

Testing for the Cointegrating Rank of a VAR Process with Level Shift at Unknown Time

by

Helmut Lütkepohl
Humboldt University Berlin

Pentti Saikkonen
University of Helsinki

and
Carsten Trenkler
Humboldt University Berlin

Address for correspondence: Helmut Lütkepohl, Institut für Statistik und Ökonometrie, Humboldt-Universität, Spandauer Str. 1, D-10178 Berlin, Germany

Abstract

A systems cointegration rank test is proposed which 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 mean 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 with a GLS procedure. Then the series are adjusted by subtracting the estimated deterministic part and 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 which does not depend on the break date. The performance of the procedure in small samples is investigated by simulations. Finally, the procedure is applied for two sets of example series to illustrate its virtue for econometric analyses.

Key words: Cointegration, structural break, vector autoregressive process, error correction model

*We are grateful to Jürgen Wolters for comments and 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.

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. A range of tests suitable for various kinds of shifts and alternative assumptions regarding the shift date are available. For example, there may be a simple one time shift in the mean, a break in the slope of a time trend or a smooth transition to a new level (see, e.g., Perron (1989, 1990), Lanne, Lütkepohl & Saikkonen (2001)). Moreover, the shift date may be known or it may be assumed to be endogenous (e.g., Evans (1989), Christiano (1992), Perron & Vogelsang (1992), Banerjee, Lumsdaine & Stock (1992), Zivot & Andrews (1992)). If these features are observed in univariate time series, they also have to be taken into account in multiple time series analyses. In particular, 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 & Hansen (1996), Campos, Ericsson & Hendry (1996), Quintos (1998), Seo (1998), Inoue (1999), Johansen & Nielsen (1993), Johansen, Mosconi & Nielsen (2000), Saikkonen & Lütkepohl (2000a) and Lütkepohl, Saikkonen & Trenkler (2000)). 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 which works in this situation. The shift is assumed to be a simple shift in the mean.

The test generalizes the one proposed by Saikkonen & Lütkepohl (2000a) (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 feasible generalized least squares (GLS) 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

to use 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 & Andrews (1992)). Therefore, being able to estimate the break date such that the asymptotic properties of the cointegration tests are unaffected is a major 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 & Lütkepohl (2001)). 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 & 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. Hence, the same tables with critical values may be used as in the case of a known break date.

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 deterministic part of the DGP 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 and applications are presented in Section 6. A summary and conclusions are given in Section 7. All proofs are deferred to Appendix A and data sources are provided in Appendix B.

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 , that is, the stochastic part of 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 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 and the rank of the matrix A are denoted by $\text{tr}(A)$ and $\text{rk}(A)$, 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 . For matrices A_1, \dots, A_s , $\text{diag}[A_1 : \dots : A_s]$ is the block-diagonal matrix with A_1, \dots, A_s on the diagonal. LS, GLS and RR are used to abbreviate least squares, generalized least squares and reduced rank, respectively. LR is short for likelihood ratio, DGP stands for data generation process and VAR and VECM abbreviate vector autoregressive and vector error correction model, respectively. 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, $y_t = (y_{1t}, \dots, y_{nt})'$ ($t = 1, \dots, T$) is assumed to be generated by a process with constant, linear trend and level shift terms,

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

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

$$d_{t\tau} = \begin{cases} 0, & t < \tau, \\ 1, & t \geq \tau, \end{cases} \quad (2.2)$$

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

$$\tau = [T\lambda] \quad \text{with} \quad 0 < \underline{\lambda} \leq \lambda \leq \bar{\lambda} < 1, \quad (2.3)$$

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 et al. (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,

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (2.4)$$

where the A_j are $(n \times n)$ coefficient matrices. For simplicity, it is assumed that $x_t = 0$ for $t \leq 0$ and $\varepsilon_t \sim \text{i.i.d.}(0, \Omega)$, 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 which does not depend on the sample size.

The VECM form of the process x_t is

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

where $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ ($j = 1, \dots, p-1$) 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

$$\Pi = \alpha \beta', \quad (2.6)$$

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 $I(0)$ processes. 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

$$x_t = C \sum_{j=1}^t \varepsilon_j + \xi_t, \quad t = 1, 2, \dots, \quad (2.7)$$

where ξ_t is a zero mean $I(0)$ process.

In the following we use the lag polynomial

$$\begin{aligned} A(L) &= I_n - A_1L - \dots - A_pL^p \\ &= I_n\Delta - \Pi L - \Gamma_1\Delta L - \dots - \Gamma_{p-1}\Delta L^{p-1}. \end{aligned} \tag{2.8}$$

Notice that the relation between the two different parameterizations is given by

$$\begin{aligned} A_1 &= I_n + \alpha\beta' + \Gamma_1, \\ A_j &= \Gamma_j - \Gamma_{j-1}, \quad j = 2, \dots, p-1, \\ A_p &= -\Gamma_{p-1}. \end{aligned} \tag{2.9}$$

Multiplying (2.1) by $A(L)$ yields

$$\begin{aligned} \Delta y_t &= \nu + \alpha(\beta'y_{t-1} - \phi(t-1) - \theta d_{t-1,\tau}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j,\tau} + \varepsilon_t, \\ &t = p+1, p+2, \dots, \end{aligned} \tag{2.10}$$

where $\nu = -\Pi\mu_0 + \Psi\mu_1$, $\phi = \beta'\mu_1$, $\theta = \beta'\delta$ and

$$\gamma_j^* = \begin{cases} \delta, & j = 0, \\ -\Gamma_j\delta, & j = 1, \dots, p-1. \end{cases}$$

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 to estimate the deterministic part of the DGP. In that procedure, first stage estimators for the parameters of the error process x_t , that is, for α , β , Γ_j ($j = 1, \dots, p-1$) and Ω will be based on (2.10). A conventional RR regression will be used although that procedure does not provide exact ML estimators because there are nonlinear restrictions between the parameters in (2.10). Given that the restrictions occur in coefficient vectors of impulse dummies only, ignoring them should not cause any great loss of efficiency. The observations will then be adjusted for deterministic terms and cointegration tests are based on the adjusted series. They will be discussed in Section 4. In the next section estimation of the parameters of the deterministic part as well as p and τ will be considered.

3 Estimation

We will begin by discussing estimators for the break date and the choice of the VAR order and then consider estimation of the parameters of the deterministic terms.

3.1 Choice of τ and p

Suppose for the moment that the VAR order p is known. Then 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,

$$y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_{t\tau}, \quad t = p+1, \dots, T, \quad (3.1)$$

which is obtained from (2.10) 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 $\gamma_j = \gamma_j^*$ ($j = 1, \dots, p-1$). Thus, a linear trend, a step dummy variable and p impulse dummies are included as deterministic terms. The error term in (3.1) is furnished with a subscript τ because τ may be any value from the parameter space $\mathcal{T} = [T\underline{\lambda}, T\bar{\lambda}]$. For a given VAR order p , the shift date is estimated as

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

where 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,

$$y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_{t\tau}^*, \quad t = p+1, \dots, T, \quad (3.3)$$

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

$$\tilde{\tau} = \arg \min_{\tau \in \mathcal{T}} \det \left(\sum_{t=p+1}^T \hat{\varepsilon}_{t\tau}^* \hat{\varepsilon}_{t\tau}^{*'} \right) \quad (3.4)$$

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 is shown in Appendix A.

Theorem 3.1. If $\delta_1 \neq 0$, then

- (i) $\hat{\tau} - \tau = o_p(T^{2\eta})$ for some η with $\frac{1}{6} < \eta \leq \frac{1}{4}$ and

(ii) $\tilde{\tau} - \tau = O_p(1)$.

Here τ denotes the true shift date. □

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, its convergence rate is better than that of $\hat{\tau}$ which is based on a correctly specified model under our assumptions. Admittedly, this may be an artefact of the proof used to establish Theorem 3.1. On the other hand, the relatively poor performance of $\hat{\tau}$ has also some intuitive plausibility. The impulse dummy variables in (3.4) may react to exceptional observations or ‘outliers’ which may occur in periods close to the true break date. If only low order moments of the innovation process ε_t exist, so that the value of η is only slightly smaller than $\frac{1}{4}$, the probability of exceptional observations can be rather large and thus there is some chance of capturing an outlier with an impulse dummy instead of finding the level shift. Thereby a value of $\hat{\tau}$ away from the shift date τ may be obtained and the convergence rate of $\hat{\tau}$ may suffer. In contrast, the estimator $\tilde{\tau}$ is not likely to be sensitive to outliers because the impulse dummies are omitted.

Although the result in Theorem 3.1 and the foregoing discussion may suggest a preference for the shift date estimator $\tilde{\tau}$, it is worth emphasizing that the result in the theorem relies on the condition $\delta_1 \neq 0$. Note, however, that $\delta_1 = 0$ may hold even if $\delta \neq 0$. In that case, neither $\hat{\tau}/T$ nor $\tilde{\tau}/T$ are consistent estimators of the sample fraction λ . 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.

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 $p = 1, \dots, p_{\max}$, 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.

3.2 Estimating the Parameters of the Deterministic Part

Once the VAR order is specified and the break date is estimated, the parameters of the deterministic part of the process may be considered. We present the procedure for some VAR order p and break date estimator $\hat{\tau}$ which may be those obtained in the procedure described in the foregoing, of course. However, instead of $\hat{\tau}$, the estimator $\tilde{\tau}$ or any other estimator satisfying Theorem 3.1(i) could be used. We shall estimate the parameters μ_i ($i = 0, 1$) and δ in (2.1) by using a feasible GLS approach as in Saikkonen & Lütkepohl (2000b). Defining

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

and multiplying (2.1) from the left by $A(L)$ gives

$$A(L)y_t = H_{0t}\mu_0 + H_{1t}\mu_1 + K_{t\tau}\delta + \varepsilon_t, \quad t = 1, 2, \dots \quad (3.5)$$

Here the y_t are set to zero for $t \leq 0$, $H_{it} = A(L)a_{it}$ ($i = 0, 1$) and $K_{t\tau} = A(L)d_{t\tau}$. The matrix

$$Q = [\Omega^{-1}\alpha(\alpha'\Omega^{-1}\alpha)^{-1/2} : \alpha_{\perp}(\alpha'_{\perp}\Omega\alpha_{\perp})^{-1/2}] \quad (3.6)$$

has the property $QQ' = \Omega^{-1}$. Therefore we premultiply (3.5) by Q' to transform the covariance matrix of the error term to an identity matrix, as required in GLS estimation. Thereby we have a transformation which results in a (multivariate) regression model with standard properties of the error term.

Of course, the above transformation is not feasible because it involves unknown parameters. For a feasible transformation we need suitable estimators of the parameters α , β , Γ_j ($j = 1, \dots, p-1$) and Ω which can be obtained by a RR regression of (2.10) using the rank which is specified in the null hypothesis to be tested in the next 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 . These estimators will be denoted by \tilde{A}_j . We define $\tilde{A}(L) = I_n - \tilde{A}_1L - \dots - \tilde{A}_pL^p$, $\tilde{H}_{it} = \tilde{A}(L)a_{it}$ ($i = 0, 1$) and $\tilde{K}_{t\tau} = \tilde{A}(L)d_{t\tau}$.

A suitable estimator of Q can be obtained by forming $\tilde{\alpha}_\perp$ from $\tilde{\alpha}$ and replacing Ω , α and α_\perp in the definition of Q by their estimators. Denoting by \tilde{Q} the resulting estimator of Q leads to the multivariate auxiliary regression model

$$\tilde{Q}'\tilde{A}(L)y_t = \tilde{Q}'\tilde{H}_{0t}\mu_0 + \tilde{Q}'\tilde{H}_{1t}\mu_1 + \tilde{Q}'\tilde{K}_{t\hat{\tau}}\delta + \eta_{t\hat{\tau}}, \quad t = 1, \dots, T. \quad (3.7)$$

The LS estimators of the parameters μ_i ($i = 0, 1$) and δ from this auxiliary model, denoted by $\hat{\mu}_i$ ($i = 0, 1$) and $\hat{\delta}$, respectively, are used in the next section to obtain tests for the cointegrating rank.

4 Cointegrating Rank Tests

We wish to test the null hypothesis

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

Once the estimators $\hat{\mu}_i$ ($i = 0, 1$) and $\hat{\delta}$ are determined as described in the previous section, a sample analog of the series x_t can be obtained as

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

or, alternatively, with $\hat{\tau}$ replaced by $\tilde{\tau}$. 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 (2.5). More precisely, the test statistic can be determined from

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

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 $\Delta\hat{X}_{t-1} = [\Delta\hat{x}'_{t-1} : \dots : \Delta\hat{x}'_{t-p+1}]'$. Denoting the resulting ordered eigenvalues by $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n$, the LR type statistic for the pair of hypotheses in (4.1) becomes

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

The limiting distribution is given in 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 (4.1) is true and the assumptions of the previous section hold,

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

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

Notice that the result of this theorem holds even if $\theta = 0$ in (2.10) 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 $r = 0$. The limiting distribution obtained in Theorem 4.1 is free of unknown nuisance parameters. It is actually 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 & 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.

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 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 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 and, from the proofs given in Appendix A and in that paper, it can be seen that the above conclusion actually holds. 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, 1995) which was also used by some other authors for investigating the properties of cointegrating rank tests (see, e.g., Hubrich, Lütkepohl & Saikkonen (2001)):

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

where $\boldsymbol{\psi} = \text{diag}(\psi_1, \dots, \psi_r)$ and Θ are $(r \times r)$ and $(r \times (n - r))$ matrices, respectively. As shown by Toda, this type of process may be regarded as a canonical form 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 which leave such tests invariant. We have used two-, three- and four-dimensional processes for simulations and report some of the results from the three-dimensional DGPs in more detail here because they are representative for other cases as well. We will analyze processes with cointegrating rank $r = 0$ and $r = 1$ because they allow over- as well as under-estimation of r . If $r = 1$, $\boldsymbol{\psi} = \psi_1$ 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 we use $\mu_i = 0$ ($i = 0, 1$) throughout. In other words, the intercept and trend terms are actually zero. Hence, $y_t = \delta d_{t\tau} + x_t$ and 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 are simulated starting with initial values of zero and discarding the first 50 values as presample observations. Thereby we average out the effects of initial values in our simulation results. Sample sizes of $T = 100$ and $T = 200$ are considered, but in general we focus on the results for $T = 100$. The number of replications is 1000. Thus the standard error of an estimator of a true rejection probability P is $s_P = \sqrt{P(1 - P)/1000}$, e.g., $s_{0.05} = 0.007$. Linear trend terms are 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 use different VAR orders p , although the true order is $p = 1$, to explore the impact of this quantity on

the estimation and testing results.

Table 1 and Figures 1-3 contain results for $T = 100$ and a shift date $\tau = 50$, that is, the shift occurs roughly in the middle of the sample. The search procedures are applied to all possible break points between the 5th and the 96th observation. Results are shown for different values of the shift parameters $\delta = (0, 0, \delta_3)'$, alternative VAR orders p , different values of Θ , and both shift date estimators $\hat{\tau}$ and $\tilde{\tau}$. For comparison purposes, we also report test results for the case where the true shift date is assumed to be known. The tests are based on asymptotic critical values for a nominal significance level of 5%.

We first consider the VAR order $p = 1$. Clearly, the actual shift date is found with greater reliability if an impulse dummy is included in addition to a shift dummy. For example, for $\psi_1 = 0.9$, $\Theta = (0, 0)$ and $\delta = (0, 0, 5)'$, $\tilde{\tau}$ finds the true shift date in 19% of the cases, whereas the relative frequency of $\hat{\tau} = 50$ is 75% (see Table 1). Obviously, it is easier for both procedures to find the correct break date if the parameter δ_3 is larger or if strong instantaneous correlation between the components is present ($(\theta_1, \theta_2) = (0.4, 0.8)$) (compare Figure 1).

The properties of the cointegrating rank tests do not deteriorate substantially if the shift date has to be estimated relative to the situation where it is assumed to be known. In general we observe an increase in the sizes of the tests if the break date is estimated (see Table 1). Using $\tilde{\tau}$, the tests are more affected by this increase. In case of strong instantaneous correlation ($\theta_1 = 0.4$, $\theta_2 = 0.8$) the power is slightly lower if the shift date is estimated by $\tilde{\tau}$ compared to the power obtained with $\hat{\tau}$, whereas the reverse result follows if correlation is absent ($\theta_1 = \theta_2 = 0$) (see Figure 1). The sizes of the cointegration tests have a tendency to decrease for increasing δ_3 . For large values of δ_3 , the search procedure finds more often the correct shift point and thus, the sizes of the tests resemble the lower size in case of a known shift date. As a consequence, the power of the tests also decreases with larger values of δ_3 .

Generally, although the search procedure including a shift dummy only is less successful in estimating the shift date ($\tilde{\tau}$), the properties of the cointegration tests based on this estimation method do not differ much from the approach which includes an impulse dummy in addition ($\hat{\tau}$). Hence, in the presently considered case where $p = 1$, if the cointegrating rank of the process is the only quantity of interest, using $\tilde{\tau}$ may not lead to significant distortions.

