August 18, 2004

Break Date Estimation and Cointegration Testing in VAR Processes with Level Shift

by

Pentti Saikkonen
University of Helsinki

Helmut Lütkepohl
European University Institute, Florence and Humboldt University Berlin

and
Carsten Trenkler
Humboldt University Berlin

Abstract

In testing for the cointegrating rank of a vector autoregressive (VAR) process it is important to take into account level shifts that have occurred in the sample period. Therefore the properties of estimators of the time period where a shift has taken place are investigated. The possible structural break is modelled as a simple shift in the level of the process. Three alternative estimators for the break date are considered and their asymptotic properties are derived under various assumptions regarding the size of the shift. In particular, properties of the shift date estimator are obtained under the assumption of an increasing or decreasing size of the shift when the sample size grows. Moreover, the implications for testing the cointegrating rank of the process are explored. A new rank test is proposed and its asymptotic properties are derived. It is shown that its asymptotic null distribution is unaffected by the level shift. The performance of the shift date estimators and the cointegration rank tests in small samples is investigated by simulations.

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

*We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) for financial support. The first author also acknowledges financial support by the Yrjö Jahnsson Foundation, the Academy of Finland and the Alexander von Humboldt Foundation under a Humboldt research award. Part of this research was done while he was visiting the Humboldt University in Berlin and part of the research was carried out while he and the third author were visiting the European University Institute, Florence.

1 Introduction

From the unit root and cointegration testing literature it is well-known that structural shifts in the time series of interest have a major impact on inference procedures. In particular, they affect the small sample and asymptotic properties of unit root and cointegrating rank tests (see, e.g., Perron (1989) for unit root testing and Lütkepohl, Saikkonen & Trenkler (2004) for cointegration rank testing). In the latter article, henceforth abbreviated as LST, it is assumed that a level shift has occurred in a system of time series variables at an unknown time. LST propose to estimate the shift date in a first step and then apply a cointegrating rank test as follows. First 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. They provide conditions under which the asymptotic null distribution of the cointegrating rank tests in this procedure is unaffected by the level shift. They also show, however, that in small samples the way the break date is estimated may have an impact on the actual properties of the cointegrating rank test. In particular, the size of the level shift is important for the small sample properties of the break date estimators and the tests.

Therefore, in this study we extend the results of LST in several directions. First of all we consider a further possible break date estimator. Second, we derive asymptotic properties of all the estimators accounting explicitly for the size of the level shift. More precisely, we make the size of the level shift dependent on the sample size and provide asymptotic results both for increasing and decreasing shift size when the sample size goes to infinity. These results provide interesting new insights in the properties of the estimators and explain simulation results of LST which are difficult to understand if a fixed shift size is considered. Under our assumptions the null distribution of the cointegrating rank tests is still unaffected by the shift or the shift size just as in the case of a fixed shift size. We also modify the cointegrating rank tests considered by LST. In their approach estimators of all parameters associated with the deterministic part of the model are estimated by the GLS procedure although the level parameters are not fully identified. In this paper we propose to estimate the identified parameters only and modify the cointegrating rank tests accordingly. Finally, we perform a more detailed and more insightful investigation of the small sample properties of three different break date estimators and the resulting cointegrating rank tests by extending the simulation design of LST.

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, we do not derive the asymptotic distribution of the estimators but focus on rates of convergence. Our results are important for investigating the properties of inference procedures such as cointegration rank tests that are based on a VAR model with estimated break date. Although Bai, Lumsdaine & Stock (1998) also discuss shift sizes that depend on the sample size, our results go beyond their analysis because we consider decreasing as well as increasing shift sizes.

The study is structured as follows. In Section 2, the modelling framework of LST is summarized because that will be the basis for our investigation. Section 3 is devoted to a discussion of the break date estimators and their asymptotic properties. The properties of cointegrating rank tests based on a model with estimated break date are considered in Section 4 and small sample simulation results of the break date estimators and the cointegrating rank tests are presented in Section 5. In Section 6, a summary and conclusions are given. The proofs of several theorems stated in the main body of the paper are given in the Appendix.

The following general notation will be used. The differencing and lag operators are denoted by Δ and L, respectively. 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 $\stackrel{d}{\rightarrow}$ and i.i.d. stands for independently, identically distributed. The symbols for boundedness and convergence in probability are as usual $O_p(\cdot)$ and $o_p(\cdot)$, respectively. Moreover, $\|\cdot\|$ denotes the Euclidean norm. The trace, determinant and rank of the matrix A are denoted by $\text{tr}(A)$, $\det(A)$ 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 $A_\perp$. The zero matrix is the orthogonal complement of a nonsingular square matrix and an identity matrix of suitable dimension is the orthogonal complement of a zero matrix. An $(n \times n)$ identity matrix is denoted by I_n. For matrices $A_1, \dots, A_s$, $\text{diag}[A_1 : \cdots : A_s]$ is the block-diagonal matrix with $A_1, \dots, A_s$ on the diagonal. LS, GLS, RR and VECM are used to abbreviate least squares, generalized least squares, reduced rank and vector error correction model, respectively. As usual, a sum is defined to be zero if the lower bound of the summation index exceeds the upper bound.

2 The Data Generation Process

We use the general setup of LST. 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,

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

Here μ_i $(i = 0, 1)$ and δ are unknown $(n \times 1)$ parameter vectors and $d_{t\tau}$ is a shift dummy variable representing a shift in period τ so that

$$d_{t\tau} = 0 \text{ for } t < \tau \text{ and } 1 \text{ for } t \geq \tau. \tag{2.2}$$

We make the following assumption for the shift date τ.

Assumption 1. Let $\underline{\lambda}$, λ and $\bar{\lambda}$ be fixed real numbers such that $0 < \underline{\lambda} \leq \lambda \leq \bar{\lambda} < 1$. The shift date τ satisfies

$$\tau = [T\lambda], \tag{2.3}$$

where $[\cdot]$ denotes the integer part of the argument. □

In other words, the shift is assumed to occur at a fixed fraction of the sample length. The shift date may not be at the very beginning or at the very end of the sample, although $\underline{\lambda}$ and $\bar{\lambda}$ may be arbitrarily close to zero and one, respectively. The condition has also been employed by Bai, Lumsdaine & Stock (1998) in models containing $I(1)$ variables. It is obviously not very restrictive.

The term $\mu_1 t$ may be dropped from (2.1), if $\mu_1 = 0$ is known to hold and, thus, the DGP does not have a deterministic linear trend. 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 without major changes to our arguments. They are not included in our basic model to avoid more complex notation.

The process x_t is assumed to be at most $I(1)$ and have a VAR(p) representation. More precisely, we make the following assumption.

Assumption 2. The process x_t is integrated of order at most $I(1)$ with cointegrating rank r and

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

where the A_j are $(n\times n)$ coefficient matrices. The initial values x_t, $t \leq 0$, are assumed to be such that the cointegration relations and Δx_t are stationary. The ε_t are i.i.d.$(0, \Omega)$ with positive definite covariance matrix Ω and existing moments of order $b > 4$. □

Under Assumption 2, the process x_t has the VECM form

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

where $\Pi = -(I_n - A_1 - \cdots - A_p)$ and $\Gamma_j = -(A_{j+1} + \cdots + A_p)$ $(j = 1, \ldots, p-1)$ are $(n\times n)$ matrices. Because the cointegrating rank is r, the matrix Π can be written as $\Pi = \alpha\beta'$, where α and β are $(n\times r)$ matrices of full column rank. As is well-known, $\beta' x_t$ and Δx_t are then zero mean $I(0)$ processes. Defining $\Psi = I_n - \Gamma_1 - \cdots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} j A_{j+1}$ and $C = \beta_\perp(\alpha'_\perp \Psi \beta_\perp)^{-1}\alpha'_\perp$, we have

$$x_t = C \sum_{j=1}^{t} \varepsilon_j + \xi_t, \quad t = 1, 2, \ldots, \tag{2.6}$$

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

Multiplying (2.1) by $A(L) = I_n - A_1 L - \cdots - A_p L^p = I_n \Delta - \Pi L - \Gamma_1 \Delta L - \cdots - \Gamma_{p-1} \Delta L^{p-1}$ yields

$$\begin{gathered}\Delta y_t = \nu + \alpha(\beta' y_{t-1} - \phi(t-1) - \theta d_{t-1,\tau}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j,\tau} + \varepsilon_t, \\ t = p+1, p+2, \ldots,\end{gathered} \tag{2.7}$$

where $\nu = -\Pi\mu_0 + \Psi\mu_1$, $\phi = \beta'\mu_1$, $\theta = \beta'\delta$, $\gamma_0^* = \delta$ and $\gamma_j^* = -\Gamma_j \delta$ for $j = 1, \ldots, p-1$. The quantity $\Delta d_{t-j,\tau}$ 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, Johansen type cointegration tests can be performed in our model framework. In the next section we will discuss three different estimators of the break date and then we will discuss cointegration tests based on a model with estimated break date in Section 4.

3 Shift Date Estimation

In the following we consider three different estimators of the shift date τ. The first one is based on estimating an unrestricted VAR model in which the cointegrating rank as well as the restrictions for the parameters related to the impulse dummies are not taken into account. The latter restrictions are accounted for by the second estimator, whereas the third estimator ignores the impulse dummies altogether. For all procedures we assume that the VAR order p is given or has been chosen by some statistical procedure in a previous step. For the time being it is assumed to be known.

3.1 Estimator Based on Unrestricted Model

Our first estimator of τ is based on the least restricted model with respect to the cointegrating rank or

$$\Delta y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \quad t = p+1, \dots, T, \tag{3.1}$$

which is obtained from (2.7) by imposing no rank restriction on Π and rearranging terms. Here $\nu_0 = \nu + \Pi\mu_1$, $\nu_1 = -\Pi\mu_1$, $\delta_1 = -\Pi\delta$, $\gamma_0 = \delta - \delta_1$, $\gamma_j = \gamma_j^*$ $(j = 1, \dots, p-1)$ and T is the sample size. The shift date is estimated as

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

where the $\hat{\varepsilon}_{t\tau}$ are LS residuals from (3.1), and $\mathcal{T} \subset \{1, \dots, T\}$ is the set of all shift dates considered. Notice that $\mathcal{T}$ cannot include all sample periods if Assumption 1 is made. Moreover, there may be nonsample information regarding the possible shift dates which makes it desirable to limit the search to a specific part of the sample period.

We will present asymptotic properties of the shift date estimators under the assumption that the size of the shift depends on the sample size and may increase or decrease when the sample size gets larger. More precisely, we make the following assumption for the parameter δ.

Assumption 3. For some fixed $(n \times 1)$ vector δ_*,

$$\delta = \delta_T = T^a \delta_*, \quad a \leq 1/2. \tag{3.3}$$

□

Thus, we allow for a decreasing, constant or increasing shift size with growing sample size, depending on a being smaller, equal to or greater than zero, respectively. In most cases there will be no need to use the subscript T and so the notation δ will usually be used instead of δ_T. The same convention applies to parameters depending on δ (e.g., δ_1) and their estimators. As mentioned earlier, break date estimation when the shift size decreases with increasing sample size has also been discussed by other authors (Bai, Lumsdaine & Stock (1998)). An increasing shift size is treated here for completeness and it turns out that it provides interesting insights in the actual behaviour of our shift date estimators, as will be seen in the simulations in Section 5. Moreover, letting the shift size increase with the sample size may provide information on problems related to large shifts. In particular, it is of interest to check whether large shifts may affect the asymptotic distribution of the cointegrating rank tests discussed in Section 4. The upper bound $a = 1/2$ for the rate of increase of the shift size is chosen for technical reasons because we need this bound in our proofs. From a practical point of view such a bound should not be a problem because there may not be a need to estimate the shift date by formal statistical methods if the shift size is very large. We can now present asymptotic properties of our estimator $\hat{\tau}$ which generalize results presented in LST.

Theorem 3.1. Suppose Assumptions 1 - 3 hold.
(i) Let $0 \leq j_0 \leq p-1$ and suppose there exists an integer j_0 such that $\gamma_{j_0} \neq 0$ and, when $j_0 < p-1$, $\gamma_j = 0$ for $j = j_0+1, \dots, p-1$. Then, if $a > 0$ and $\delta_1 \neq 0$ or $a > 1/b$,

$$Pr\{\tau - p + 1 + j_0 \leq \hat{\tau} \leq \tau\} \to 1.$$

In particular, $\hat{\tau} \xrightarrow{p} \tau$ if $\gamma_{p-1} \neq 0$. If $\gamma_j = 0$ for all $j = 0, \dots, p-1$, the preceding convergence result holds with $j_0 = -1$.
(ii) If $a \leq 0$ and $\delta_1 \neq 0$, then

$$\hat{\tau} - \tau = O_p(T^{-2a/(1-2\eta)}),$$

where $\frac{1}{b} < \eta < \frac{1}{4}$. In particular, if $a > \eta - 1/2$,

$$\hat{\lambda} - \lambda = o_p(1),$$

where $\hat{\lambda} = \hat{\tau}/T$. □

For $\delta_1 \neq 0$ and $a = 0$, LST have shown that $\hat{\tau} - \tau = O_p(1)$ which is obviously a special case of our theorem. In fact, Theorem 3.1(i) shows that when the size of the break is sufficiently large, that is, $a > 1/b$ or $a > 0$ and $\delta_1 \neq 0$, the break date can be estimated accurately. More precisely, asymptotically the break date can then be located at the true break date or just a few time points before the true break date. Estimating the break date larger than the true one cannot occur in large samples. However, consistent estimation of the break date is not possible without an additional assumption for the parameters related to the impulse dummies in model (3.1). The required assumption $\gamma_{p-1} \neq 0$ can be seen as an identification condition for the break date. Indeed, if $\gamma_{p-1} = 0$ and $\gamma_{p-2} \neq 0$, Theorem 3.1(i) only tells us that asymptotically the break date estimator will take a value which is either the true break date or the preceding time point. The intuition for this is that one of the $p-1$ impulse dummies in (3.1) can be used to allow for such an incorrect estimation of the break date. In particular, even if we choose a break date one smaller than the true one we can still obtain a correct model specification with white noise errors. A similar situation occurs when more than one of the parameters γ_i at the largest lags are zero. Notice also that $\gamma_j = 0$ for all $j = 0, \ldots, p-1$ can only occur if $\delta_1 \neq 0$ because $\delta \neq 0$ and $\gamma_0 = \delta - \delta_1$.

The above discussion implies that an overspecification of the VAR order will always make the break date estimation $\hat{\tau}$ inconsistent. This observation explains some of the small sample results of LST. These authors fitted VAR(3) models to VAR(1) DGPs and found that $\hat{\tau}$ often underestimated the true break date. In principle the same phenomenon can occur also in other situations where $\gamma_{p-1} = 0$. However, since γ_0 is always nonzero when $\delta \neq 0$ (and $p \geq 1$) a reasoning similar to that used above explains why the break date will asymptotically not be estimated larger than the true one.

