}_\perp(\tilde{\alpha}'_\perp\tilde{\Psi}\tilde{\beta}_\perp)^{-1}\tilde{\alpha}'_\perp$. The matrix E_T is defined by

$$E_T = \begin{bmatrix} 1 & (\frac{T-T_1}{T})^{1/2} & T^{-3/2}(T-T_1)^{-1/2}\sum_{t=T_1+1}^{T}(t-1) \\ (\frac{T-T_1}{T})^{1/2} & 1 & T^{-2}\sum_{t=1}^{T}(t-1) \\ T^{-3/2}(T-T_1)^{-1/2}\sum_{t=T_1+1}^{T}(t-1) & T^{-2}\sum_{t=1}^{T}(t-1) & T^{-3}\sum_{t=1}^{T}(t-1)^2 \end{bmatrix},$$

while the vector e_T is the second column of E_T, that is,

$$e_T = \left[\left(\frac{T-T_1}{T}\right)^{1/2} : 1 : T^{-2}\sum_{t=1}^{T}(t-1)\right]'.$$

It is not difficult to check that the matrix E_T is nonsingular and has a nonsingular limit. From the definitions it also follows that $e'_T E_T^{-1} = [0 : 1 : 0]$ and $e'_T E_T^{-1} e_T = 1$. Using these facts and Lemma 3.1 it is straightforward to show that the matrix in $(A.27)$ converges in probability to a positive definite limit. Hence we can conclude that when $a_1 < 1$ in (2.8) the standardized moment matrix of the auxiliary regression model $(A.23)$ converges in probability to a positive definite limit which in conjunction with $(A.25)$ and $(A.26)$ shows that $\hat{\underline{\gamma}}_1$ and $\hat{\underline{\gamma}}_2$, the LS estimators obtained from $(A.23)$, satisfy

$$\hat{\underline{\gamma}}_1 = \underline{\gamma}_1 + O_p(1) \quad \text{and} \quad \Upsilon_T(\hat{\underline{\gamma}}_2 - \underline{\gamma}_2) = O_p(1). \tag{A.28}$$

Now consider the case where $a_1 = 1$ in (2.8) and notice that the reasoning leading to $(A.27)$ is still valid. However, now the first row of the matrix E_T is of the form $[1 : o(1) : o(1)]$ and the result $\Upsilon_T^{-1}\sum_{t=1}^{T}\tilde{F}'_{2t}\tilde{F}_{1t} = o_p(1)$ does not necessarily hold. (It still holds if $T - T_1 \to \infty$). Anyway, it follows from the above that the standardized moment matrix of the regressors in $(A.23)$ converges in probability and the limit can fail to be nonsingular only

due to linear dependencies between $\tilde{F}_{2t}$ and $\tilde{F}_{1t}$. From the definitions one can readily check that this is not possible. Thus, since $(A.25)$ and $(A.26)$ also hold in the present context we again have $(A.28)$.

Since the LS estimates $\underline{\hat{\gamma}}_1$ and $\underline{\hat{\gamma}}_2$ are obtained by replacing the parameters μ_i and δ_i in the definitions of $\underline{\gamma}_1$ and $\underline{\gamma}_2$ by the LS estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ respectively, it immediately follows from $(A.28)$ that $\hat{\mu}_i = O_p(1)$ and $\hat{\delta}_i = O_p(1)$ $(i = 0, 1)$. From this and Lemma 3.1 one obtains (3.4) - (3.9) in the same way as in the proof of Theorem 1 of S&L. For example, (3.7) follows from

$$\beta'(\hat{\delta}_1 - \delta_1) = \tilde{\beta}'(\hat{\delta}_1 - \delta_1) - (\tilde{\beta} - \beta)'(\hat{\delta}_1 - \delta_1) = O_p((T - T_1)^{-1/2}) + O_p(T^{-1}).$$