Table 1: Estimated Frequency Distributions for Three-Dimensional DGP with Cointegrating Rank $r = 1$, $\theta_1 = \theta_2 = 0$, $\psi_1 = 0.9$, Sample Size $T = 100$, Break Point $\tau = 50$ and Nominal Significance Level 5% in each Individual Test Based on 1000 Replications

VAR order	p	true shift date	$\delta = 1\iota^a$	$\delta = 2\iota$	$\delta = 3\iota$	$\delta = 5\iota$	$\delta = 7\iota$	$\delta = 10\iota$
1			shift date estimator $\hat{\tau}$ (without impulse dummies)					
	< 50		0.456	0.455	0.450	0.407	0.328	0.184
	= 50		0.013	0.035	0.061	0.193	0.384	0.692
	> 50		0.531	0.510	0.489	0.400	0.288	0.124
	$r = 0$	0.078	0.116	0.113	0.113	0.107	0.102	0.095
	$r = 1$	0.009	0.018	0.018	0.020	0.019	0.018	0.016
	$r \geq 2$	0.005	0.009	0.008	0.008	0.007	0.009	0.005
3			shift date estimator $\hat{\tau}$ (without impulse dummies)					
	< 50		0.455	0.455	0.444	0.383	0.301	0.164
	= 50		0.021	0.034	0.064	0.199	0.396	0.684
	> 50		0.524	0.511	0.492	0.418	0.303	0.152
	$r = 0$	0.082	0.141	0.138	0.142	0.128	0.119	0.106
	$r = 1$	0.010	0.020	0.020	0.020	0.022	0.015	0.014
	$r \geq 2$	0.003	0.005	0.009	0.009	0.005	0.005	0.006
1			shift date estimator $\hat{\tau}$ (with impulse dummies)					
	< 50		0.472	0.450	0.373	0.132	0.010	0.000
	= 50		0.021	0.083	0.238	0.748	0.982	1.000
	> 50		0.507	0.467	0.389	0.120	0.008	0.000
	$r = 0$	0.078	0.097	0.093	0.087	0.081	0.079	0.078
	$r = 1$	0.009	0.017	0.018	0.019	0.012	0.009	0.009
	$r \geq 2$	0.005	0.005	0.006	0.079	0.005	0.005	0.005
3			shift date estimator $\hat{\tau}$ (with impulse dummies)					
	< 50		0.470	0.481	0.510	0.610	0.634	0.643
	= 50		0.014	0.043	0.098	0.243	0.348	0.357
	> 50		0.516	0.476	0.392	0.147	0.018	0.000
	$r = 0$	0.082	0.138	0.131	0.125	0.105	0.091	0.092
	$r = 1$	0.010	0.025	0.020	0.017	0.017	0.013	0.014
	$r \geq 2$	0.003	0.009	0.006	0.005	0.003	0.002	0.004

$${}^a\iota = (0, 0, 1)'$$

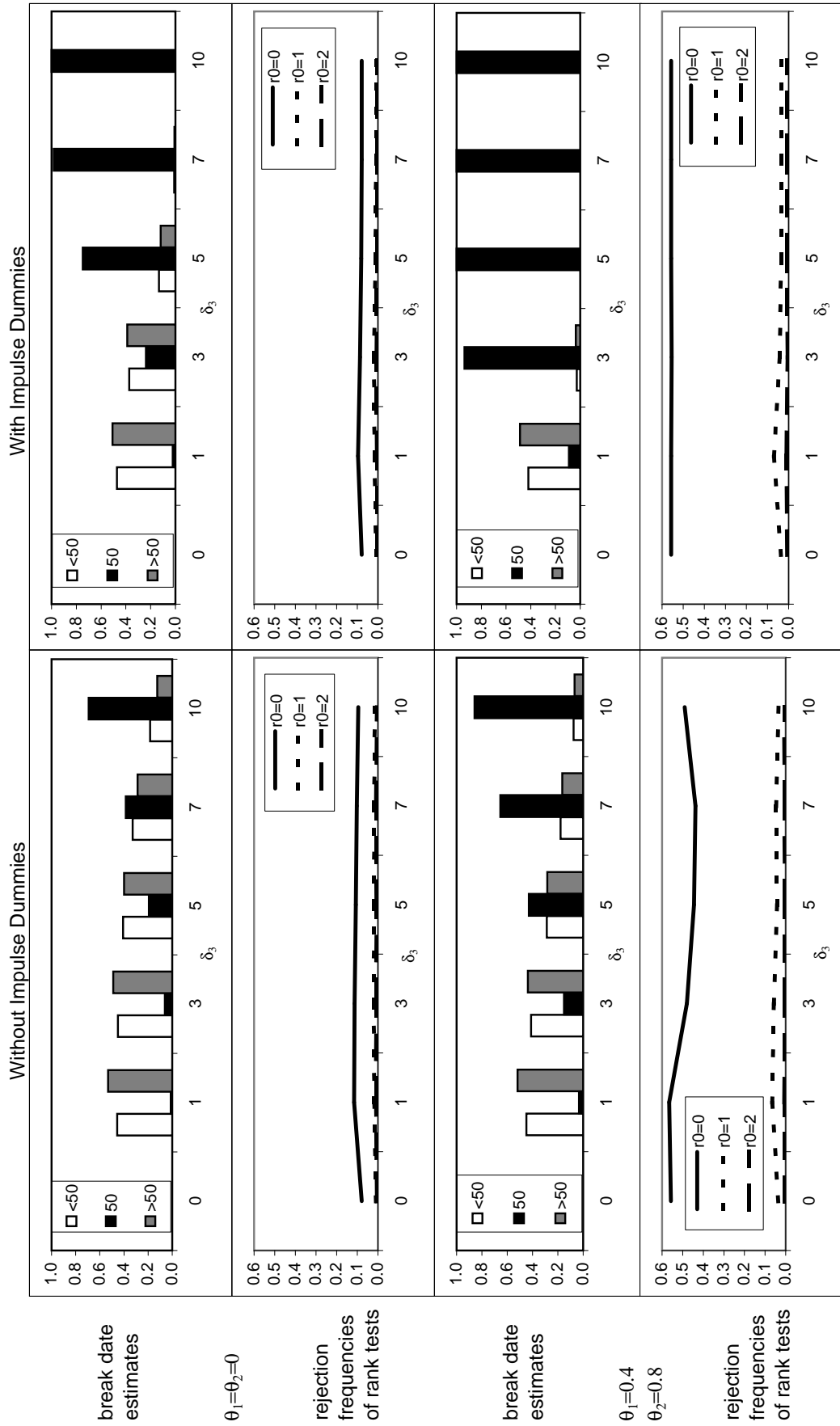


Figure 1: Break point estimates and rejection frequencies for three-dimensional DGPs with $r = 1$, $T = 100$ and a VAR order $p = 1$.

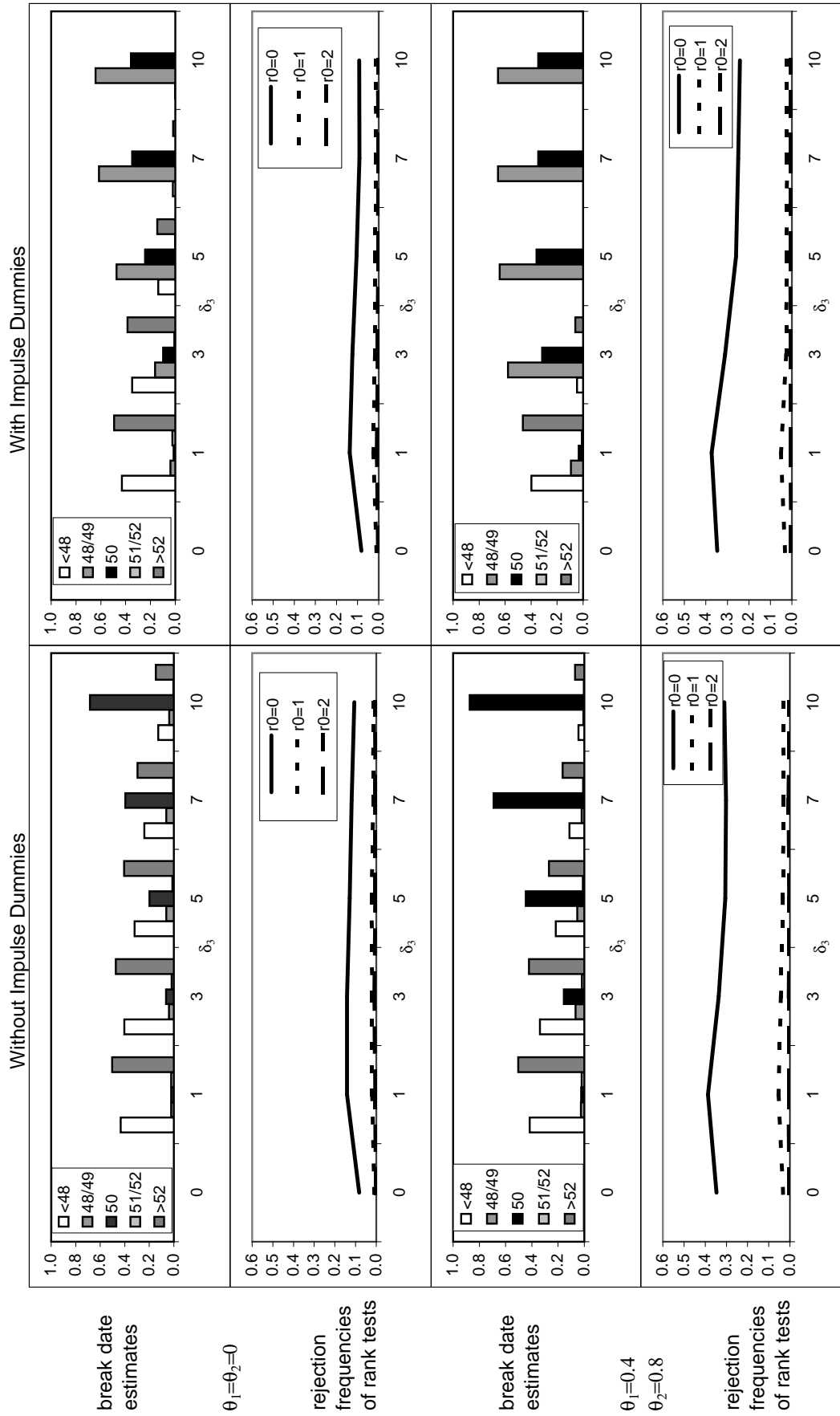


Figure 2: Break point estimates and rejection frequencies for three-dimensional DGPs with $r = 1, T = 100$ and a VAR order $p = 3$.

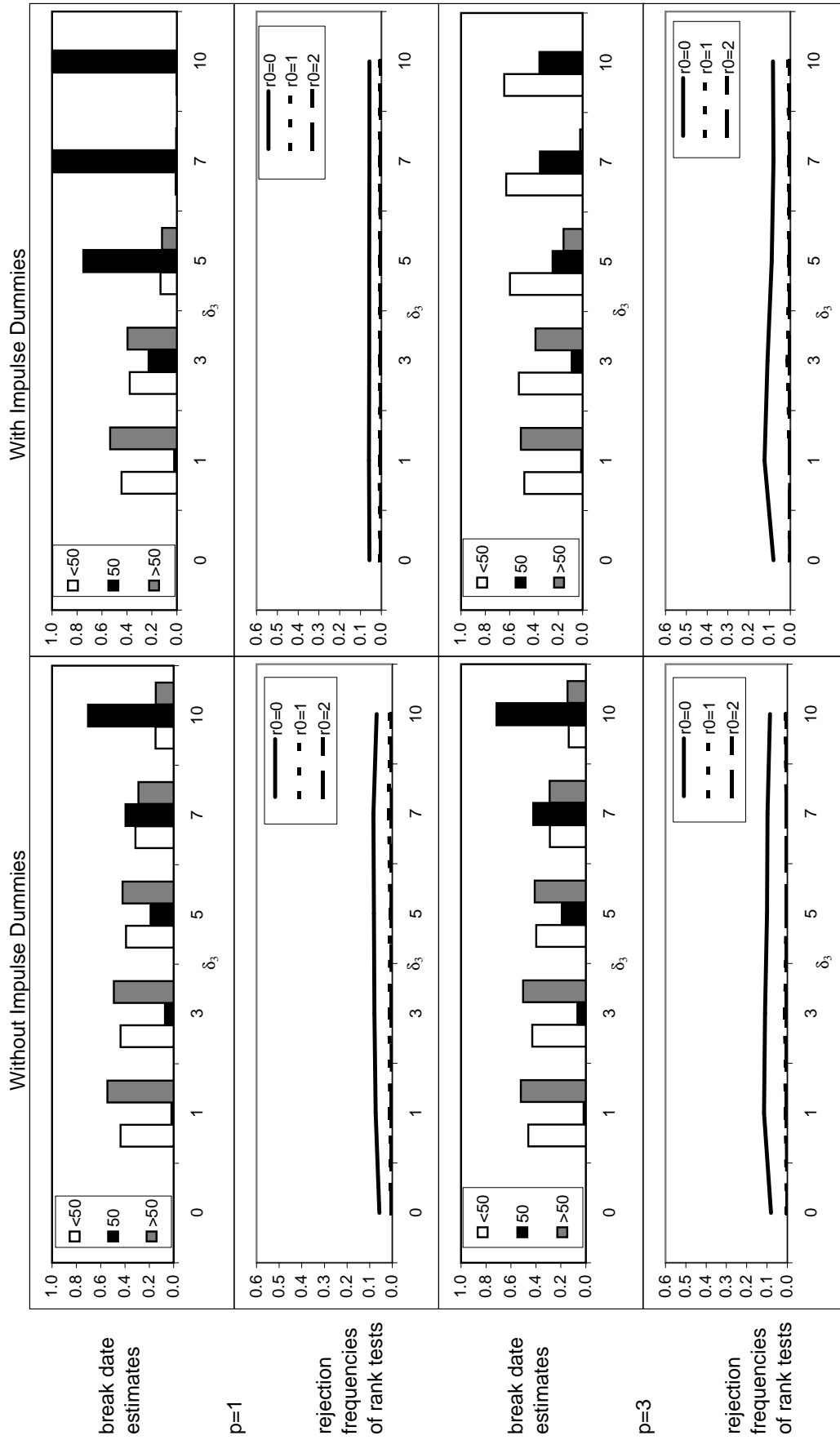


Figure 3: Break point estimates and rejection frequencies for three-dimensional DGPs with $r = 0$, $T = 100$ and a VAR order $p = 1$.

Obviously, the VAR order has some impact on estimating the break point if impulse dummies are included. For $p = 3$, the shift date is found too early compared to the true break date (see Table 1 and Figure 2). Most of the wrong estimations are located within the $p - 1$ periods before the actual break as can be seen in Figure 2. 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. This also means that $\tilde{\tau}$ is more reliable for larger values of δ_3 if the VAR order is $p = 3$.

In line with the findings in other studies, overstating the VAR order leads to a loss of power (e.g., Lütkepohl & Saikkonen (1999)). Surprisingly, the poor performance of $\hat{\tau}$ when $p = 3$, affects the power only for strong error term correlation ($\theta_1 = 0.4$, $\theta_2 = 0.8$) if $\psi_1 = 0.9$ ($r = 1$) (see Figure 2). In these situations we observe a lower small sample power for large values of δ_3 compared to the cases where the true shift date or the estimator $\hat{\tau}$ is used. However, for $\psi_1 = 0.9$ and $\theta_1 = \theta_2 = 0$ the power for $p = 3$ is in fact slightly larger than the power for the true order $p = 1$ (see Table 1 or Figure 1). Concerning the sizes of the cointegration tests, it is also interesting to note that overstating the VAR order leads to some size distortion (overrejection) if the true cointegrating rank r is zero (Figure 3). This behavior is observed for both shift date estimators. Again, $\tilde{\tau}$ clearly finds the true break date more often if $p = 3$, whereas $\hat{\tau}$ is preferable for $p = 1$.

We have studied several other issues related to our DGP as well. In the following we comment on some of the results without providing detailed figures or tables. Variations in the parameter ψ_1 have a similar impact on the sizes and powers of the cointegration tests for both break date estimators. For instance, setting $\psi_1 = 0.7$ increases the power of the tests and improves the size properties in a similar way for both break date estimators. Moreover, setting $\psi_1 = 0.7$ does not change the frequencies of finding the true break date in any important way.

We have also considered a sample length of $T = 200$ and a shift date $\tau = 100$. The frequency of finding the correct break date drops somewhat compared to the sample length $T = 100$ if $\tilde{\tau}$ and $p = 1$ is used because the relative importance of the shift declines and

there are more alternative periods over which the search is performed. In contrast, using $\hat{\tau}$ produces similar figures as for $T = 100$. However, the small sample properties of the cointegration tests improve for both break date estimators. Again, the overall performance of the tests is almost invariant to the choice of break date estimator.

We further studied different shift dates. Using a sample size $T = 100$ and setting $\tau = 80$, we obtained virtually the same results as for $\tau = 50$. However, if the break date was close to the beginning or the end of the sample, say in the first or last 10% of the observations, it was much more difficult to detect the correct shift date. Nevertheless, the small sample power of the cointegration tests deteriorated only slightly.

Finally, we have allowed for breaks in all three components of the process, that is, all three components of δ were nonzero. In the absence of innovation correlation the correct shift date was found much more often compared to the case of a shift in only one component. By contrast, in case of high innovation correlation ($\Theta = (0.4, 0.8)$) the frequency of finding the correct break point was clearly lower. However, this did not affect the small sample properties of the cointegration tests.

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 tests. Therefore, both estimators, $\hat{\tau}$ and $\tilde{\tau}$, may be used and our procedure appears to be suitable for empirical work. We will discuss applications in the next section.

6 Applications

We have applied our testing procedure to two different systems of economic time series. The first one consists of exchange rates of the Polish zloty versus a set of currencies. In this system the shift date is actually known. We still consider this system here to check that our procedure works for real data sets. For this purpose we compare the results for known break dates to those obtained from our method without assuming knowledge of the shift date. The second system describes the price-wage dynamics in the Polish economy for which a break is likely at the time when the socialist system changed to a market economy. In this case it is not quite clear where the shift actually has occurred. We begin with the exchange rate system.

Table 2: Cointegration Tests for Polish Crawling Peg System

H_0	$LR_{\hat{\tau}=123}$	$LR_{\hat{\tau}=125}$	LR_{GLS}	critical values		
				10%	5%	1%
$r_0 = 0$	81.57***	125.15***	39.72	61.81	65.69	73.57
$r_0 = 1$	27.85	28.74	27.76	42.03	45.12	51.27
$r_0 = 2$	10.62	10.63	11.26	25.90	28.47	33.54
$r_0 = 3$	4.52	5.67	5.50	13.89	15.92	20.37
$r_0 = 4$	0.25	1.65	2.44	5.43	6.83	10.19

Note: *** denotes significance at a 1% level. Critical values are taken from Lütkepohl & Saikkonen (2000, Table 1).

6.1 Polish Crawling Peg System

Trenkler (2000) has analyzed the relationship between the Polish zloty and the basket currencies within the crawling peg system Poland established on October 15, 1991. In this system the zloty is tied to a currency basket and devalued with a preannounced monthly rate (rate of crawl). The currencies comprising the basket were the U.S. dollar (45% weight), German mark (35%), British pound (10%), French franc (5%) and Swiss franc (5%). Assuming that the monetary authorities can keep the actual zloty rate in line with the theoretical rate determined by the crawling peg, Trenkler (2000) derives a long-run relationship between the logarithms of the zloty - U.S. dollar rate (zl_t) and the basket currencies denominated in U.S. dollars (gm_t, bp_t, ff_t, sf_t), where the variables denote the logarithms of the exchange rates of the German mark, the British pound, the French franc and the Swiss franc to the U.S. dollar, respectively. He tests for this relationship within the cointegration framework using daily data from March 1, 1993 to February 28, 1994. This period includes a discrete step devaluation of the Polish zloty by 7.4% against the basket on August 27, 1993, which can be seen in Figure 4. This level shift is modeled with a shift dummy within a VAR system of order $p = 3$ consisting of $zl_t, gm_t, bp_t, ff_t, sf_t$, a constant and a linear trend. Applying the test suggested by S&L for the case of a known break date and a cointegration test suggested by Johansen & Nielsen (1993), Trenkler (2000) finds a cointegrating rank of one confirming the existence of the postulated long-run relationship.

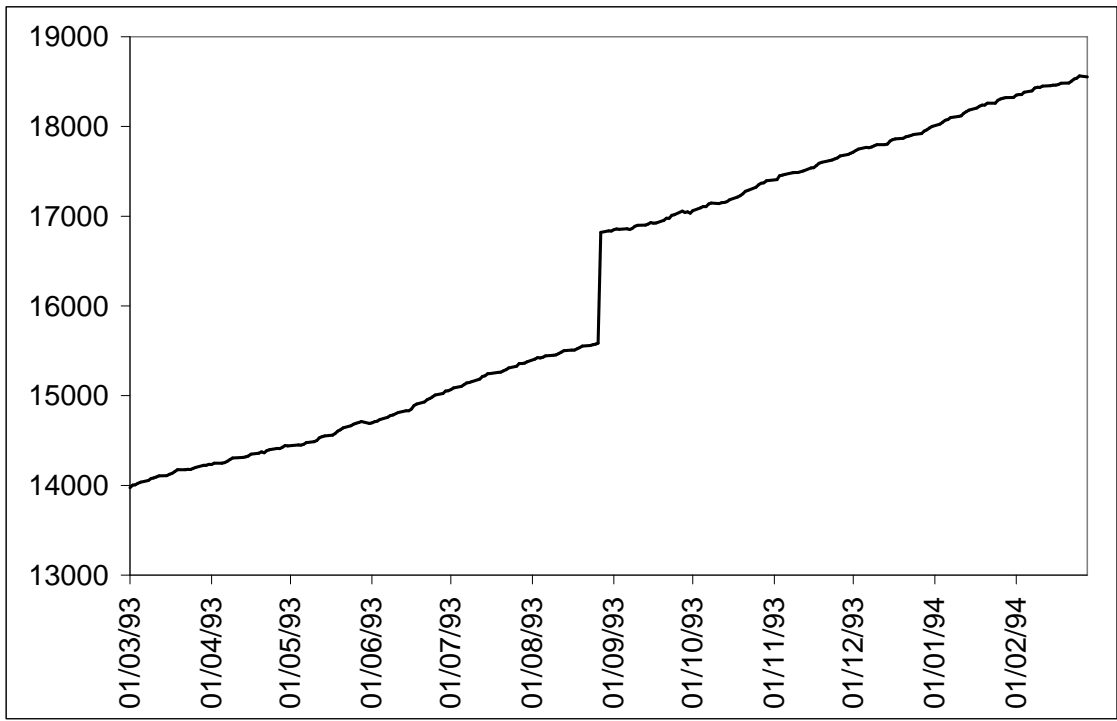


Figure 4: Zloty per unit of currency basket

We estimated the break date although it is actually known to see how our procedure works in practice. Hence, we fit a VAR process of order $p = 3$ with the same variables and deterministic terms as in Trenkler (2000). In this case we get a break date estimate $\hat{\tau} = 123$ which corresponds to August 25, 1993, two days before the devaluation, whereas $\tilde{\tau} = 125$ estimates the break date correctly. This outcome is in line with our simulation evidence and asymptotic theory, which suggest that $\tilde{\tau}$ may find the true break date more often for VAR orders greater than 1. Notice, however, that the cointegration test finds one cointegrating vector at a 1% level for both break dates, although the values of the test statistics differ (see Table 2). In Table 2 we also report a cointegration test (LR_{GLS}) suggested by Saikkonen & Lütkepohl (2000b) which does not allow for structural shifts but proceeds otherwise in the same way as the tests presented in the current study. It also has the same asymptotic null distribution so that the critical values in the table may be used. Clearly, this test does not find a cointegration relation, which shows the benefit of applying tests accommodating level shifts.

