Econometrica, Vol. 72, No. 2 (March, 2004), 647–662

TESTING FOR THE COINTEGRATING RANK OF A VAR PROCESS WITH LEVEL SHIFT AT UNKNOWN TIME

BY HELMUT LÜTKEPOHL, PENTTI SAIKKONEN, AND CARSTEN TRENKLER[1]

A systems cointegration rank test is proposed that is applicable for vector autoregressive (VAR) processes with a structural shift at unknown time. The structural shift is modeled as a simple shift in the level of the process. It is proposed to estimate the break date first on the basis of a full unrestricted VAR model. Two alternative estimators are considered and their asymptotic properties are derived. In the next step the deterministic part of the process including the shift size is estimated and the series are adjusted by subtracting the estimated deterministic part. A Johansen type test for the cointegrating rank is applied to the adjusted series. The test statistic is shown to have a well-known asymptotic null distribution that does not depend on the break date. The performance of the procedure in small samples is investigated by simulations.

KEYWORDS: Cointegration, structural break, vector autoregressive process, error correction model.

1. INTRODUCTION

THE UNIT ROOT LITERATURE has reacted to the observation that many economic time series have level shifts by developing unit root tests that take such shifts into account (see, e.g., Perron (1989, 1990)). A range of tests suitable for various kinds of shifts and alternative assumptions regarding the shift date are available. In particular, the shift date may be known or it may be assumed to be endogenous (e.g., Perron and Vogelsang (1992), Banerjee, Lumsdaine, and Stock (1992), and Zivot and Andrews (1992)). If these features are observed in univariate time series, they also have to be taken into account in multiple time series analyses. Thus, they have to be allowed for in testing for cointegration.

A number of articles address testing for cointegration in the presence of structural shifts (e.g., Hansen (1992), Gregory and Hansen (1996), Campos, Ericsson, and Hendry (1996), Seo (1998), Inoue (1999), Johansen and Nielsen (1993), Johansen, Mosconi, and Nielsen (2000), Arranz and Escribano (2000), Saikkonen and Lütkepohl (2000), and Lütkepohl, Saikkonen, and Trenkler (2003)). In these studies, single equation as well as systems cointegration tests are considered. None of the proposed tests is suitable for testing the cointegrating rank of a system when the break date is unknown, however. Because it has been argued in the unit root literature that the case of an unknown break date is of particular importance, we will present a cointegration rank test in the following that works in this situation. The shift is assumed to be a simple shift in the mean.

[1]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. The second author also acknowledges financial support by the Yrjö Jahnsson Foundation and the Alexander von Humboldt Foundation under a Humboldt research award. Part of this research was done while he was visiting the Humboldt University in Berlin and while the third author was visiting the European University Institute as a Jean Monnet Fellow. We thank three referees and a co-editor for helpful comments that resulted in substantial changes of the presentation and contents of an earlier version of the paper.

The test generalizes the one proposed by Saikkonen and Lütkepohl (2000) (henceforth S&L) for the case of a known break date. It proceeds by estimating the break date in a first step based on a full vector autoregressive (VAR) process in levels of the variables. Then the parameters of the deterministic part of the data generation process (DGP) are estimated by a suitable procedure. Using these estimators, the original series is adjusted for deterministic terms including the structural shift and a cointegrating rank test of the Johansen likelihood ratio (LR) type is applied to the adjusted series. The advantage of this procedure is that the asymptotic distribution of the test statistic under the null hypothesis is the same as in the case of a known break date and does not depend on the break date. Clearly, this property is very convenient in practice because it allows use of critical values readily available in the literature. Notice, that even in the univariate unit root testing literature some procedures for dealing with an unknown break date have a considerable effect on the properties of the corresponding unit root tests (see, e.g., Zivot and Andrews (1992)). Therefore, being able to estimate the break date such that the asymptotic properties of the cointegration tests are unaffected is an advantage.

It may be worth mentioning that dealing with an unknown shift date in a cointegrated system is considerably more difficult than in unit root testing for single time series. If our approach is used in the latter context, a one time shift in the mean reduces to a single outlier for the differenced series if the unit root null hypothesis holds (see Saikkonen and Lütkepohl (2002)). In contrast, for a nonzero cointegrating rank, the shift cannot be eliminated easily even under the null hypothesis. Unlike in the univariate case, a consistent estimator of the break date is therefore required in our procedure for testing for the cointegrating rank.

Estimating the break date in a system of $I(1)$ variables has also been considered by Bai, Lumsdaine, and Stock (1998). These authors consider the asymptotic distribution of a pseudo maximum likelihood (ML) estimator of the break date. Although we use a similar estimator, for our purposes it is sufficient to have a consistent estimator for the shift date because in that case we can show that estimating the shift date does not affect the asymptotic null distribution of the test statistic for the cointegrating rank.

The structure of this study is as follows. In Section 2, the DGP is specified and our basic assumptions are presented. Section 3 deals with the estimation of the break date and the convergence rates of two different break date estimators are given. Section 4 considers the tests for the cointegrating rank. Simulation results in Section 5 support the virtue of the tests for applied work. A summary and conclusions are given in Section 6. A sketch of the proofs is deferred to the Appendix.

We will use the following general notation. The differencing and lag operators are denoted by Δ and L, respectively, that is, for a time series or stochastic process y_t we define $\Delta y_t = y_t - y_{t-1}$ and $Ly_t = y_{t-1}$. The symbol $I(d)$ denotes an integrated process of order d. Convergence in distribution is signified by $\xrightarrow{d}$ and i.i.d. stands for independently, identically distributed. The symbols for boundedness and convergence in probability are as usual $O_p(\cdot)$ and $o_p(\cdot)$, respectively. Moreover, $\|\cdot\|$ denotes the Euclidean norm. The trace, determinant, and rank of the matrix A are denoted by $\mathrm{tr}A$, $\det A$ and $\mathrm{rk}A$, respectively. 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. VECM stands for vector error correction model. If the lower bound of the summation index of a sum exceeds the upper bound, the sum is defined to be zero.

2. BASIC ASSUMPTIONS AND DATA GENERATION PROCESS

We use the setup of S&L except that the shift date is unknown and an impulse dummy variable is not included. Hence, $y_t = (y_{1t}, \ldots, y_{nt})'$ $(t = 1, \ldots, T)$ is assumed to be generated by a process with constant, linear trend and level shift terms,

$$y_t = \mu_0 + \mu_1 t + \delta d_{t\tau} + x_t \qquad (t = 1, 2, \ldots), \tag{2.1}$$

where μ_i $(i = 0, 1)$ and δ are unknown $(n \times 1)$ parameter vectors and $d_{t\tau}$ is a dummy variable defined by $d_{t\tau} = 0$ for $t < \tau$ and $d_{t\tau} = 1$ for $t \geq \tau$, that is, $d_{t\tau}$ is a step dummy variable representing a shift in period τ. It is assumed that the actual value of τ is unknown and depends on the sample size such that the shift occurs at a fixed fraction of the sample size. More precisely, it is assumed that

$$\tau = [T\lambda] \quad \text{with} \quad 0 < \underline{\lambda} \leq \lambda \leq \bar{\lambda} < 1, \tag{2.2}$$

where $\underline{\lambda}$ and $\bar{\lambda}$ are specified real numbers and $[\cdot]$ denotes the integer part of the argument. In other words, the shift date may not be at the very beginning or at the very end of the sample. Note that $\underline{\lambda}$ and $\bar{\lambda}$ may be arbitrarily close to zero and one, respectively. Therefore our assumption regarding the break date is not very restrictive. The condition has also been employed by Bai, Lumsdaine, and Stock (1998) in models containing $I(1)$ variables.