The second part of the Theorem 3.1 deals with the asymptotic behavior of the estimator $\hat{\tau}$ when the size of the break is "small". In this case we need to assume that $\delta_1 \neq 0$ or that there is actually a level shift in model (3.1) and not just some exceptional observations which can be handled with impulse dummies. This assumption is not needed in the first part of the theorem where the size of the break is "large" ($a > 1/b$) because then even the impulse dummies can be used to estimate the break date accurately. However, even though consistent estimation of the break date is not possible in the case of Theorem 3.1(ii), consistent estimation of the sample fraction λ is still possible provided the size of the break is not "too small". The result obtained in this context is weaker than its previous counterparts in Bai (1994) which, instead of $a > \eta - 1/2$, only require $a > -1/2$ (see, e.g., Proposition 3 of Bai (1994)). Complications caused by the presence of impulse dummies in model (3.1) are the reason for our weaker result. In any case, our assumption $a > \eta - 1/2$ is equivalent to $-2a/(1-2\eta) < 1$ which is clearly not very restrictive because $\hat{\tau} - \tau$ cannot be larger than T and is hence necessarily $O_p(T)$.

3.2 Constrained Estimation of τ

We shall now consider the constrained estimation of the break date in which the restrictions between the autoregressive parameters and coefficients related to the dummies are taken into account. Instead of (3.1) it is now convenient to start with the specification

$$\Delta y_t = \nu_0 + \nu_1 t + \delta_1 d_{t-1,\tau} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j,\tau} + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \quad t = p+1, \dots, T, \tag{3.4}$$

where $\delta_1 = -\Pi\delta$, as before, and the γ_j^* are as in (2.7). Thus, we can write (3.4) as

$$\Delta y_t = \nu_0 + \nu_1 t + \left(I_n \Delta d_{t,\tau} - \sum_{j=1}^{p-1} \Gamma_j \Delta d_{t-j,\tau} - \Pi d_{t-1,\tau} \right) \delta + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \; t = p+1, \dots, T. \tag{3.5}$$

Unlike the unrestricted model (3.1), the impulse dummies do not appear separately anymore in the representation (3.5) but are included in the term which also involves the shift dummy. Thus the restrictions imply that a single parameter vector δ is associated with all the dummy variables. A consequence is that the break date can be estimated more precisely, as we will see in the next theorem.

For any given value of the break date τ the parameters ν_0, ν_1, δ, Π and $\Gamma_1, \dots, \Gamma_{p-1}$ can be estimated from (3.5) by nonlinear LS. The estimator of the break date is then obtained by minimizing an analog of (3.2) with $\hat{\varepsilon}_{t\tau}$ replaced by residuals from this nonlinear LS estimation. The following theorem presents asymptotic properties of this break date estimator denoted by $\hat{\tau}_R$.

Theorem 3.2. Let Assumptions 1 - 3 hold and suppose that $\delta \neq 0$.
(i) If $a > 0$ and $\delta_1 \neq 0$ or $a > 1/b$, then

$$\hat{\tau}_R - \tau = o_p(1)$$

(ii) If $a \leq 0$ and $\delta_1 \neq 0$, then

$$\hat{\tau}_R - \tau = O_p(T^{-2a/(1-2\eta)}),$$

where $\frac{1}{b} < \eta < \frac{1}{4}$. □

The first part of the theorem shows that taking the restrictions into account is beneficial. Unlike in Theorem 3.1(i) consistency now obtains without any additional assumptions about coefficients. The second part of the theorem, which deals with the case of a "small" break size, is similar to its previous counterpart, however. The constrained estimator $\hat{\tau}_R$ is computationally more demanding than its unconstrained counterpart $\hat{\tau}$. One way to reduce the amount of computation is to apply both estimates. First, one can use $\hat{\tau}$ to locate the potential break date roughly and then apply $\hat{\tau}_R$ for restricted values of τ based on the value of $\hat{\tau}$ (e.g., $\hat{\tau}_R$ can be computed such that $|\hat{\tau} - \hat{\tau}_R| \leq p$ or $|\hat{\tau} - \hat{\tau}_R| \leq 2p$). Another possibility to reduce the amount of computation is to use a two-step estimator. Specifically, one can first estimate the parameters Π and $\Gamma_1, \ldots, \Gamma_{p-1}$ from (3.4) without constraints and use the resulting estimators to replace their theoretical counterparts in the term in parentheses in (3.5). Then LS can be applied to the resulting version of (3.5). These possibilities will be explored further in the Monte Carlo study in Section 5.

3.3 Ignoring Dummies in Estimating τ

Our third break date estimator was also considered by LST. Because the impulse dummies in (3.1) eliminate the observations where they assume a value of one, they may make it more difficult to locate the true break date. Therefore, LST consider estimating the break date from a VAR model without impulse dummy variables,

$$\Delta y_t = \nu_0 + \nu_1 t + \delta_1 d_{t\tau} + \Pi y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t^*, \quad t = p+1, \ldots, T, \tag{3.6}$$

where $\varepsilon_t^* = \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j,\tau} + \varepsilon_t$. We act as if ε_t^* had the same white noise properties as ε_t although this is not quite the case. Thus the estimator of the break date considered in this context is defined as

$$\tilde{\tau} = \arg\min_{\tau \in \mathcal{T}} \det \left(\sum_{t=p+1}^{T} \hat{\varepsilon}_{t\tau}^* \hat{\varepsilon}_{t\tau}^{*\prime} \right), \tag{3.7}$$

where the $\hat{\varepsilon}_{t\tau}^*$ are the LS residuals from (3.6). The following theorem again generalizes a result of LST by allowing the shift size to depend on the sample size. It shows that the estimator $\tilde{\tau}$ works well, provided that $\delta_1 \neq 0$.

Theorem 3.3. Suppose Assumptions 1 - 3 hold and $\delta_1 \neq 0$.
(i) If $a > 0$, then

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

(ii) If $a \leq 0$, then

$$\tilde{\tau} - \tau = O_p(T^{-2a/(1-2\eta)}),$$

where $\eta > 0$. □

Again this result is an obvious generalization of one obtained by LST who show that for $\delta_1 \neq 0$ and $a = 0$, $\tilde{\tau} - \tau = O_p(1)$. Although $\tilde{\tau}$ is based on a misspecified model, its convergence rate is equally good as that of the other two estimators, provided $\delta_1 \neq 0$ and $a \leq 0$. Clearly, $\delta_1 = -\alpha\beta'\delta = 0$ may hold even if $\delta \neq 0$. In fact, $\delta_1 = 0$ always holds if the cointegrating rank is zero. If $\delta_1 = 0$, there is co-breaking and the process $\beta' y_t$ has no break. For such processes, $\tilde{\tau}$ can find the shift date only by chance, whereas $\hat{\tau}$ and $\hat{\tau}_R$ can still find the true break date with some likelihood in large samples, if the shift size is large. Thus, using only the estimator $\tilde{\tau}$ may be problematic, unless the case $\delta_1 = 0$ can be ruled out.

As a final remark on our three break date estimators we mention that, if the DGP is known to have no deterministic linear trend ($\tau_1 = 0$), the corresponding terms in (3.1), (3.4) and (3.6) may be dropped without changing the convergence rates of our break date estimators. In the next section we consider the consequences of using a model with estimated break date for testing the cointegrating rank of a system of time series variables.

4 Testing the Cointegrating Rank

For given VAR order p and some estimator of the shift date, the cointegrating rank of the DGP can be tested as discussed by LST. They propose to use the tests suggested by Saikkonen & Lütkepohl (2000). In their 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 are determined by RR regression applied to (2.7). Using these estimators, LST apply a feasible GLS procedure to (2.1) to estimate all the parameters of the deterministic part. The observations are then adjusted for deterministic terms and cointegration tests are based on the adjusted series. Because the levels parameter μ_0 is not identified in the direction of $\beta_\perp$ one may wish to avoid its estimation. Therefore, in the following we shall consider an approach in which only the parameters μ_1 and δ in the deterministic part are estimated. The effect of the level parameter will be taken into account when the test is performed. The estimators of the parameters μ_1 and δ and their asymptotic properties will be discussed first and then the cointegration tests will be presented.

4.1 Estimating the Parameters of the Deterministic Part

We present the estimation procedure of the parameters μ_1 and δ for a given VAR order p, cointegration rank r and break date τ. First consider the estimation of the parameter μ_1. Recall the identity $\nu = -\Pi\mu_0 + \Psi\mu_1$ which can be written as

$$\nu = -\Pi\mu_0 + \Psi\beta(\beta'\beta)^{-1}\beta'\mu_1 + \Psi\beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'\mu_1$$

or, more briefly,

$$\nu = -\Pi\mu_0 + \Psi_\beta\phi + \Psi_{\beta_\perp}\phi_*,$$

where $\phi = \beta'\mu_1$, $\phi_* = \beta_\perp'\mu_1$, $\Psi_\beta = \Psi\beta(\beta'\beta)^{-1}$ and $\Psi_{\beta_\perp} = \Psi\beta_\perp(\beta_\perp'\beta_\perp)^{-1}$. Because $\alpha_\perp'\Pi = \alpha_\perp'\alpha\beta' = 0$, a multiplication of this identity from the left by $\alpha_\perp'$ yields $\alpha_\perp'(\nu - \Psi_\beta\phi) = \alpha_\perp'\Psi_{\beta_\perp}\phi_*$. The matrix $\alpha_\perp'\Psi_{\beta_\perp}$ is nonsingular and its inverse is $(\alpha_\perp'\Psi_{\beta_\perp})^{-1} = \beta_\perp'\beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}$. Thus, we can solve for ϕ_* as follows:

$$\phi_* = \beta_\perp' C(\nu - \Psi_\beta\phi),$$

where $C = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp'$ as before. Thus, if $\tilde{C}$ and $\tilde{\Psi}_\beta$ are sample analogs of C and Ψ_β, respectively, based on the RR estimation of (2.7), an estimator of ϕ_* is given by

$$\tilde{\phi}_* = \tilde{\beta}_\perp' \tilde{C}(\tilde{\nu} - \tilde{\Psi}_\beta \tilde{\phi}).$$

Here $\tilde{\nu}$, $\tilde{\phi}$ and $\tilde{\beta}_\perp$ are also based on the RR estimation of (2.7). Using the estimators $\tilde{\phi}$ and $\tilde{\phi}_*$ together we can form an estimator for μ_1 as

$$\tilde{\mu}_1 = \tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\phi} + \tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\phi}_*.$$

The parameter δ can be estimated in a similar way. From the definitions we find that

$$\begin{bmatrix} \gamma_0^* \\ \gamma_1^* \\ \vdots \\ \gamma_{p-1}^* \end{bmatrix} = \begin{bmatrix} I_n \\ -\Gamma_1 \\ \vdots \\ -\Gamma_{p-1} \end{bmatrix} \delta.$$

Multiplying this equation from the left by the matrix $[\alpha_\perp' : \cdots : \alpha_\perp']$ yields

$$\alpha_\perp' \sum_{j=0}^{p-1} \gamma_j^* = \alpha_\perp'\Psi\delta = \alpha_\perp'\Psi_\beta\theta + \alpha_\perp'\Psi_{\beta_\perp}\theta_*,$$

where $\theta_* = \beta_\perp'\delta$ and $\theta = \beta'\delta$ as in (2.7). From the foregoing equation we can solve for θ_* in the same way as for ϕ_* above. The result is

$$\theta_* = \beta_\perp' C \left(\sum_{j=0}^{p-1} \gamma_j^* - \Psi_\beta\theta \right)$$

from which we form an estimator for θ_* as

$$\tilde{\theta}_* = \tilde{\beta}_\perp' \tilde{C} \left(\sum_{j=0}^{p-1} \tilde{\gamma}_j^* - \tilde{\Psi}_\beta\tilde{\theta} \right).$$

Here $\tilde{\gamma}_j^*$ and $\tilde{\theta}$ are again based on the RR estimation of (2.7). Thus, an estimator of δ is obtained as

$$\tilde{\delta} = \tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\theta} + \tilde{\beta}_\perp(\tilde{\beta}_\perp'\tilde{\beta}_\perp)^{-1}\tilde{\theta}_*.$$

We shall now consider asymptotic properties of the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ by assuming that the break date τ in (2.7) is replaced by any one of the estimators $\hat{\tau}$, $\hat{\tau}_R$ or $\tilde{\tau}$ introduced in Section 3. For simplicity, other estimators based on the VECM (2.7) with τ replaced by an estimator will be denoted as before (e.g., $\tilde{\beta}$, $\tilde{\alpha}, \dots$), that is, the estimated τ will not be explicitly indicated. The following result is shown in the Appendix.

Lemma 4.1. Consider the four cases **(i)** $\delta_1 \neq 0$, $a \leq 1/2$, **(ii)** $\delta_1 = 0$, $a \leq 1/b$, **(iii)** $\delta = 0$ and **(iv)** $\delta_1 = 0$, $a > 1/b$. If **(i)**, **(ii)** or **(iii)** holds and the break date is estimated by $\hat{\tau}$, $\hat{\tau}_R$ or $\tilde{\tau}$ or if **(iv)** holds and τ is estimated by $\hat{\tau}$ or $\hat{\tau}_R$, then the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ have the following properties.

$$\beta'(\tilde{\mu}_1 - \mu_1) = O_p(T^{\eta-3/2}) \tag{4.1}$$

$$T^{1/2}\beta_\perp'(\tilde{\mu}_1 - \mu_1) \xrightarrow{d} N(0, \beta_\perp' C\Omega C'\beta_\perp) \tag{4.2}$$

$$\beta'(\tilde{\delta} - \delta) = O_p(T^{\eta-1/2}) \tag{4.3}$$

$$\beta_\perp'(\tilde{\delta} - \delta) = o_p(T^{\eta}) \tag{4.4}$$

where $\frac{1}{b} < \eta < \frac{1}{4}$. □

The result of Lemma 4.1 is not the best possible in that improvements in the convergence rates in (4.1), (4.3) and (4.4) can be obtained in some of the cases **(i)**-**(iv)**. For ease of exposition and because the given results suffice for our purposes we have preferred not to go into details in this matter. Note that if $\hat{\tau}$ or $\hat{\tau}_R$ are used, the results of Lemma 4.1 hold for all δ and a permitted by Assumption 3. On the other hand, for $\tilde{\tau}$ we exclude the case where $\delta_1 = 0$, $\delta \neq 0$ and $a > 1/b$ because the case $\delta_1 = 0$ is not considered in Theorem 3.3 and for this case we do not have a proof of the properties of the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ stated in Lemma 4.1. It may be worth emphasizing, however, that if $\delta = 0$ so that there is no break (and, hence, $\delta_1 = 0$) the results of the lemma hold also for $\tilde{\tau}$ because the lemma is valid for all three break date estimators if **(iii)** holds. In other words, the results not only hold under the assumptions of Theorems 3.1, 3.2 and 3.3 but also when $\delta = 0$ so that there is no break. In the following we will use the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ in constructing cointegrating rank tests.

4.2 Cointegration Tests

We wish to construct a test of the null hypothesis

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

The test will be based on series which are adjusted for the deterministic trend and the shift term.