6.2 Price-Wage Dynamics in Poland

In a series of papers Golinelli & Orsi (1994, 1998, 2000) have studied the price-wage dynamics of economies in transition from a centrally planned to a market economy. Poland is one of the countries under consideration. Marcellino & Mizon (2000) also studied the wage-price relationship in Poland and other European countries.

Referring to Golinelli & Orsi (1998) we have built a model for Poland consisting of cpi_t , the consumer price index, w_t , the nominal wage, ex_t , the exchange rate against the U.S. dollar, lp_t , labor productivity, and $epop_t$, the employment-population ratio. All variables are in natural logs. By including the exchange rate, the impact of this variable on inflation and employment can be captured. From an economic point of view, at least two cointegrating relationships describing a price and a wage equation are expected to exist. The theoretical basis for these equations is a simple mark-up pricing system where price movements are linked to cost movements (Golinelli & Orsi, 1994). Also, a third cointegrating relationship explaining the labor productivity may exist (Golinelli & Orsi, 1998). In our analysis we used quarterly, seasonally non-adjusted data for the period 1970:1 - 1998:4 (see Figure 5).

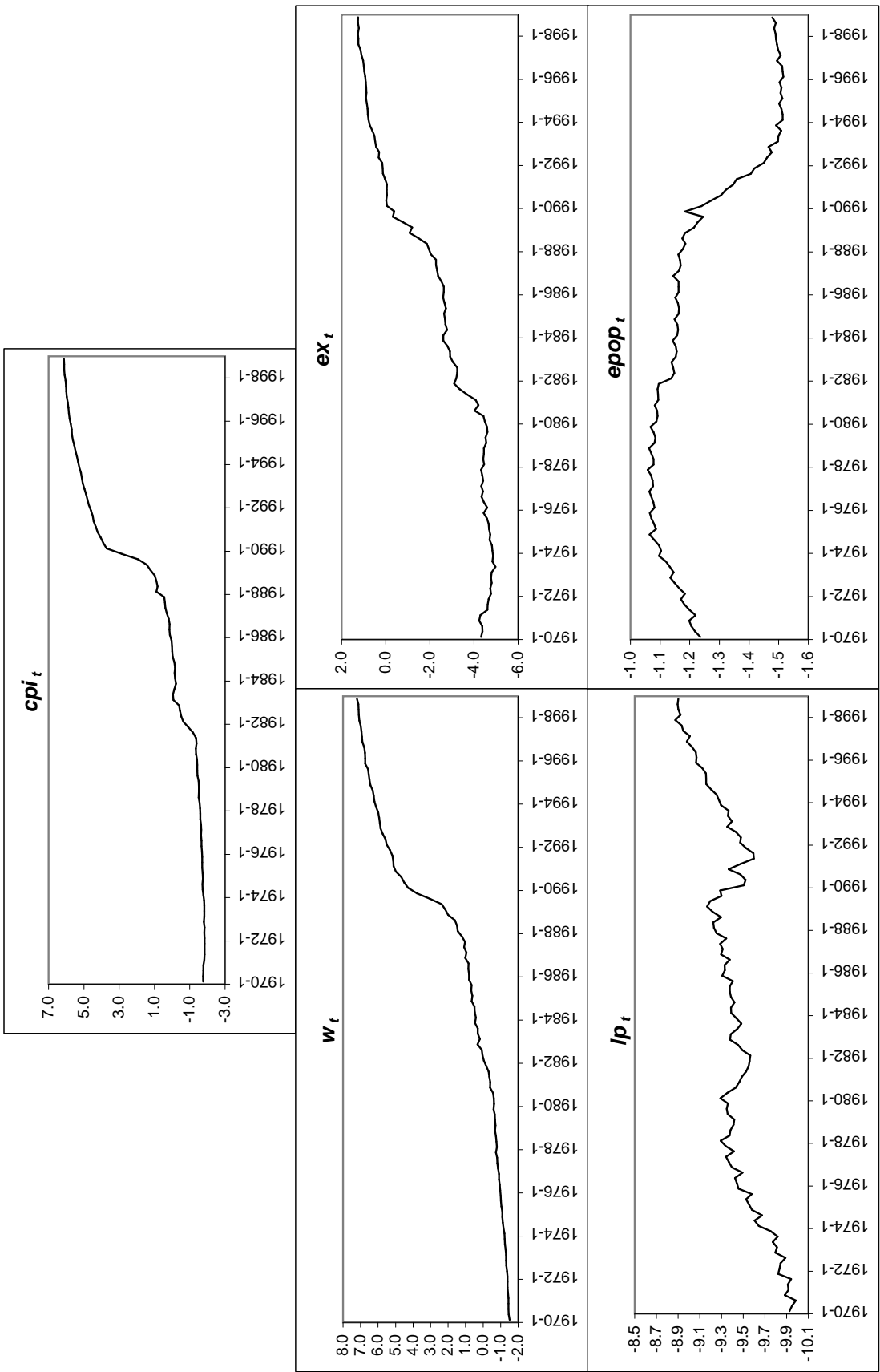


Figure 5: Time series used in wage-price model for Poland

Table 3: Dependence of Break Date Estimations on the VAR Order p

	shift estimation without seasonals		shift estimation with seasonals	
	$\hat{\tau}$	$\tilde{\tau}$	$\hat{\tau}_s$	$\tilde{\tau}_s$
$p = 1$	79 (1989:3)	79 (1989:3)	80 (1989:4)	79 (1989:3)
$p = 2$	79 (1989:3)	79 (1989:3)	79 (1989:3)	79 (1989:3)
$p = 3$	79 (1989:3)	79 (1989:3)	79 (1989:3)	79 (1989:3)
$p = 4$	79 (1989:3)	82 (1990:2)	79 (1989:3)	79 (1989:3)
$p = 5$	77 (1989:1)	80 (1989:4)	77 (1989:1)	80 (1989:4)
$p = 6$	73 (1988:1)	80 (1989:4)	73 (1988:1)	80 (1989:4)
$p = 7$	73 (1988:1)	80 (1989:4)	73 (1988:1)	80 (1989:3)
$p = 8$	73 (1988:1)	80 (1989:4)	73 (1988:1)	79 (1989:3)

One of the main problems in the econometric modeling process is the occurrence of structural breaks in the Polish time series. As can be seen in Figure 5, there exist two potential periods of changes, a first one in the early 1980s and a second one during 1989 and 1990, the beginning of the transition process. Marcellino & Mizon (2000) used annual data until 1989 for a similar system and assumed a break after 1980. Obviously, the exact time points of the breaks are difficult to see from the graphs. Therefore we have applied our shift date estimation procedures.

In Table 3 we present the shift date estimates for different VAR orders. We obtained $\hat{\tau} = 79$ (1989:3) for VAR orders $p = 1, \dots, 4$, $\hat{\tau} = 77$ (1989:1) for $p = 5$, and $\hat{\tau} = 73$ (1988:1) for $p = 6, \dots, 8$. If $\tilde{\tau}$ is used, we obtained an estimate $\tilde{\tau} = 79$ (1989:3) for an order up to $p = 3$, $\tilde{\tau} = 82$ (1990:2) for $p = 4$, and $\tilde{\tau} = 80$ (1989:4) for $p = 5, \dots, 8$. The differences in the break date estimates with respect to the higher VAR orders are in line with our simulation evidence. Our simulations in the previous section have shown that $\hat{\tau}$ tends to estimate the break date too early if higher values of p are applied. The results in Table 3 reflect this property of the break date estimator $\hat{\tau}$.

The shift date estimations reported in the first two columns of Table 3 were conducted without including seasonal dummy variables in (3.1) or (3.3) although the time series are not seasonally adjusted. Therefore, we also incorporated seasonal dummy variables into both

Table 4: Joint Determination of the Break Date and the VAR Order p

criteria	$\hat{\tau}$ (without seasonals)	$\tilde{\tau}$ (without seasonals)
AIC	$p = 8$ and $\hat{\tau} = 73$	$p = 8$ and $\tilde{\tau} = 80$
SC	$p = 2$ and $\hat{\tau} = 79$	$p = 2$ and $\tilde{\tau} = 79$
HQ	$p = 7$ and $\hat{\tau} = 73$	$p = 2$ and $\tilde{\tau} = 79$
	$\hat{\tau}_s$ (with seasonals)	$\tilde{\tau}_s$ (with seasonals)
AIC	$p = 8$ and $\hat{\tau}_s = 73$	$p = 7$ and $\tilde{\tau}_s = 80$
SC	$p = 1$ and $\hat{\tau}_s = 80$	$p = 2$ and $\tilde{\tau}_s = 79$
HQ	$p = 8$ and $\hat{\tau}_s = 73$	$p = 2$ and $\tilde{\tau}_s = 79$

estimation procedures and obtained the estimates shown in the last two columns of Table 3. Obviously, the inclusion of seasonal dummies does not have an important impact on the break date estimation.

So far, we have not discussed the choice of the VAR order p . As suggested in Section 3.1, one possible approach is to use the VAR order that minimizes a standard model selection criterion given the estimated break date for the respective order p . Specifically, we computed the values of information criteria for each order $p = 1, \dots, 8$, applying the respective break date that has been estimated for the considered order. This was done for the break date estimates emerging from both estimation procedures with and without seasonal dummy variables, that is, the computations are conducted for the four sets of break date estimates given in Table 3. Thus, we may obtain different pairs of VAR orders and break dates depending on the estimation procedure used to determine the shift date. The computations of the information criteria were performed on the basis of the model (3.1) augmented by seasonal dummy variables because this is the model underlying the cointegration tests applied later. We used the information criteria AIC, SC and HQ (see Lütkepohl (1991, Chapter 4) for definitions). The results are shown in Table 4.

Whether or not seasonal dummies are included in estimating the break dates does not make much difference whereas the choice of information criterion is important. As usual, the SC criterion prefers smaller lag orders than AIC and HQ. We also applied single equation LM tests for autocorrelated residuals and a vector error autocorrelation test (see Doornik &

Table 5: Cointegration Tests for the Polish Price-Wage Model

H_0	$p = 8$		$p = 7$		$p = 6$		critical values		
	$LR_{\hat{\tau}=73}$	$LR_{\tilde{\tau}=80}$	$LR_{\hat{\tau}=73}$	$LR_{\tilde{\tau}=80}$	$LR_{\hat{\tau}=73}$	$LR_{\tilde{\tau}=80}$	10%	5%	1%
$r_0 = 0$	74.06***	80.97***	56.86	90.35***	69.41**	67.74**	61.81	65.69	73.57
$r_0 = 1$	42.13*	41.40	39.86	45.33**	57.04***	45.46**	42.03	45.12	51.27
$r_0 = 2$	26.27*	13.77	19.48	15.09	26.45*	20.93	25.90	28.47	33.54
$r_0 = 3$	5.76	5.50	5.64	4.65	9.87	4.91	13.89	15.92	20.37
$r_0 = 4$	0.13	0.05	0.00	0.02	0.63	0.10	5.43	6.83	10.19

Note: *, **, *** denote significance at a 10%, 5% and 1% level respectively. Critical values are taken from Lütkepohl & Saikkonen (2000, Table 1).

Hendry, 1997). These tests indicated that $p = 2$ leaves some autocorrelation in the estimation residuals. Therefore higher VAR orders are used in the following analysis.

In Table 5 we have collected the results of the cointegration tests using different VAR orders and corresponding break date estimates. Obviously, there is some evidence for at least two cointegrating relations. According to our simulation results, estimating the break date by $\hat{\tau}$ and thereby dating the break date possibly too early does not necessarily lead to a loss of power of the cointegrating rank tests. In fact, for $p = 6$ and $p = 8$, using a break date estimate $\hat{\tau} = 73$ leads to stronger evidence for two or, depending on the significance level, three cointegrating relations than with $\tilde{\tau} = 80$. Only for $p = 7$, the cointegration tests based on $\hat{\tau} = 73$ do not even reject a zero cointegrating rank whereas $\tilde{\tau} = 80$ leads to clear evidence for two cointegrating relations.

Golinelli & Orsi (1998) used data for the period 1970:1-1995:3 and applied Johansen LR tests including shift and impulse dummies but using critical values for the case without level shifts, that is, they have used inadequate critical values. In their specification they implemented breaks in 1981:1 and 1989:4. Golinelli & Orsi do not explain how they find these shift dates. Our previous analysis has shown that modeling of a shift in 1981 is not necessary to get rid of error term autocorrelation. Also, the residuals of our higher order models used in the cointegration tests do not show a break in the early 1980s. Nevertheless, we have explored the possibility of an additional break by including a second shift term and searching over the whole sample again. Due to multicollinearity of the dummy variables one

has to exclude the observation of the first break date estimate, of course. In case of $\hat{\tau}$ the preceding p periods also have to be excluded because of the use of impulse dummy variables. We have applied our search procedures with and without seasonal dummies. The search procedures suggested shift dates in the late 1980s close to the ones modeled already. This result supports the visual impression from Figure 5 that the first period at the beginning of the 1980s is less likely to contain a shift.

7 Conclusions

We have proposed a procedure for testing the cointegrating rank of VAR processes with a structural shift in the mean 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 which 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 which 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 has still advantages in small samples and may therefore be useful in practice.

Once the break date has been estimated and a suitable lag order is determined on the basis of standard model selection criteria, the deterministic part of the process including the shift size is estimated with a feasible GLS procedure. Then the series are adjusted by subtracting the estimated deterministic part and it is proposed to use a Johansen LR type test for the cointegrating rank applied to the adjusted series. The 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. Therefore it is not necessary to get new critical values, e.g., if new data become available for an extended sample period. As usual in unit root and cointegration tests, the asymptotic distributions of the test statistics depend on whether or not a deterministic trend term is present in the DGP, however. We have derived asymptotic results for DGPs with and without deterministic linear trend terms 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, although the break date estimator which ignores the impulse dummies is more successful in locating the true break date, the choice of break date estimator does not have a substantial impact on the properties of the performance of the cointegrating rank tests. We have also applied the procedure to two sets of example series to illustrate its performance in practice.

There are a number of possible extensions which may be of interest in applied work. For example, as shown in one of the examples, it may be useful to allow for more than one shift and it may also be of interest to include further impulse dummy variables. Moreover, dummies which 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. 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.

Appendix A. Proofs

In the proofs it is assumed for convenience that the initial values of the DGP are such that $\beta'x_t$ and Δx_t are stationary. Also, it is occasionally helpful to be more explicit about the true parameter values. In these cases they will be indicated with a subscript ‘ o ’ (e.g., ν_{0o} , ν_{1o} , τ_o etc.). Moreover, we use the following notation. Right hand side and left hand side will be abbreviated by r.h.s. and l.h.s., respectively. The effective sample size used for estimation is denoted by N , i.e., $N = T - p$, and $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix. The symbol \Rightarrow signifies weak convergence in a product space of $D([\underline{\lambda}, \bar{\lambda}])$ or $D([0, 1])$. The former is relevant for random functions depending on the parameter λ , whereas the latter is used when the weak limit is a Brownian motion. Unless otherwise stated, all limits assume that $T \rightarrow \infty$. When obtaining weak convergences in a product space of $D([\underline{\lambda}, \bar{\lambda}])$ we frequently make use of results given in Appendix A.1 of Gregory & Hansen (1996). It is straightforward to check that these results are applicable despite the

differences in assumptions. The subsequent results assume that $\varepsilon_t \sim \text{i.i.d.}(0, \Omega)$ with the components of ε_t having finite moments of order b , where $b > 4$, as set out in Section 2. We will now first prove Theorem 3.1 and then turn to Theorem 4.1.

A.1 Proof of Theorem 3.1

We begin by establishing the result for the estimator $\hat{\tau}$. Instead of the representation (3.1) it will be convenient to use the corresponding VECM form

$$\Delta y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} + \Pi y_{t-1} + \sum_{j=0}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_{t\tau}, \quad t = p+1, \dots, T. \quad (\text{A.1})$$

Using an (infeasible) linear reparameterization we can transform y_t to $y_t - \mu_{0o} - \mu_{1o}t$. This means that we can proceed by assuming that $\mu_{0o} = \mu_{1o} = 0$ and, consequently, $y_t = \delta_o d_{t\tau_o} + x_t$ and $\nu_{0o} = \nu_{1o} = 0$. A further (infeasible) reparameterization is obtained by writing

$$\Pi y_{t-1} = \alpha u_{t-1} + \rho v_{t-1},$$

where $u_t = \beta' y_t$, $v_t = \beta'_\perp y_t$ and $\rho = \Pi \beta_\perp (\beta'_\perp \beta_\perp)^{-1}$. Notice that here parameters refer to their true values so that $\rho_o = 0$. Thus, instead of the preceding error correction form we employ its infeasible counterpart

$$\Delta y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} + \alpha u_{t-1} + \rho v_{t-1} + \sum_{j=0}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_{t\tau}, \quad t = p+1, \dots, T. \quad (\text{A.2})$$

To write (A.2) in a more compact form, denote $w_t = [1, t, v'_{t-1}, u'_{t-1}, \Delta y'_{t-1}, \dots, \Delta y'_{t-p+1}]'$, $q_{t\tau} = [d_{t\tau}, \Delta d_{t\tau}, \dots, \Delta d_{t-p+1,\tau}]'$ and set $z_{t\tau} = [w'_t, q'_{t\tau}]'$. Then (A.2) can be written as

$$\Delta y_t = \Phi z_{t\tau} + \varepsilon_{t\tau}, \quad t = p+1, \dots, T, \quad (\text{A.3})$$

where $\Phi = [\Phi_1 : \Phi_2]$ with $\Phi_1 = [\nu_0 : \nu_1 : \rho : \alpha : \Gamma_1 : \dots : \Gamma_{p-1}]$ and $\Phi_2 = [\delta_1, \gamma_0, \dots, \gamma_{p-1}]$.

For every $\tau \in \mathcal{T}$ we can estimate the parameter matrix Φ in (A.3) by LS. The corresponding estimator obtained for the error covariance matrix is obviously the same as in the case of the original model (3.1). Thus, we can use the infeasible model (A.3) for theoretical purposes. Since the regressors in the levels VAR form (3.1) are the same in each of the n equations the shift date estimator defined by (3.2) can equivalently be defined by

$$\hat{\tau} = \arg \min_{\tau \in \mathcal{T}} \text{tr} \left(\sum_{t=p+1}^T \hat{\varepsilon}_{t\tau} \hat{\varepsilon}'_{t\tau} \right). \quad (\text{A.4})$$

In what follows this definition of the estimator $\hat{\tau}$ is used because it appears convenient for mathematical derivations. In our analysis we employ a theoretical counterpart of the sum in (A.4) and express it as a sum whose components depend on the specified break date τ and its true value τ_o . An explicit treatment will only be given in the case $\tau < \tau_o$ because the case $\tau \geq \tau_o$ can be treated similarly (cf. Bai et al. (1998), proof of Theorem 3).