If it is known that the DGP does not have a linear trend term, that is, $\mu_1 = 0$, this term may be dropped in (2.1). The necessary adjustments in the following analysis are straightforward and we will comment on this situation as we go along. Also seasonal dummies may be added. They leave the asymptotics of the cointegration tests unaffected and are not included in our basic model to avoid more complex notation.

The process x_t is assumed to have a VAR(p) representation,

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t \qquad (t = 1, 2, \ldots), \tag{2.3}$$

where the A_j are $(n \times n)$ coefficient matrices. For simplicity, it is assumed that $x_t = 0$ for $t \leq 0$ and $\varepsilon_t \sim$ i.i.d.$(0, \Omega)$, that is, the ε_t are i.i.d. vectors with zero mean and covariance matrix Ω. We also assume that all moments of ε_t of order b exist, where b is a number greater than 4. The initial value assumption could be replaced by the assumption that the initial values are from a fixed probability distribution that does not depend on the sample size.

The VECM form of the process x_t is

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t \qquad (t = 1, 2, \ldots), \tag{2.4}$$

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.

The process x_t is assumed to be at most $I(1)$ and cointegrated with cointegrating rank r. Hence, the matrix Π can be written as $\Pi = \alpha\beta'$, 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 $I(0)$ processes. Defining $\Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} jA_{j+1}$ and $C = \beta_\perp(\alpha_\perp' \Psi \beta_\perp)^{-1}\alpha_\perp'$, we have $x_t = C\sum_{j=1}^{t} \varepsilon_j + \xi_t$, $(t = 1, 2, \ldots)$, where ξ_t is a zero mean $I(0)$ process.

In the following we use the lag polynomial

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

Notice that the relation between the two different parameterizations is given by $A_1 = I_n + \alpha\beta' + \Gamma_1$, $A_j = \Gamma_j - \Gamma_{j-1}$ $(j = 2, \ldots, p-1)$, and $A_p = -\Gamma_{p-1}$. Multiplying (2.1) by $A(L)$ yields

$$\begin{aligned} \Delta y_t = \nu + \alpha(\beta' y_{t-1} - \phi(t-1) - \theta d_{t-1,\tau}) \\ + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j,\tau} + \varepsilon_t \qquad (t = p+1, p+2, \ldots), \end{aligned} \tag{2.5}$$

where $\nu = -\Pi\mu_0 + \Psi\mu_1$, $\phi = \beta'\mu_1$, $\theta = \beta'\delta$, and $\gamma_j^* = \delta$, for $j = 0$ and $-\Gamma_j\delta$, for $j = 1, \ldots, p-1$. Here $\Delta d_{t-j,\tau}$ is an impulse dummy with value one in period $t = \tau + j$ and zero elsewhere.

For given values of the VAR order p and the shift date τ, our formulation of the model allows estimation of the deterministic part of the DGP as in S&L. In that procedure, 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 are based on (2.5). A conventional RR regression of Δy_t on $(y_{t-1}, t-1, d_{t-1,\tau})$ corrected for $(1, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}, \Delta d_{t,\tau}, \ldots, \Delta d_{t-p+1,\tau})$ may be used although that procedure does not provide exact ML estimators because there are nonlinear restrictions between the parameters in (2.5). Given that the restrictions occur in coefficient vectors of impulse dummies only, ignoring them should not cause any great loss of efficiency. The observations may then be adjusted for deterministic terms and cointegration tests may be based on the adjusted series. These tests will be discussed in Section 4. In the next section estimation of the break date τ will be considered.

3. ESTIMATION OF THE SHIFT DATE

For a given VAR order p, τ can be estimated on the basis of the levels VAR form, that is, on the basis of the least restricted model with respect to the cointegrating rank,

$$\begin{aligned} y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} \\ + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t \qquad (t = p+1, \ldots, T), \end{aligned} \tag{3.1}$$

which is obtained from (2.5) by imposing no rank restriction on Π and rearranging terms. Here $\nu_0 = \nu + \Pi\mu_1$, $\nu_1 = -\Pi\mu_1$, $\delta_1 = -\Pi\delta$, $\gamma_0 = \delta - \delta_1$, and $\gamma_j = \gamma_j^*$

($j = 1, \ldots, p-1$). Thus, a linear trend, a step dummy variable, and p impulse dummies are included as deterministic terms. The shift date is estimated as

$$\hat{\tau} = \arg\min_{\tau \in \mathcal{T}} \det\left(\sum_{t=p+1}^{T} \hat{\varepsilon}_{t\tau} \hat{\varepsilon}_{t\tau}' \right), \tag{3.2}$$

where $\mathcal{T} = [T\underline{\lambda}, T\bar{\lambda}]$ is the space of τ and the $\hat{\varepsilon}_{t\tau}$ are the LS residuals from (3.1).

It is possible that the p impulse dummies in (3.1) make it difficult to determine the true break date in this way because they eliminate the information in the observations associated with the periods where they assume a value of one. Therefore one may consider estimating the shift date from a VAR model without impulse dummy variables,

$$\begin{aligned} y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} \\ + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_{t\tau}^* \qquad (t = p+1, \ldots, T), \end{aligned} \tag{3.3}$$

where $\varepsilon_{t\tau}^* = \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} + \varepsilon_t$. The resulting estimator of the shift date will be denoted by $\tilde{\tau}$, that is,

$$\tilde{\tau} = \arg\min_{\tau \in \mathcal{T}} \det\left(\sum_{t=p+1}^{T} \hat{\varepsilon}_{t\tau}^* \hat{\varepsilon}_{t\tau}^{*\prime} \right) \tag{3.4}$$

and the $\hat{\varepsilon}_{t\tau}^*$ are LS residuals from (3.3). It turns out that both ways of estimating the shift date result in consistent estimators of the sample fraction λ where the shift occurs, provided $\delta_1 \neq 0$. More precisely, the following result holds.

THEOREM 3.1: *If* $\delta_1 \neq 0$, *then* (i) $\hat{\tau} - \tau = O_p(1)$ *and* (ii) $\tilde{\tau} - \tau = O_p(1)$. *Here* τ *denotes the true shift date*.

A proof of this theorem is outlined in the Appendix. Obviously, the theorem implies that $\hat{\tau}/T$ and $\tilde{\tau}/T$ are consistent estimators of the sample fraction λ. Interestingly, even though the estimator $\tilde{\tau}$ is based on a misspecified model, under our assumptions its convergence rate is equally good as that of $\hat{\tau}$ which is based on a correctly specified model.