Recall that $y_t = \mu_0 + \mu_1 t + \delta d_{t\tau} + x_t$, where x_t has a (possibly) cointegrated VAR(p) representation (see (2.4)-(2.5)). Because estimators for τ, μ_1 and δ are available for the cointegrating rank r_0 specified in the null hypothesis, we can form the series

$$\begin{aligned} \tilde{y}_t^{(0)} &= y_t - \tilde{\mu}_1 t - \tilde{\delta} d_{t\hat{\tau}} \\ &= \mu_0 + x_t - (\tilde{\mu}_1 - \mu_1)t - \tilde{\delta} d_{t\hat{\tau}} + \delta d_{t\tau}. \end{aligned} \tag{4.5}$$

Thus, apart from estimation errors we have $\tilde{y}_t^{(0)} \sim \mu_0 + x_t$. This suggests that we can base a test on this approximation or on the auxiliary model

$$\Delta \tilde{y}_t^{(0)} = \Pi^+ \tilde{y}_{t-1}^{(+)} + \sum_{j=1}^{p-1} \Gamma_j \Delta \tilde{y}_{t-j}^{(0)} + e_{t\hat{\tau}}, \tag{4.6}$$

where $\tilde{y}_{t-1}^{(+)} = [\tilde{y}_{t-1}^{(0)\prime}, 1]'$ and Π^+ is defined by adding an extra column to the matrix Π in (2.5). This auxiliary model can be treated as a true model and a LR test statistic for a specified cointegrating rank can be formed in the usual way by solving the related generalized eigenvalue problem (see Johansen (1995, Theorem 6.3) for the resulting test statistic). We will denote the LR statistic for the null hypothesis $\mathrm{rk}(\Pi) = r_0$ by $LR(r_0)$ in the following. Its limiting distribution differs from that given in Theorem 6.3 of Johansen (1995) for the corresponding LR test statistic. We have the following result which is also proven in the Appendix.

Theorem 4.1. Suppose that the assumptions of Lemma 4.1 hold and moreover that in the case $a = 1/2$ the employed break date estimator is consistent. Then, if $H_0(r_0)$ is true,

$$LR(r_0) \xrightarrow{d} \mathrm{tr}\left\{ \left(\int_0^1 B_+(s) dB_*(s)' \right)' \left(\int_0^1 B_+(s) B_+(s)' ds \right)^{-1} \left(\int_0^1 B_+(s) dB_*(s)' \right) \right\},$$

where $B_*(s) = B(s) - sB(1)$ is an $(n - r_0)$-dimensional Brownian bridge, $B_+(s) = [B_*(s)', 1]'$ and $dB_*(s) = dB(s) - dsB(1)$, that is, $\int_0^1 B_+(s) dB_*(s)'$ abbreviates $\int_0^1 B_+(s) dB(s)' - \int_0^1 B_+(s) ds B(1)'$. □

Several remarks are worth making regarding this theorem. First, a similar result for their break date estimator and cointegrating rank test was obtained by LST under more restrictive assumptions regarding the break size. The limiting distribution in Theorem 4.1

differs from its earlier counterpart in LST in that the process $B_+(s)$ appears in place of the Brownian bridge $B_*(s)$. The reason is of course that here an intercept term is included in the auxiliary model on which the test is based. On the other hand, the limiting distribution is formally similar to its counterpart in Theorem 6.3 of Johansen (1995) where a standard Brownian motion appears in place of the Brownian bridge in Theorem 4.1. Notice that the term $\int_0^1 B_+(s)dB_*(s)'$ consists of two components. The first one is

$$\int_0^1 B_*(s)dB_*(s)' = \int_0^1 B(s)dB(s)' - B(1)\int_0^1 sdB(s)' - \int_0^1 B(s)dsB(1)' + \frac{1}{2}B(1)B(1)'$$

and the second one is

$$\int_0^1 1dB_*(s)' = \int_0^1 dB(s)' - \int_0^1 dsB(1)' = 0.$$

Second, in the case when there is no trend in the model, that is, $\mu_1 = 0$ a priori and hence $\tilde{\mu}_1 = 0$, the processes $B_+(s)$ and $B_*(s)$ can be replaced by $[B(s)', 1]'$ and $B(s)$, respectively. Then the limiting distribution of the test statistic $LR(r_0)$ is the same as the limiting distribution of the LR test statistic in Theorem 6.3 of Johansen (1995). This result can be seen by a careful examination of the proof of Theorem 4.1 in the Appendix and the proof of the corresponding result in Saikkonen & Lütkepohl (2000).

Third, note that, although Theorem 4.1 applies to the break date estimator $\hat{\tau}_R$ irrespectively of the value a in Assumption 3 and whether δ_1 or δ is nonzero or not, this is not the case for the estimators $\hat{\tau}$ and $\tilde{\tau}$. The reason is that by Theorem 3.1 $\hat{\tau}$ is not necessarily consistent for $a = 1/2$. Moreover, the consistency of $\tilde{\tau}$ does not follow from Theorem 3.3 if $\delta_1 = 0$. In the latter theorem we only give asymptotic results for the case $\delta_1 \neq 0$ and, therefore, Theorem 4.1 is not justified when $\tilde{\tau}$ is used and $\delta_1 = 0$ while $\delta \neq 0$.

Fourth, from the proof of Theorem 4.1 it is apparent that the same limiting distribution is obtained if the shift date is assumed known or if it is known that there is no shift in the process. In the latter case $\delta = 0$ and only μ_1 is estimated in the first step of the procedure. Thus, in our framework, including a shift dummy in the model and estimating its coefficients and the shift date as described in the foregoing has no effect on the limiting distribution of the cointegration tests. The same result was obtained by LST for their cointegrating rank tests in a more limited model framework. It may be worth emphasizing that such a result will not be obtained if instead of our estimation procedure for the deterministic parameters, the Johansen (1995) ML approach is applied to a model with estimated shift date (see also

Johansen, Mosconi & Nielsen (2000) for a discussion of the case when the break date is known).

Extensions of our results in different directions are conceivable. If the cointegrating rank tests considered by LST are used instead of our new ones, we expect that the same results as in LST can be obtained for the limiting distributions under the present or similar assumptions for the break size. Moreover, it seems likely that our results can be extended 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 Saikkonen & Lütkepohl (2000). The result in Theorem 4.1 remains valid with additional dummies if the corresponding shift dates are known and the parameters of the additional deterministic terms are estimated in a similar way as μ_1 or δ. If the dates of further shifts are unknown, it may be more difficult to construct suitable shift date estimators. This issue may be an interesting project for future research.

To apply the cointegration rank tests we need critical values for the limiting distributions in Theorem 4.1. They will be presented next.

4.3 Simulation of Percentiles

The limiting distribution given in Theorem 4.1 is simulated numerically by approximating the standard Brownian motions with T-step random walks of the same dimension $n - r_0$. We use $T = 1000$ as the sample length. Then, discrete counterparts for the Brownian bridge $\mathbf{B}_*(s)$ and the functions of $\mathbf{B}_*(s)$ can be formed using the random walks. The percentiles in Table 1 are derived from 100,000 replications of the simulation experiment by means of a program written in GAUSS V5.

The generated random walks are based on independent standard normal variates which have been derived from the Monster-KISS random number generator. The Monster-KISS algorithm was suggested by Marsaglia (2000) and is implemented in GAUSS V5. We use independent realizations of the standard normal variates for each dimension $n - r_0$.

In case of $\mu_1 = 0$ the limiting distribution of $LR(r_0)$ is the same as the distribution of the LR test statistic in Theorem 6.3 of Johansen (1995). Therefore, one can apply the corresponding percentiles given in Table 15.2 in Johansen (1995) for the setup of no linear trend.

Table 1: Percentiles of Limiting Distribution of $LR(r_0)$

$n - r_0$	50%	75%	80%	85%	90%	95%	97.5%	99%
1	3.578	5.356	5.893	6.576	7.509	9.046	10.589	12.645
2	11.694	14.658	15.498	16.508	17.855	20.010	22.073	24.623
3	23.712	27.857	28.972	30.316	32.125	34.897	37.431	40.447
4	39.569	44.895	46.320	47.955	50.121	53.612	56.690	60.570
5	59.341	65.776	67.457	69.473	72.080	76.015	79.667	84.117
6	83.090	90.760	92.704	95.025	98.069	102.705	106.916	112.106
7	110.856	119.613	121.884	124.552	128.014	133.253	137.840	143.404
8	142.276	152.287	154.833	157.881	161.719	167.556	172.820	179.112
9	177.780	188.799	191.638	194.971	199.236	205.784	211.621	218.775
10	217.039	229.419	232.616	236.300	241.029	248.043	254.424	262.249
11	260.208	273.643	277.038	281.156	286.353	294.106	300.790	309.092
12	307.017	321.719	325.492	329.900	335.460	343.999	351.124	359.944
13	358.218	373.905	377.893	382.515	388.495	397.416	405.240	414.683
14	412.647	429.672	433.969	438.982	445.361	454.694	462.861	472.893
15	471.304	489.298	493.765	499.239	506.088	516.412	525.570	536.449

In the next section we will discuss small sample properties of the break date estimators and cointegration tests.

5 Monte Carlo Simulations

A Monte Carlo experiment was performed to compare our break date estimators and to explore the finite sample properties of the corresponding test procedures. Furthermore, we compare the cointegration test suggested by LST with the test proposal in this paper. The simulations are based on the following x_t process from Toda (1994) which was also used by a number of other authors for investigating the properties of cointegrating rank tests (see, e.g., Hubrich, Lütkepohl & Saikkonen (2001)):

$$x_t = A_1 x_{t-1} + \varepsilon_t = \begin{bmatrix} \boldsymbol{\psi} & 0 \\ 0 & I_{n-r} \end{bmatrix} x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{ i.i.d. } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{n-r} \end{bmatrix} \right), \quad (5.1)$$

where $\boldsymbol{\psi} = \text{diag}(\psi_1, \ldots, \psi_r)$ and Θ are $(r \times r)$ and $(r \times (n-r))$ matrices, respectively. As shown by Toda, this type of process is useful for investigating the properties of LR tests for the cointegrating rank because other cointegrated VAR(1) processes of interest can be obtained from (5.1) by linear transformations which leave such tests invariant. Obviously, if $|\psi_i| < 1$ $(i = 1, \ldots, r)$ we have r stationary series and, thus, the cointegrating rank is equal to r. Hence, Θ describes the contemporaneous error term correlation between the stationary and nonstationary components. We have used three- and four-dimensional processes for simulations and report some of the results in more detail here. For given VAR order p and break date τ, the test results are invariant to the parameter values of the constant and trend because we allow for a linear trend in our tests. Therefore we use $\mu_i = 0$ $(i = 0, 1)$ as parameter values throughout without loss of generality. In other words, the intercept and trend terms are actually zero although we take such terms into account and thereby we pretend that this information is unknown to the analyst. Hence, $y_t = \delta d_{t\tau} + x_t$ and we have performed simulations with different δ vectors. Rewriting x_t in VECM form (2.5) shows that $\Pi = -(I_n - A_1) = \text{diag}(\boldsymbol{\psi} - I_r : 0)$ and, thus, $\delta_1 = -\Pi\delta$ can only be nonzero if level shifts occur in stationary components of the DGP.

Samples are simulated starting with initial values of zero. We have also used other initial conditions for some simulations and obtained qualitatively the same results. Because the tests used in LST may have a disadvantage for nonzero initial values and because we want to perform a comparison with these tests we present only results for zero initial conditions to simplify an overall comparison. We have considered a sample size of $T = 100$. 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., $s_{0.05} = 0.007$. Moreover, we use different VAR orders p, although the true order is $p = 1$, in order to explore the impact of this quantity on the estimation and testing results. In all simulations the search procedures are applied to all possible break points τ from the 5th up to and including the 96th observation.

In Section 3 the break date estimators $\hat{\tau}$, $\hat{\tau}_R$, and $\tilde{\tau}$ have been introduced. Only $\hat{\tau}_R$ takes account of the nonlinear restrictions between the parameters in (3.5). The estimator

$\hat{\tau}$ disregards these restrictions which involve the model parameters related to the dummy variables and the autoregressive terms, and $\tilde{\tau}$ ignores the dummy variables completely when estimating the shift date. To compute $\hat{\tau}_R$ we use a nonlinear LS estimation method by applying the Gauss-Newton algorithm in order to minimize the sum of squared residuals corresponding to (3.5). The iterations of the algorithm stop if the change in $D_{\hat{\tau}_R} = \det[(T - p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_{t\tau}^R \hat{\varepsilon}_{t\tau}^{R\prime}]$ from iteration i to $i+1$ is less than $(T-p)^{-n}$, where $\hat{\varepsilon}_{t\tau}^R$ $(t = p+1, \ldots, T)$ are the residual vectors from the nonlinear estimation of (3.5). Thus, the precision is about 10^{-6} for a three-dimensional process. In addition, the maximum number of iterations is set to 25. We have also worked with smaller values of our stopping criterion and higher maximum numbers of iterations for a subset of our simulation experiments but did not obtain different results. Of course, the application of the Gauss-Newton algorithm is computationally rather demanding. Therefore, as suggested earlier, we use $\hat{\tau}$ first in order to locate the shift date roughly. Then, we only apply $\hat{\tau}_R$ for restricted values of τ such that $|\hat{\tau}_R - \hat{\tau}| \leq 2p$. The resulting estimator will be abbreviated as $\hat{\tau}_{R,p}$. By restricting the range of possible break points the computation time is reduced to 15-25% of the time for the full range depending on the order and the dimension of the process. Even more time can be saved if the two-step estimator is used. Here, (3.4) is estimated first ignoring the nonlinear restrictions. Then, the estimators for Π and $\Gamma_1, \ldots, \Gamma_{p-1}$ are used to replace their theoretical counterparts in the expression given in parentheses in (3.5) and the resulting model corresponding to (3.5) is reestimated by LS. This procedure is repeated for the whole range of possible break points. The corresponding two-step break date estimator is denoted by $\hat{\tau}_R^{(2)}$. Within our simulation study we analyze the effects of the different ways of computing the constrained estimator.

The interpretation of the simulation results is done in three steps. First, we analyze the ability of the shift date estimators to locate the true break point. Secondly, we discuss the small sample properties of the corresponding cointegration tests based on these estimators. Finally, we compare the type of cointegration tests suggested by LST and the test proposed in this paper.