The relevant theoretical counterpart of the trace in (A.4) is defined by

$$Q_T(\tau, \Phi) = \text{tr} \left(\sum_{t=p+1}^T \varepsilon_{t\tau}(\Phi) \varepsilon_{t\tau}(\Phi)' \right), \quad (\text{A.5})$$

where $\varepsilon_{t\tau}(\Phi) = \Delta y_t - \Phi z_{t\tau}$. Roughly speaking, we will first show that in large samples the infimum of $Q_T(\tau, \Phi)$ is obtained for $\tau = \tau_o$ with probability approaching one. For this purpose, assuming $\tau < \tau_o$, we split the function $Q_T(\tau, \Phi)$ into a sum of four components, the first one is defined in terms of

$$\varepsilon_t(\Psi_1) = \Delta y_t - \Psi_1 w_t, \quad t = p+1, \dots, \tau-1, \quad (\text{A.6})$$

where $\Psi_1 = [\nu_0 : \nu_1 : \rho : \alpha : \Gamma_1 : \dots : \Gamma_{p-1}]$. The true value of Ψ_1 is $\Psi_{1o} = [0 : 0 : 0 : \alpha_o : \Gamma_{1o} : \dots : \Gamma_{p-1,o}]$ and, obviously, $\varepsilon_t(\Psi_{1o}) = \varepsilon_t$ ($t = p+1, \dots, \tau-1$).

Next set $\mathbf{d}_{t\tau} = [\Delta d_{t\tau}, \dots, \Delta d_{t-p+1,\tau}]$, $\boldsymbol{\gamma} = [\gamma_0, \dots, \gamma_{p-1}]$ and define

$$\varepsilon_{t\tau}(\Psi_2) = \Delta y_t - \Psi_2 \mathbf{z}_{t\tau}, \quad t = \tau, \dots, \tau_o - 1, \quad (\text{A.7})$$

where $\mathbf{z}_{t\tau} = [w'_t : \mathbf{d}'_{t\tau}]'$ and $\Psi_2 = [\nu_0 + \delta_1 : \nu_1 : \rho : \alpha : \Gamma_1 : \dots : \Gamma_{p-1} : \boldsymbol{\gamma}]$. Here the corresponding true parameter value is defined as $\Psi_{2o} = [\Psi_{1o} : 0]$ because then $\varepsilon_{t\tau}(\Psi_{2o}) = \varepsilon_t$ for all $t = \tau, \dots, \tau_o - 1$.

For $t \geq \tau_o > \tau$ the vectors $\mathbf{d}_{t\tau}$ and $\mathbf{d}_{t\tau_o}$ are nonzero only when $\tau_o \leq t \leq \tau + p - 1$ and $\tau_o \leq t \leq \tau_o + p - 1$, respectively. Thus, we consider the time points $t = \tau_o, \dots, \tau_o + p - 1$ separately. For these time points it is still relevant to consider

$$\varepsilon_{t\tau}(\Psi_2) = \Delta y_t - \Psi_2 \mathbf{z}_{t\tau}, \quad t = \tau_o, \dots, \tau_o + p - 1. \quad (\text{A.8})$$

However, unlike in (A.7) and with a slight abuse of notation, it is now reasonable to define the true parameter value as $\Psi_{2o} = [\delta_{1o} : 0 : 0 : \alpha_o : \Gamma_{1o} : \dots : \Gamma_{p-1,o} : \boldsymbol{\gamma}_o]$. Then we have $\varepsilon_{t\tau}(\Psi_{2o}) = \varepsilon_t + \boldsymbol{\gamma}_o(\mathbf{d}_{t\tau} - \mathbf{d}_{t\tau_o})$, $t = \tau_o, \dots, \tau_o + p - 1$.

For the remaining values of t we define

$$\varepsilon_t(\Psi_3) = \Delta y_t - \Psi_3 w_t, \quad t = \tau_o + p, \dots, T, \quad (\text{A.9})$$

where $\Psi_3 = [\nu_0 + \delta_1 : \nu_1 : \rho : \alpha : \Gamma_1 : \cdots : \Gamma_{p-1}]$ with the true value $\Psi_{3o} = [\delta_{1o} : 0 : 0 : \alpha_o : \Gamma_{1o} : \cdots : \Gamma_{p-1,o}]$. Obviously, we again have $\varepsilon_t(\Psi_{3o}) = \varepsilon_t$, $t = \tau_o + p, \dots, T$.

Using (A.6) – (A.9) we can now decompose the function $Q_T(\tau, \Phi)$ defined in (A.5) as

$$\begin{aligned} Q_T(\tau, \Phi) &= \text{tr} \left(\sum_{t=p+1}^{\tau-1} \varepsilon_t(\Psi_1) \varepsilon_t(\Psi_1)' \right) + \text{tr} \left(\sum_{t=\tau}^{\tau_o-1} \varepsilon_{t\tau}(\Psi_2) \varepsilon_{t\tau}(\Psi_2)' \right) \\ &\quad + \text{tr} \left(\sum_{t=\tau_o}^{\tau_o+p-1} \varepsilon_{t\tau}(\Psi_2) \varepsilon_{t\tau}(\Psi_2)' \right) + \text{tr} \left(\sum_{t=\tau_o+p}^T \varepsilon_t(\Psi_3) \varepsilon_t(\Psi_3)' \right) \\ &\stackrel{\text{def}}{=} Q_{1,\tau-1}(\Psi_1) + Q_{2,\tau_o-1}(\tau, \Psi_2) + Q_{2,\tau_o+p-1}(\tau, \Psi_2) + Q_{3T}(\Psi_3). \end{aligned}$$

We shall analyze the four components of $Q_T(\tau, \Phi)$ separately. The first one is considered in the following lemma where

$$D_T = \text{diag}[N^{1/2} : N^{3/2} : NI_{n-r} : N^{1/2}I_r : N^{1/2}I_{n(p-1)}].$$

Lemma A.1. There exists a real number $c_1 > 0$ such that, with probability approaching one,

$$Q_{1,\tau-1}(\Psi_1) - Q_{1,\tau-1}(\Psi_{1o}) \geq -a_{1T} \|(\Psi_1 - \Psi_{1o})D_T\| + c_1 \|(\Psi_1 - \Psi_{1o})D_T\|^2,$$

where $a_{1T} \geq 0$ and $a_{1T} = O_p(1)$ uniformly in $[T\lambda] \leq \tau < \tau_o$ and Φ .

Proof: Because $\varepsilon_t(\Psi_{1o}) = \varepsilon_t$ and $\varepsilon_t(\Psi_1) - \varepsilon_t(\Psi_{1o}) = -(\Psi_1 - \Psi_{1o})w_t$ we have

$$\begin{aligned} &Q_{1,\tau-1}(\Psi_1) - Q_{1,\tau-1}(\Psi_{1o}) \\ &= -2\text{tr} \left((\Psi_1 - \Psi_{1o})D_T(D_T^{-1} \sum_{t=p+1}^{\tau-1} w_t \varepsilon_t') \right) \\ &\quad + \text{tr} \left((\Psi_1 - \Psi_{1o})D_T(D_T^{-1} \sum_{t=p+1}^{\tau-1} w_t w_t' D_T^{-1}) D_T (\Psi_1 - \Psi_{1o})' \right). \end{aligned} \tag{A.10}$$

Furthermore, $\tau \geq [T\lambda]$ implies

$$\lambda_{\min} \left(D_T^{-1} \sum_{t=p+1}^{\tau-1} w_t w_t' D_T^{-1} \right) \geq \lambda_{\min} \left(D_T^{-1} \sum_{t=p+1}^{[T\lambda]-1} w_t w_t' D_T^{-1} \right) \geq c_1, \tag{A.11}$$

where $c_1 > 0$ and the latter inequality holds with probability approaching one. To justify this, note that here we have $t < \tau_o$ so that $y_t = x_t$. Thus, by well-known properties of stationary and integrated processes, the matrix in the second expression of (A.11) converges weakly to a limit which is positive definite (a.s.). This implies the latter inequality in (A.11).

Assuming (A.11) it follows from (A.10) and the Cauchy-Schwarz inequality that

$$Q_{1,\tau-1}(\Psi_1) - Q_{1,\tau-1}(\Psi_{1o}) \geq -2 \|(\Psi_1 - \Psi_{1o})D_T\| \left\| D_T^{-1} \sum_{t=p+1}^{\tau-1} w_t \varepsilon_t' \right\| + c_1 \|(\Psi_1 - \Psi_{1o})D_T\|^2.$$

To complete the proof it suffices to show that

$$D_T^{-1} \sum_{t=p+1}^{[T\lambda]-1} w_t \varepsilon'_t = O_p(1) \quad (A.12)$$

uniformly in $\underline{\lambda} \leq \lambda \leq \lambda_o$. For the third block on the l.h.s. defined by $v_{t-1} \varepsilon'_t$ this follows from Theorem 2.1 of Hansen (1992). For the remaining blocks the desired result is obtained because an invariance principle holds in these cases. To see that an invariance principle holds, observe that $w_t \varepsilon'_t$ is a martingale difference sequence and that it is straightforward to verify that for the considered blocks the conditions required for the application of Theorem 27.14 of Davidson (1994) hold. This completes the proof. \square

Of the four components of $Q_T(\tau, \Phi)$ the second one requires most effort. Therefore, we shall first consider the third and fourth components. For the third component we need the following result.

Lemma A.2.

$$Q_{2, \tau_o + p - 1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau_o}^{\tau_o + p - 1} \varepsilon_t \varepsilon'_t \right) \geq -a_{2T} \|(\Psi_3 - \Psi_{3o}) D_T\| - a_{3T} \|\gamma\| - a_{4T},$$

where $a_{iT} \geq 0$ ($i = 2, 3, 4$) and $a_{2T} = o_p(1)$, $a_{3T} = O_p(1)$ and $a_{4T} = O_p(1)$ uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$ and Φ .

Proof: First note that here Ψ_2 is defined as in (A.8) so that $\Psi_2 = [\Psi_3 : \gamma]$, $\Psi_{2o} = [\Psi_{3o} : \gamma_o]$ and $\varepsilon_t(\Psi_2) = \varepsilon_t - (\Psi_3 - \Psi_{3o}) w_t - \gamma \mathbf{d}_{t\tau} + \gamma_o \mathbf{d}_{t\tau_o}$. Thus, expanding the difference on the l.h.s. of the stated inequality in the same way as in (A.10) and omitting the nonnegative terms on the r.h.s. yields

$$\begin{aligned} Q_{2, \tau_o + p - 1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau_o}^{\tau_o + p - 1} \varepsilon_t \varepsilon'_t \right) &\geq -2 \text{tr} \left((\Psi_3 - \Psi_{3o}) D_T (D_T^{-1} \sum_{t=\tau_o}^{\tau_o + p - 1} w_t \varepsilon'_t) \right) \\ &\quad - 2 \text{tr} \left(\sum_{t=\tau_o}^{\tau_o + p - 1} (\gamma \mathbf{d}_{t\tau} - \gamma_o \mathbf{d}_{t\tau_o}) \varepsilon'_t \right). \end{aligned}$$

Next, write the difference in the last expression as $(\gamma - \gamma_o) \mathbf{d}_{t\tau} + \gamma_o (\mathbf{d}_{t\tau} - \mathbf{d}_{t\tau_o})$ and use the Cauchy-Schwarz inequality as well as the triangle inequality to conclude that

$$\begin{aligned} Q_{2, \tau_o + p - 1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau_o}^{\tau_o + p - 1} \varepsilon_t \varepsilon'_t \right) &\geq -2 \|(\Psi_3 - \Psi_{3o}) D_T\| \|D_T^{-1} \sum_{t=\tau_o}^{\tau_o + p - 1} w_t \varepsilon'_t\| \\ &\quad - 2 (\|\gamma\| + \|\gamma_o\|) \left\| \sum_{t=\tau_o}^{\tau_o + p - 1} \mathbf{d}_{t\tau} \varepsilon'_t \right\| \\ &\quad - 2 \|\gamma_o\| \left\| \sum_{t=\tau_o}^{\tau_o + p - 1} (\mathbf{d}_{t\tau} - \mathbf{d}_{t\tau_o}) \varepsilon'_t \right\|. \end{aligned}$$

The result of the lemma is easily obtained from this inequality because the sums on the r.h.s. only contain a fixed finite number of terms. \square

Lemma A.3. There exists a real number $c_2 > 0$ such that, with probability approaching one,

$$Q_{3T}(\Psi_3) - Q_{3T}(\Psi_{3o}) \geq -a_{5T} \|(\Psi_3 - \Psi_{3o})D_T\| + c_2 \|(\Psi_3 - \Psi_{3o})D_T\|^2,$$

where $a_{5T} \geq 0$ and $a_{5T} = O_p(1)$ uniformly in $[T\lambda] \leq \tau < \tau_o$ and Φ .

Proof: In the same way as in (A.10) we have

$$\begin{aligned} Q_{3T}(\Psi_3) - Q_{3T}(\Psi_{3o}) &\geq -2\text{tr} \left((\Psi_3 - \Psi_{3o})D_T(D_T^{-1} \sum_{t=\tau_o+p}^T w_t \varepsilon'_t) \right) \\ &\quad + \text{tr} \left((\Psi_3 - \Psi_{3o})D_T(D_T^{-1} \sum_{t=\tau_o+p}^T w_t w'_t D_T^{-1})D_T(\Psi_3 - \Psi_{3o})' \right). \end{aligned}$$

Because the r.h.s. is independent of τ , uniformity with respect to this parameter is of no concern. Moreover, the stated result is obviously obtained by establishing analogs of (A.11) and (A.12) with the range of summation changed to conform to the present case. This can be done by using arguments similar to those used for (A.11) and (A.12). Notice, however, that now the level shift affects and we have $y_t = \delta_{1o} d_{t\tau_o} + x_t$. It is still straightforward to use well-known results of stationary and integrated processes and establish the desired result. \square

For the function $Q_{2,\tau_o-1}(\Psi_3)$ we need two different results. The following lemma contains the first one.

Lemma A.4. For every $\underline{\lambda} \leq \lambda < \lambda_o$ there exists a real number $c_3 > 0$ such that, with probability approaching one,

$$\begin{aligned} Q_{2,\tau_o-1}(\tau, \Psi_2) - Q_{2,\tau_o-1}(\tau, \Psi_{2o}) &\geq -a_{6T} \|(\Psi_3 - \Psi_{1o})D_T\| - a_{7T} \|\gamma\| \\ &\quad + c_3 \|(\Psi_3 - \Psi_{1o})D_T\|^2 + c_3 \|\gamma\|^2, \end{aligned}$$

where $a_{iT} \geq 0$ ($i = 6, 7$), $a_{6T} = O_p(1)$ and $a_{7T} = o_p(T^\eta)$, $\frac{1}{b} < \eta \leq \frac{1}{4}$, uniformly in $[T\underline{\lambda}] \leq \tau \leq [T\lambda] < \tau_o$ and Φ .

Proof: In the same way as in (A.10) we first have

$$\begin{aligned}
& Q_{2,\tau_o-1}(\tau, \Psi_2) - Q_{2,\tau_o-1}(\tau, \Psi_{2o}) \\
&= -2\text{tr} \left((\Psi_2 - \Psi_{2o}) \underline{D}_T (\underline{D}_T^{-1} \sum_{t=\tau}^{\tau_o-1} \mathbf{z}_{t\tau} \varepsilon'_t) \right) \\
&\quad + \text{tr} \left((\Psi_2 - \Psi_{2o}) \underline{D}_T (\underline{D}_T^{-1} \sum_{t=\tau}^{\tau_o-1} \mathbf{z}_{t\tau} \mathbf{z}'_{t\tau} \underline{D}_T^{-1}) \underline{D}_T (\Psi_2 - \Psi_{2o})' \right),
\end{aligned} \tag{A.13}$$

where $\underline{D}_T = \text{diag}[D_T : I_p]$. We shall show next that

$$D_T^{-1} \sum_{t=\tau}^{\tau_o-1} w_t \Delta d_{t-i,\tau} = o_p(T^{\eta-1/2}), \quad i = 0, \dots, p-1, \tag{A.14}$$

uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$. This implies that the matrix in the middle of the latter expression on the r.h.s. of (A.13) is asymptotically block diagonal. For the first and second components of w_t (A.14) is obvious and the order is actually $O(N^{-1/2})$. For the third component the order is $O_p(N^{-1/2})$ because $N^{-1/2} \max_{1 \leq t \leq \tau_o} \|v_{t-1}\| \leq N^{-1/2} \max_{1 \leq t \leq T} \|\beta'_\perp x_{t-1}\| = O_p(1)$ due to the fact that $N^{-1/2} \beta'_\perp x_{[Ts]}$ obeys an invariance principle. For the fourth component of w_t we can use an argument similar to that for (A.14) of Saikkonen & Lütkepohl (2001) and conclude that $\max_{1 \leq t \leq \tau_o} \|u_{t-1}\| \leq \max_{1 \leq t \leq T} \|\beta' x_{t-1}\| = o_p(T^\eta)$, which gives the desired result. The same reasoning applies to the remaining components of w_t , that is, to Δy_{t-j} ($j = 1, \dots, p$), and, hence, we have established (A.14).

From (A.14) it follows that

$$\underline{D}_T^{-1} \sum_{t=\tau}^{\tau_o-1} \mathbf{z}_{t\tau} \mathbf{z}'_{t\tau} \underline{D}_T^{-1} = \text{diag} \left[D_T^{-1} \sum_{t=\tau}^{\tau_o-1} w_t w'_t D_T^{-1} : I_p \right] + o_p(T^{\eta-1/2}). \tag{A.15}$$

The next step is to show that

$$D_T^{-1} \sum_{t=[T\underline{\lambda}]}^{[T\underline{\lambda}_o]-1} w_t w'_t D_T^{-1} \Rightarrow M_{ww}(\lambda_o) - M_{ww}(\lambda), \quad \underline{\lambda} \leq \lambda \leq \lambda_o,$$

where the matrix $M_{ww}(\lambda)$ is the weak limit of $D_T^{-1} \sum_{t=p+1}^{[T\underline{\lambda}]-1} w_t w'_t D_T^{-1}$ which is (a.s.) positive definite and continuous and such that the difference $M_{ww}(\lambda_o) - M_{ww}(\lambda)$ is positive definite when $\underline{\lambda} \leq \lambda < \lambda_o$. First consider the lower right hand corner of the matrix $D_T^{-1} \sum_{t=p+1}^{[T\underline{\lambda}]-1} w_t w'_t D_T^{-1}$ which is determined by $u_{t-1} = \beta' x_{t-1}$ and $\Delta y_{t-j} = \Delta x_{t-j}$ ($j = 1, \dots, p-1$). It is straightforward to check that, after subtracting expectations, the quantities in this corner obey an invariance principle, which implies the stated weak convergence with the limit determined by the expectations. For the upper left hand corner determined by 1, t and v_{t-1} the stated weak convergence is obtained from an analog of the result in (A.3)

of Gregory & Hansen (1996) which extends to the case involving the deterministic terms 1 and t . Finally, a combination of (A.4) of Gregory & Hansen (1996) and an application of the invariance principle shows the remaining weak convergence and that the limit is actually zero in this case. Because the required continuity and positive definiteness claims are straightforward to establish, we have completed the proof of this step.