It is perhaps worth emphasizing that the result in Theorem 3.1 excludes the case $\delta_1 = -\alpha\beta'\delta = 0$, which may hold even if $\delta \neq 0$. In that case the process $\beta' y_t$ has no break and, thus, we have a case of co-breaking. For such processes, neither $\hat{\tau}/T$ nor $\tilde{\tau}/T$ are consistent estimators of the sample fraction λ in our framework. However, at least when the level shift is relatively large, the chances for the impulse dummies to react to it may be quite large and, hence, there is at least a possibility for $\hat{\tau}$ to find the shift date. In contrast, the estimator $\tilde{\tau}$ can find the shift date by accident only in this situation. Thus, unless the case $\delta_1 = 0$ can be ruled out, using only the estimator $\tilde{\tau}$ may be problematic. In this context it may be worth noting that $\delta_1 = 0$ always holds if the cointegrating rank is zero.

Of course, other possibilities exist for estimating the break date. Notably an estimator which takes into account the restrictions for the γ_j in (3.1) could be considered. From a practical point of view such an estimator has the disadvantage, however, that the restrictions are nonlinear and, hence, taking them into account increases the

computational burden. In addition, they are related to parameters attached to impulse dummy variables only, which cannot be estimated consistently. Another possibility may be to consider estimating τ under the cointegrating rank specified in the null hypothesis of a subsequent cointegration test. Such estimators may be interesting to explore in future work. We focus on $\hat{\tau}$ and $\tilde{\tau}$ because of their simplicity in practice.

So far we have assumed a given VAR order p. In practice p is also unknown, of course. In that case one may estimate the shift date first on the basis of some VAR order possibly greater than the true one. Alternatively, a shift date may be estimated for a range of orders, say $p = 1, \ldots, p_{\max}$, where $p_{\max}$ is some prespecified upper bound for the VAR order and the VAR order to be used in the tests is chosen such that a standard model selection criterion like AIC applied to model (3.1) is minimized.

Of course, if the DGP is known to have no linear trend, the corresponding terms in (3.1) and (3.3) may be dropped. In the next section testing the cointegrating rank is discussed based on the break date estimators $\hat{\tau}$ and $\tilde{\tau}$.

4. COINTEGRATING RANK TESTS

We wish to test the null hypothesis

$$H_0(r_0): \operatorname{rk}(\Pi) = r_0 \quad \text{vs.} \quad H_1(r_0): \operatorname{rk}(\Pi) > r_0. \tag{4.1}$$

For a given break date, S&L propose to estimate the parameters of the deterministic part first. Denoting the estimators of μ_0, μ_1, and δ by $\hat{\mu}_0$, $\hat{\mu}_1$, and $\hat{\delta}$, respectively, the test is based on a sample analog of the series x_t obtained as

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

where $\hat{\tau}$ may be replaced by any estimator that satisfies $\hat{\tau} = \tau + O_p(1)$ including, of course, $\tilde{\tau}$. The series $\hat{x}_t$ can be used to compute LR type test statistics for the null hypothesis $H_0(r_0)$ in the same way as the usual LR test statistic based on the VECM (2.4). More precisely, the test statistic can be determined from

$$\Delta \hat{x}_t = \Pi \hat{x}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta \hat{x}_{t-j} + e_{t\hat{\tau}} \qquad (t = p+1, \ldots, T), \tag{4.2}$$

by solving the generalized eigenvalue problem $\det(\hat{\Pi} \hat{M}_T \hat{\Pi}' - \lambda \hat{\Omega}) = 0$, where $\hat{\Pi}$ is the LS estimator of Π obtained from (4.2), $\hat{\Omega}$ 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 $\Delta\hat{X}_{t-1} = [\Delta\hat{x}'_{t-1} : \cdots : \Delta\hat{x}'_{t-p+1}]'$. Denoting the resulting ordered eigenvalues by $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$, the LR type statistic for the pair of hypotheses in (4.1) becomes

$$LR(r_0) = \sum_{j=r_0+1}^{n} \log(1+\hat{\lambda}_j). \tag{4.3}$$

Following a similar line of arguments as in S&L it can be shown that a convenient null distribution of this test is obtained if the parameter estimators of the deterministic part satisfy the following conditions:

ASUMPTION 1: *$\beta'(\hat{\mu}_0 - \mu_0) = O_p(T^{-1/2})$, $\beta'_\perp(\hat{\mu}_0 - \mu_0) = O_p(1)$, $\beta'(\hat{\delta} - \delta) = O_p(T^{-1/2})$, $\beta'_\perp(\hat{\delta} - \delta) = O_p(1)$, $\beta'(\hat{\mu}_1 - \mu_1) = O_p(T^{-3/2})$, $T^{1/2}\beta'_\perp(\hat{\mu}_1 - \mu_1) \xrightarrow{d} N(0, \beta'_\perp C\Omega C'\beta_\perp)$, and all quantities converge jointly in distribution upon appropriate standardization. Here $C = \beta_\perp(\alpha'_\perp\Psi\beta_\perp)^{-1}\alpha'_\perp$, as before.*

Notice that in our model (2.1), μ_0 is only identified in the direction of β and, in the direction of $\beta_\perp$, δ is identified by the impulse dummies only for which the information does not increase with the sample size. Therefore we do not assume that $\beta'_\perp\mu_0$ and $\beta'_\perp\delta$ are estimated consistently. For estimators $\hat{\mu}_0$, $\hat{\mu}_1$, and $\hat{\delta}$ satisfying Assumption 1, the limiting distribution of the test statistic $LR(r_0)$ is given in the following theorem, where $\mathbf{B}(s)$ is an $(n-r_0)$-dimensional standard Brownian motion. In the theorem it is understood that all relevant quantities converge jointly upon appropriate standardization. In particular, all the estimators of the parameters associated with the deterministic part of the model are based on the same data as the cointegration test. A brief outline of the proof is given in the Appendix.

THEOREM 4.1: *Suppose the DGP is described by the model* (2.1)–(2.3) *with all variables at most* $I(1)$ *and* $\varepsilon_t \sim$ i.i.d.$(0, \Omega)$ *has moments of order* $b > 4$. *Furthermore, let* $\hat{\tau}$ *be a break date estimator such that* $\hat{\tau} = \tau + O_p(1)$ *if* $\delta_1 \neq 0$. *Then, if Assumption* 1 *holds and* $H_0(r_0)$ *in* (4.1) *is true,*

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

where $\mathbf{B}_*(s) = \mathbf{B}(s) - s\mathbf{B}(1)$ *is an* $(n-r_0)$*-dimensional Brownian bridge and* $d\mathbf{B}_*(s) = d\mathbf{B}(s) - ds\mathbf{B}(1)$, *that is,* $\int_0^1 \mathbf{B}_*(s)\, d\mathbf{B}_*(s)'$ *abbreviates* $\int_0^1 \mathbf{B}(s)\, d\mathbf{B}(s)' - \mathbf{B}(1)\int_0^1 s\, d\mathbf{B}(s)' - \int_0^1 \mathbf{B}(s)\, ds\mathbf{B}(1)' + \frac{1}{2}\mathbf{B}(1)\mathbf{B}(1)'$.