To complete the proof, we have to establish (3.10). If (2.8) holds with $a_1 = 1$ then the LS estimator $\underline{\hat{\gamma}}_{24}$ is asymptotically equivalent to the GLS estimator of $\underline{\gamma}_{24}$ studied in S&L. The reason is that then the standardized moment matrix of $(A.23)$ is asymptotically block diagonal between $[\tilde{F}'_{22t} : \tilde{F}'_{23t} : \tilde{F}'_{24t}]'$ and the other regressors (see the justification of $(A.28)$). Thus it suffices to consider the case $a_1 < 1$. Then the standardized moment matrix between $\tilde{F}_{1t}$ and $\tilde{F}_{2t}$ is asymptotically block diagonal and from our previous derivation it follows that

$$\Upsilon_T(\underline{\hat{\gamma}}_2 - \underline{\gamma}_2) = \left(\Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{F}'_{2t} \tilde{F}_{2t} \Upsilon_T^{-1} \right)^{-1} \Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{F}'_{2t} \eta_t + o_p(1).$$

We have to calculate the last $n - 1$ rows of the inverse here. For this purpose we apply the well-known formula for the partitioned inverse to the last matrix in $(A.27)$. Since $e'_T E_T^{-1} = [0 : 1 : 0]$ and $e'_T E_T^{-1} e_T = 1$ this yields (cf. S&L)

$$\left(\Upsilon_T^{-1} \sum_{t=1}^{T} \tilde{F}'_{2t} \tilde{F}_{2t} \Upsilon_T^{-1} \right)^{-1} = \begin{bmatrix} * & * & * & * \\ 0 & -\tilde{B}\tilde{A}_{21}\tilde{A}_{11}^{-1} & 0 & \tilde{B} \end{bmatrix} + o_p(1)$$

where the blocks denoted by "$*$" are not needed and the partition on the r.h.s. conforms to the four estimators $\underline{\hat{\gamma}}_{2i}$ $(i = 1, ..., 4)$. From the definitions and the above equations it now follows that we have again reduced the problem to that considered in S&L, proof of Lemma A.2, so that (3.10) is obtained from that paper and thus the proof is complete. □

A.3 Proof of Theorem 4.1

We shall only give an outline of the proof because details are similar to those in S&L and L&S. First note the identity

$$\hat{x}_t = x_t - (\hat{\mu}_0 - \mu_0) - (\hat{\mu}_1 - \mu_1)t - (\hat{\delta}_0 - \delta_0)d_{0t} - (\hat{\delta}_1 - \delta_1)d_{1t} \qquad (A.29)$$

which, in conjunction with the assumed consistency properties of the estimators $\tilde{\beta}$, $\tilde{\alpha}$ and $\tilde{\Omega}$ as well as the consistency results of the estimators $\hat{\mu}_i$ and $\hat{\delta}_i$ $(i = 0, 1)$ obtained from Theorem 3.1, will be central in the subsequent derivations. Using these arguments one can show that

$$N^{-1} \sum_{t=p+1}^{T} \hat{z}_t \hat{z}_t' = E(z_t z_t') + o_p(1), \tag{A.30}$$

$$N^{-3/2} \sum_{t=p+1}^{T} \hat{z}_t \hat{v}_{t-1}' = o_p(1) \tag{A.31}$$

and

$$\begin{aligned} &N^{-2} \textstyle\sum_{t=p+1}^{T} \hat{v}_{t-1} \hat{v}_{t-1}' \\ &= T^{-2} \textstyle\sum_{t=p+1}^{T} [\beta_\perp' x_{t-1} - \beta_\perp'(\hat{\mu}_1 - \mu_1)(t-1)][\beta_\perp' x_{t-1} - \beta_\perp'(\hat{\mu}_1 - \mu_1)(t-1)]' + o_p(1), \end{aligned} \tag{A.32}$$

where $z_t = [x_{t-1}'\beta : \Delta x_{t-1}' : \ldots : \Delta x_{t-p+1}']'$ and the first term on the r.h.s. of $(A.32)$ converges weakly to $\int_0^1 \mathbf{B}_*(s)\mathbf{B}_*(s)'ds$. These results can be justified in a straightforward manner by using $(A.29)$ and the consistency properties of the associated estimators. To give a heuristic explanation note that L&S and S&L obtain $(A.30)$ - $(A.32)$ in a model without any dummy variables. In that case the effect of the estimator $\hat{\mu}_0$ on the l.h.s. of $(A.30)$ - $(A.32)$ is asymptotically negligible. Thus, when $a_1 < 1$ in (2.8) the properties of the estimator $\hat{\delta}_1$ are similar to those of $\hat{\mu}_0$ so that it should be clear that the effect of the estimator $\hat{\delta}_1$ on the l.h.s. of $(A.30)$ - $(A.32)$ is also asymptotically negligible. For the impulse dummy d_{0t} this is clear in any case because $\hat{\delta}_0 - \delta_0 = O_p(1)$. When $a_1 = 1$ in (2.8) the step dummy d_{1t} behaves very similarly to the impulse dummy d_{0t} which explains $(A.30)$ - $(A.32)$ in this case. A formal justification is obtained by using arguments similar to those used to justify $(A.17)$.