As a basis for the comparison of the shift date estimators we start with a three-dimensional DGP with $r = 1$ $(\psi_1 = 0.9)$, $\Theta = (0.4, 0.8)$ and $\tau = 50$. Afterwards, we comment on the importance of the value of τ and the innovation correlation. In a next step we turn to a four-dimensional DGP with two cointegration relations in order to study the properties

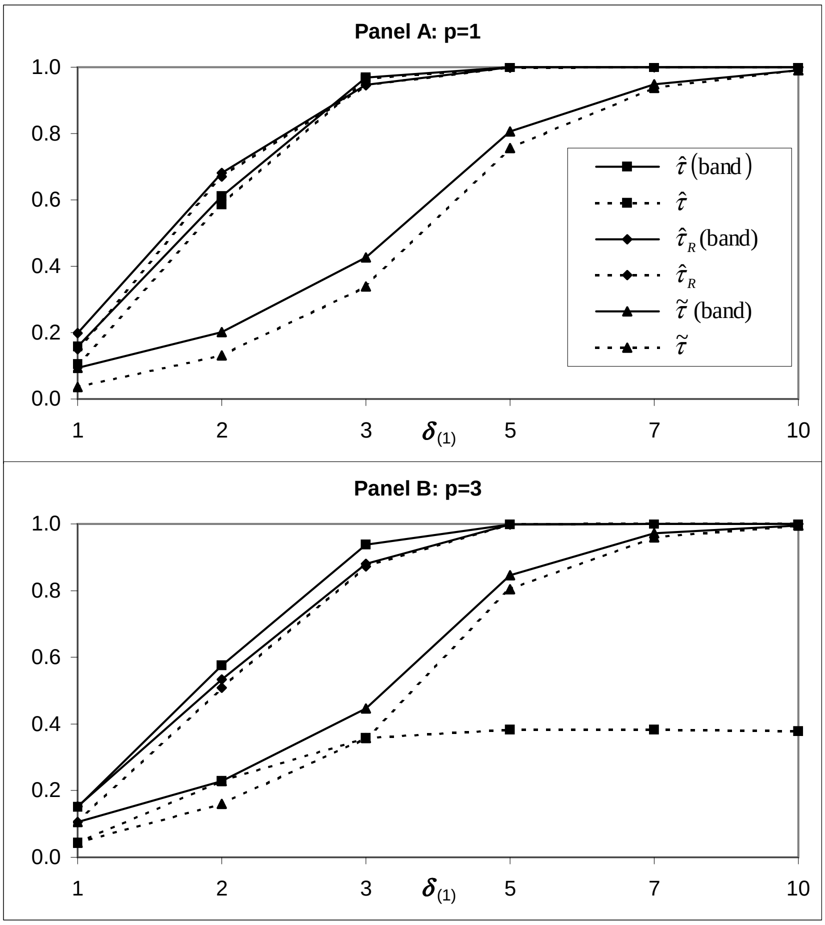

Figure 1: Relative frequency of true break point estimates ($\hat{\tau}$, $\hat{\tau}_R$, $\tilde{\tau}$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}$ (band), $\hat{\tau}_R$ (band), $\tilde{\tau}$ (band)) for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

of the procedures in case of more complicated processes. Finally, we examine situations where $\delta_1 = -\Pi\delta = 0$ and, hence, asymptotically consistent estimation of τ requires stronger conditions or is even not possible in case of $\tilde{\tau}$.

The break date estimates with respect to our three-dimensional basis DGP with $r = 1$ and a VAR order $p = 1$ are reported in Table 2 and Panel A of Figure 1. We consider a shift $\delta = (\delta_{(1)}, \delta_{(2)}, \delta_{(3)})'$ with $\delta_{(1)}$ ranging from 1 to 10 and $\delta_{(2)} = \delta_{(3)} = 0$. Hence, the shift occurs in the first component of the DGP which is stationary according to (5.1). Thus, as discussed above, we have $\delta_1 = \Pi\delta = \alpha\beta'\delta \neq 0$ in (3.1) and, hence, $\theta = \beta'\delta \neq 0$ in (2.7).

It can be seen that $\hat{\tau}_R$ is clearly most successful in finding the correct break date for small

Table 2: Break Date Estimates and Rejection Frequencies for Three-Dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $p = 1$, $\Theta = (0.4, 0.8)$, Sample Size $T = 100$, VAR order $p = 1$, True Break Point $\tau = 50$, Nominal Significance Level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

		$\delta_{(1)}=1$	$\delta_{(1)}=2$	$\delta_{(1)}=3$	$\delta_{(1)}=5$	$\delta_{(1)}=7$	$\delta_{(1)}=10$
		$\hat{\tau}$ (Ignoring Nonlinear Restrictions)					
Break date estimates	< 48	0.423	0.200	0.022	0.000	0.000	0.000
	$= 48/49$	0.038	0.023	0.005	0.000	0.000	0.000
	$= 50$	0.104	0.586	0.964	1.000	1.000	1.000
	$= 51/52$	0.016	0.002	0.000	0.000	0.000	0.000
	> 52	0.419	0.189	0.009	0.000	0.000	0.000
Rejection frequencies	$r_0 = 0$	0.631	0.620	0.624	0.630	0.630	0.630
	$r_0 = 1$	0.135	0.103	0.079	0.075	0.075	0.075
	$r_0 = 2$	0.017	0.017	0.009	0.009	0.009	0.009
		$\hat{\tau}_R$ (Considering Nonlinear Restrictions)					
Break date estimates	< 48	0.373	0.145	0.025	0.001	0.000	0.000
	$= 48/49$	0.028	0.006	0.001	0.000	0.000	0.000
	$= 50$	0.150	0.671	0.946	0.999	1.000	1.000
	$= 51/52$	0.020	0.005	0.000	0.000	0.000	0.000
	> 52	0.429	0.173	0.028	0.000	0.000	0.000
Rejection frequencies	$r_0 = 0$	0.699	0.649	0.633	0.630	0.630	0.630
	$r_0 = 1$	0.118	0.094	0.080	0.076	0.075	0.075
	$r_0 = 2$	0.015	0.015	0.009	0.009	0.009	0.009
		$\tilde{\tau}$ (Ignoring Impulse Dummies)					
Break date estimates	< 48	0.459	0.399	0.296	0.112	0.028	0.007
	$= 48/49$	0.039	0.063	0.082	0.049	0.012	0.000
	$= 50$	0.036	0.131	0.339	0.757	0.937	0.991
	$= 51/52$	0.018	0.008	0.006	0.001	0.000	0.000
	> 52	0.448	0.399	0.277	0.081	0.023	0.002
Rejection frequencies	$r_0 = 0$	0.643	0.614	0.547	0.558	0.602	0.625
	$r_0 = 1$	0.126	0.122	0.106	0.080	0.076	0.075
	$r_0 = 2$	0.024	0.018	0.019	0.012	0.009	0.009

shift magnitudes. Only if $\delta_{(1)} = 3$, $\hat{\tau}$ performs slightly better. For large values of $\delta_{(1)}$ both estimators perform identically. In fact, the case $\delta_{(1)} = 3$ is one of the few exceptions in all our simulation experiments where $\hat{\tau}$ outperforms $\hat{\tau}_R$. Clearly, $\tilde{\tau}$ is inferior compared to the other break date estimators. These observations also hold if one considers the small band $[\tau - 2; \tau + 2]$ instead of τ to evaluate the break date estimator. The number of estimates in this band which are different from τ is rather small. Only with respect to the estimator

Table 3: Break Date Estimates and Rejection Frequencies for Three-Dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $p = 3$, $\Theta = (0.4, 0.8)$, Sample Size $T = 100$, VAR order $p = 3$, True Break Point $\tau = 50$, Nominal Significance Level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

		$\delta_{(1)}=1$	$\delta_{(1)}=2$	$\delta_{(1)}=3$	$\delta_{(1)}=5$	$\delta_{(1)}=7$	$\delta_{(1)}=10$
		$\hat{\tau}$ (Ignoring Nonlinear Restrictions)					
Break date estimates	< 48	0.412	0.215	0.039	0.001	0.000	0.000
	$= 48/49$	0.091	0.343	0.580	0.616	0.617	0.622
	$= 50$	0.044	0.229	0.358	0.383	0.383	0.378
	$= 51/52$	0.016	0.004	0.000	0.000	0.000	0.000
	> 52	0.437	0.209	0.023	0.000	0.000	0.000
Rejection frequencies	$r_0 = 0$	0.482	0.463	0.436	0.411	0.396	0.382
	$r_0 = 1$	0.109	0.109	0.092	0.077	0.072	0.075
	$r_0 = 2$	0.017	0.018	0.017	0.015	0.017	0.016
		$\hat{\tau}_R$ (Considering Nonlinear Restrictions)					
Break date estimates	< 48	0.406	0.234	0.057	0.001	0.000	0.000
	$= 48/49$	0.024	0.006	0.001	0.000	0.000	0.000
	$= 50$	0.105	0.509	0.873	0.998	1.000	1.000
	$= 51/52$	0.023	0.019	0.007	0.000	0.000	0.000
	> 52	0.442	0.232	0.062	0.001	0.000	0.000
Rejection frequencies	$r_0 = 0$	0.533	0.496	0.440	0.423	0.422	0.422
	$r_0 = 1$	0.114	0.109	0.093	0.086	0.085	0.085
	$r_0 = 2$	0.015	0.013	0.010	0.011	0.011	0.011
		$\tilde{\tau}$ (Ignoring Impulse Dummies)					
Break date estimates	< 48	0.440	0.369	0.268	0.084	0.014	0.003
	$= 48/49$	0.042	0.067	0.085	0.040	0.011	0.001
	$= 50$	0.044	0.160	0.358	0.804	0.960	0.995
	$= 51/52$	0.019	0.011	0.004	0.002	0.001	0.000
	> 52	0.455	0.393	0.285	0.070	0.014	0.001
Rejection frequencies	$r_0 = 0$	0.531	0.502	0.466	0.429	0.428	0.422
	$r_0 = 1$	0.129	0.113	0.108	0.088	0.083	0.085
	$r_0 = 2$	0.023	0.020	0.013	0.013	0.012	0.011

$\tilde{\tau}$ these estimates have some relevance for small values of $\delta_{(1)}$ (see Panel A of Figure 1). Obviously, the frequency of finding τ increases for larger shift magnitudes. This result is not surprising given the asymptotic properties of the estimators and the fact that $\delta_1 \neq 0$ in the present situation. Because T is fixed, changing $\delta_{(1)}$ from one to ten may be interpreted as changing a or δ_* in (3.3) accordingly.

Next, we have fitted a VAR(3) model although the true DGP has only an order $p = 1$.

In this case $j_0 = 0$ in Theorem 3.1(i) because $\gamma_1 = \gamma_2 = 0$. Thus, we obtain from Theorem 3.1(i), $Pr\{48 \leq \hat{\tau} \leq 50\} \to 1$ in the present context. In line with this result it is clear that $\hat{\tau}$ does not necessarily find $\tau = 50$ with probability 1 for increasing $\delta_{(1)}$'s. In fact, we observe in Table 3 and Panel B of Figure 1 that in about two thirds of the replications the break date is located too early. However, the estimates converge to the stated range for $\hat{\tau}$, in line with our asymptotic results. Interestingly, with respect to the band $[\tau - 2; \tau + 2]$, $\hat{\tau}$ is slightly more successful than $\hat{\tau}_R$ for values of $\delta_{(1)}$ between two and five. Otherwise, the general outcomes regarding $\hat{\tau}_R$ and $\tilde{\tau}$ do not change. The frequency of detecting $\tau = 50$ reduces somewhat for $\hat{\tau}_R$, but increases slightly for $\tilde{\tau}$.

So far we have only considered the constrained estimator $\hat{\tau}_R$ based on the Gauss-Newton algorithm. Figure 2 presents the results also for $\hat{\tau}_{R,p}$ and $\hat{\tau}_R^{(2)}$. It can be seen that $\hat{\tau}_{R,p}$ is always outperformed by at least one of the other constrained estimators in the sense that it never locates the true break date more often than both $\hat{\tau}_R$ and $\hat{\tau}_R^{(2)}$. The situations of $p = 3$ with $\delta_{(1)} = 3$ and $\delta_{(1)} = 5$ belong to the rare cases where we observe that $\hat{\tau}_{R,p}$ is more successful than one of its constrained competitors. The findings do not change if we evaluate the estimators' ability to locate the break point within the band $[\tau - 2, \tau + 2]$. Hence, $\hat{\tau}_R$ and $\hat{\tau}_R^{(2)}$ are in general superior although their advantage is often not very strong, as seen in Figure 2. Nevertheless, it does not pay to use $\hat{\tau}$ first and apply constrained estimation only to observations around the pre-estimated date. Therefore, we recommend to use either $\hat{\tau}_R$ or $\hat{\tau}_R^{(2)}$ if one wants to apply a constrained estimator.

These two estimators perform rather similarly for $p = 1$. In case of $p = 3$, $\hat{\tau}_R^{(2)}$ is in fact superior to $\hat{\tau}_R$ for values of $\delta_{(1)}$ from one to five. We find a similar effect regarding the VAR order p also for other processes. A reason for this finding could be that fitting a VAR(3) model increases the number of parameters importantly compared to a VAR(1) model. Within a three-dimensional framework 18 additional parameters have to be estimated. This larger number may make it more difficult for the Gauss-Newton algorithm to find the global minimum of $D_{\hat{\tau}_R}$ when estimating (3.5). Indeed, we have examined some of the simulation repetitions in more detail and could observe that the algorithm can get stuck in a local minimum in situations where $\hat{\tau}_R^{(2)}$ finds $\tau = 50$ but $\hat{\tau}_R$ does not.

The relative outcomes for $\hat{\tau}_R$, $\hat{\tau}_{R,p}$, and $\hat{\tau}_R^{(2)}$ also hold for the other DGPs considered. Accordingly, we do not present detailed results for all constrained estimators in the following

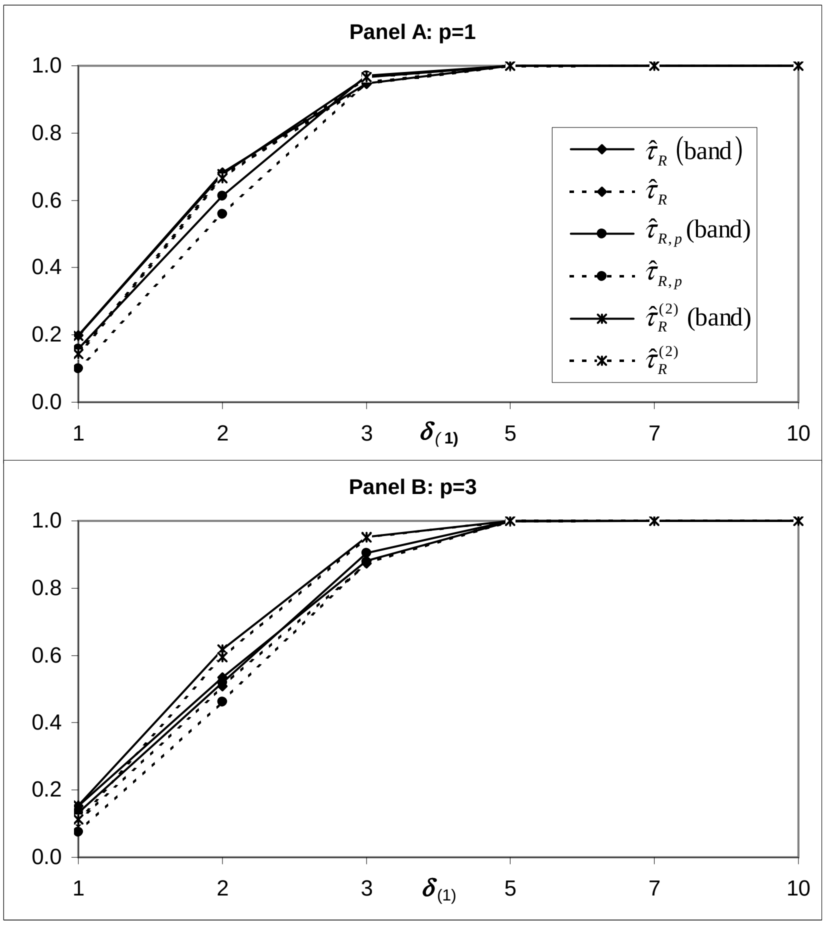

Figure 2: Relative frequency of true break point estimates ($\hat{\tau}_R$, $\hat{\tau}_{R,p}$, $\hat{\tau}_R^{(2)}$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}_R$ (band), $\hat{\tau}_{R,p}$ (band), $\tilde{\tau}_R^{(2)}$ (band)) of constrained estimators for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

but only focus on $\hat{\tau}_R^{(2)}$.