It can be shown that the GLS estimators proposed by S&L satisfy Assumption 1.[2] Hence, they may be used in the test to obtain the asymptotic properties of the $LR(r_0)$ statistic claimed in Theorem 4.1. Notice that the result of this theorem holds even if

[2]See the accompanying material to this paper at http://www.iue.it/Personal/Luetkepohl/Welcome.html.

$\delta_1 = 0$ and, hence, $\theta = \beta'\delta = 0$ in (2.5) so that the shift dummy does not appear in the model. In that case the shift date cannot be estimated consistently. This situation occurs, e.g., whenever $r = 0$. In this case a consistent estimator of τ is not needed, however, to obtain the stated asymptotic distribution of the test statistic because under H_0 the step dummy becomes an impulse dummy, the effect of which vanishes asymptotically. The limiting distribution in Theorem 4.1 is free of unknown nuisance parameters. It is the same as the one obtained by S&L in a model with a shift at a known point in time τ. Critical values are given in Table 1 of Lütkepohl and Saikkonen (2000). Thus, in our framework, including a step dummy in the model and estimating its coefficients and the shift date has no effect on the limiting distribution of the cointegration tests. Technically the reason for this result is that the process $\beta_\perp'\hat{x}_t$ is divided by $T^{1/2}$ and, moreover, $\beta_\perp'(\hat{\delta} - \delta) = o_p(T^{1/2})$. Therefore all effects from the shift disappear asymptotically.

It is also possible to extend our results by including more than one shift dummy or other dummy variables in model (2.1). In fact, an additional impulse dummy and seasonal dummies were considered by S&L. The result in Theorem 4.1 remains valid with additional dummies. If additional shift dummies are included and the shift date is unknown, it may be more difficult to construct suitable shift date estimators. We leave this issue for future research.

The preceding discussion also suggests that if the a priori restriction $\mu_1 = 0$ is employed in (2.1) and the GLS estimation of S&L 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 is obtained by replacing the Brownian bridge $\mathbf{B}_*(s)$ in Theorem 4.1 by the Brownian motion $\mathbf{B}(s)$. This situation was also studied by S&L. In the next section we will discuss small sample properties of the tests.

5. MONTE CARLO SIMULATIONS

A small Monte Carlo experiment was performed to explore the finite sample properties of our procedures and to compare the two break date estimators $\hat{\tau}$ and $\tilde{\tau}$. The simulations are based on the following x_t process from Toda (1994), which was also used by some other authors for investigating the properties of cointegrating rank tests (see, e.g., Hubrich, Lütkepohl, and Saikkonen (2001)):

$$x_t = \begin{bmatrix} \boldsymbol{\psi} & 0 \\ 0 & I_{n-r} \end{bmatrix} x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{n-r} \end{bmatrix} \right), \tag{5.1}$$

where $\boldsymbol{\psi} = \text{diag}(\psi_1, \ldots, \psi_r)$ and Θ are $(r \times r)$ and $(r \times (n-r))$ matrices, respectively. This type of process represents a large class of relevant processes for investigating the properties of LR tests for the cointegrating rank because other VAR(1) processes of interest can be obtained from (5.1) by linear transformations that leave such tests invariant. We report some representative results for three-dimensional DGP's with cointegrating rank $r = 1$ in more detail here. In this case, $\boldsymbol{\psi} = \psi_1$ with $|\psi_1| < 1$ and $\Theta = (\theta_1, \theta_2)$ represents the instantaneous dependency between the stationary and nonstationary components. For given VAR order p and break date τ, the test results are invariant to the parameter values of the deterministic component. Therefore the intercept and trend terms are set to zero throughout ($\mu_i = 0\ (i = 0, 1)$) so that $y_t = \delta d_{t\tau} + x_t$.

We have performed simulations with different δ vectors because the sizes of the shifts are expected to have an impact on the estimators $\hat{\tau}$ and $\tilde{\tau}$ which in turn may affect the small sample properties of the tests for the cointegrating rank.

Samples of size $T = 100$ were simulated starting with initial values of zero. The number of replications was 1000. Thus the standard error of an estimator of a true rejection probability P is $s_P = \sqrt{P(1-P)/1000}$, e.g., $s_{.05} = .007$. Linear trend terms were included in the test regressions although they are actually zero, that is, we pretend that this information is not available to the analyst. Moreover, we used different VAR orders p, although the "true" order is $p = 1$, to explore the impact of this quantity on the estimation and testing results.

Figure 1 contains results for a shift date $\tau = 50$. The search procedures were applied to all possible break points between the 5th and the 96th observation. Results are shown for a range of different $\delta_{(1)}$ using a shift parameter $\delta = (\delta_{(1)}, 0, 0)$. Notice that in this case the shift does not vanish from the cointegration relation so that $\delta_1 \neq 0$ in (3.1) and (3.3). The tests are based on asymptotic critical values for a nominal significance level of 5%.

For a VAR order $p = 1$, the actual shift date is obviously found with greater reliability if an impulse dummy is included in addition to a shift dummy. Clearly, it is easier for both procedures to find the correct break date if the parameter $\delta_{(1)}$ is larger. The power is slightly lower if the shift date is estimated by $\tilde{\tau}$ compared to the power obtained with $\hat{\tau}$. Nevertheless, the properties of the cointegration tests based on $\tilde{\tau}$ do not differ much from the approach that includes an impulse dummy in addition ($\hat{\tau}$), although the search procedure including a shift dummy only ($\tilde{\tau}$) is less successful in estimating the shift date.

The VAR order has some impact on estimating the break point if impulse dummies are included. For $p = 3$, $\hat{\tau}$ tends to estimate the shift date too early compared to the true break date. Most of the incorrect estimations are located within the $p - 1$ periods before the actual break, as can be seen in Figure 1. This fact can be explained by the inclusion of the lagged impulse dummies according to (3.1). The impulse dummies delete the impact of the observations corresponding to the periods where they are nonzero so that the shift dummy becomes effective only p periods after date $\hat{\tau}$. In contrast, if just a shift dummy is included ($\tilde{\tau}$), the higher VAR order does not make it more difficult to find the actual shift date. As can be seen in Figure 1, the small sample power of the cointegration test based on $\hat{\tau}$ falls for increasing shift magnitudes $\delta_{(1)}$ if $p = 3$.

We have also studied several other issues related to our DGP. For example, we considered a shift in the nonstationary part of the DGP using $\delta = (0, 0, \delta_{(3)})$. In this case δ_1 in (3.1) is zero and $\hat{\tau}$ and $\tilde{\tau}$ are not consistent estimators of the relative break point. Interestingly, only with respect to $\tilde{\tau}$ the probability of finding the true shift date is reduced compared to the previous setup. However, these deteriorations in the break date estimates do not affect the performance of the cointegration tests in any important way.

Finally, we have analyzed $\psi_1 = .7$, different shift dates, a sample length of $T = 200$, the case $r = 0$, and we have considered the procedures excluding a linear trend ($\mu_1 = 0$). Moreover, we have considered processes of larger dimension. By and large, the relative performance of the break date estimators has not changed. The same is true regarding the cointegration tests. Their small sample properties are more affected by the specification of the DGP than by the fact that the shift date has to be estimated.