The above discussion implies that, with probability approaching one, the smallest eigenvalue of the matrix on the l.h.s. of (A.15) is bounded away from zero if $\underline{\lambda} \leq \lambda \leq \lambda_o$. Thus, since now $\Psi_2 = [\Psi_3 : \gamma]$ and $\Psi_{2o} = [\Psi_{1o} : 0]$, it follows from (A.13) that, with probability approaching one,

$$\begin{aligned} Q_{2,\tau_o-1}(\Psi_2) - Q_{2,\tau_o-1}(\Psi_{2o}) &\geq -2\text{tr}((\Psi_3 - \Psi_{1o})D_T(D_T^{-1} \sum_{t=\tau}^{\tau_o-1} w_t \varepsilon'_t)) \\ &\quad - 2\text{tr}(\gamma \sum_{t=\tau}^{\tau_o-1} \mathbf{d}_{t\tau} \varepsilon'_t) \\ &\quad + c_3 \|(\Psi_3 - \Psi_{1o})D_T\|^2 + c_3 \|\gamma\|^2, \end{aligned} \tag{A.16}$$

where $c_3 > 0$ is a suitably chosen (small) constant. The argument used to prove (A.14) in the case of the component u_{t-1} shows that the sum in the second term on the r.h.s. of (A.16) is of order $o_p(T^\eta)$ uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$. This fact and an application of the Cauchy-Schwarz inequality implies that the second term on the r.h.s. of (A.16) can be replaced by $a_{7T}\|\gamma\|$ with a_{7T} as stated in the formulation of the lemma. Similar arguments applied to the first term on the r.h.s. of (A.16) show that the result of the lemma follows if

$$D_T^{-1} \sum_{t=\tau}^{\tau_o-1} w_t \varepsilon'_t = O_p(1)$$

uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$. This, however, can be established in the same way as the analogous result in the proof of Lemma A.1 (see (A.12)). This completes the proof. \square

Now we can prove the following result.

Lemma A.5. If $\delta_{1o} \neq 0$, then $\hat{\tau} - \tau_o = o_p(T)$.

Proof: Let $\epsilon > 0$ and define the set

$$B_{1T} = \{(\tau, \Phi) : [T\underline{\lambda}] \leq \tau \leq \tau_o - [T\epsilon]\}.$$

We use Lemmas A.1 - A.4 to show that

$$Pr \left\{ \inf_{(\tau, \Phi) \in B_{1T}} Q_T(\tau, \Phi) - Q_T(\tau_o, \Phi_o) > 0 \right\} \rightarrow 1. \quad (A.17)$$

Because a minimum of the function $Q_T(\tau, \Phi)$ exists for every τ larger than some constant and analogs of Lemmas A.1 - A.4 can be proved for $\tau \geq \tau_o$ one can similarly show that (A.17) holds with the inequalities in the definition of B_{1T} replaced by $\tau_o + [T\epsilon] \leq \tau \leq [T\bar{\lambda}]$.

The result of the lemma follows from these facts.

Thus, assume that $(\tau, \Phi) \in B_{1T}$ and note that $Q_T(\tau_o, \Phi_o) = \text{tr} \left(\sum_{t=p+1}^T \varepsilon_t \varepsilon_t' \right)$. From Lemmas A.1 - A.4 we then find that, with probability approaching one,

$$\begin{aligned} Q_T(\tau, \Phi) - Q_T(\tau_o, \Phi_o) &\geq -a_{1T} \|(\Psi_1 - \Psi_{1o})D_T\| + c_1 \|(\Psi_1 - \Psi_{1o})D_T\|^2 \\ &\quad - (a_{2T} + a_{5T}) \|(\Psi_3 - \Psi_{3o})D_T\| + c_2 \|(\Psi_3 - \Psi_{3o})D_T\|^2 \\ &\quad - a_{6T} \|(\Psi_3 - \Psi_{1o})D_T\| + c_3 \|(\Psi_3 - \Psi_{1o})D_T\|^2 \\ &\quad - (a_{3T} + a_{7T}) \|\gamma\| + c_3 \|\gamma\|^2 - a_{4T}, \end{aligned} \quad (A.18)$$

where the quantities a_{iT} ($i = 1, \dots, 7$) and c_i ($i = 1, 2, 3$) are independent of τ and Φ .

Next note that, for any $a_0, a_1 \geq 0$ and $a_2 > 0$, we have the elementary inequality

$$a_2 x^2 - a_1 x + a_0 \geq -\frac{a_1^2}{4a_2} + a_0, \quad x > 0. \quad (A.19)$$

Thus, because $a_{3T} + a_{7T} = o_p(T^\eta)$, $a_{4T} = O_p(1)$ and $c_3 > 0$, the sum of the last three terms on the r.h.s. of (A.18) divided by N can be replaced by $o_p(1)$. Because $a_{1T} = O_p(1)$ and $c_1 > 0$, the same argument shows that this is also the case for the sum of the first two terms on the r.h.s. of (A.18). Thus, with probability approaching one,

$$\begin{aligned} N^{-1}Q_T(\tau, \Phi) - N^{-1}Q_T(\tau_o, \Phi_o) &\geq -N^{-1/2}(a_{2T} + a_{5T}) \|(\Psi_3 - \Psi_{3o})D_T N^{-1/2}\| \\ &\quad + c_2 \|(\Psi_3 - \Psi_{3o})D_T N^{-1/2}\|^2 \\ &\quad - N^{-1/2}a_{6T} \|(\Psi_3 - \Psi_{1o})D_T N^{-1/2}\| \\ &\quad + c_3 \|(\Psi_3 - \Psi_{1o})D_T N^{-1/2}\|^2 + o_p(1), \end{aligned} \quad (A.20)$$

where the term $o_p(1)$ is independent of τ and Φ .

We shall analyze the r.h.s. of (A.20) in two parts and to this end define the sets

$$B_{2T} = B_{1T} \cap \left\{ (\tau, \Phi) : \|(\Psi_3 - \Psi_{3o})D_T N^{-1/2}\| \leq \frac{1}{2} \|\delta_{1o}\| \right\}$$

and

$$B_{3T} = B_{1T} \cap \left\{ (\tau, \Phi) : \|(\Psi_3 - \Psi_{3o})D_T N^{-1/2}\| \geq \frac{1}{2} \|\delta_{1o}\| \right\}.$$

By (A.19) and $a_{6T} = O_p(1)$ we can replace the sum of the third and fourth terms on the r.h.s. of (A.20) by $o_p(1)$ and conclude that, with probability approaching one,

$$\begin{aligned}
& \inf_{(\tau, \Phi) \in B_{3T}} N^{-1} Q_T(\tau, \Phi) - N^{-1} Q_T(\tau_o, \Phi_o) \\
& \geq \inf_{(\tau, \Phi) \in B_{3T}} c_2 \|(\Psi_3 - \Psi_{3o}) D_T N^{-1/2}\|^2 \left(1 - \frac{a_{2T} + a_{5T}}{c_2 N^{1/2} \|(\Psi_3 - \Psi_{3o}) D_T N^{-1/2}\|} \right) + o_p(1) \\
& \geq \frac{c_2}{4} \|\delta_{1o}\|^2 \left(1 - \frac{2(a_{2T} + a_{5T})}{c_2 N^{1/2} \|\delta_{1o}\|} \right) + o_p(1) \\
& = \frac{1}{4} c_2 \|\delta_{1o}\|^2 + o_p(1).
\end{aligned}$$

Here the inequality follows because $a_{2T} + a_{5T} = O_p(1)$ by Lemmas A.2 and A.3. This shows that (A.17) holds when the infimum of $Q_T(\tau, \Phi)$ is taken over B_{3T} .

It remains to study the infimum of $Q_T(\tau, \Phi)$ over B_{2T} . First observe that $\Psi_3 - \Psi_{3o} = \Psi_3 - \Psi_{1o} - [\delta_{1o} : 0]$ and hence

$$\|(\Psi_3 - \Psi_{1o}) D_T N^{-1/2}\| \geq \|\delta_{1o}\| - \|(\Psi_3 - \Psi_{3o}) D_T N^{-1/2}\|$$

by the triangle inequality. On B_{2T} the square of the r.h.s. is bounded from below by $\frac{1}{4} \|\delta_{1o}\|^2$. This and the inequality $\|(\Psi_3 - \Psi_{1o}) D_T N^{-1/2}\| \leq \|(\Psi_3 - \Psi_{3o}) D_T N^{-1/2}\| + \|\delta_{1o}\|$ can be used to show that on B_{2T} we can replace (A.20) by

$$\begin{aligned}
& N^{-1} Q_T(\tau, \Phi) - N^{-1} Q_T(\tau_o, \Phi_o) \\
& \geq -N^{-1/2} (a_{2T} + a_{5T} + a_{6T}) \|(\Psi_3 - \Psi_{3o}) D_T N^{-1/2}\| \\
& \quad + c_2 \|(\Psi_3 - \Psi_{3o}) D_T N^{-1/2}\|^2 - N^{-1/2} a_{6T} \|\delta_{1o}\| + \frac{c_3}{4} \|\delta_{1o}\|^2 + o_p(1).
\end{aligned}$$

By (A.19) and $a_{2T} + a_{5T} = O_p(1)$ the sum of the first two terms on the r.h.s. can further be replaced by $o_p(1)$ and hence we can conclude that, with probability approaching one,

$$\inf_{(\tau, \Phi) \in B_{2T}} N^{-1} Q_T(\tau, \Phi) - N^{-1} Q_T(\tau_o, \Phi_o) \geq \frac{c_3}{4} \|\delta_{1o}\|^2 \left(1 - \frac{4a_{6T}}{c_3 N^{1/2} \|\delta_{1o}\|} \right) + o_p(1) \geq \frac{c_3}{4} \|\delta_{1o}\|^2 + o_p(1).$$

Thus (A.17) also holds with the infimum of $Q_T(\tau, \Phi)$ taken over B_{2T} and we have altogether established (A.17). This completes the proof of Lemma A.5. \square

Lemma A.5 shows that $\hat{\tau}/N$ is a consistent estimator of the break date fraction λ . Unfortunately, this consistency result is not sufficient to prove Theorem 4.1 although it is a useful intermediate step for obtaining a sufficient improvement. To this end, we have to replace Lemma A.4 by another result. In this replacement it will be convenient to also

consider $Q_{2,\tau_0+p-1}(\tau, \Psi_2)$ and hence replace both Lemma A.2 and A.4 by the following result.

Lemma A.6.

$$\begin{aligned} & Q_{2,\tau_0-1}(\tau, \Psi_2) + Q_{2,\tau_0+p-1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau}^{\tau_0+p-1} \varepsilon_t \varepsilon_t' \right) \\ & \geq -a_{8T} \|(\Psi_3 - \Psi_{3o}) D_T\| - a_{9T} \|(\Psi_3 - \Psi_{3o}) D_T\|^2 \\ & \quad - \left(b_{1T} + \left(\frac{\tau_0 - \tau}{N} \right)^{1/2} b_{2T} \|(\Psi_3 - \Psi_{3o}) D_T\| \right) (\tau_0 - \tau)^{1/2} \|\delta_{1o}\| + (\tau_0 - \tau) \|\delta_{1o}\|^2 - a_{10T} \end{aligned}$$

where $a_{iT} \geq 0$ ($i = 8, 9, 10$), $b_{iT} \geq 0$ ($i = 1, 2$) and $a_{8T} = o_p(T^\eta)$, $a_{9T} = o_p(1)$, $a_{10T} = o_p(T^{2\eta})$, $b_{1T} = O_p(1)$ and $b_{2T} = O_p(1)$ uniformly in $[T\lambda] \leq \tau < \tau_0$ and Φ .

Proof: First consider $Q_{2,\tau_0-1}(\tau, \Psi_2)$ for which $\Psi_2 - \Psi_{2o} = [\Psi_3 - \Psi_{1o} : \gamma]$. Thus, we can write (A.13) as

$$\begin{aligned} Q_{2,\tau_0-1}(\tau, \Psi_2) - Q_{2,\tau_0-1}(\tau, \Psi_{2o}) &= Q_{2,\tau_0-1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau}^{\tau_0-1} \varepsilon_t \varepsilon_t' \right) \\ &= -2\text{tr} \left((\Psi_3 - \Psi_{1o}) \sum_{t=\tau}^{\tau_0-1} w_t \varepsilon_t' \right) - 2\text{tr} \left(\gamma \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau} \varepsilon_t' \right) \\ &\quad + \text{tr} \left((\Psi_3 - \Psi_{1o}) \sum_{t=\tau}^{\tau_0-1} w_t w_t' (\Psi_3 - \Psi_{1o})' \right) \\ &\quad + 2\text{tr} \left((\Psi_3 - \Psi_{1o}) \sum_{t=\tau}^{\tau_0-1} w_t \mathbf{d}_{t\tau}' \gamma' \right) \\ &\quad + \text{tr} \left(\gamma \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau} \mathbf{d}_{t\tau}' \gamma' \right). \end{aligned}$$

Because $\Psi_3 - \Psi_{1o} = \Psi_3 - \Psi_{3o} + [\delta_{1o} : 0]$ this yields further

$$\begin{aligned} & Q_{2,\tau_0-1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau}^{\tau_0-1} \varepsilon_t \varepsilon_t' \right) \\ &= -2\text{tr} \left((\Psi_3 - \Psi_{3o}) \sum_{t=\tau}^{\tau_0-1} w_t \varepsilon_t' \right) - 2\text{tr} \left(\delta_{1o} \sum_{t=\tau}^{\tau_0-1} \varepsilon_t' \right) \\ &\quad - 2\text{tr} \left(\gamma \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau} \varepsilon_t' \right) + \text{tr} \left((\Psi_3 - \Psi_{3o}) \sum_{t=\tau}^{\tau_0-1} w_t w_t' (\Psi_3 - \Psi_{3o})' \right) \\ &\quad + 2\text{tr} \left((\Psi_3 - \Psi_{3o}) \sum_{t=\tau}^{\tau_0-1} w_t \delta_{1o}' \right) + (\tau_0 - \tau) \|\delta_{1o}\|^2 \\ &\quad + 2\text{tr} \left((\Psi_3 - \Psi_{3o}) \sum_{t=\tau}^{\tau_0-1} w_t \mathbf{d}_{t\tau}' \gamma' \right) + 2\text{tr} \left(\delta_{1o} \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau}' \gamma' \right) + \text{tr} \left(\gamma \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau} \mathbf{d}_{t\tau}' \gamma' \right). \end{aligned}$$

For simplicity, we assume $\tau_0 - \tau \geq p$ and note that by redefining $\gamma \mathbf{d}_{t\tau}$, the subsequent analysis can readily be modified to deal with the case $\tau_0 - p < \tau < \tau_0$. Then the last term on the r.h.s. of the above equality is bounded from below by $\|\gamma\|^2$ whereas the fourth term is nonnegative. Using these facts and the Cauchy-Schwarz inequality we thus find that

$$\begin{aligned} & Q_{2,\tau_0-1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau}^{\tau_0-1} \varepsilon_t \varepsilon_t' \right) \\ & \geq -2 \|(\Psi_3 - \Psi_{3o}) D_T\| \|D_T^{-1} \sum_{t=\tau}^{\tau_0-1} w_t \varepsilon_t'\| + (\tau_0 - \tau) \|\delta_{1o}\|^2 \\ & \quad - 2 \|\delta_{1o}\| \left(\left\| \sum_{t=\tau}^{\tau_0-1} \varepsilon_t \right\| + \|(\Psi_3 - \Psi_{3o}) D_T\| \|D_T^{-1} \sum_{t=\tau}^{\tau_0-1} w_t\| \right) \\ & \quad - 2 \|\gamma\| \left(\left\| \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau} \varepsilon_t' \right\| + \|\delta_{1o}\| \left\| \sum_{t=\tau}^{\tau_0-1} \mathbf{d}_{t\tau} \right\| \right) \\ & \quad - 2 \|\gamma\| \|(\Psi_3 - \Psi_{3o}) D_T\| \|D_T^{-1} \sum_{t=\tau}^{\tau_0-1} w_t \mathbf{d}_{t\tau}'\| + \|\gamma\|^2 \end{aligned}$$

The latter norm in the first term on the r.h.s. is dominated by $\frac{1}{2}a_{6T}$ (see the end of the proof of Lemma A.4). Thus, defining

$$b_{1T}(\tau) = 2 \left\| (\tau_o - \tau)^{-1/2} \sum_{t=\tau}^{\tau_o-1} \varepsilon_t \right\|,$$

$$b_{2T}(\tau) = 2 \left\| \left(\frac{N^{1/2}}{\tau_o - \tau} \right) D_T^{-1} \sum_{t=\tau}^{\tau_o-1} w_t \right\|$$

and

$$b_{3T}(\tau) = 2 \left\| D_T^{-1} \sum_{t=\tau}^{\tau_o-1} w_t \mathbf{d}_{t\tau} \right\|,$$

we can combine the above inequality with that in Lemma A.2 and obtain

$$\begin{aligned} & Q_{2,\tau_o-1}(\tau, \Psi_2) + Q_{2,\tau_o+p-1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau}^{\tau_o+p-1} \varepsilon_t \varepsilon_t' \right) \\ & \geq -(a_{2T} + a_{6T}) \|(\Psi_3 - \Psi_{3o}) D_T\| + (\tau_o - \tau) \|\delta_{1o}\|^2 \\ & \quad - \left(b_{1T}(\tau) + \left(\frac{\tau_o - \tau}{N} \right)^{1/2} b_{2T}(\tau) \|(\Psi_3 - \Psi_{3o}) D_T\| \right) (\tau_o - \tau)^{1/2} \|\delta_{1o}\| \\ & \quad - (b_{4T}(\tau) + b_{3T}(\tau) \|(\Psi_3 - \Psi_{3o}) D_T\|) \|\gamma\| + \|\gamma\|^2 - a_{4T}, \end{aligned} \tag{A.21}$$

where

$$b_{4T}(\tau) = 2 \left\| \sum_{t=\tau}^{\tau_o-1} \mathbf{d}_{t\tau} \varepsilon_t' \right\| + 2 \|\delta_{1o}\| \left\| \sum_{t=\tau}^{\tau_o-1} \mathbf{d}_{t\tau} \right\| + a_{3T}.$$

Notice that here we have $a_{2T} = o_p(1)$, $a_{3T} = O_p(1)$ and $a_{4T} = O_p(1)$ uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$ and Φ (see Lemma A.2).

Now define $b_{iT} = \max\{b_{iT}(\tau) : [T\underline{\lambda}] \leq \tau < \tau_o\}$, $i = 1, \dots, 4$. Then, $b_{3T} = o_p(1)$ by (A.14) and from the arguments used to show that result we also find that the first term in the definition of $b_{4T}(\tau)$ is of order $o_p(T^\eta)$ uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$. Thus, $b_{4T} = o_p(T^\eta)$ whereas the Hájek-Rényi inequality given in Proposition 1 of Bai (1994) implies that $b_{1T} = O_p(1)$. From the Hájek-Rényi inequality, the previously used fact that $N^{-1/2} \max_{1 \leq t < \tau_o} \|v_{t-1}\| = O_p(1)$ and the definition of w_t we finally obtain $b_{2T} = O_p(1)$. Thus, b_{1T} and b_{2T} are as stated in the Lemma.

By the definition of b_{3T} and b_{4T} and (A.19) the sum of the last three terms on the r.h.s. of (A.21) is bounded from below by

$$-\frac{1}{4}b_{4T}^2 - \frac{1}{2}b_{3T}b_{4T} \|(\Psi_3 - \Psi_{3o}) D_T\| - \frac{1}{4}b_{3T}^2 \|(\Psi_3 - \Psi_{3o}) D_T\|^2 - a_{4T}.$$

Inserting this lower bound into (A.21) yields

$$\begin{aligned}
& Q_{2,\tau_o-1}(\tau, \Psi_2) + Q_{2,\tau_o+p-1}(\tau, \Psi_2) - \text{tr} \left(\sum_{t=\tau}^{\tau_o+p-1} \varepsilon_t \varepsilon_t' \right) \\
& \geq -(a_{2T} + a_{6T} + \frac{1}{2}b_{3T}b_{4T}) \|(\Psi_3 - \Psi_{3o})D_T\| - \frac{1}{4}b_{3T}^2 \|(\Psi_3 - \Psi_{3o})D_T\|^2 \\
& \quad + (\tau_o - \tau) \|\delta_{1o}\|^2 - \left(b_{1T} + \left(\frac{\tau_o - \tau}{N} \right)^{1/2} b_{2T} \|(\Psi_3 - \Psi_{3o})D_T\| \right) (\tau_o - \tau)^{1/2} \|\delta_{1o}\| - \left(\frac{1}{4}b_{4T}^2 + a_{4T} \right)
\end{aligned}$$

The stated result in Lemma A.6 follows from this inequality by defining a_{8T} , a_{9T} and a_{10T} in an obvious way. \square

Now we are ready to improve Lemma A.5 and thereby establish the first assertion of Theorem 3.1.