Previously, we have only applied a true break point $\tau = 50$. To analyze possible effects of the location of τ we have also studied the break points $\tau = 10$, $\tau = 25$, $\tau = 75$, and $\tau = 90$ using the same three-dimensional DGP as before. In Figure 3 we present some findings for $\delta_{(1)} = 2$ (Panels A and B) and $\delta_{(1)} = 7$ (Panels C and D) representing small and large shift magnitudes. With respect to $\delta_{(1)} = 2$ we observe that it is only slightly more difficult to detect the more extreme break points. In some situations it seems to be even easier for the estimators to find the true break date. In case of large shift magnitudes ($\delta_{(1)} = 7$) the location of the break date becomes even less important for the estimation results. These

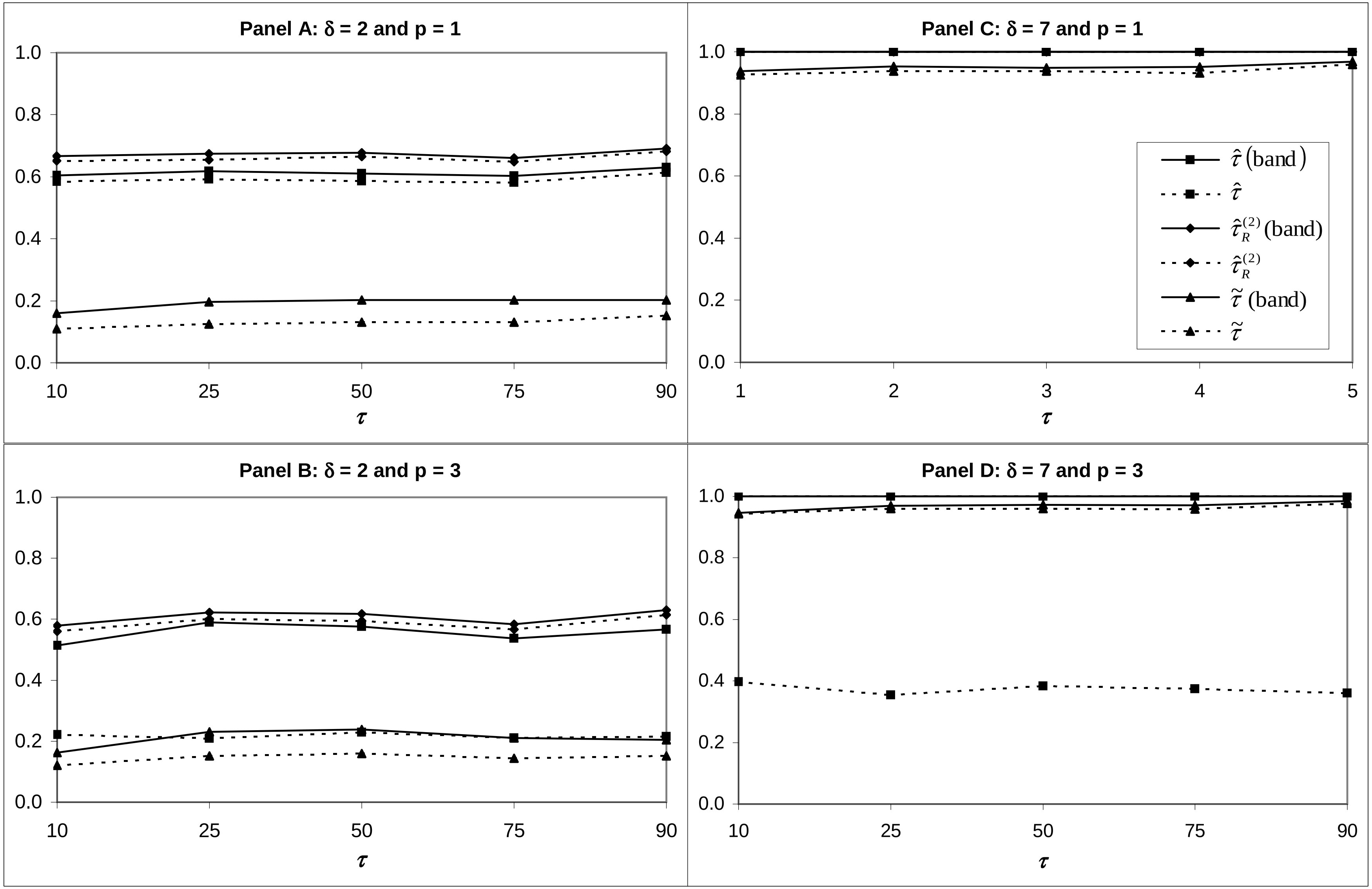

Figure 3: Impact of location of the the break point on break date estimation for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4\ , 0.8)$, sample size $T = 100$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

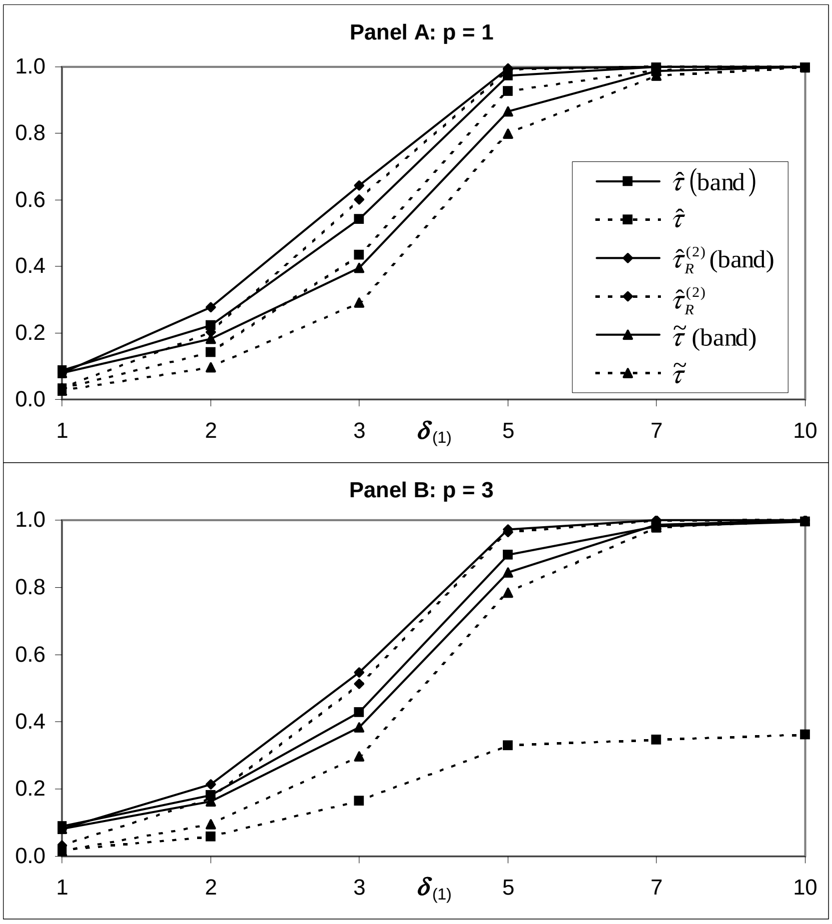

Figure 4: Relative frequency of true break point estimates ($\hat{\tau}$, $\hat{\tau}_R^{(2)}$, $\tilde{\tau}$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}$ (band), $\hat{\tau}_R^{(2)}$ (band), $\tilde{\tau}$ (band)) for four-dimensional DGP with $r = 2$ ($\psi_1 = \psi_2 = 0.7$), $\Theta = ([0.4 : 0.4]' : [0.4 : 0.4]')$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = \delta_{(4)} = 0$.

observations are made for both fitted VAR orders of $p = 1$ and $p = 3$ and also hold in case of the constrained estimators $\hat{\tau}_R$ and $\hat{\tau}_{R,p}$ for which the results are not given here.

Next, we have studied the effect of the error term correlation between the stationary and nonstationary components by considering a three-dimensional DGP as before but with $\Theta = (0, 0)$ and comparing the outcomes with the previous findings. We do not present detailed results but just summarize them. The absence of instantaneous error term correlation makes it more difficult for all estimators to locate the true break point no matter whether the order $p = 1$ or $p = 3$ is used. This outcome can be explained by the fact that we consider a shift only in one of the three components so that a weaker link of the components owing to $\Theta = (0, 0)$ complicates the break date search. However, $\hat{\tau}_R$ and $\hat{\tau}_R^{(2)}$ are now always the most successful procedures and usually their advantage is even more pronounced than in case of $\Theta = (0.4, 0.8)$. Otherwise the relative performance of the estimators is as before. The estimators $\tilde{\tau}$ and $\hat{\tau}$ are outperformed by the other procedures if $p = 1$ and $p = 3$, respectively.

The break date estimates with respect to the more complicated four-dimensional DGP with a cointegrating rank $r = 2$ ($\psi_1 = \psi_2 = 0.7$) and $\Theta = ([0.4 : 0.4]' : [0.4 : 0.4]')$ are reported in Figure 4. In the present setup the shift vector has the form $\delta = (\delta_{(1)}, \delta_{(2)}, \delta_{(3)}, \delta_{(4)})'$ with $\delta_{(1)}$ ranging again from 1 to 10 and $\delta_{(2)} = \delta_{(3)} = \delta_{(4)} = 0$. Clearly, the performance of the break date estimators deteriorates for the more complex four-dimensional DGP in case of smaller level shifts. Although this inferior performance is especially marked for $\hat{\tau}$ and $\hat{\tau}_R^{(2)}$, the relative ranking of the estimation procedures does not change.

Finally, we examine two DGPs for which $\delta_1 = \Pi\delta = 0$. For this situation, Theorems 3.1 and 3.2 state that compared to the case $\delta_1 \neq 0$ "larger" shift magnitudes are needed to ensure that $\hat{\tau}$ and $\hat{\tau}_R$ can estimate the break date consistently. For the estimator $\tilde{\tau}$ this situation is not covered in Theorem 3.3. First, we consider a three-dimensional process as in the base case but with $\delta_{(3)}$ ranging from one to ten and $\delta_{(1)} = \delta_{(2)} = 0$. Since the shift occurs in the third component which is nonstationary the level shift is orthogonal to the cointegration space in line with our DGP design (5.1). Thus, we simulate a case of co-breaking. Second, we use a three-dimensional process with $\psi_1 = 1$ so that the cointegrating rank is $r = 0$. In case of $r = 0$, all components of the DGP are nonstationary and therefore no error term correlation is present because Θ vanishes.

The results for the shift date estimators are depicted in Figures 5 and 6. Clearly, it is now more difficult for all procedures to locate τ. In line with the asymptotic results the difficulties are especially pronounced for $\tilde{\tau}$. Nevertheless, for very large shift magnitudes this estimator is able to find τ with a relatively high frequency. However, $\tilde{\tau}$ is outperformed by all other procedures except in case of $p = 3$, for which $\hat{\tau}$ is still not very successful in locating the correct break date. But even for $p = 3$, $\hat{\tau}$ is superior to $\tilde{\tau}$ if small shift magnitudes are considered. We see that the situation of no cointegration is much more difficult to deal with than co-breaking. When $r = 0$, $\hat{\tau}_R^{(2)}$ and $\hat{\tau}$ always locate τ correctly (or within the band $[\tau-2, \tau+2]$) only if $\delta_{(1)} = 10$. With respect to co-breaking, by contrast, this outcome already

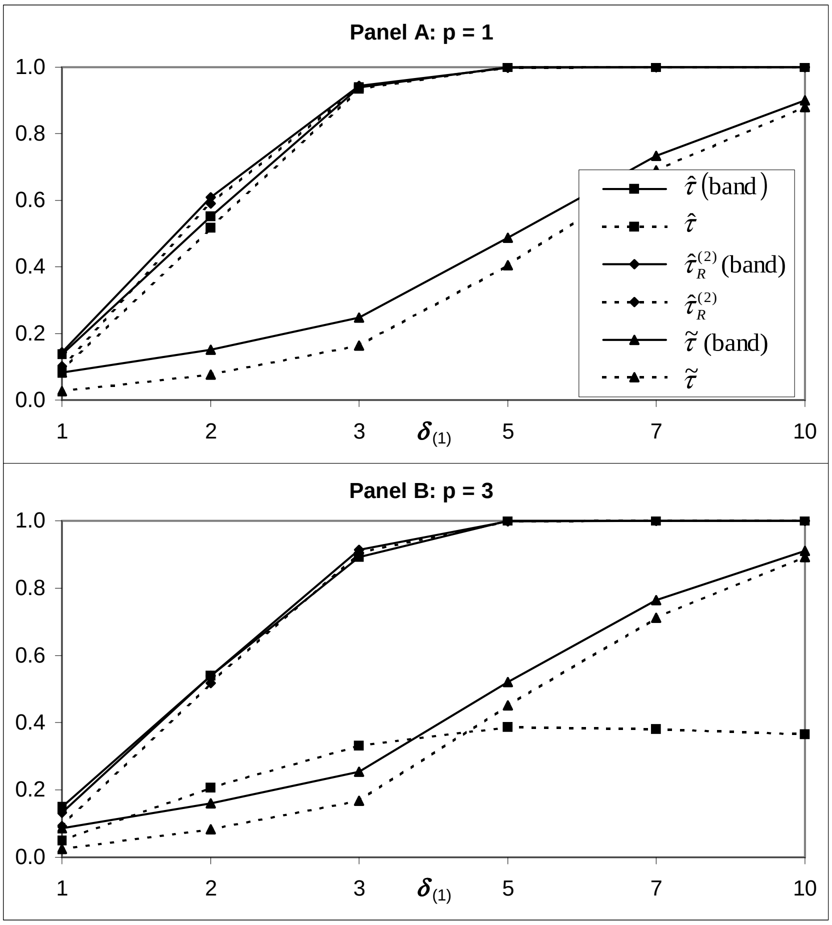

Figure 5: Relative frequency of true break point estimates ($\hat{\tau}$, $\hat{\tau}_R^{(2)}$, $\tilde{\tau}$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}$ (band), $\hat{\tau}_R^{(2)}$ (band), $\tilde{\tau}$ (band)) for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(1)} = \delta_{(2)} = 0$.

occurs for shift magnitudes of five or seven. In any case, the relatively poor performance of these procedures for small shift sizes relative to DGPs with $\delta_1 \neq 0$ is in accordance with the finding in Section 3 that precise estimation in the presence of $\delta_1 = 0$ requires large shift magnitudes.