Summarizing our simulation results, we can conclude that having to estimate the shift date does not lead to substantial changes in the properties of the cointegration

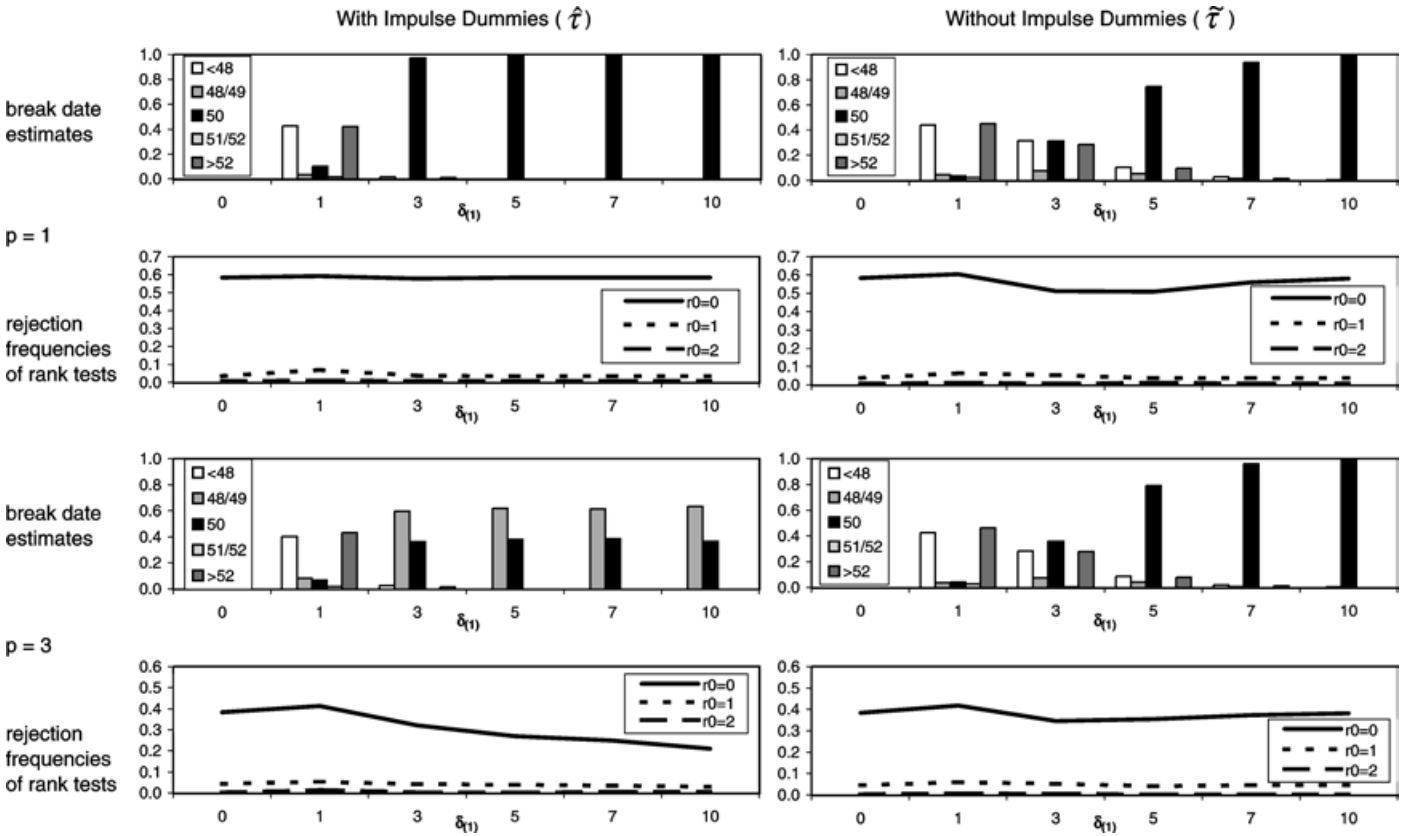


FIGURE 1.—Break point estimates and rejection frequencies for three-dimensional DGP's with true cointegrating rank $r = 1$, $\psi = .9$, $\Theta = (.4, .8)$, and $T = 100$.

tests although the shift date estimators may give quite different results. Therefore, both estimators, $\hat{\tau}$ and $\tilde{\tau}$, may be used and our procedure appears to be suitable for empirical work.

6. CONCLUSIONS

We have proposed a procedure for testing the cointegrating rank of VAR processes with a structural shift in the level at unknown time. In this procedure the shift date is estimated in a first step. Then the deterministic terms are estimated, and finally a cointegration test is performed on the series that is adjusted for deterministic terms.

Two possible estimators of the break date based on a full unrestricted VAR model are explored. One estimator takes into account a set of impulse dummies that results from rewriting the DGP in levels VAR form whereas the other estimator ignores the impulse dummies and is therefore actually based on a misspecified model. It is found that ignoring the impulse dummies still has advantages in small samples and may therefore be useful in practice.

The resulting cointegration rank test statistic is shown to have a limiting null distribution for which critical values exist in the literature. It does not depend on the break date. As usual in cointegration tests, the asymptotic distributions of the test statistics depend on whether or not a deterministic trend term in the DGP is accounted for. Asymptotic results for DGP's with and without deterministic linear trend terms are given and we have also discussed the consequences of including seasonal dummy variables.

In addition to providing asymptotic results we have also investigated the small sample properties of the procedure using a Monte Carlo simulation experiment. It turns out that the break date estimator that ignores the impulse dummies may be more successful in locating the true break date, if the VAR order is greater than one. The choice of break date estimator may have some impact on the properties of the performance of the cointegrating rank tests in small samples.

There are a number of possible extensions that may be of interest in future work. For example, other estimators of the break date may be considered. Moreover, it may be useful to allow for more than one shift and it may also be of interest to include further impulse dummy variables. In addition, dummies that take on a value of one for a specified number of periods and are zero otherwise may be of interest to accommodate special effects that have been of importance for a limited time only. Furthermore, one may desire to model a shift in the slope of the linear trend term. In principle, additional shift, impulse, and other dummy variables can be treated in a similar way as in our theoretical derivations if break date estimators with suitable properties can be found. The derivation of such estimators is likely to cause theoretical as well as practical problems, however. Nevertheless, the framework developed in the foregoing may be a useful point of departure for overcoming such problems.

Economics Dept., European University Institute, Via della Piazzuola 43, I-50133 Florence, Italy; helmut.luetkepohl@iue.it,

Dept. of Statistics, University of Helsinki, P.O. Box 54, FIN-00014 Helsinki, Finland; saikkone@mappi.helsinki.fi,

and

Institute for Statistics and Econometrics, School of Business and Economics, Humboldt University Berlin, Spandauer Str. 1, D-10178 Berlin, Germany; trenkler@wiwi.hu-berlin.de.

Manuscript received August, 2001; final revision received February, 2003.