Next note that the error term e_t in (4.1) has the representation

$$\begin{aligned} e_t &= \varepsilon_t - \alpha\beta'(\hat{x}_{t-1} - x_{t-1}) + \Delta\hat{x}_t - \Delta x_t - \textstyle\sum_{j=1}^{p-1} \Gamma_j(\Delta\hat{x}_{t-j} - \Delta x_{t-j}) \\ &= \varepsilon_t + \alpha\beta'(\hat{\mu}_0 - \mu_0) + \alpha\beta'(\hat{\mu}_1 - \mu_1)(t-1) - \Psi(\hat{\mu}_1 - \mu_1) \\ &\quad + \alpha\beta'(\hat{\delta}_0 - \delta_0)d_{0,t-1} + (\hat{\delta}_0 - \delta_0)\Delta d_{0t} - \textstyle\sum_{j=1}^{p} \Gamma_j(\hat{\delta}_0 - \delta_0)\Delta d_{0,t-j} \\ &\quad + \alpha\beta'(\hat{\delta}_1 - \delta_1)d_{1,t-1} + (\hat{\delta}_1 - \delta_1)\Delta d_{1t} - \textstyle\sum_{j=1}^{p} \Gamma_j(\hat{\delta}_1 - \delta_1)\Delta d_{1,t-j}. \end{aligned}$$

The part of the last expression not involving dummy variables appeared in L&S and S&L

where it was shown that in this case

$$N^{-1}\sum_{t=p+1}^{T}\hat{v}_{t-1}e_t^{*'}\tilde{\alpha}_\perp$$
$$= N^{-1}\sum_{t=p+1}^{T}[\beta_\perp' x_{t-1} - \beta_\perp'(\hat{\mu}_1-\mu_1)(t-1)][\varepsilon_t'\alpha_\perp C'\beta_\perp - (\hat{\mu}_1-\mu_1)'\beta_\perp](\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\Psi'\alpha_\perp$$
$$+o_p(1) \qquad (A.33)$$

Using $(A.29)$ and the consistency properties of the involved estimators it can similarly be shown that $(A.33)$ holds in the present context. A heuristic explanation can again be obtained by observing that the effect of the estimator $\hat{\mu}_0$ on the l.h.s. of $(A.33)$ is asymptotically negligible so that, given the properties of the estimators $\hat{\delta}_0$ and $\hat{\delta}_1$, the same happens also when the impulse dummy d_{0t} and the step dummy d_{1t} are included in the model.

Transforming the above expressions of e_t in the same way as in $(A.31)$ of L&S and arguing as in $(A.30)$ - $(A.33)$ it can also be shown that

$$N^{-1/2}\sum_{t=p+1}^{T}\hat{Z}_t e_t^{*'}\tilde{\alpha}_\perp = O_p(1). \qquad (A.34)$$

Thus, from $(A.30)$ - $(A.34)$ it follows that we have reduced the problem to that in L&S and S&L so that the stated limiting distribution can be obtained in the same way as in these papers. □

To see that the RR regression estimators of α, β and Ω based on (4.1) have the consistency properties stated in Lemma 3.1, we first note that results entirely similar to those in Lemma A.3 and A.4 of S&L also hold in the present context. In the former case we have to show that the asymptotic behaviour of the second sample moments of $\beta'\hat{x}_t$, $\Delta\hat{x}_t$ and $\beta_\perp'\hat{x}_t$ are similar to those of $\beta' x_t$, Δx_t and $\beta_\perp' x_t$ except that the weak limit of the matrix of second sample moments of $\beta_\perp'\hat{x}_t$ is the same as was obtained in $(A.32)$. Using $(A.29)$ and Theorem 3.1 it is straightforward to show that this is the case and, since the estimation of μ_0 did not affect the previous results, this is also fairly obvious by the heuristic argument used above. In the case of Lemma A.4 of S&L we have to show that the second sample moments between $[\hat{x}_t'\beta : \Delta\hat{x}_{t-i}']$ and e_t in (4.1) are of order $O_p(T^{-1/2})$ $(i = 1, ..., p)$ while the second sample moments between $\beta_\perp'\hat{x}_{t-1}$ and e_t are of order $O_p(1)$. Using the representation of e_t given above and the same arguments as in the preceding case it can be seen that also this result holds in the present context. After this we can show that the RR regression estimators of α,