Lemma A.7. If $\delta_{1o} \neq 0$, then $\hat{\tau} - \tau_o = o_p(T^{2\eta})$.

Proof: As before, we consider the case $\tau < \tau_o$. By Lemmas A.1, A.3 and A.6 we have, with probability approaching one,

$$\begin{aligned}
& N^{-2\eta}Q_T(\tau, \Phi) - N^{-2\eta}Q_T(\tau_o, \Phi_o) \\
& \geq -N^{-\eta}a_{1T} \|(\Psi_1 - \Psi_{1o})D_T N^{-\eta}\| + c_1 \|(\Psi_1 - \Psi_{1o})D_T N^{-\eta}\|^2 \\
& \quad - N^{-\eta}(a_{5T} + a_{8T}) \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| + (c_2 - a_{9T}) \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|^2 \quad (A.22) \\
& \quad - \left(N^{-\eta}b_{1T} + \left(\frac{\tau_o - \tau}{N} \right)^{1/2} b_{2T} \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| \right) \left(\frac{\tau_o - \tau}{N^{2\eta}} \right)^{1/2} \|\delta_{1o}\| \\
& \quad + \left(\frac{\tau_o - \tau}{N^{2\eta}} \right) \|\delta_{1o}\|^2 - N^{-2\eta}a_{10T}.
\end{aligned}$$

We shall show first that $\|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|$ can be restricted arbitrarily small. To this end, use (A.19) to conclude that the sum of the first two terms on the r.h.s. of (A.22) can be bounded from below by a term which is of order $o_p(1)$ uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$ and Φ . In the same way, the sum of the last three terms can be bounded from below by

$$-\frac{1}{4} \left(N^{-\eta}b_{1T} + \left(\frac{\tau_o - \tau}{N} \right)^{1/2} b_{2T} \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| \right)^2 + o_p(1).$$

Expanding the square we can then replace (A.22) by

$$\begin{aligned}
& N^{-2\eta}Q_T(\tau, \Phi) - N^{-2\eta}Q_T(\tau_o, \Phi_o) \\
& \geq - \left(N^{-\eta}(a_{5T} + a_{8T}) + \frac{1}{2}N^{-\eta}b_{1T}b_{2T} \left(\frac{\tau_o - \tau}{N} \right)^{1/2} \right) \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| \\
& \quad + (c_2 - a_{9T} - \left(\frac{\tau_o - \tau}{N} \right) b_{2T}^2) \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|^2 + o_p(1).
\end{aligned}$$

Now, for $\epsilon^* > 0$, define the set

$$B_{4T} = \{(\tau, \Phi) : \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| \geq \epsilon^*, [T\underline{\lambda}] \leq \tau \leq \tau_o - [T\epsilon]\}$$

where, by Lemma A.5, $\epsilon > 0$ can be chosen arbitrarily small. In particular, since $a_{9T} = o_p(1)$ and $b_{2T} = O_p(1)$, we can choose ϵ such that $c_2 - a_{9T} - N^{-1}(\tau_o - \tau)b_{2T}^2 > \frac{\epsilon^2}{2}$ with probability approaching one. Thus, since the term in front of $\|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|$ in the above inequality is of order $o_p(1)$ uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$ and Φ , we have, with probability approaching one, that

$$\begin{aligned} & \inf_{(\tau, \Phi) \in B_{4T}} N^{-2\eta} Q_T(\tau, \Phi) - N^{-2\eta} Q_T(\tau_o, \Phi_o) \\ & \geq \inf_{(\tau, \Phi) \in B_{4T}} \left[\frac{c_2}{2} \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|^2 \left(1 - \frac{o_p(1)}{c_2 \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|} \right) \right] + o_p(1) \quad (A.23) \\ & \geq \frac{c_2}{2} \epsilon^{*2} \left(1 - \frac{o_p(1)}{c_2 \epsilon^*} \right) + o_p(1) > 0. \end{aligned}$$

Because this holds for every $\epsilon^* > 0$ it follows that $\|(\hat{\Psi}_3 - \Psi_{3o})D_T N^{-\eta}\| = o_p(1)$ and we can proceed by restricting Ψ_3 in the way described above.

Now consider inequality (A.22) again and note that, by (A.19), we can replace the first four terms by an appropriate uniform lower bound of order $o_p(1)$. A similar replacement can be done for the last term, so that we can replace (A.22) by

$$\begin{aligned} & N^{-2\eta} Q_T(\tau, \Phi) - N^{-2\eta} Q_T(\tau_o, \Phi_o) \\ & \geq - \left[N^{-\eta} b_{1T} + \left(\frac{\tau_o - \tau}{N} \right)^{1/2} b_{2T} \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| \right] \left(\frac{\tau_o - \tau}{N^{2\eta}} \right)^{1/2} \|\delta_{1o}\| \\ & \quad + \left(\frac{\tau_o - \tau}{N^{2\eta}} \right) \|\delta_{1o}\|^2 + o_p(1) \quad (A.24) \\ & \geq \left(\frac{\tau_o - \tau}{N^{2\eta}} \right) \|\delta_{1o}\|^2 \left[1 - \frac{N^{-\eta} b_{1T} + b_{2T} \|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|}{\left(\frac{\tau_o - \tau}{N^{2\eta}} \right)^{1/2} \|\delta_{1o}\|} \right] + o_p(1). \end{aligned}$$

Now let $\epsilon > 0$ and suppose that $\tau_o - \tau \geq [N^{2\eta}\epsilon]$. Since we can assume that $\|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\|$ is arbitrarily small it follows that, for T large enough, the term in the brackets in the last expression can be replaced by a term which is positive with positive probability. This implies that asymptotically, the function $Q_T(\tau, \Phi)$ cannot attain its minimum value in a set such that $\tau_o - \tau \geq [N^{2\eta}\epsilon]$. Because a similar analysis can be used to exclude the case $\tau - \tau_o \geq [N^{2\eta}\epsilon]$, the result of the lemma follows. \square

We still have to prove the second part of Theorem 3.1. Therefore we now consider the estimator $\tilde{\tau}$ and we shall prove the following result.

Lemma A.8. If $\delta_{1o} \neq 0$, then $\tilde{\tau} - \tau_o = O_p(1)$.

Proof: The break date estimator $\tilde{\tau}$ is based on a model without impulse dummies or on

$$\Delta y_t = \nu_{0o} + \nu_{1o}t + \delta_{1o}d_{t\tau_o} + \Pi_o y_{t-1} + \sum_{j=1}^{p-1} \Gamma_{jo} \Delta y_{t-j} + e_t, \quad t = p+1, \dots, T, \quad (A.25)$$

where $e_t = \varepsilon_t + \sum_{j=0}^{p-1} \gamma_{jo} \Delta d_{t-j, \tau_o} = \varepsilon_t + \boldsymbol{\gamma}_o \mathbf{d}_{t\tau_o}$. Because here parameters refer to their true values, there is no need to include the break date in the notation e_t . Now the break date estimator $\tilde{\tau}$ is defined by replacing $\hat{\varepsilon}_t$ in (A.4) by LS residuals from (A.25). The same residuals can be obtained from the (infeasible) model (A.2) or (A.3) by imposing the restriction $\gamma_0 = \dots = \gamma_{p-1} = 0$. Note that this restriction means that the error term ε_t is replaced by e_t .

The consistency of the estimator $\tilde{\tau}$ can be proven in a straightforward way by making appropriate modifications to the proof of Lemma A.7 and the preceding lemmas used in its proof. Instead of the function $Q_T(\tau, \Phi)$ we now consider

$$Q_T^+(\tau, \Phi_1, \delta_1) = \text{tr} \left(\sum_{t=p+1}^T e_{t\tau}(\Phi_1, \delta_1) e_{t\tau}(\Phi_1, \delta_1)' \right),$$

where $e_{t\tau}(\Phi_1, \delta_1) = \Delta y_t - \Phi_1 w_t - \delta_1 d_{t\tau}$. Similarly to $Q_T(\tau, \Phi)$ also $Q_T^+(\tau, \Phi_1, \delta_1)$ can be expressed as a sum of four components and for each component a counterpart of Lemmas A.1 - A.4 and A.6 can be obtained. We shall not go into details because the main points can be illustrated quite well by looking at the differences between the present case and the previous one. First recall that in place of the previous error term ε_t we now have e_t . However, e_t differs from ε_t only by additive impulse dummies which take nonzero values for $t = \tau_o, \dots, \tau_o + p - 1$. Therefore this difference only affects the counterpart of the function $Q_{2, \tau_o + p - 1}(\tau, \Psi_2)$ and even there the effect is asymptotically of no significance. The only other feature which makes the present case different from the previous one is that the parameter $\boldsymbol{\gamma}$ does not appear in the objective function. However, the effect of this can be illustrated by looking at the Lemmas A.1 - A.4 and A.6 without the parameter $\boldsymbol{\gamma}$ and quantities due to it. This means that we can treat the present case by modifying Lemmas

A.2 and A.4 in such a way that the terms $-a_{3T}\|\gamma\| - a_{4T}$ and $-a_{7T}\|\gamma\| + c_3\|\gamma\|^2$, respectively, are removed, whereas the corresponding modification of Lemma A.6 amounts to requiring that $a_{8T} = O_p(1)$, $a_{9T} = 0$ and $a_{10T} = O_p(1)$. Taking these considerations into account and checking the proof of Lemma A.5 it can be seen that $\tilde{\tau} - \tau_o = o_p(T)$, that is, an analog of Lemma A.5 holds for $\tilde{\tau}$.

Now consider the proof of Lemma A.7. Setting $\eta = 0$ and modifying the given arguments based on Lemmas A.1, A.3 and A.6 in the way described above, it can first be seen that (A.23) holds with $\eta = 0$ and the terms $o_p(1)$ replaced by $O_p(1)$. Choosing ϵ^* large enough it then follows that the previous result $\|(\Psi_3 - \Psi_{3o})D_T N^{-\eta}\| = o_p(1)$ can be strengthened to $\|(\Psi_3 - \Psi_{3o})D_T\| = O_p(1)$. Hence, $\|(\Psi_3 - \Psi_{3o})D_T\|$ can be restricted to a bounded set after which we can consider an analog of (A.24) with $\eta = 0$ and $o_p(1)$ replaced by $O_p(1)$. Since $\|(\Psi_3 - \Psi_{3o})D_T\|$ can be assumed bounded by a (large) constant it now follows that, for T large enough, the r.h.s. of the analog of (A.24) is positive with positive probability if $\tau_o - \tau$ is chosen large. The conclusion $\tilde{\tau} - \tau_o = O_p(1)$ of Lemma A.8 then follows in the same way as that of Lemma A.7.

A.2 Proof of Theorem 4.1

In the following proof we use the break date estimator $\hat{\tau}$. The same result can be obtained with any estimator for which Theorem 3.1(i) holds. Because we do not need values from \mathcal{T} other than $\hat{\tau}$ and the true value in the following, we denote the latter again by τ and hence we drop the subscript ‘ o ’ for convenience.

We shall first show that the RR estimators of the parameters α , β , Γ_j ($j = 1, \dots, p-1$) and Ω discussed in Section 3.2 satisfy appropriate consistency properties (cf. Lemma 2.1 of S&L). These estimators are based on equation (2.10) with the unknown break date τ replaced by the estimator $\hat{\tau}$. This replacement changes the error term ε_t to

$$\varepsilon_{t\hat{\tau}} = \varepsilon_t + \Pi\delta(d_{t-1,\hat{\tau}} - d_{t-1,\tau}) - \sum_{j=0}^{p-1} \gamma_j^* (\Delta d_{t-j,\hat{\tau}} - \Delta d_{t-j,\tau}). \quad (\text{A.26})$$

By a linear reparameterization we can transform y_t in equation (2.10) to

$$y_{t\hat{\tau}} = y_t - \mu_0 - \mu_1 t - \delta d_{t\hat{\tau}} = x_t - \delta(d_{t\hat{\tau}} - d_{t\tau}). \quad (\text{A.27})$$

Thus, instead of (2.10), we can consider its infeasible counterpart

$$\Delta y_{t\hat{\tau}} = \nu + \alpha(\beta' y_{t-1, \hat{\tau}} - \phi(t-1) - \theta d_{t-1, \hat{\tau}}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j, \hat{\tau}} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j, \hat{\tau}} + \varepsilon_{t\hat{\tau}}, \quad t = p+1, p+2, \dots, \quad (\text{A.28})$$

where $\nu = 0$, $\phi = 0$, $\theta = 0$ and $\gamma_j^* = 0$ ($j = 0, \dots, p-1$). It will be shown that using the estimator $\hat{\tau}$ in (A.28) instead of the true break date τ has no effect on the asymptotic properties of the RR estimators. When $\delta_1 \neq 0$ the proof makes use of the following general results.

Lemma A.9. Let J_{tT} ($t = 1, \dots, T$) be a (possibly) random vector such that $\max_{1 \leq t \leq T} \|J_{tT}\| = O_p(1)$ and J_t a vector valued stochastic process satisfying $\sup_t E\|J_t\| < \infty$. Then, if $\hat{\tau} = \tau + o_p(T^{2\eta})$, $\frac{1}{b} < \eta \leq \frac{1}{4}$,

(i) $N^{-1/2} \sum_{t=p+1}^T J_{tT}(d_{t\hat{\tau}} - d_{t\tau}) = o_p(1)$

(ii) $N^{-1/2} \sum_{t=p+1}^T J_t(d_{t\hat{\tau}} - d_{t\tau}) = o_p(1)$.

Proof: The first result is an immediate consequence of the assumptions which imply that the l.h.s. of the equality is dominated by $O_p(N^{-1/2})|\hat{\tau} - \tau| = o_p(1)$. To prove the second result, let $\mathbf{I}(\cdot)$ stand for the indicator function and, for $\epsilon > 0$, write

$$\begin{aligned} N^{-1/2} \sum_{t=p+1}^T J_t(d_{t\hat{\tau}} - d_{t\tau}) &= N^{-1/2} \sum_{t=p+1}^T J_t(d_{t\hat{\tau}} - d_{t\tau}) \mathbf{I}(|\hat{\tau} - \tau| > T^{2\eta}\epsilon) \\ &\quad + N^{-1/2} \sum_{t=p+1}^T J_t(d_{t\hat{\tau}} - d_{t\tau}) \mathbf{I}(|\hat{\tau} - \tau| \leq T^{2\eta}\epsilon) \\ &\stackrel{def}{=} A_{1T} + A_{2T}. \end{aligned}$$

Now, since $\hat{\tau} - \tau = o_p(T^{2\eta})$ we have for every $\epsilon > 0$,

$$Pr\{\|A_{1T}\| > \epsilon\} \leq Pr\{|\hat{\tau} - \tau| > T^{2\eta}\epsilon\} \rightarrow 0.$$

Thus, $A_{1T} = o_p(1)$. As for A_{2T} , notice that for every $\epsilon > 0$,

$$\|A_{2T}\| \leq N^{-1/2} \sum_{t=p+1}^T \|J_t\| |d_{t\hat{\tau}} - d_{t\tau}| \mathbf{I}(|\hat{\tau} - \tau| \leq T^{2\eta}\epsilon) \leq N^{-1/2} \sum_{|t-\tau| \leq [T^{2\eta}\epsilon]} \|J_t\|.$$

Hence,

$$E\|A_{2T}\| \leq \text{const. } T^{2\eta-1/2}\epsilon \rightarrow 0.$$

This implies that $A_{2T} = o_p(1)$ and the stated result follows. \square

Lemma A.9 will 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. The first result can be used with $J_{t\tau}$ given by $N^{-1/2}\beta'_\perp y_{t-1,\hat{\tau}}$ and $N^{-1}(t-1)$, whereas in the second result typical choices of J_t are $\Delta y_{t-j,\hat{\tau}}$, $\beta' y_{t-1,\hat{\tau}}$ and $d_{t-1,\hat{\tau}}$ (see (A.27)). Now we can prove the necessary consistency properties of the RR estimators discussed above.

Lemma A.10. 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, $\tilde{\beta}_\xi = \beta + O_p(T^{-1})$, $\tilde{\alpha}_\xi = \alpha + O_p(T^{-1/2})$, $\tilde{\Gamma}_j = \Gamma_j + O_p(T^{-1/2})$ ($j = 1, \dots, p-1$) and $\tilde{\Omega} = \Omega + O_p(T^{-1/2})$.

Proof: First we introduce some notation. Define $\mathbf{x}_{t\hat{\tau}} = [y'_{t-1,\hat{\tau}} : (t-1) : d_{t-1,\hat{\tau}}]'$ and $\mathbf{p}_{1t\hat{\tau}} = [1 : \Delta y'_{t-1,\hat{\tau}} : \dots : \Delta y'_{t-p+1,\hat{\tau}}]'$. Then equation (A.28) can be expressed as

$$\Delta y_{t\hat{\tau}} = \alpha \psi' \mathbf{x}_{t\hat{\tau}} + \Xi \mathbf{p}_{t\hat{\tau}} + \varepsilon_{t\hat{\tau}}, \quad t = p+1, p+2, \dots, \quad (\text{A.29})$$

where $\mathbf{p}_{t\hat{\tau}} = [\mathbf{p}'_{1t\hat{\tau}} : \mathbf{d}'_{t\hat{\tau}}]'$, $\psi' = [\beta' : -\phi : -\theta]$ and $\Xi = [\Xi_1 : \gamma^*]$ with $\Xi_1 = [\nu : \Gamma_1 : \dots : \Gamma_{p-1}]$. Here $\mathbf{d}_{t\tau} = [\Delta d_{t\tau}, \dots, \Delta d_{t-p+1,\tau}]$ and $\gamma^* = [\gamma_0^*, \dots, \gamma_{p-1}^*]$. The RR estimators of α , ψ and Ω can be obtained as follows. Define

$$S_{00\hat{\tau}} = N^{-1} \sum_{t=p+1}^T \Delta y_{t\hat{\tau}} \Delta y'_{t\hat{\tau}} - N^{-1} \sum_{t=p+1}^T \Delta y_{t\hat{\tau}} \mathbf{p}'_{t\hat{\tau}} \left(\sum_{t=p+1}^T \mathbf{p}_{t\hat{\tau}} \mathbf{p}'_{t\hat{\tau}} \right)^{-1} \sum_{t=p+1}^T \mathbf{p}_{t\hat{\tau}} \Delta y'_{t\hat{\tau}},$$

$$S_{01\hat{\tau}} = S'_{10\hat{\tau}} = N^{-1} \sum_{t=p+1}^T \Delta y_{t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}} - N^{-1} \sum_{t=p+1}^T \Delta y_{t\hat{\tau}} \mathbf{p}'_{t\hat{\tau}} \left(\sum_{t=p+1}^T \mathbf{p}_{t\hat{\tau}} \mathbf{p}'_{t\hat{\tau}} \right)^{-1} \sum_{t=p+1}^T \mathbf{p}_{t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}}$$

and

$$S_{11\hat{\tau}} = N^{-1} \sum_{t=p+1}^T \mathbf{x}_{t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}} - N^{-1} \sum_{t=p+1}^T \mathbf{x}_{t\hat{\tau}} \mathbf{p}'_{t\hat{\tau}} \left(\sum_{t=p+1}^T \mathbf{p}_{t\hat{\tau}} \mathbf{p}'_{t\hat{\tau}} \right)^{-1} \sum_{t=p+1}^T \mathbf{p}_{t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}}.$$

As is well-known, the RR estimator of ψ is based on the eigenvectors corresponding to the r largest eigenvalues of the determinantal equation

$$\det(\lambda S_{11\hat{\tau}} - S_{10\hat{\tau}} S_{00\hat{\tau}}^{-1} S_{01\hat{\tau}}) = 0. \quad (\text{A.30})$$

When the RR estimator of ψ is available, those of α and Ω can be obtained by replacing ψ by its estimator in the formulas $S_{01\hat{\tau}} \psi (\psi' S_{11\hat{\tau}} \psi)^{-1}$ and $S_{00\hat{\tau}} - S_{01\hat{\tau}} \psi (\psi' S_{11\hat{\tau}} \psi)^{-1} \psi' S_{10\hat{\tau}}$, respectively.