Hence, we can summarize our results as follows. The constrained estimators $\hat{\tau}_R$ and $\hat{\tau}_R^{(2)}$ are usually superior to all other procedures. Apart from a few exceptions no other estimator performs better than these procedures in terms of locating the true shift date. In general the performance of $\hat{\tau}_R$ and $\hat{\tau}_R^{(2)}$ is similar, but the simple two-step estimator $\hat{\tau}_R^{(2)}$ may have some more pronounced advantages if we consider VAR models with higher orders, i.e. models with

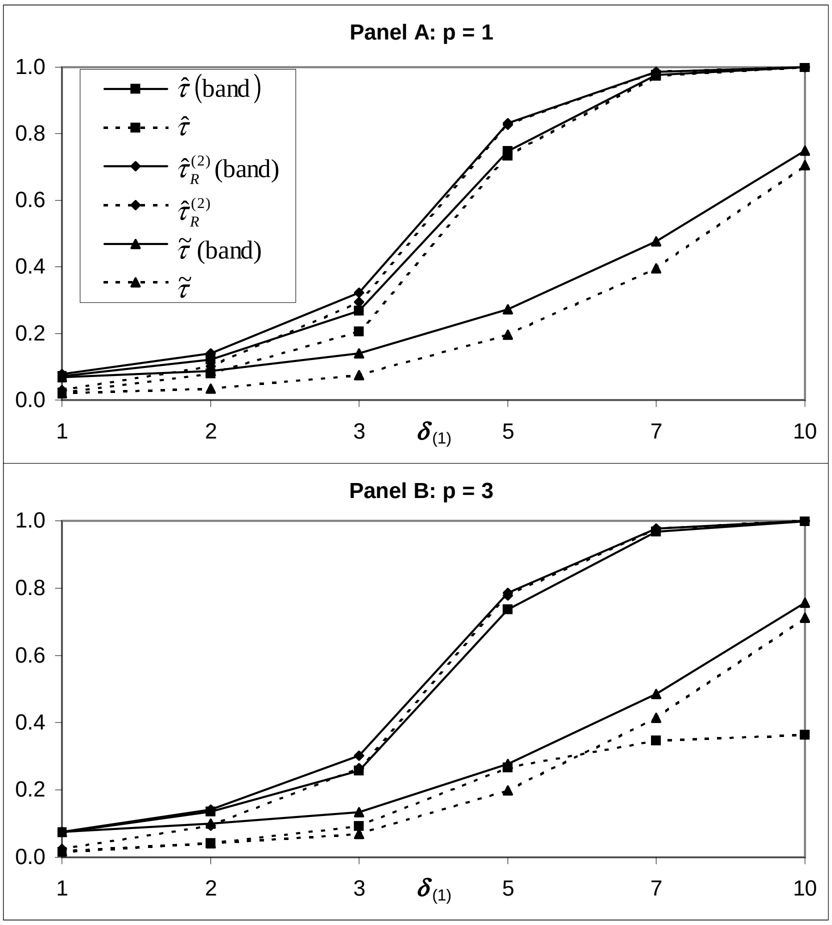

Figure 6: Relative frequency of true break point estimates ($\hat{\tau}$, $\hat{\tau}_R^{(2)}$, $\tilde{\tau}$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}$ (band), $\hat{\tau}_R^{(2)}$ (band), $\tilde{\tau}$ (band)) for three-dimensional DGP with $r = 0$ ($\psi_1 = 1$), $\Theta = (0, 0)$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

many parameters to be estimated. In this case the Gauss-Newton algorithm can fail to detect the global minimum of the respective determinant of the residual covariance matrix so that a wrong break point is detected. Given the fact that $\hat{\tau}_R^{(2)}$ is much faster in terms of computation time one may have a preference for this estimator. Applying nonlinear LS to a subset of observations around a pre-break date estimate only does not pay, as the performance of $\hat{\tau}_{R,p}$ has shown. The estimator $\hat{\tau}$ which does not take the nonlinear restrictions involving the model's impulse dummy variables into account is outperformed by all other procedures when fitting a VAR order $p = 3$. The small sample results for $\hat{\tau}$ are in line with our asymptotic derivations which say that this procedure can estimate the break date too early when the

VAR model is overspecified. However, if the frequency of estimates within a small band around the true shift date is considered, $\hat{\tau}$ may outperform the constrained estimators in case of $p = 3$.

So far we have just analyzed the small sample properties of the break date estimators in terms of their ability to locate the true shift date. If one is primarily interested in the cointegrating rank of the system the focus should be on the small sample properties of the cointegration tests based on these different estimators. Our main conclusion is that the tests' small sample size and power generally differ less than the results of the break date estimators. Therefore, we only discuss the outcomes for our three-dimensional base DGP and the processes with $\delta_1 = \Pi\delta = 0$.

The results with respect to the three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\delta_{(2)} = \delta_{(3)} = 0$ and $\delta_{(1)}$ ranging from one to ten are also given in the Tables 2 and 3. To be precise, we present the rejection frequencies for the null hypothesis $H_0 : r = r_0$ when the test is applied to a process with estimated shift date. The size and power values for the case of a known break date can be read from the situations where the procedures find the correct shift date $\tau = 50$ in all simulation repetitions (100%). The rejection frequencies for the case $r_0 = 1$ should give an indication of the tests' sizes in small samples. Therefore we use the term *size* in the following when we refer to this case. In Table 2 we see that the tests' sizes are clearly higher in cases of small shift magnitudes for which we obtain many incorrect break date locations. However, for increasing shift magnitudes the sizes approach the values for a known shift date in line with the greater success of the estimators to locate τ. Regarding the small sample power we observe an increase for small values of $\delta_{(1)}$ in case of $\hat{\tau}_R$ and a stronger drop for $\delta_{(1)} = 3$ and $\delta_{(1)} = 5$ if $\tilde{\tau}$ is used. However, the increase in power is relatively minor compared to the increase in size. Recall that the tests are based on asymptotic critical values and no adjustment for the larger actual size is made. The power of the test based on $\hat{\tau}$ is unaffected by the estimation of the shift date.

However, the situation is a bit different with respect to $\hat{\tau}$ when a VAR(3) model is fitted (compare Table 3). Here, the sizes and powers of the corresponding cointegration tests fall below the values for a known shift date when $\delta_{(1)}$ is equal to seven or ten. Obviously, the effect of the wrong locations on the small sample properties becomes important if the shift magnitude is large. As in the case of a VAR(1) model, the sizes and the powers of the tests

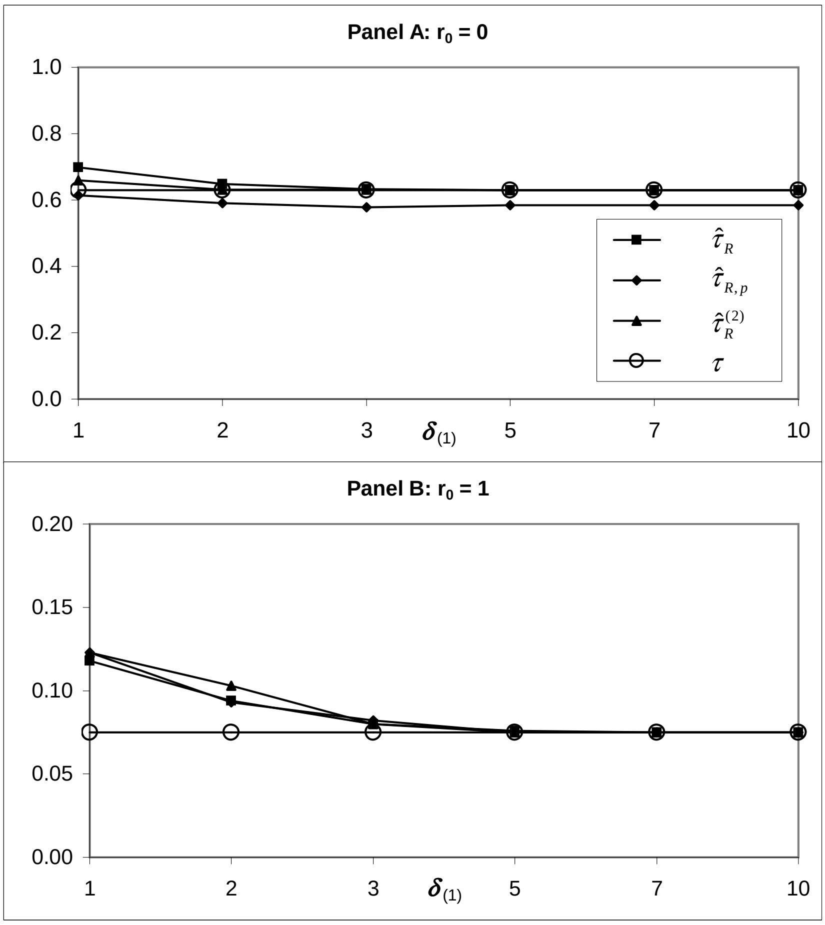

Figure 7: Relative rejection frequencies of cointegration rank tests based on constrained estimators for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, $p = 1$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

based on $\hat{\tau}_R$ are higher for small values of $\delta_{(1)}$ whereas we do not observe a fall in the power for $\tilde{\tau}$. Note, that the tests' small sample powers are clearly lower when fitting a VAR(3) instead of a VAR(1) model even if the true shift date is known.

Furthermore, we have found that there are no important differences between the small sample properties of the cointegration tests referring to the three constrained estimators $\hat{\tau}_R$, $\hat{\tau}_{R,p}$, and $\hat{\tau}_R^{(2)}$ for both VAR orders used. As an example the results for $p = 1$ are shown in Figure 7. The rejection frequencies for the null hypotheses $r = 0$ and $r = 1$ represent the small sample powers and sizes, respectively. The lines regarding τ show the outcomes for the situation of a known shift date. Thus, the differences with respect to the location of the break point do not carry over to the cointegration tests. Therefore, we do not present more

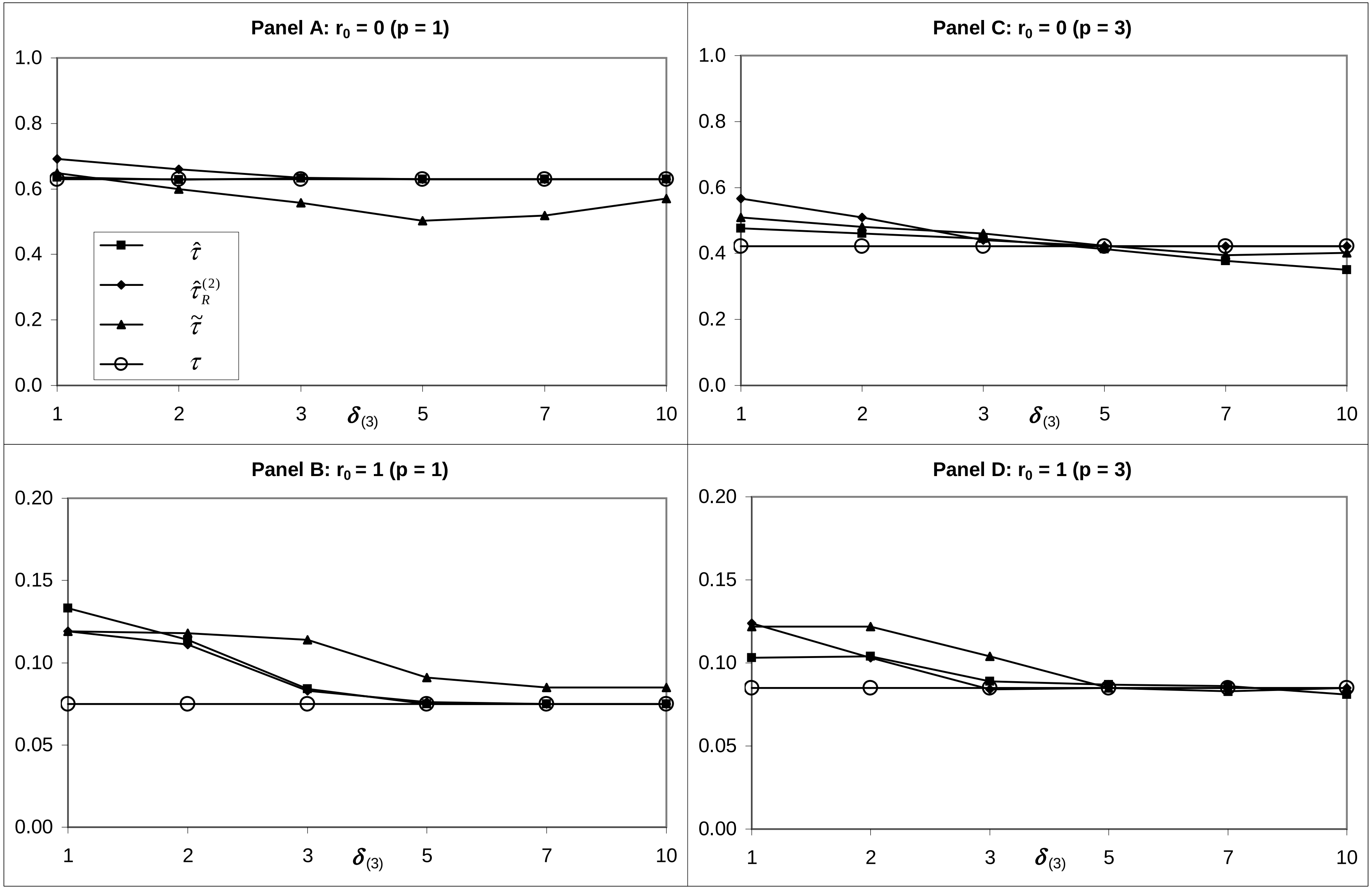

Figure 8: Relative rejection frequencies of cointegration rank tests for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(1)} = \delta_{(2)} = 0$.