β and Ω based on (4.1) have the desired properties. The argument is the same as in Lemma A.5 of S&L or Lemmas 13.1 and 13.2 of Johansen (1995) (or in the proof of Theorem 3.1).

Appendix B. Data Sources

Seasonally unadjusted quarterly data for the period from the first quarter of 1975 to the fourth quarter of 1996 (88 observations) were used for the following variables taken from the given sources. The data refer to West Germany until 1990(2) and to the unified Germany afterwards.

Price index: GNP deflator (1991 = 100) from *Deutsches Institut für Wirtschaftsforschung, Volkswirtschaftliche Gesamtrechnung.* The variable p is the logarithm of the price index.

M3: nominal monthly values from *Monatsberichte der Deutschen Bundesbank*; the quarterly values are the values of the last month of each quarter. The variable m is log M3 $-p$.

GNP: real 'Bruttosozialprodukt' quarterly values from *Deutsches Institut für Wirtschaftsforschung, Volkswirtschaftliche Gesamtrechnung.* The variable y is log GNP $-p$.

References

Amsler, C. & J. Lee (1995), An LM test for a unit root in the presence of a structural change, *Econometric Theory*, 11, 359 - 368.

Banerjee, A., R.L. Lumsdaine & J.H. Stock (1992), Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence, *Journal of Business & Economic Statistics*, 10, 271 - 287.

Campos, J., N.R. Ericsson & D.F. Hendry (1996), Cointegration tests in the presence of structural breaks, *Journal of Econometrics*, 70, 187 - 220.

Ghysels, E. & P. Perron (1996), The effects of linear filters on dynamic time series with structural change, *Journal of Econometrics*, 70, 69 - 97.

Gregory, A.W. & B.E. Hansen (1996), Residual-based tests for cointegration in models with regime shifts, *Journal of Econometrics*, 70, 99 - 126.

Hackl, P. & A.H. Westlund (1989), Statistical analysis of "Structural Change": An annotated bibliography, *Empirical Economics*, 14, 167 - 192.

Hansen, B.E. (1992), Tests for parameter instability in regressions with I(1) processes, *Journal of Business & Economic Statistics*, 10, 321 - 335.

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

Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.

Johansen, S. & B. Nielsen (1993), Manual for the simulation program DisCo. Institute of Mathematical Statistics, University of Copenhagen.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.

Lütkepohl, H. & P. Saikkonen (1997), Testing for the cointegrating rank of a VAR process with a time trend, Discussion Paper 79, SFB 373, Humboldt-Universität zu Berlin.

Lütkepohl, H. & J. Wolters (1998), A money demand system for M3 in the Unified Germany, *Empirical Economics*, 23, 371-386.

Perron, P. (1989), The great crash, the oil price shock and the unit root hypothesis, *Econometrica*, 57, 1361 - 1401.

Perron, P. (1990), Testing for a unit root in a time series with a changing mean, *Journal of Business & Economic Statistics*, 8, 153 - 162.

Perron, P. & T.J. Vogelsang (1992), Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business & Economic Statistics*, 10, 301 - 320.

Rappoport, P. & L. Reichlin (1989), Segmented trends and non-stationary time series, *Economic Journal*, 99, 168 - 177.