APPENDIX: OUTLINE OF PROOFS

We only outline the proofs here. A detailed version is available in the internet at http://www.iue.it/Personal/Luetkepohl/Welcome.html.

PROOF OF THEOREM 3.1: Unless otherwise stated, all limits assume that $T \to \infty$. Moreover, B^c denotes the complement of the set B. The true DGP is one specific process from our model class. It is occasionally helpful to be more explicit about its particular parameter values. In these cases they will be indicated with a subscript "o" (e.g., μ_{0o}, μ_{1o}, τ_o, etc.). Instead of the series y_t it will be convenient to use the mean adjusted series

$$x_t = y_t - \mu_{0o} - \mu_{1o}t - \delta_o d_{t\tau_o} \qquad (t = 1, 2, \ldots).$$

Solving the above equation for y_t and inserting the result into (3.1) yields

$$\begin{aligned}(\text{A.1}) \qquad \Delta x_t = {} & \nu_0^{(0)} + \nu_1^{(0)} t + \delta_1 d_{t\tau} + \underline{\gamma}\underline{d}_{t\tau} - \delta_1^{(0)} d_{t\tau_o} - \underline{\gamma}^{(0)}\underline{d}_{t\tau_o} \\ & + \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t \qquad (t = p+1, p+2, \ldots).\end{aligned}$$

Here $\nu_0^{(0)} = \nu_0 + \Pi\mu_{0o} - \Psi\mu_{1o} - \Pi\mu_{1o}$, $\nu_1^{(0)} = \nu_1 + \Pi\mu_{1o}$, $\underline{\gamma} = [\gamma_0 : \cdots : \gamma_{p-1}]$, $\underline{d}_{t\tau} = [\Delta d_{t\tau} : \cdots : \Delta d_{t-p+1,\tau}]'$, $\delta_1^{(0)} = -\Pi\delta_o$, and $\underline{\gamma}^{(0)} = [\gamma_0^{(0)} : \cdots : \gamma_{p-1}^{(0)}]$ with $\gamma_j^{(0)} = \delta_o - \delta_1^{(0)}$ for $j = 0$ and $\gamma_j^{(0)} = -\Gamma_j\delta_o$ for $j = 1, \ldots, p-1$. Note that the true values of $\nu_0^{(0)}$ and $\nu_1^{(0)}$ are zero.

It will also be convenient to use the transformation $\Pi x_t = \alpha^{(0)} u_{t-1}^{(0)} + \rho^{(0)} v_{t-1}^{(0)}$, where $u_{t-1}^{(0)} = \beta_o' x_{t-1}$, $v_{t-1}^{(0)} = \beta_{o\perp}' x_{t-1}$, $\alpha^{(0)} = \alpha\beta'\beta_o(\beta_o'\beta_o)^{-1}$, and $\rho^{(0)} = \alpha\beta'\beta_{o\perp}(\beta_{o\perp}'\beta_{o\perp})^{-1}$. Clearly, the true values of $\alpha^{(0)}$ and $\rho^{(0)}$ are α_o and zero, respectively. With this transformation the preceding error correction form can be expressed as

$$\begin{aligned}(\text{A.2}) \qquad \Delta x_t = {} & \nu_0^{(0)} + \nu_1^{(0)} t + \delta_1 d_{t\tau} + \underline{\gamma}\underline{d}_{t\tau} - \delta_1^{(0)} d_{t\tau_o} - \underline{\gamma}^{(0)}\underline{d}_{t\tau_o} \\ & + \alpha^{(0)} u_{t-1}^{(0)} + \rho^{(0)} v_{t-1}^{(0)} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t \qquad (t = p+1, p+2, \ldots).\end{aligned}$$

Denote

$$w_t^{(0)} = \left[1 : \frac{t}{T} : T^{-1/2} v_{t-1}^{(0)\prime} : u_{t-1}^{(0)\prime} : \Delta x_{t-1}' : \cdots : \Delta x_{t-p+1}'\right]'$$

and $q_{t\tau} = [d_{t\tau} : \underline{d}_{t\tau}']'$. With this notation (A.2) becomes

$$(\text{A.3}) \qquad \Delta x_t = \Phi w_t^{(0)} + \Xi q_{t\tau} - \Xi^{(0)} q_{t\tau_o} + \varepsilon_t \qquad (t = p+1, p+2, \ldots),$$

where

$$\Phi = \left[\nu_0^{(0)} : T\nu_1^{(0)} : T^{1/2}\rho^{(0)} : \alpha^{(0)} : \Gamma_1 : \cdots : \Gamma_{p-1}\right],$$

$$\Xi = \left[\delta_1 : \underline{\gamma}\right], \quad \text{and} \quad \Xi^{(0)} = \left[\delta_1^{(0)} : \underline{\gamma}^{(0)}\right].$$

Let $\Theta = [\Phi : \Xi]$ contain the freely varying parameters in (A.3) or (A.2) and set

$$\varepsilon_{t\tau}(\Theta) = \Delta x_t - \Phi w_t^{(0)} - \Xi q_{t\tau} + \Xi^{(0)} q_{t\tau_o}.$$

Then

$$l_T(\Theta, \tau, \Omega) = (T-p)\log\det\Omega + \operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{t\tau}(\Theta)\varepsilon_{t\tau}(\Theta)'\right)$$

is -2 times the (conditional) Gaussian log-likelihood function of the parameters in (A.3). Minimizing this function yields Gaussian ML estimators of the parameters Θ, τ, and Ω. It is not difficult to see that the resulting estimators of Θ and τ can alternatively be obtained by minimizing the concentrated counterpart of $l_T(\Theta, \tau, \Omega)$, that is, $l_T^{(c)}(\Theta, \tau) = (T-p) \times \log\det(\sum_{t=p+1}^{T}\varepsilon_{t\tau}(\Theta)\varepsilon_{t\tau}(\Theta)')$. The definition of $\varepsilon_{t\tau}(\Theta)$ (and the fact that $\Xi^{(0)}$ is not a freely varying parameter) makes it clear that the value of τ that minimizes the function $l_T^{(c)}(\Theta, \tau)$ is identical to $\hat{\tau}$ defined by (3.2). Thus, (asymptotic) properties of $\hat{\tau}$ can be studied by using the Gaussian ML estimator of τ discussed above. The foregoing discussion also shows that a minimizer of $l_T(\Theta, \tau, \Omega)$ exists (for every T larger than some constant).

The proof of Theorem 3.1 consists of several steps. In the first one we consider a subset of the parameter space of (Θ, Ω) defined by

$$0 < \underline{\omega} \le \lambda_{\min}(\Omega) \le \lambda_{\max}(\Omega) \le \bar{\omega} < \infty \tag{A.4}$$

and

$$\|\Phi\|^2 + \|\delta_1 - \delta_1^{(0)}\|^2 \le \bar{M} < \infty. \tag{A.5}$$

Note that here $\bar{M}$ does not depend on T although Φ does. Denoting by $B_1 = B_1(\bar{M}, \underline{\omega}, \bar{\omega})$ the part of the parameter space of (Θ, τ, Ω) in which conditions (A.4) and (A.5) hold, it can be shown that there exist choices of $\bar{M}$, $\underline{\omega}$, and $\bar{\omega}$ such that

$$\inf_{(\Theta,\tau,\Omega)\in B_1^c} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) > 0 \tag{A.6}$$

with probability approaching one. Hence, a minimizer of $l_T(\Theta, \tau, \Omega)$ will asymptotically satisfy inequality restrictions of the form (A.4) and (A.5). In what follows, the set B_1 is always assumed to be defined in such a way that (A.6) holds.