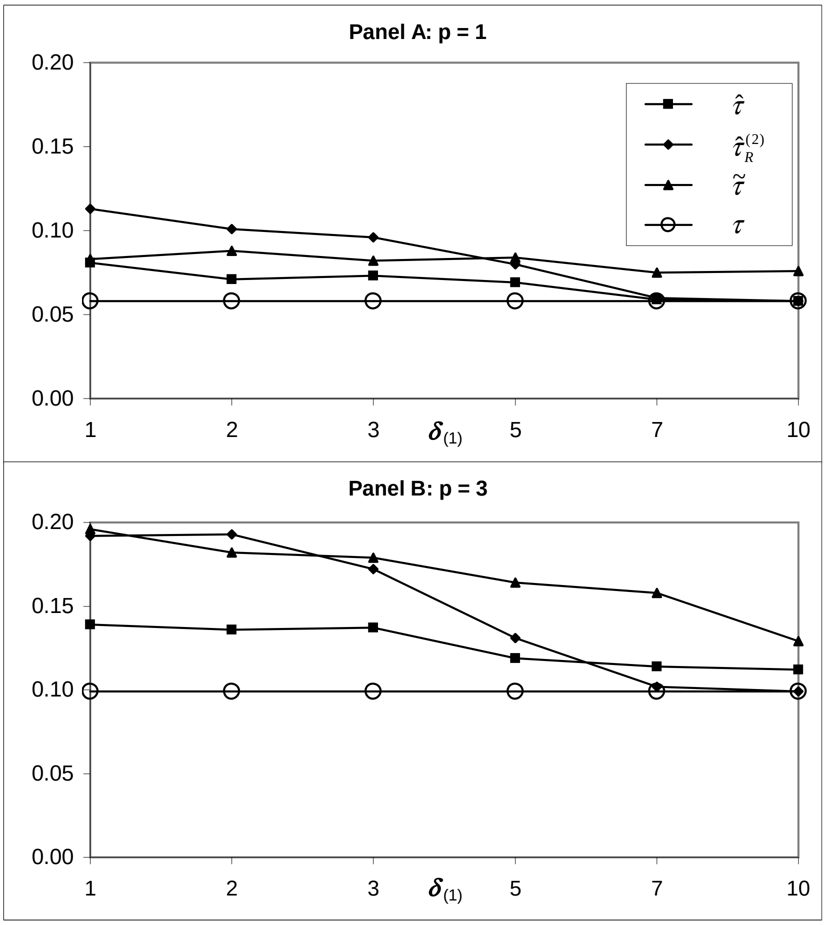

Figure 9: Relative rejection frequencies of cointegration rank tests with respect to $r_0 = 0$ for three-dimensional DGP with $r = 0$ ($\psi_1 = 1$), $\Theta = (0, 0)$, sample size $T = 100$, true break point $\tau = 50$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

detailed results. Instead, we focus on the cointegration tests based on $\hat{\tau}_R^{(2)}$ in the following as we have done when evaluating the ability of the estimators to locate the true shift date.

We now consider some small sample properties of the cointegration tests for the processes with $\delta_1 = \Pi\delta = 0$. Figure 8 contains the results for the three-dimensional DGP with $r = 1$ and different values of $\delta_{(3)}$ with respect to the orders $p = 1$ and $p = 3$. We see that the cointegration test based on $\tilde{\tau}$ suffers from power losses in small samples in case of $p = 1$. However, the loss is relatively small compared to the poor performance of the estimator $\tilde{\tau}$ for the current setup (compare Figure 5, Panels A and B). In case of $p = 3$ the power reduction for large shift magnitudes can be neglected. The size in small samples tends to be higher for all estimators than the size in case of the correct shift date if we consider small values of

$\delta_{(3)}$. The size increase seems to be most important with respect to $\tilde{\tau}$.

Figure 9 presents the rejection frequencies of the cointegration tests for the three-dimensional DGP with $r = 0$. All tests have a higher size in small samples for small shift magnitudes compared to a cointegration test using the true shift date and reject the null hypothesis too often given the significance level of 5%. Obviously, the size increase is stronger for $p = 3$. The test based on $\hat{\tau}_R^{(2)}$ has the largest size for small values of $\delta_{(1)}$ but its size is closer to the values for a known shift date in case of larger shift magnitudes. But again, the overall small sample properties of the different cointegration procedures do not differ much. This was also observed for the three-dimensional DGP with $r = 1$ and no innovation correlation and the four-dimensional process we have studied.

Thus, we have seen that the differences with respect to the cointegration tests are less marked than concerning the location of the break point. However, we observe some reduction in power in case of overspecified VAR models if $\hat{\tau}$ is used. Hence, if one runs the risk of overspecifying the model it is not recommended to apply this break date estimator. Furthermore, $\tilde{\tau}$ can also induce power losses in some situations. Therefore, we suggest to use the constrained estimators $\hat{\tau}_R$, $\hat{\tau}_R^{(2)}$, and $\hat{\tau}_{R,p}$ if the primary interest is the determination of the cointegrating rank. Since $\hat{\tau}_R$ and $\hat{\tau}_R^{(2)}$ are in addition the most successful procedures to find the true shift date we have a clear preference for these estimators. Hence, we conclude that taking account of the nonlinear restrictions is beneficial to both the location of the shift date and the testing for the cointegrating rank. In fact, estimating the shift date does not worsen the small sample properties of the cointegration tests much relative to the case of a known break point if an appropriate shift date estimator is used.

As mentioned in Section 4, LST have suggested a cointegration rank test based on GLS estimation of all the deterministic terms including μ_0 and adjusting y_t accordingly. In contrast, the level term is considered in the second stage only when setting up the LR tests treated so far. Now we compare the small sample properties of these two cointegration test variants for known and unknown break date. In the latter case we use all three break date estimators $\hat{\tau}$, $\hat{\tau}_R^{(2)}$, and $\tilde{\tau}$. In LST theoretical results are only given for fixed shift size and $\hat{\tau}_R$ is not considered explicitly. Because $\hat{\tau}_R$ satisfies the conditions of Theorem 4.1 of LST for the case $a = 0$, using this estimator here as well is justified, however. The results for our basic three-dimensional DGP with fitted VAR orders $p = 1$ and $p = 3$ are summarized in

Table 4: Relative Rejection Frequencies of Cointegration Rank Tests for Three-Dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $p = 1$, $\Theta = (0.4, 0.8)$, Sample Size $T = 100$, VAR Order $p = 1$, True Break Point $\tau = 50$, Nominal Significance Level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

		Known break date	$\delta_{(1)} = 1$	$\delta_{(1)} = 2$	$\delta_{(1)} = 3$	$\delta_{(1)} = 5$	$\delta_{(1)} = 7$	$\delta_{(1)} = 10$
		New cointegration rank test based on $\hat{\tau}$						
Rejection frequencies	$r_0 = 0$	0.630	0.631	0.620	0.624	0.630	0.630	0.630
	$r_0 = 1$	0.075	0.135	0.103	0.079	0.075	0.075	0.075
	$r_0 = 2$	0.009	0.017	0.017	0.009	0.009	0.009	0.009
		LST cointegration rank test based on $\hat{\tau}$						
Rejection frequencies	$r_0 = 0$	0.594	0.593	0.592	0.591	0.594	0.594	0.594
	$r_0 = 1$	0.040	0.066	0.053	0.045	0.040	0.040	0.040
	$r_0 = 2$	0.008	0.008	0.008	0.008	0.008	0.008	0.008
		New cointegration rank test based on $\hat{\tau}_R^{(2)}$						
Rejection frequencies	$r_0 = 0$	0.630	0.696	0.653	0.635	0.630	0.630	0.630
	$r_0 = 1$	0.075	0.123	0.103	0.080	0.075	0.075	0.075
	$r_0 = 2$	0.009	0.015	0.016	0.010	0.009	0.009	0.009
		LST cointegration rank test based on $\hat{\tau}_R^{(2)}$						
Rejection frequencies	$r_0 = 0$	0.594	0.649	0.618	0.600	0.594	0.594	0.594
	$r_0 = 1$	0.040	0.064	0.060	0.044	0.040	0.040	0.040
	$r_0 = 2$	0.008	0.009	0.009	0.007	0.008	0.008	0.008
		New cointegration rank test based on $\tilde{\tau}$						
Rejection frequencies	$r_0 = 0$	0.630	0.643	0.614	0.547	0.558	0.602	0.625
	$r_0 = 1$	0.075	0.126	0.122	0.106	0.080	0.076	0.075
	$r_0 = 2$	0.009	0.024	0.018	0.019	0.012	0.009	0.009
		LST cointegration rank test based on $\tilde{\tau}$						
Rejection frequencies	$r_0 = 0$	0.594	0.590	0.545	0.506	0.531	0.562	0.588
	$r_0 = 1$	0.040	0.060	0.059	0.058	0.047	0.044	0.041
	$r_0 = 2$	0.008	0.013	0.011	0.007	0.007	0.007	0.008

Tables 4 and 5, respectively.

The rejection frequencies for the case of a known break date are given in the first column of the Tables 4 and 5. Obviously, the new tests reject somewhat too often if the null hypothesis is true ($r = 1$). These higher rejection frequencies were also found for other DGPs. This may lead to a substantial size distortion in some situations. Thus, it may be worth exploring small sample corrections for the new tests in future work.

Let us now turn to the case of an unknown break date where we use the estimators $\hat{\tau}$,

Table 5: Relative Rejection Frequencies of Cointegration Rank Tests for Three-Dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $p = 1$, $\Theta = (0.4, 0.8)$, Sample Size $T = 100$, VAR Order $p = 3$, True Break Point $\tau = 50$, Nominal Significance Level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$.

		Known break date	$\delta_{(1)} = 1$	$\delta_{(1)} = 2$	$\delta_{(1)} = 3$	$\delta_{(1)} = 5$	$\delta_{(1)} = 7$	$\delta_{(1)} = 10$
		New cointegration rank test based on $\hat{\tau}$						
Rejection frequencies	$r_0 = 0$	0.422	0.482	0.463	0.436	0.411	0.396	0.382
	$r_0 = 1$	0.085	0.109	0.109	0.092	0.077	0.072	0.075
	$r_0 = 2$	0.011	0.017	0.018	0.017	0.015	0.017	0.016
		LST cointegration rank test based on $\hat{\tau}$						
Rejection frequencies	$r_0 = 0$	0.392	0.406	0.382	0.350	0.289	0.246	0.221
	$r_0 = 1$	0.046	0.050	0.043	0.044	0.045	0.035	0.034
	$r_0 = 2$	0.008	0.010	0.008	0.006	0.004	0.004	0.002
		New cointegration rank test based on $\hat{\tau}_R^{(2)}$						
Rejection frequencies	$r_0 = 0$	0.422	0.557	0.491	0.435	0.422	0.422	0.422
	$r_0 = 1$	0.085	0.124	0.118	0.090	0.085	0.085	0.085
	$r_0 = 2$	0.011	0.018	0.018	0.012	0.011	0.011	0.011
		LST cointegration rank test based on $\hat{\tau}_R^{(2)}$						
Rejection frequencies	$r_0 = 0$	0.392	0.474	0.425	0.396	0.392	0.392	0.392
	$r_0 = 1$	0.046	0.073	0.054	0.048	0.046	0.046	0.046
	$r_0 = 2$	0.008	0.011	0.011	0.008	0.008	0.008	0.008
		New cointegration rank test based on $\tilde{\tau}$						
Rejection frequencies	$r_0 = 0$	0.422	0.531	0.502	0.466	0.429	0.428	0.422
	$r_0 = 1$	0.085	0.129	0.113	0.108	0.088	0.083	0.085
	$r_0 = 2$	0.011	0.023	0.020	0.013	0.013	0.012	0.011
		LST cointegration rank test based on $\tilde{\tau}$						
Rejection frequencies	$r_0 = 0$	0.392	0.424	0.407	0.388	0.370	0.384	0.390
	$r_0 = 1$	0.046	0.064	0.057	0.050	0.045	0.047	0.046
	$r_0 = 2$	0.008	0.010	0.012	0.012	0.006	0.008	0.008

$\hat{\tau}_R^{(2)}$, and $\tilde{\tau}$. Interestingly, the relative performance of the old and new tests based on these estimators is in general similar with respect to an increasing shift magnitude. An exception is the case of a fitted VAR order $p = 3$ when $\hat{\tau}$ is used. For increasing shift sizes, the small sample power of the LST test falls clearly below the power in case of a known break date. We also observe a drop in small sample power for the new test procedure when the shift magnitude is large but the drop is relatively smaller. Obviously, our new test proposal is less affected by the incorrect break date estimates.

6 Conclusions

We have analyzed the asymptotic properties of three estimators for the shift date in a cointegrated VAR process with level shift. The shift is modelled by a simple shift dummy variable. The first estimator is based on an unrestricted VAR model, the second one is obtained by taking into account the relation between the parameters of the stochastic and deterministic parts of the model and, finally, the third estimator is based on a misspecified model, ignoring impulse dummy variables that are present in our model setup. Asymptotic properties of all three estimators are given under the assumption that the shift may depend on the sample size. Both, a growing and a declining shift size when the sample size tends to infinity are considered. These results extend previous results of LST who consider two of the three shift date estimators assuming a fixed shift size. Our results shed new light on previously unexplained small sample phenomena. We have also considered the implications of using models with estimated instead of true shift dates in testing for the cointegrating rank and we have proposed new variants of cointegration rank tests. These tests differ from those considered by LST in that they avoid estimating the nonidentified part of the levels parameter and proceed otherwise in a similar manner. More precisely, the trend and shift parameters are estimated in a first step and then rank tests of the LR type are applied to adjusted series. The asymptotic distribution of the new tests is derived.

In addition to providing asymptotic results, we have also investigated the small sample properties of the procedures using a Monte Carlo simulation experiment. It is found that the estimator that takes the restrictions into account is overall the most successful one in locating the true shift date. A computationally efficient variant that does not require computer intensive iterative optimization algorithms is shown to work as well as an estimator based on a full optimization of the nonlinear objective function. Although a superior break date estimator tends to improve the small sample properties of subsequent cointegration tests, such improvements are relatively small because the differences between the break date estimators are small when the shift size is large and, hence, the shift is important. Generally it pays to account for a shift in testing for the cointegrating rank of a system of variables when such a shift is actually present.

A comparison of the tests considered by LST and the new tests of the present paper shows, however, that the latter tend to reject a true null hypothesis more often than the

LST tests. Generally the new tests tend to reject true null hypotheses too often and, hence, in future research it may be of interest to develop small sample corrections to ensure a test size close to the nominal level.

Appendix A. Proofs

Some parts of the proofs are similar to those of the corresponding results stated in LST under more restrictive conditions. Because these authors provide brief sketches of the proofs only, we also present more detailed and more complete versions of the similar parts here.

The following notational conventions are used in addition to the notation defined earlier. Right hand side and left hand side will be abbreviated by r.h.s. and l.h.s., respectively. The smallest and largest eigenvalues of a matrix are denoted by $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$, respectively. The complement of a set B is signified by B^c. The dependence of quantities on the sample size T is not indicated. 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 $T \to \infty$. 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.

In the proofs we assume the model and conditions described in Sections 2 and 3, where the parameters $\mu_0, \mu_1, \delta_* \in \mathbf{R}^n$ and the true α, β, Π and Γ_j $(j = 1, \dots, p-1)$ satisfy the restrictions which ensure that the observed variables are at most $I(1)$ whereas these restrictions are not imposed in the estimation.

The true DGP is one specific process from our model class. It is occasionally helpful to be more explicit about its particular parameter values. In these cases they will be indicated with a subscript 'o' (e.g., μ_{0o}, μ_{1o}, τ_o etc.). For the break date we assume for convenience that $\underline{\lambda} < \lambda_o < \bar{\lambda}$ and $\mathcal{T} = [T\underline{\lambda}, T\bar{\lambda}]$. We begin by proving Theorem 3.1.