First consider the case $\delta_1 \neq 0$. The idea of the following proof is to show that the consistency properties of the RR estimators of Ψ , α and Ω discussed above are the same as

in the case where the true value of the break date is used instead of the estimator $\hat{\tau}$. Once this has been shown, the desired results follow from S&L.

In the same way as in S&L, we follow the proof of Lemma 13.1 of Johansen (1995) and transform equation (A.30) to

$$\det(\lambda A'_T S_{11\hat{\tau}} A_T - A'_T S_{10\hat{\tau}} S_{00\hat{\tau}}^{-1} S_{01\hat{\tau}} A_T) = 0, \quad (\text{A.31})$$

where

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

and, consequently,

$$A'_T \mathbf{x}_{t\hat{\tau}} = \begin{bmatrix} \beta' y_{t-1,\hat{\tau}} \\ T^{-1/2} (\beta'_{\perp} \beta_{\perp})^{-1} \beta'_{\perp} y_{t-1,\hat{\tau}} \\ T^{-1} (t-1) \\ (\frac{T}{T-\tau})^{1/2} d_{t-1,\hat{\tau}} \end{bmatrix}.$$

Using (A.14), (A.27) and Lemma A.9 we can now show that

$$S_{00\hat{\tau}} = S_{00\tau} + o_p(1), \quad (\text{A.32})$$

$$S_{01\hat{\tau}} A_T = S_{01\tau} A_T + o_p(1) \quad (\text{A.33})$$

and

$$A'_T S_{11\hat{\tau}} A_T = A'_T S_{11\tau} A_T + o_p(1). \quad (\text{A.34})$$

Details are straightforward but somewhat tedious and will be omitted.

From (A.32) – (A.34) and the proof of Lemma 2.1 of S&L we can conclude that the estimators $\tilde{\beta}_{\xi}$, $\tilde{\phi}$ and $\tilde{\theta}$ are consistent of orders $o_p(T^{-1/2})$, $o_p(T^{-1})$ and $o_p(1)$, respectively. Because $\psi' \mathbf{x}_{t\hat{\tau}} = \beta' y_{t-1,\hat{\tau}}$, arguments similar to those used to obtain (A.32) and (A.33) combined with these consistency results also readily show the consistency of the estimators $\tilde{\alpha}_{\xi}$, $\tilde{\Gamma}_j$ ($j = 0, \dots, p-1$) and $\tilde{\Omega}$.

To obtain the stated orders of consistency, we follow S&L and note that the first order conditions for $\tilde{\psi}$ can be expressed as

$$0 = \tilde{\alpha}'_{\xi} \tilde{\Omega}^{-1} (S_{\varepsilon 1\hat{\tau}} B_T - \tilde{\alpha}_{\xi} U'_T [T^{-1} B'_T S_{11\hat{\tau}} B_T] - (\tilde{\alpha}_{\xi} - \alpha) \psi' S_{11\hat{\tau}} B_T), \quad (\text{A.35})$$

where $S_{\varepsilon 1\hat{\tau}} = S_{01\hat{\tau}} - \alpha\psi'S_{11\hat{\tau}}$,

$$U_T = \begin{bmatrix} T\beta'_\perp\tilde{\beta}_\xi \\ T^{3/2}\tilde{\phi}' \\ (T-\tau)^{1/2}\tilde{\theta}' \end{bmatrix}$$

and

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

Now, notice that B_T is formed by the last $(n-r+2)$ columns of $T^{-1/2}A_T$ and recall that $\psi'\mathbf{x}_{t\hat{\tau}} = \beta'y_{t-1,\hat{\tau}}$. Using these facts, (A.14), (A.27) and Lemma A.9, it is straightforward to proceed in the same way as in the above consistency proof and show that replacing $\hat{\tau}$ by τ on the r.h.s. of (A.35) causes an error of order $o_p(1)$. Thus, from the proof of Lemma 2.1 of S&L we can conclude that $U_T = O_p(1)$ and furthermore that $\tilde{\beta}_\xi = \beta + O_p(T^{-1})$.

To obtain the remaining orders of consistency, notice that $U_T = O_p(1)$ also implies that $\tilde{\psi}$, the RR estimator of ψ , satisfies $A_T^{-1}(\tilde{\psi} - \psi) = O_p(T^{-1/2})$ (cf. S&L). Using this fact, the expression for the error term $\varepsilon_{t\hat{\tau}}$ given in (A.26) and arguments used in connection with equations (A.32) – (A.35) it can be seen that replacing the true break date in (A.11) of S&L by the estimator $\hat{\tau}$ has no effect on the orders of consistency obtained for the estimators $\tilde{\alpha}_\xi$, $\tilde{\Gamma}_j$ ($j = 0, \dots, p-1$), and $\tilde{\Omega}$. Thus, the desired result follows from Lemma 2.1 of S&L. This completes the proof when $\delta_1 \neq 0$.

Now consider the case $\delta_1 = 0$ for which we cannot use Lemma A.9 because $\hat{\tau} = \tau + o_p(T^{2\eta})$ does not hold and even $\hat{\tau}/N$ is not a consistent estimator of the sample fraction λ . However, we still have (A.32) and (A.33), as we shall now demonstrate. The former is simple because, by (A.27) and the definition of $\mathbf{p}_{t\hat{\tau}}$, the matrix $S_{00\hat{\tau}}$ only depends on the impulse dummies $\Delta d_{t-j,\hat{\tau}}$ and $\Delta d_{t-j,\tau}$ ($j = 0, \dots, p-1$) but not on the corresponding step dummies. Thus, because (A.14) still holds, we can proceed in the same way as in the proof of Lemma 2.1 of S&L and show that replacing $\mathbf{p}_{t\hat{\tau}}$ in the definition of $S_{00\hat{\tau}}$ first by $\mathbf{p}_{1t\hat{\tau}}$ and then further by $\mathbf{p}_{1t\tau}$ causes an error of order $o_p(1)$. Because in $S_{00\tau}$ the vector $\mathbf{p}_{t\tau}$ can similarly be replaced by $\mathbf{p}_{1t\tau}$ (see S&L), (A.32) follows.

Now consider (A.33). Using (A.14) and (A.27) we can again readily show that replacing $\mathbf{p}_{t\hat{\tau}}$ and $\mathbf{p}_{t\tau}$ in the definitions of $S_{01\hat{\tau}}A_T$ and $S_{01\tau}A_T$ by $\mathbf{p}_{1t\hat{\tau}}$ and $\mathbf{p}_{1t\tau}$, respectively, causes an

error of order $o_p(1)$. Next, an application of (A.14) and (A.27) gives

$$N^{-1} \sum_{t=p+1}^T \mathbf{p}_{1t\hat{\tau}} \mathbf{p}'_{1t\hat{\tau}} = N^{-1} \sum_{t=p+1}^T \mathbf{p}_{1t\tau} \mathbf{p}'_{1t\tau} + o_p(1) \quad (\text{A.36})$$

and

$$N^{-1} \sum_{t=p+1}^T \Delta y_{t\hat{\tau}} \mathbf{p}'_{1t\hat{\tau}} = N^{-1} \sum_{t=p+1}^T \Delta y_{t\tau} \mathbf{p}'_{1t\tau} + o_p(1). \quad (\text{A.37})$$

In addition to these results, we have

$$N^{-1} \sum_{t=p+1}^T \Delta y_{t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}} A_T = N^{-1} \sum_{t=p+1}^T \Delta y_{t\tau} \mathbf{x}'_{t\tau} A_T + o_p(1). \quad (\text{A.38})$$

To justify this relation, notice that now $\beta' y_{t-1,\hat{\tau}} = \beta' x_{t-1}$ which shows that (A.38) holds for the first r columns of the involved matrices. For the next $n - r$ columns the result is simple because the sums are divided by $NT^{1/2}$. The same can be said about the $(n + 1)$ th column. That the stated result holds for the last column can be seen from

$$\begin{aligned} N^{-1} \sum_{t=p+1}^T \Delta y_{t-j,\hat{\tau}} d_{t-1,\tau} &= N^{-1} \sum_{t=\tau+1}^T \Delta x_{t-j} + o_p(1) \\ &= o_p(1) \quad (j = 0, \dots, p-1) \end{aligned} \quad (\text{A.39})$$

uniformly in $[T\underline{\lambda}] \leq \tau \leq [T\bar{\lambda}]$. Here the former equality is an immediate consequence of (A.27) and the latter follows because an invariance principle applies to partial sums of Δx_t .

In addition to (A.36) - (A.38), we also need to consider the matrix $N^{-1} \sum_{t=p+1}^T \mathbf{p}_{1t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}} A_T$. Arguments used above show that replacing $\hat{\tau}$ here by τ causes an error of order $o_p(1)$ except for the element in the first row and last column which is

$$\left(\frac{T}{T-\tau} \right)^{1/2} N^{-1} \sum_{t=p+1}^T d_{t-1,\hat{\tau}} = O_p(1).$$

However, the contribution of this element to the matrix on the l.h.s. of (A.33) is of order $o_p(1)$. Because in the definitions of $S_{01\hat{\tau}}$ and $S_{01\tau}$ we can replace $\mathbf{p}_{t\hat{\tau}}$ and $\mathbf{p}_{t\tau}$ by $\mathbf{p}_{1t\hat{\tau}}$ and $\mathbf{p}_{1t\tau}$, respectively, this follows from the following two facts. (i) The first matrix on the r.h.s. of (A.36) is asymptotically block diagonal with blocks defined after the first row and first column. (ii) The first column of the first matrix on the r.h.s. of (A.37) is of order $o_p(1)$. Both of these results follow from (A.27) and the fact that Δx_t obeys a weak law of large numbers. Taking these results together, we can thus conclude that (A.33) also holds when $\delta_1 = 0$.

Since (A.32) and (A.33) hold we can write

$$A'_T S_{10\hat{\tau}} S_{00\hat{\tau}}^{-1} S_{01\hat{\tau}} A_T = A'_T S_{10\tau} S_{00\tau}^{-1} S_{01\tau} A_T + o_p(1) = \begin{bmatrix} \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix} + o_p(1), \quad (\text{A.40})$$

where the latter equality is justified by (A.5) of S&L and the notation is as explained therein. The partition is after the first r rows and columns.

Now consider the matrix $A'_T S_{11\hat{\tau}} A_T$. In the present case we do not have (A.34) which, given (A.40), would be sufficient for the consistency of the RR estimator of ψ . However, (A.34) is not necessary. Write

$$A'_T S_{11\hat{\tau}} A_T = \begin{bmatrix} \bar{S}_{11\hat{\tau}}^{11} & \bar{S}_{11\hat{\tau}}^{12} \\ \bar{S}_{11\hat{\tau}}^{21} & \bar{S}_{11\hat{\tau}}^{22} \end{bmatrix},$$

where the partition is after the first r rows and columns. Because $\beta' y_{t-1, \hat{\tau}} = \beta' x_{t-1}$ by (A.27), we can show that $\bar{S}_{11\hat{\tau}}^{11} = \Sigma_{\beta\beta} + o_p(1)$ with $\Sigma_{\beta\beta}$ as in S&L and $\bar{S}_{11\hat{\tau}}^{12} = o_p(1)$. The required arguments are based on (A.14) and (A.27) in the same way as in the case of (A.32), (A.33) and (A.39). Without going into the lengthy details we note that, although the estimator $\hat{\tau}$ in the last component of $A'_T \mathbf{x}_{t\hat{\tau}}$ cannot be replaced by τ , this feature does not affect the above results (cf. the above justification of (A.33)). Thus, because we also have $\bar{S}_{11\hat{\tau}}^{21} = o_p(1)$, the above discussion and (A.40) show that equation (A.31) is to order $o_p(1)$ identical to

$$\det(\lambda \Sigma_{\beta\beta} - \Sigma_{\beta 0} \Sigma_{00}^{-1} \Sigma_{0\beta}) \det(\lambda \bar{S}_{11\hat{\tau}}^{22}) = 0. \quad (\text{A.41})$$

This implies that the consistency proof given in Johansen (1995, pp. 180 - 181) applies if, with probability approaching one, $\lambda_{\min}(\bar{S}_{11\hat{\tau}}^{22}) \geq \epsilon$ for some $\epsilon > 0$. This, however, is the case because the argument used below (A.15) shows that $\bar{S}_{11[T\lambda]}^{22}$ converges weakly in $D([\underline{\lambda}, \bar{\lambda}])$ to a (a.s.) positive definite limit. In fact, arguments used to arrive at (A.41) show that $\bar{S}_{11[T\lambda]}^{22}$ is to order $o_p(1)$ identical to a demeaned version of the matrix of second sample moments formed from the last $n - r + 2$ components of $A'_T \mathbf{x}_{t[T\lambda]}$.

Thus, in the same way as in the case $\delta_1 \neq 0$ we can conclude that the estimators $\tilde{\beta}_\xi, \tilde{\phi}$ and $\tilde{\theta}$ are consistent of orders $o_p(T^{-1/2}), o_p(T^{-1})$ and $o_p(1)$, respectively. After this, the consistency of the estimators $\tilde{\alpha}_\xi, \tilde{\Gamma}_j$ ($j = 1, \dots, p - 1$) and $\tilde{\Omega}$ can be proved in the same way as in S&L. Since the second form in the definition of $\varepsilon_{t\hat{\tau}}$ in (A.26) now vanishes the employed arguments consist of similar analyses of second sample moments already carried out at previous steps of the proof. Details are therefore omitted.

To obtain the orders of consistency, consider the first order conditions (A.35) and notice that now $\psi' \mathbf{x}_{t\hat{\tau}} = \beta' x_{t-1}$. Using this fact it can be seen that $\psi' S_{11\hat{\tau}} B_T = O_p(1)$ and $S_{\varepsilon 1\hat{\tau}} B_T = O_p(1)$. The employed arguments are similar to those used in the above consistency proof and in the proof of Theorem 3.1. To give an example of the latter, note that in order to prove $\psi' S_{11\hat{\tau}} B_T = O_p(1)$ we need to show that

$$(T - \tau)^{-1/2} \sum_{t=p+1}^T \beta' x_{t-1} d_{t-1,\hat{\tau}} = (T - \tau)^{-1/2} \sum_{t=\hat{\tau}+1}^T \beta' x_{t-1} = O_p(1).$$

This, however, follows from the relevant part of (A.12). Omitting other details we need to note that $\lambda_{\min}(T^{-1} B_T S_{11\hat{\tau}} B_T) = \lambda_{\min}(\bar{S}_{11\hat{\tau}}^{22})$, which is asymptotically bounded away from zero, as noticed after (A.41). From (A.35) and what has been said above we now find that $U_T = O_p(1)$. In the same way as in the case $\delta_1 \neq 0$ this implies $\tilde{\beta}_\xi = \beta + O_p(T^{-1})$ and more generally $A_T^{-1}(\tilde{\psi} - \psi) = O_p(T^{-1/2})$. After this the remaining orders of consistency can be obtained in the same way as in S&L. Details are again based on arguments already employed in the foregoing. This completes the proof. \square

Next we prove a counterpart of Theorem 2.1 of S&L.

Lemma A.11.

$$\beta'(\hat{\mu}_0 - \mu_0) = O_p(T^{-1/2}), \quad (\text{A.42})$$

$$\beta'_\perp(\hat{\mu}_0 - \mu_0) = O_p(1), \quad (\text{A.43})$$

$$\beta'(\hat{\delta} - \delta) = O_p(T^{-1/2}), \quad (\text{A.44})$$

$$\beta'_\perp(\hat{\delta} - \delta) = O_p(1), \quad (\text{A.45})$$

$$\beta'(\hat{\mu}_1 - \mu_1) = O_p(T^{-3/2}), \quad (\text{A.46})$$

$$T^{1/2} \beta'_\perp(\hat{\mu}_1 - \mu_1) \xrightarrow{d} N(0, \beta'_\perp C \Omega C' \beta_\perp). \quad (\text{A.47})$$

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

Proof: The idea is to show how the stated results can be obtained from the proof of Theorem 2.1 of S&L. In doing this we ignore the fact that S&L had additional impulse dummies in their model because this feature has no essential effect on the arguments.

We shall first introduce a reparameterized form of the auxiliary regression model (3.7) which is a counterpart of equation (A.12) of S&L. To this end, define

$$k_1 = \begin{bmatrix} k_{11} \\ k_{12} \end{bmatrix} = \begin{bmatrix} \tilde{\beta}'_{\perp} \delta \\ \tilde{\beta}'_{\perp} \mu_0 \end{bmatrix} \quad \text{and} \quad k_2 = \begin{bmatrix} k_{21} \\ k_{22} \\ k_{23} \\ k_{24} \end{bmatrix} = \begin{bmatrix} \tilde{\beta}' \delta \\ \tilde{\beta}' \mu_0 \\ \tilde{\beta}' \mu_1 \\ \tilde{\beta}'_{\perp} \mu_1 \end{bmatrix}.$$

Next transform the regressors in (3.7) accordingly as $\tilde{F}_{11t} = \tilde{Q}' \tilde{K}_{t\hat{\tau}} \tilde{\beta}_{\perp} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1}$, $\tilde{F}_{12t} = \tilde{Q}' \tilde{H}_{0t} \tilde{\beta}_{\perp} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1}$ and $\tilde{F}_{21t} = \tilde{Q}' \tilde{K}_{t\hat{\tau}} \tilde{\beta} (\tilde{\beta}' \tilde{\beta})^{-1}$, $\tilde{F}_{22t} = \tilde{Q}' \tilde{H}_{0t} \tilde{\beta} (\tilde{\beta}' \tilde{\beta})^{-1}$, $\tilde{F}_{23t} = \tilde{Q}' \tilde{H}_{1t} \tilde{\beta} (\tilde{\beta}' \tilde{\beta})^{-1}$, $\tilde{F}_{24t} = \tilde{Q}' \tilde{H}_{1t} \tilde{\beta}_{\perp} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1}$. Then, setting $\tilde{F}_{1t} = [\tilde{F}_{11t} : \tilde{F}_{12t}]$ and $\tilde{F}_{2t} = [\tilde{F}_{21t} : \tilde{F}_{22t} : \tilde{F}_{23t} : \tilde{F}_{24t}]$, we can write (3.7) as

$$\tilde{Q}' \tilde{A}(L) y_t = \tilde{F}_{1t} k_1 + \tilde{F}_{2t} k_2 + \eta_{t\hat{\tau}}, \quad t = 1, \dots, T. \quad (\text{A.48})$$

As in S&L, \tilde{F}_{1t} takes nonzero values only for a fixed number of time indices t .