We shall now proceed in the same way as in Saikkonen (2001) and express the function $l_T(\Theta, \tau, \Omega)$ as a sum of two components. To this end, define

$$w_{1t}^{(0)} = \left[1 : \frac{t}{T} : T^{-1/2} v_{t-1}^{(0)\prime}\right]' \quad \text{and} \quad w_{2t}^{(0)} = \left[u_{t-1}^{(0)\prime} : \Delta x_{t-1}' : \cdots : \Delta x_{t-p+1}'\right]'.$$

Then $w_t^{(0)} = [w_{1t}^{(0)\prime} : w_{2t}^{(0)\prime}]'$ and we also partition the parameter matrix Φ conformably as $\Phi = [\Phi_1 : \Phi_2]$, where $\Phi_1 = [\nu_0^{(0)} : T\nu_1^{(0)} : T^{1/2}\rho^{(0)}]$ and $\Phi_2 = [\alpha^{(0)} : \Gamma_1 : \cdots : \Gamma_{p-1}]$. With these definitions, $\varepsilon_{t\tau}(\Theta) = \varepsilon_{1t\tau}(\Theta) + \varepsilon_{2t}(\Phi_2)$, where $\varepsilon_{1t\tau}(\Theta) = -\Phi_1 w_{1t}^{(0)} - \Xi q_{t\tau} + \Xi^{(0)} q_{t\tau_o}$ and $\varepsilon_{2t}(\Phi_2) = \Delta x_t - \Phi_2 w_{2t}^{(0)}$. Clearly, $\varepsilon_{1t\tau_o}(\Theta_o) = 0$ and $l_T(\Theta, \tau, \Omega) = l_{1T}(\Theta, \tau, \Omega) + l_{2T}(\Phi_2, \Omega)$, where

$$l_{1T}(\Theta, \tau, \Omega) = \operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{1t\tau}(\Theta)\varepsilon_{1t\tau}(\Theta)'\right) + 2\operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{1t\tau}(\Theta)\varepsilon_{2t}(\Phi_2)'\right)$$

and

$$l_{2T}(\Phi_2, \Omega) = (T-p)\log\det\Omega + \operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{2t}(\Phi_2)\varepsilon_{2t}(\Phi_2)'\right).$$

The idea is to study the two components of $l_T(\Theta, \tau, \Omega)$ separately. For $l_{2T}(\Phi_2, \Omega)$ we have

$$\inf_{(\Phi_2,\Omega)} l_{2T}(\Phi_2, \Omega) - l_{2T}(\Phi_{2o}, \Omega_o) = O_p(1), \tag{A.7}$$

where the infimum is over unrestricted values of Φ_2 and $\Omega > 0$. As for $l_{1T}(\Theta, \tau, \Omega)$, our treatment is divided into several steps in which the time index t is suitably restricted (cf. Bai, Lumsdaine, and Stock (1998)). This means considering the function $l_{1T}(\Theta, \tau, \Omega)$ with the sample size T replaced by appropriate quantities smaller than T. The purpose is to obtain lower bounds for components of the function $l_{1T}(\Theta, \tau, \Omega)$ that can be used to restrict the parameter space further from B_1. These components are of the form $l_{1T_2}(\Theta, \tau, \Omega) - l_{1T_1}(\Theta, \tau, \Omega)$, where $T_1 < T_2 \le T$ and $l_{1T_1}(\Theta, \tau, \Omega)$ can be zero.

Of the several lower bounds obtained for the components of $l_{1T}(\Theta, \tau, \Omega)$ we here only mention one. It makes use of the notation $\Psi_2 = \Phi_1 + [\delta_1 - \delta_1^{(0)} : 0]$ and $\zeta_{t\tau}^{(0)} = (d_{t\tau} - d_{t\tau_o})\delta_1^{(0)} + \underline{\gamma} \underline{d}_{t-\tau} - \underline{\gamma}^{(0)} \underline{d}_{t-\tau_o}$ and shows the existence of a constant $c_5 > 0$ such that with probability approaching one and uniformly in $[T\underline{\lambda}] \le \tau \le \tau_o - 1$ and $(\Theta, \tau, \Omega) \in B_1$,

$$\begin{aligned} l_{1,\tau_o-1}(\Theta, \tau, \Omega) - l_{1,\tau-1}(\Theta, \tau, \Omega) \\ \ge c_5 \sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 - \left(a_{7T}(\tau_o - \tau)^{\eta} + a_{8T}\left(\frac{\tau_o - \tau}{T}\right)^{1/2} \|T^{1/2}\Psi_2\| \right) \left(\sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} \\ - a_{9T}\|T^{1/2}\Psi_2\|, \end{aligned} \tag{A.8}$$

where $\frac{1}{b} < \eta < \frac{1}{4}$, $a_{iT} \ge 0$, and $a_{iT} = O_p(1)$ $(i = 7, 8, 9)$.

Now, let $\epsilon > 0$ and $B_2 = \{(\Theta, \tau, \Omega) : \|T^{1/2-\eta}\Phi_1\|^2 + \|T^{1/2-\eta}\Psi_2\|^2 \le \epsilon^2\}$ and $B_3 = \{(\Theta, \tau, \Omega) : \|T^{1/2-\eta}(\Phi_2 - \Phi_{2o})\| \le \epsilon\}$ where η is the same as in (A.8). Then, using (A.7) we first find that

$$\begin{aligned} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) &= l_{1T}(\Theta, \tau, \Omega) - l_{2T}(\Phi_2, \Omega) - l_{2T}(\Phi_{2o}, \Omega_o) \\ &\ge l_{1T}(\Theta, \tau, \Omega) + O_p(1) \end{aligned}$$

which in conjunction with (A.8) and the other lower bounds obtained for the components of $l_{1T}(\Theta, \tau, \Omega)$ can be used to show that

$$\inf_{(\Theta,\tau,\Omega)\in B_i^c} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) > 0 \qquad (i = 2, 3) \tag{A.9}$$

with probability approaching one and uniformly in $[T\underline{\lambda}] \le \tau \le \tau_o - 1$.