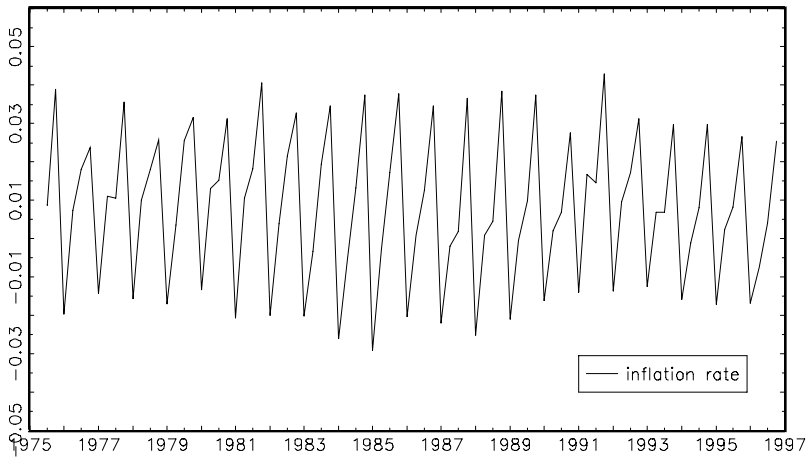
Saikkonen, P. & H. Lütkepohl (1997), Trend adjustment prior to testing for the cointegrating rank of a VAR process, Discussion Paper 84, SFB 373, Humboldt-Universität zu Berlin.

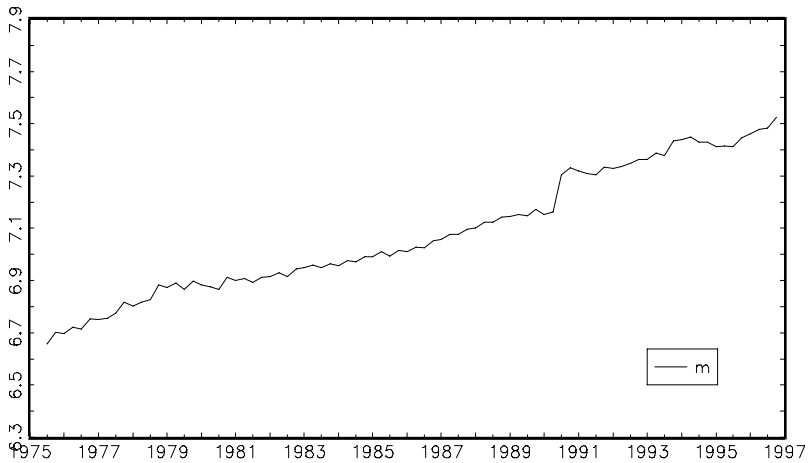
Saikkonen, P. & H. Lütkepohl (1998), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory*, forthcoming.

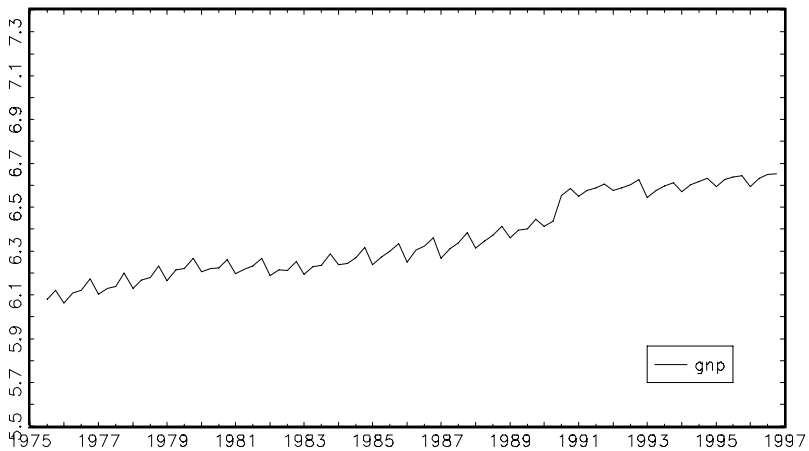
Saikkonen, P. & R. Luukkonen (1997), Testing cointegration in infinite order vector autoregressive processes, *Journal of Econometrics*, 81, 93 - 126.

Wolters, J., T. Teräsvirta & H. Lütkepohl (1998), Modelling the Demand for M3 in the Unified Germany, *Review of Economics and Statistics*, 80, 399 - 409.

Zivot E. & D.W.K. Andrews (1992), Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis, *Journal of Business & Economic Statistics*, 10, 251 - 270.


0.05
0.03
0.01
−0.01
−0.03
−0.05
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
inflation rate


7.9
7.7
7.5
7.3
7.1
6.9
6.7
6.5
6.3
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
m


7.3
7.1
6.9
6.7
6.5
6.3
6.1
5.9
5.7
5.5
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
gnp