Because (A.48) is obtained from (2.1) by first replacing $d_{t\tau}$ by $d_{t\hat{\tau}}$ and then premultiplying by $\tilde{Q}' \tilde{A}(L)$, the error term $\eta_{t\hat{\tau}}$ differs from its counterpart η_t in S&L by the additive term $-\tilde{Q}' \tilde{A}(L) \delta (d_{t\hat{\tau}} - d_{t\tau})$. Thus, using (2.8) we have

$$\begin{aligned} \eta_{t\hat{\tau}} &= \eta_{t\tau} + \tilde{Q}' \tilde{\alpha} \tilde{\beta}' \delta (d_{t-1,\hat{\tau}} - d_{t-1,\tau}) - \tilde{Q}' \delta (\Delta d_{t\hat{\tau}} - \Delta d_{t\tau}) \\ &\quad + \tilde{Q}' \sum_{j=1}^{p-1} \tilde{\Gamma}_j \delta (\Delta d_{t-j,\hat{\tau}} - \Delta d_{t-j,\tau}), \end{aligned} \quad (\text{A.49})$$

where $\eta_{t\tau} = \tilde{Q}' \varepsilon_t - \tilde{Q}' \tilde{\alpha} (\tilde{\beta} - \beta)' x_{t-1} - \tilde{Q}' (\tilde{\alpha} - \alpha) \beta' x_{t-1} - \tilde{Q}' \sum_{j=1}^{p-1} (\tilde{\Gamma}_j - \Gamma_j) \Delta x_{t-j}$ is identical to η_t defined in (A.13) of S&L.¹

Now assume that $\delta_1 \neq 0$. Then Lemmas A.9 and A.10 in conjunction with the arguments used in S&L to justify (A.14) and (A.15) of that paper show that

$$\sum_{t=1}^T \tilde{F}'_{1t} \eta_{t\hat{\tau}} = \sum_{t=1}^T \tilde{F}'_{1t} \eta_{t\tau} + O_p(1) = O_p(1) \quad (\text{A.50})$$

and

$$\Upsilon_T^{-1} \sum_{t=1}^T \tilde{F}'_{2t} \eta_{t\hat{\tau}} = \Upsilon_T^{-1} \sum_{t=1}^T \tilde{F}'_{2t} \eta_{t\tau} + o_p(1) = O_p(1), \quad (\text{A.51})$$

where $\Upsilon_T^{-1} = \text{diag}[(T - \tau)^{-1/2} I_r : T^{-1/2} I_r : T^{-3/2} I_r : T^{-1/2} I_{n-r}]$. Standardizing the moment matrix of the auxiliary regression model (A.48) to conform to the standardizations of the

¹Note that there is a typo in (A.13) of S&L because in the third term on the r.h.s. there is $\tilde{\beta}$ in place of β .

sums in (A.50) and (A.51) and using Lemmas A.9 and A.10 shows that the resulting matrix is asymptotically equivalent to its counterpart obtained when the break date is known. This combined with (A.50) and (A.51) implies that we have reduced the problem to that treated in Theorem 2.1 of S&L. Because we can write $O_p(T^{-1/2}) = O_p((T - \tau)^{-1/2})$ in (A.44) this gives the stated results when $\delta_1 \neq 0$.

To complete the proof, assume that $\delta_1 = 0$ and observe that in the second term on the r.h.s. of (A.49) we have $\tilde{\alpha}\tilde{\beta}'\delta = \tilde{\alpha}_\xi(\tilde{\beta}_\xi - \beta)'\delta = O_p(T^{-1})$ by Lemma A.10. Using this fact, it is easy to check that (A.50) also holds in the present case. Regarding subsequent derivations it appears convenient to replace Υ_T in (A.51) by $\hat{\Upsilon}_T$ obtained by using $\hat{\tau}$ in place of τ . This change does not affect rates of convergence because $[T\lambda] \leq \hat{\tau} \leq [T\bar{\lambda}]$ by assumption. The arguments used to justify (A.50) combined with Lemma A.10 now straightforwardly show that (A.51) holds with Υ_T^{-1} replaced by $\hat{\Upsilon}_T^{-1}$.

Next consider the standardized moment matrix related to the LS estimation of the parameters k_1 and k_2 in (A.48). Here we use arguments similar to those used in the discussion paper version of S&L to prove Theorem 2.1 therein. First note that arguments already used above give $\hat{\Upsilon}_T^{-1} \sum_{t=1}^T \tilde{F}'_{2t} \tilde{F}_{1t} = o_p(1)$ so that the standardized moment matrix is asymptotically block diagonal. From the definition of \tilde{F}_{2t} it can further be shown that

$$\begin{aligned} \hat{\Upsilon}_T^{-1} \sum_{t=1}^T \tilde{F}'_{2t} \tilde{F}_{2t} \hat{\Upsilon}_T^{-1} &= \sum_{t=1}^T \begin{pmatrix} -(T - \hat{\tau})^{-1/2} \tilde{\alpha}' \tilde{Q} d_{t\hat{\tau}} \\ -T^{-1/2} \tilde{\alpha}' \tilde{Q} \\ -T^{-3/2} (t - 1) \tilde{\alpha}' \tilde{Q} \\ -T^{-1/2} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1} \tilde{\beta}'_{\perp} \tilde{\Psi}' \tilde{Q} \end{pmatrix} \begin{pmatrix} -(T - \hat{\tau})^{-1/2} \tilde{\alpha}' \tilde{Q} d_{t\hat{\tau}} \\ -T^{-1/2} \tilde{\alpha}' \tilde{Q} \\ -T^{-3/2} (t - 1) \tilde{\alpha}' \tilde{Q} \\ -T^{-1/2} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1} \tilde{\beta}'_{\perp} \tilde{\Psi}' \tilde{Q} \end{pmatrix}' + o_p(1) \\ &= \begin{bmatrix} \hat{E}_T \otimes \tilde{A}_{11} & \hat{e}_T \otimes \tilde{A}_{12} \\ \hat{e}'_T \otimes \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix} + o_p(1). \end{aligned} \quad (\text{A.52})$$

Here the notation is defined as follows: $\tilde{A}_{11} = \tilde{\alpha}' \tilde{\Omega} \tilde{\alpha}$, $\tilde{A}_{12} = \tilde{A}'_{21} = \tilde{\alpha}' \tilde{\Omega}^{-1} \tilde{\Psi} \tilde{\beta}_{\perp} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1}$ and $\tilde{A}_{22} = (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1} \tilde{\beta}'_{\perp} \tilde{\Psi}' \tilde{\Omega}^{-1} \tilde{\Psi} \tilde{\beta}_{\perp} (\tilde{\beta}'_{\perp} \tilde{\beta}_{\perp})^{-1} = \tilde{A}_{21} \tilde{A}_{11}^{-1} \tilde{A}_{12} + \tilde{B}^{-1}$, where

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

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

$$\hat{E}_T = \begin{bmatrix} 1 & \left(\frac{T - \hat{\tau}}{T}\right)^{1/2} & T^{-3/2} (T - \hat{\tau})^{-1/2} \sum_{t=\hat{\tau}+1}^T (t - 1) \\ \left(\frac{T - \hat{\tau}}{T}\right)^{1/2} & 1 & T^{-2} \sum_{t=1}^T (t - 1) \\ T^{-3/2} (T - \hat{\tau})^{-1/2} \sum_{t=\hat{\tau}+1}^T (t - 1) & T^{-2} \sum_{t=1}^T (t - 1) & T^{-3} \sum_{t=1}^T (t - 1)^2 \end{bmatrix},$$

while the vector \hat{e}_T is the second column of \hat{E}_T , that is,

$$\hat{e}_T = \left[\left(\frac{T - \hat{\tau}}{T} \right)^{1/2} : 1 : T^{-2} \sum_{t=1}^T (t-1) \right]'$$

Because now $\hat{\tau}/T$ does not converge, the same is true for \hat{E}_T and \hat{e}_T and, consequently, for the first term in the last expression of (A.52). However, using the definitions and Lemma A.10, it can be shown that the smallest eigenvalue of this matrix is asymptotically bounded away from zero. Because it is easy to check that this is also the case for the smallest eigenvalue of the matrix $\sum_{t=1}^T \tilde{F}'_{1t} \tilde{F}_{1t}$, we can conclude that the smallest eigenvalue of the standardized moment matrix related to the LS estimation of k_1 and k_2 in (A.48) is similarly asymptotically bounded away from zero. Thus, because we also have (A.50) and (A.51) with Υ_T replaced by $\hat{\Upsilon}_T$, the LS estimators obtained from (A.48) satisfy $\hat{k}_1 = k_1 + O_p(1)$ and $\hat{\Upsilon}_T(\hat{k}_2 - k_2) = O_p(1)$. Because here $\hat{\Upsilon}_T$ can be replaced by Υ_T without changing rates of convergence, we can conclude in the same way as in the proof of Theorem 2.1 of S&L that the results of the lemma hold except for the weak limit in (A.47).

To see that (A.47) holds, notice that, according to our previous derivations, we can write

$$\hat{\Upsilon}_T(\hat{k}_2 - k_2) = \left(\hat{\Upsilon}_T^{-1} \sum_{t=1}^T \tilde{F}'_{2t} \tilde{F}_{2t} \hat{\Upsilon}_T^{-1} \right)^{-1} \hat{\Upsilon}_T \sum_{t=1}^T \tilde{F}'_{2t} \eta_{t\tau} + o_p(1).$$

We have to consider the last $n - r$ components of the vector on the r.h.s. and therefore calculate the last $n - r$ rows of the inverse on the r.h.s.. For this purpose we apply the well-known formula for the partitioned inverse to the first term in the last expression of (A.52). Because it is easy to check that $\hat{e}'_T \hat{E}_T^{-1} = [0 : 1 : 0]$ and $\hat{e}'_T \hat{E}_T^{-1} \hat{e}_T = 1$, this yields

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

where the blocks denoted by ‘*’ are not needed and the partition on the r.h.s. conforms to the four components of the estimator vector \hat{k}_2 . Thus, we can write

$$T^{1/2}(\hat{k}_{24} - k_{24}) = -\tilde{B} \tilde{A}_{21} \tilde{A}_{22}^{-1} T^{-1/2} \sum_{t=1}^T \tilde{F}'_{22t} \eta_{t\tau} + \tilde{B} T^{-1/2} \sum_{t=1}^T \tilde{F}'_{24t} \eta_{t\tau} + o_p(1).$$

This is asymptotically equivalent to what was obtained in S&L so that (A.45) follows from that paper. \square

Now we can prove Theorem 4.1. First note that

$$\hat{x}_t = x_t - (\hat{\mu}_0 - \mu_0) - (\hat{\mu}_1 - \mu_1)t - (\hat{\delta} - \delta)d_{t\tau} - \hat{\delta}(d_{t\hat{\tau}} - d_{t\tau}). \quad (A.53)$$

This expression differs from its counterpart in the proof of Theorem 3.1 of S&L by the last term if we again ignore the unessential impulse dummies in S&L. This implies that the error term $e_{t\hat{\tau}}$ in (4.2) can be written as

$$e_{t\hat{\tau}} = e_{t\tau} - \Pi\hat{\delta}(d_{t\hat{\tau}} - d_{t\tau}) - \sum_{j=0}^{p-1} \gamma_j(\Delta d_{t-j,\hat{\tau}} - \Delta d_{t-j,\tau}), \quad (A.54)$$

where $e_{t\tau}$ equals the error term e_t defined in the proof of Theorem 3.1 of S&L except that it is based on different estimators and some unessential impulse dummies are ignored. The difference in estimators has no consequences, however, because the required asymptotic properties given in Lemma A.11 and Theorem 2.1 of S&L are the same. Thus, when $\delta_1 \neq 0$, we can use Lemmas A.9 and A.11 in conjunction with (A.53) and (A.54) to conclude that the standardized second sample moments of $\beta'\hat{x}_{t-1}$, $\beta'_\perp\hat{x}_{t-1}$ and $\Delta\hat{x}_{t-j}$ ($j = 1, \dots, p-1$) are asymptotically equivalent to their counterparts in S&L. By similar arguments it can be further shown that the standardized second sample moments between $e_{t\hat{\tau}}$ and $[\hat{x}'_{t-1}\beta : \Delta\hat{x}'_{t-1} : \dots : \Delta\hat{x}'_{t-p+1}]'$ as well as between $e_{t\hat{\tau}}$ and $\beta'_\perp\hat{x}_{t-1}$ are asymptotically equivalent to their counterparts in S&L. The discussion given at the end of the proof of Theorem 3.1 of S&L then implies that the RR estimators of α , β and Ω based on (4.2) have the same consistency properties as the corresponding estimators in Lemma A.10. This means that we can proceed in the same way as in the proof of Theorem 3.1 of S&L with $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\Omega}$ interpreted as the RR estimators based on (4.2).

It can be seen analogously that the same conclusion also holds when $\delta_1 = 0$, except that instead of Lemma A.9 we now use the fact that $\beta'\hat{\delta} = O_p(T^{-1/2})$ by (A.44). In this case it should be noted that the standardized second sample moments between $e_{t\hat{\tau}}$ and $[\hat{x}'_{t-1}\beta : \Delta\hat{x}'_{t-1} : \dots : \Delta\hat{x}'_{t-p+1}]'$ and between $e_{t\hat{\tau}}$ and $\beta'_\perp\hat{x}_{t-1}$ are not asymptotically equivalent to their counterparts in S&L. However, they are of order $O_p(1)$ which is sufficient for the RR estimators of α , β and Ω based on (4.2) to have the same consistency properties as the corresponding estimators in Lemma A.10. Thus we can proceed in the same way as in the proof of Theorem 3.1 of S&L with $\tilde{\alpha}$, $\tilde{\beta}$ and $\tilde{\Omega}$ interpreted as the RR estimators based on (4.2).

The proof can be completed by showing that the problem can be reduced to that treated in the proof of Theorem 3.1 of S&L. This means showing that the last term on the r.h.s. of (A.53) and the two last terms on the r.h.s. of (A.54) have no effect on the limiting distribution of the test statistic or on the asymptotic behavior of the counterparts of the quantities defined in (A.18) - (A.22) of S&L. This can be done by using arguments already used in previous proofs in conjunction with Lemma A.9 when $\delta_1 \neq 0$ and the fact $\tilde{\beta}'_x \hat{\delta} = \beta' \hat{\delta} + O_p(T^{-1}) = O_p(T^{-1/2})$ when $\delta_1 = 0$. Details are straightforward but tedious and will be omitted.

Appendix B. Data Sources

B.1 Polish Crawling Peg System

Daily data for the period from March 1, 1993 to February 28, 1994 were used.

Polish zloty-US dollar rate: official exchange rate announced by the National Bank of Poland (via http://www.nbp.pl/asps/Arch_en.asp). The variable zl_t is the logarithm of the zloty-US dollar rate.

Exchange rates of basket currencies: rates of German mark, British pound, French franc and Swiss franc against the U.S. dollar are noon buying rates in New York for cable transfers certified by the Federal Reserve Bank (FRB) of New York (via <http://www.federal-reserve.gov/releases/H10/hist/>). The variables gm_t , bp_t , ff_t , and sf_t are the logarithms of these rates, respectively.

In case of holidays at just one exchange market the respective data for the market are deleted. Therefore 249 observations are available.

B.2 Price-Wage Dynamics in Poland

Quarterly, seasonally non-adjusted data for the period from the first quarter of 1970 to the fourth quarter of 1998 (116 observations) were used for the following variables. If not otherwise stated, data have been provided by the Macroeconomic and Financial Data Centre of the University of Gdansk, Poland (via http://bg.univ.gda.pl/pub/m&fdc/lam_data/) with the sources given below.

Consumer Price Index: base 1985 = 100 (source: Polish Central Statistical Office (CSO)).

The variable cpi_t is the logarithm of the Consumer Price Index.

Nominal Wage: gross monthly per capita wage in socialist sector up to fourth quarter in 1989 and in national economy afterwards (source: CSO). The variable w_t is the logarithm of the nominal wage.

Exchange rate: US dollar market exchange rate against the Polish Zloty (sources: Picks Currency Yearbook, Currency Alert, Polish newspapers, CSO). The variable ex_t is the logarithm of the exchange rate.

Industrial Production Index: base 1985 = 100 (source: CSO).

Employment: average paid employment in socialist sector up to the fourth quarter in 1989 and in national economy afterwards (source: CSO). The variable lp_t is the difference of the logarithms of the Industrial Production Index and the employment figures.

Population: in thousands, taken from the *International Financial Statistics, IMF*. The variable $epop_t$ is the difference of the logarithms of the employment and the population figures.

References

- Bai, J. (1994), Least squares estimation of a shift in linear processes, *Journal of Time Series Analysis*, 15, 453 - 472.
- Bai, J., R.L. Lumsdaine & 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 & 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.

- Christiano, L.J. (1992), Searching for a break in GNP, *Journal of Business & Economic Statistics*, 10, 237 - 250.
- Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press.
- Evans, G.W. (1989), Output and unemployment dynamics in the United States: 1950 - 1985, *Journal of Applied Econometrics*, 4, 213 - 237.
- Gonzalo, J. & J.-Y. Pitarakis (1999), Dimensionality effects in cointegration analysis, in: R.F. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, pp. 212 - 229.
- Golinelli, R. & R. Orsi (1994), Price-Wage Dynamics in a transition economy, *Economics of Planning*, 27, 293-313.
- Golinelli, R. & R. Orsi (1998), Exchange rate, inflation and unemployment in East European economies: the case of Poland and Hungary, *Economics of Planning*, 31, 29-55.
- Golinelli, R. & R. Orsi (2000), Testing for structural change in cointegrated relationships: analysis of price-wages models for Poland and Hungary, *Economics of Planning*, 33, 19 - 51.
- Gregory, A.W. & 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 & Economic Statistics*, 10, 321 - 335.
- Hubrich, K., H. Lütkepohl & P. Saikkonen (2001), A review of systems cointegration tests, *Econometric Reviews*, forthcoming.
- 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 & B. Nielsen (2000), Cointegration analysis in the presence of structural breaks in the deterministic trend, *Econometrics Journal*, 3, 216 - 249.

- Johansen, S. & B. Nielsen (1993), Manual for the simulation program DisCo. Institute of Mathematical Statistics, University of Copenhagen: <http://math.ku.dk/~sjo/disco/disco.ps>.
- Lanne, M., H. Lütkepohl & P. Saikkonen (2001), Comparison of unit root tests for time series with level shifts, *Journal of Time Series Analysis*, forthcoming. Discussion Paper No. 88, 1999, SFB 373, Humboldt-Universität zu Berlin, <http://sfb.wiwi.hu-berlin.de>.
- Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Berlin: Springer-Verlag.
- Lütkepohl, H. & P. Saikkonen (1999), Order selection in testing for the cointegrating rank of a VAR process, in: R. F. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford University Press, Oxford, pp. 168–199.
- Lütkepohl, H. & 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 & C. Trenkler (2000), Comparison of tests for the cointegrating rank of a VAR process with a structural shift, Discussion Paper No. 10, 2000, SFB 373, Humboldt-Universität zu Berlin, <http://sfb.wiwi.hu-berlin.de>.
- Marcellino, M. & G.E. Mizon (2000), Modelling shifts in the wage-price and unemployment-inflation relationship in Italy, Poland and the UK, *Economic Modelling*, 17, 387 - 413.
- 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.
- Quintos, C.E. (1998), Stability tests in error correction models, *Journal of Econometrics*, 82, 289 - 315.

- Saikkonen, P. & H. Lütkepohl (1999), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory*, 15, 50 - 78.
- Saikkonen, P. & H. Lütkepohl (2000a), Testing for the cointegrating rank of a VAR process with structural shifts, *Journal of Business & Economic Statistics*, 18, 451 - 464.
- Saikkonen, P. & H. Lütkepohl (2000b), Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process, *Journal of Time Series Analysis*, 21, 435 - 456.
- Saikkonen, P. & H. Lütkepohl (2001), Testing for a unit root in a time series with a level shift at unknown time, *Econometric Theory*, forthcoming. Discussion Paper No. 72, 1999, SFB 373, Humboldt-Universität zu Berlin, <http://sfb.wiwi.hu-berlin.de>.
- 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.
- Toda, H.Y. (1995), Finite sample performance of likelihood ratio tests for cointegration ranks in vector autoregressions, *Econometric Theory*, 11, 1015-1032.
- Trenkler, C. (2000), Cointegration and the Polish crawling peg system, Discussion Paper No. 71, 2000, SFB 373, Humboldt-Universität zu Berlin, <http://sfb.wiwi.hu-berlin.de>.