Let $B_4 = \{(\Theta, \tau, \Omega) : (\tau_o - \tau)^{-2\eta} \sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \le M^2\}$, where $\tau < \tau_o$. Then, the next step is to prove that there exists a real number $M_0 > 0$ such that, for all $M \ge M_0$,

$$\inf_{(\Theta,\tau,\Omega)\in B_4^c} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) > 0 \tag{A.10}$$

with probability approaching one and uniformly in $[T\underline{\lambda}] \le \tau \le \tau_o$. The proof of this again makes use of previous intermediate results and particularly (A.7) and (A.8) of which the former implies that we need to establish the existence of a real number $M_0 > 0$ such that $\inf_{(\Theta,\tau,\Omega)\in B_4^c} l_{1T}(\Theta, \tau, \Omega) > M_1$ for all $M \ge M_0$ and any $M_1 > 0$. This in turn readily follows because it can be shown that

$$l_{1T}(\Theta, \tau, \Omega) \ge c^* \xi_\tau^2 - a_T^*(\tau_o - \tau)^{\eta} \xi_\tau + O_p(1) = c^* \xi_\tau^2 \left(1 - \frac{a_T^*(\tau_o - \tau)^{\eta}}{c^* \xi_\tau} \right) + O_p(1),$$

where $\xi_\tau^2 = \sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2$ and the stated relations hold uniformly in $B_1 \cap B_2 \cap B_3 \cap B_4^c$ and $[T\underline{\lambda}] \le \tau \le \tau_o - 1$.

For those results that are formulated for $\tau \leq \tau_o$, analogous versions also hold for $\tau \geq \tau_o$. Now, recall that $\delta_1^{(0)} = -\Pi\delta_o = -\alpha\beta'\delta_o$. Thus, $\delta_{1o} \neq 0$ if and only if $\beta_o'\delta_o \neq 0$. Let $M > 0$. Assume that $\delta_{1o} \neq 0$ and define $B_5 = \{(\Theta, \tau, \Omega) : (|\tau_o - \tau| - p)\|\delta_{1o}\|^{2/(1-2\eta)} \leq M\}$. Then there exists a real number $M_0 > 0$ such that, for all $M \geq M_0$,

$$\text{(A.11)} \qquad \inf_{(\Theta,\tau,\Omega)\in B_5^c} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) > 0$$

with probability approaching one. (We use the convention that the infimum over an empty set is ∞.) When $\tau_o - p > \tau$, this is straightforward to conclude from (A.10) because $\sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \geq \sum_{t=\tau+p}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 = (\tau_o - \tau - p)\|\delta_{1o}\|$. An analogous proof applies in the case $\tau_o - p < \tau$. Thus, the result of Theorem 3.1(i) follows from (A.11).

The second part of Theorem 3.1 follows in a similar way by noting that $\tilde{\tau}$ is obtained from the same model as $\hat{\tau}$ by restricting the parameter space such that the $\gamma_j = 0$ $(j = 0, \ldots, p-1)$. *Q.E.D.*

PROOF OF THEOREM 4.1: The central result in the proof of Theorem 4.1 is the following lemma.

LEMMA A.1: *Let* J_{tT} $(t = 1, \ldots, T)$ *be a* (*possibly*) *random vector such that* $\max_{1\leq t\leq T} \|J_{tT}\| = O_p(1)$ *and* J_t *is a vector valued stochastic process satisfying* $\sup_t E\|J_t\| < \infty$. *Then, if* $\hat{\tau} = \tau + O_p(1)$, $T^{-1/2}\sum_{t=p+1}^{T} J_{tT}(d_{t\hat{\tau}} - d_{t\tau}) = o_p(1)$ *and* $T^{-1/2}\sum_{t=p+1}^{T} J_t(d_{t\hat{\tau}} - d_{t\tau}) = o_p(1)$.

Lemma A.1 can be used to show that, when $\delta_1 \neq 0$, the replacement of the true break date τ by the estimator $\hat{\tau}$ has no effect on the asymptotic properties of various second sample moments. This in turn implies the following properties of the parameter estimators: Let $\tilde{\beta}_\xi = \tilde{\beta}(\xi'\tilde{\beta})^{-1}$ 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 $\tilde{\Omega} = \Omega + O_p(T^{-1/2})$. When $\delta_1 = 0$ these results are obtained by different arguments. Then use is made of the fact that the step dummy in equation (2.5) becomes redundant and the effect of the impulse dummies vanishes asymptotically. Based on these results, Theorem 4.1 can be shown in the same way as Theorem 3.1 in S&L, using also arguments from Johansen (1995). *Q.E.D.*

REFERENCES

ARRANZ, M. A., AND A. ESCRIBANO (2000): "Cointegration Testing Under Structural Breaks: A Robust Extended Error Correction Model," *Oxford Bulletin of Economics and Statistics*, 62, 23–52.

BAI, J., R. L. LUMSDAINE, AND J. H. STOCK (1998): "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432.

BANERJEE, A., R. L. LUMSDAINE, AND J. H. STOCK (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.

CAMPOS, J., N. R. ERICSSON, AND D. F. HENDRY (1996): "Cointegration Tests in the Presence of Structural Breaks," *Journal of Econometrics*, 70, 187–220.

GREGORY, A. W., AND B. E. HANSEN (1996): "Residual-Based Tests for Cointegration in Models With Regime Shifts," *Journal of Econometrics*, 70, 99–126.

HANSEN, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

HUBRICH, K., H. LÜTKEPOHL, AND P. SAIKKONEN (2001): "A Review of Systems Cointegration Tests," *Econometric Reviews*, 20, 247–318.

INOUE, A. (1999): "Tests of Cointegrating Rank with a Trend-Break," *Journal of Econometrics*, 90, 215–237.

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

JOHANSEN, S., R. MOSCONI, AND B. NIELSEN (2000): "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend," *Econometrics Journal*, 3, 216–249.

JOHANSEN, S., AND B. NIELSEN (1993): "Manual for the Simulation Program DisCo," Institute of Mathematical Statistics, University of Copenhagen. Available at http://math.ku.dk/~sjo/disco/disco.ps.

LÜTKEPOHL, H., AND P. SAIKKONEN (2000): "Testing for the Cointegrating Rank of a VAR Process with a Time Trend," *Journal of Econometrics*, 95, 177–198.

LÜTKEPOHL, H., P. SAIKKONEN, AND C. TRENKLER (2003): "Comparison of Tests for the Cointegrating Rank of a VAR Process with a Structural Shift," *Journal of Econometrics*, 113, 201–229.

PERRON, P. (1989): "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.

——— (1990): "Testing for a Unit Root in a Time Series with a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153–162.

PERRON, P., AND T. J. VOGELSANG (1992): "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics*, 10, 301–320.

SAIKKONEN, P. (2001): "Consistent Estimation in Cointegrated Vector Autoregressive Models with Nonlinear Time Trends in Cointegrating Relations," *Econometric Theory*, 17, 296–326.

SAIKKONEN, P., AND H. LÜTKEPOHL (2000): "Testing for the Cointegrating Rank of a VAR Process with Structural Shifts," *Journal of Business and Economic Statistics*, 18, 451–464.

——— (2002): "Testing for a Unit Root in a Time Series with a Level Shift at Unknown Time," *Econometric Theory*, 18, 313–348.

SEO, B. (1998): "Tests for Structural Change in Cointegrated Systems," *Econometric Theory*, 14, 222–259.

TODA, H. Y. (1994): "Finite Sample Properties of Likelihood Ratio Tests for Cointegrating Ranks when Linear Trends are Present," *Review of Economics and Statistics*, 76, 66–79.

ZIVOT, E., AND D. W. K. ANDREWS (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.