A.1 Proof of Theorem 3.1

Instead of the series y_t it will be convenient to use the mean adjusted series

$$x_t = y_t - \mu_{0o} - \mu_{1o}t - \delta_o d_{t\tau_o}, \quad t = 1, 2, \ldots$$

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

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

Here

$$\begin{aligned}
\nu_0^{(0)} &= \nu_0 + \Pi\mu_{0o} - \Psi\mu_{1o} - \Pi\mu_{1o} \\
\nu_1^{(0)} &= \nu_1 + \Pi\mu_{1o} \\
\underline{\gamma} &= [\gamma_0 : \cdots : \gamma_{p-1}] \\
\underline{d}_{t\tau} &= [\Delta d_{t\tau} : \cdots : \Delta d_{t-p+1,\tau}]' \\
\delta_1^{(0)} &= -\Pi\delta_o
\end{aligned}$$

and

$$\underline{\gamma}^{(0)} = [\gamma_0^{(0)} : \cdots : \gamma_{p-1}^{(0)}] \quad \text{with} \quad \gamma_j^{(0)} = \begin{cases} \delta_o - \delta_1^{(0)}, & j = 0 \\ -\Gamma_j \delta_o, & j = 1, \ldots, p-1. \end{cases}$$

Note that the true values of $\nu_0^{(0)}$ and $\nu_1^{(0)}$ are zero.

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

$$\Delta x_t = \nu_0^{(0)} + \nu_1^{(0)} t + \delta_1 d_{t\tau} + \underline{\gamma}\,\underline{d}_{t\tau} - \delta_1^{(0)} d_{t\tau_o} - \underline{\gamma}^{(0)}\underline{d}_{t\tau_o} + \alpha^{(0)} u_{t-1}^{(0)} + \rho^{(0)} v_{t-1}^{(0)} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t,$$

$$t = p+1, p+2, \ldots \tag{A.2}$$

Denote $q_{t\tau} = [d_{t\tau} : \underline{d}_{t\tau}']'$ and

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

With this notation $(A.2)$ becomes

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

where $\Phi = [\nu_0^{(0)} : T\nu_1^{(0)} : T^{1/2}\rho^{(0)} : \alpha^{(0)} : \Gamma_1 : \cdots : \Gamma_{p-1}]$, $\Xi = [\delta_1 : \underline{\gamma}]$ and $\Xi^{(0)} = [\delta_1^{(0)} : \underline{\gamma}^{(0)}]$.

Let $\Theta = [\Phi : \Xi]$ contain the freely varying parameters in $(A.3)$ or $(A.2)$. ($\Xi^{(0)}$ is not a freely varying parameter because it is determined by $\alpha^{(0)}$, $\rho^{(0)}$ and $\Gamma_1, \ldots, \Gamma_{p-1}$.) Set

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

Then

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

is -2 times the (conditional) Gaussian log-likelihood function of the parameters in $(A.3)$. Minimizing this function yields Gaussian ML estimators of the parameters Θ, τ and Ω. It is not difficult to see that the resulting estimators of Θ and τ can alternatively be obtained by minimizing the concentrated counterpart of $l_T(\Theta, \tau, \Omega)$, that is,

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

The definition of $\varepsilon_{t\tau}(\Theta)$ (and the fact that $\Xi^{(0)}$ is not a freely varying parameter) makes it clear that the value of τ that minimizes the function $l_T^{(c)}(\Theta, \tau)$ is identical to $\hat{\tau}$ defined by (3.2). Thus, (asymptotic) properties of $\hat{\tau}$ can be studied by using the Gaussian ML estimator of τ discussed above. Before turning to this issue we note that the above discussion also makes clear that a minimizer of $l_T(\Theta, \tau, \Omega)$ exists (for every T larger than some constant).

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

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

and

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

Note that here $\bar{M}$ does not depend on T although Φ and $\delta_1^{(0)}$ do. We now prove

Lemma A.1. Let $B_1 = B_1(\bar{M}, \underline{\omega}, \bar{\omega})$ be the part of the parameter space of (Θ, τ, Ω) in which conditions $(A.4)$ and $(A.5)$ hold. Then there exist choices of $\bar{M}$, $\underline{\omega}$ and $\bar{\omega}$ such that

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

with probability approaching one.

Proof: First note that

$$T^{-1} l_T(\Theta_o, \tau_o, \Omega_o) = \left(1 - \frac{p}{T}\right) \log\det\Omega_o + \operatorname{tr}\left(\Omega_o^{-1} T^{-1} \sum_{t=p+1}^{T} \varepsilon_t \varepsilon_t{}'\right) = O_p(1), \qquad (A.6)$$

where the latter equality is justified by the weak law of large numbers.

Next, since $[T\underline{\lambda}] \leq \tau, \tau_o \leq [T\bar{\lambda}]$, we find from the definitions that

$$\varepsilon_{t\tau}(\Theta) = \Delta x_t - \Phi w_t^{(0)}, \quad t = p+1, \ldots, [T\underline{\lambda}] - 1,$$

and

$$\varepsilon_{t\tau}(\Theta) = \Delta x_t - \Phi w_t^{(0)} - (\delta_1 - \delta_1^{(0)}), \quad t = [T\bar{\lambda}] + p, \ldots, T.$$

Hence,

$$\begin{aligned} T^{-1} l_T(\Theta, \tau, \Omega) \;\; \geq \;\; & \left(1 - \tfrac{p}{T}\right) \log\det\Omega \\ & + \operatorname{tr}\left(\Omega^{-1} T^{-1} \textstyle\sum_{t=p+1}^{[T\underline{\lambda}]-1} [\Delta x_t - \Phi w_t^{(0)}][\Delta x_t - \Phi w_t^{(0)}]'\right) \\ & + \operatorname{tr}\left(\Omega^{-1} T^{-1} \textstyle\sum_{t=[T\bar{\lambda}]+p}^{T} [\Delta x_t - \Phi^{(0)} w_t^{(0)}][\Delta x_t - \Phi^{(0)} w_t^{(0)}]'\right), \end{aligned} \qquad (A.7)$$

where $\Phi^{(0)} = \Phi + [\delta_1 - \delta_1^{(0)} : 0]$. Here

$$\lambda_{\min}\left(T^{-1} \sum_{t=p+1}^{[T\underline{\lambda}]-1} \begin{bmatrix} \Delta x_t \\ w_t^{(0)} \end{bmatrix} \begin{bmatrix} \Delta x_t' : w_t^{(0)\prime} \end{bmatrix}\right) \geq \epsilon_* \qquad (A.8)$$

where $\epsilon_* > 0$ is a suitable real number and the inequality holds with probability approaching one. This fact can be justified in the same way as Lemma A.4 of Saikkonen (2001). A similar result is also obtained by changing the range of summation on the l.h.s. of $(A.8)$ to $t = [T\bar{\lambda}] + p, \ldots, T$. When these two eigenvalue conditions are assumed arguments entirely similar to those in Saikkonen (2001, pp. 320-321) show that, with suitable choices of $\bar{M}$, $\underline{\omega}$ and $\bar{\omega}$, the r.h.s of $(A.7)$ can be made arbitrarily large whenever $(\Theta, \tau, \Omega) \notin B_1(\bar{M}, \underline{\omega}, \bar{\omega})$. The assertion of the lemma follows from this and $(A.6)$. □

Lemma A.1 implies that a minimizer of $l_T(\Theta, \tau, \Omega)$ will asymptotically satisfy inequality restrictions of the form $(A.4)$ and $(A.5)$. In what follows, the set B_1 is always assumed to be defined in such a way that the conclusion of Lemma A.1 holds. We shall now proceed in the same way as in Saikkonen (2001) and express the function $l_T(\Theta, \tau, \Omega)$ as a sum of two components. To this end, define

$$w_{1t}^{(0)} = \left[1 : \frac{t}{T} : T^{-1/2} v_{t-1}^{(0)\,\prime}\right]'$$

and

$$w_{2t}^{(0)} = \left[u_{t-1}^{(0)\,\prime} : \Delta x_{t-1}' : \cdots : \Delta x_{t-p+1}'\right]'.$$

Then $w_t^{(0)} = [w_{1t}^{(0)\prime} : w_{2t}^{(0)\prime}]'$ and we also partition the parameter matrix Φ conformably as $\Phi = [\Phi_1 : \Phi_2]$ where $\Phi_1 = [\nu_0^{(0)} : T\nu_1^{(0)} : T^{1/2}\rho^{(0)}]$ and $\Phi_2 = [\alpha^{(0)} : \Gamma_1 : \cdots : \Gamma_{p-1}]$. With these definitions,

$$\varepsilon_{t\tau}(\Theta) = \varepsilon_{1t\tau}(\Theta) + \varepsilon_{2t}(\Phi_2),$$

where $\varepsilon_{1t\tau}(\Theta) = -\Phi_1 w_{1t}^{(0)} - \Xi q_{t\tau} + \Xi^{(0)} q_{t\tau_o}$ and $\varepsilon_{2t}(\Phi_2) = \Delta x_t - \Phi_2 w_{2t}^{(0)}$. Clearly, $\varepsilon_{1t\tau_o}(\Theta_o) = 0$ and

$$l_T(\Theta, \tau, \Omega) = l_{1T}(\Theta, \tau, \Omega) + l_{2T}(\Phi_2, \Omega),$$

where

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

and

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

For $l_{2T}(\Phi_2, \Omega)$ we have the following result.

Lemma A.2.

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

where the infimum is over unrestricted values of Φ_2 and $\Omega > 0$.

Proof: Because we can treat Δx_t as a zero mean stationary process and because $l_{2T}(\Phi_2, \Omega)$ can be interpreted as -2 times the logarithm of the Gaussian likelihood function associated with the regression model $\Delta x_t = \Phi_2 w_{2t}^{(0)} + \varepsilon_t$, the stated result follows from standard regression theory (cf. Saikkonen (2001, p. 321)). □

Next consider the function $l_{1T}(\Theta, \tau, \Omega)$. Our treatment will be divided into several steps in which the time index t is suitably restricted. This means considering the function $l_{1T}(\Theta, \tau, \Omega)$ with the sample size T replaced by appropriate quantities smaller than T. Most of the subsequent results will explicitly be formulated for $\tau \leq \tau_o$ and only briefly discussed in the case $\tau \geq \tau_o$. Due to the occurrence of impulse dummies the situation is in this respect somewhat more complicated than in previous cases where the break date parameter is not affected by impulse dummies (e.g., Bai, Lumsdaine & Stock (1998)).

In the following results about the function $l_{1T}(\Theta, \tau, \Omega)$, $c_1, c_2, \ldots$ denote positive constants and $a_{1T}, a_{2T}, \ldots$ are nonnegative random variables which depend on the sample size but not on the parameters Θ, τ or Ω. First we prove

Lemma A.3. There exists a constant $c_1 > 0$ such that, with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o$ and $(\Theta, \tau, \Omega) \in B_1$,

$$l_{1,\tau-1}(\Theta, \tau, \Omega) \geq c_1 \|T^{1/2}\Phi_1\|^2 - a_{1T}\|T^{1/2}\Phi_1\|,$$

where $a_{1T} \geq 0$ and $a_{1T} = O_p(1)$.

Proof: For $t \leq \tau - 1$, $\varepsilon_{1t\tau}(\Theta) = -\Phi_1 w_{1t}^{(0)}$ and, consequently,

$$l_{1,\tau-1}(\Theta, \tau, \Omega) = \mathrm{tr}\left(\Omega^{-1}\Phi_1 \sum_{t=p+1}^{\tau-1} w_{1t}^{(0)} w_{1t}^{(0)\prime} \Phi_1'\right) - 2\mathrm{tr}\left(\Omega^{-1}\Phi_1 \sum_{t=p+1}^{\tau-1} w_{1t}^{(0)} \varepsilon_{2t}(\Phi_2)'\right) \stackrel{def}{=} L_1 + L_2.$$

For L_1 we have

$$L_1 \geq \lambda_{\min}(\Omega^{-1})\mathrm{tr}\left(\Phi_1 \sum_{t=p+1}^{\tau-1} w_{1t}^{(0)} w_{1t}^{(0)\prime} {\Phi_1}'\right) \geq \lambda_{\min}(\Omega^{-1})\lambda_{\min}\left(T^{-1}\sum_{t=p+1}^{\tau-1} w_{1t}^{(0)} w_{1t}^{(0)\prime}\right)\|T^{1/2}\Phi_1\|^2.$$

For $(\Theta, \tau, \Omega) \in B_1$, the first eigenvalue in the last expression is bounded away from zero. That the same holds with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o$ for the second eigenvalue, can be seen by using an analog of (A.3) of Gregory & Hansen (1996, p. 118). Thus, we have shown that $L_1 \geq c_1\|\Phi_1\|^2$, $c_1 \geq 0$, with probability approaching one.

It remains to show that $L_2 \geq -a_{1T}\|T^{1/2}\Phi_1\|$ with a_{1T} having the properties stated in the lemma. To demonstrate this, notice that

$$\begin{aligned} |L_2| &\leq 2\|\Omega^{-1}\| \left\| \Phi_1 \sum_{t=p+1}^{\tau-1} w_{1t}^{(0)}[\Delta x_t - \Phi_2 w_{2t}^{(0)}]' \right\| \\ &\leq 2\|\Omega^{-1}\| \left\| T^{-1/2} \sum_{t=p+1}^{\tau-1} w_{1t}^{(0)}[\Delta x_t - \Phi_2 w_{2t}^{(0)}]' \right\| \|T^{1/2}\Phi_1\|. \end{aligned}$$

Here we have used the definition of $\varepsilon_{2t}(\Phi_2)$, the Cauchy-Schwarz inequality and the norm inequality. By an analog of (A.4) of Gregory & Hansen (1996, p. 118), the norm in the middle of the last expression is of order $O_p(1)$ uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_0$ and for any fixed value of Φ_2. Thus, because the parameters Φ_2 and Ω belong to bounded sets when $(\Theta, \tau, \Omega) \in B_1$, it can similarly be shown that the last expression as a whole has an upper bound $a_{1T}\|T^{1/2}\Phi_1\|$ with a_{1T} as required. This completes the proof. □

Our next result deals with the contribution of $l_{1,\tau_o-1}(\Theta,\tau,\Omega) - l_{1,\tau-1}(\Theta,\tau,\Omega)$ to $l_{1T}(\Theta,\tau,\Omega)$. Here the relevant expression of $\varepsilon_{1t\tau}(\Theta)$ is

$$\varepsilon_{1t\tau}(\Theta) = -\Psi_1 w_{1t}^{(0)} - \underline{\gamma}\underline{d}_{t\tau}, \quad t = \tau, \ldots, \tau_o - 1,$$

where $\Psi_1 = \Phi_1 + [\delta_1 : 0]$.

Lemma A.4. Let ϵ be any real number with the property $0 < \epsilon < \lambda_o - \underline{\lambda}$. Then, for $\underline{\lambda} \leq \lambda \leq \lambda_o - \epsilon$ there exists a constant $c_2 > 0$ such that, with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq [T(\lambda_o - \epsilon)]$ and $(\Theta, \tau, \Omega) \in B_1$,

$$l_{1,\tau_0-1}(\Theta,\tau,\Omega) - l_{1,\tau-1}(\Theta,\tau,\Omega) \geq c_2\|T^{1/2}\Psi_1\|^2 + c_2\|\underline{\gamma}\|^2 - a_{2T}\|T^{1/2}\Psi_1\| - a_{3T}\|\underline{\gamma}\|,$$

where $a_{iT} \geq 0$ $(i = 2, 3)$, $a_{2T} = O_p(1)$ and $a_{3T} = o_p(T^\eta)$ with $\frac{1}{b} < \eta < \frac{1}{4}$.

Proof: By the definitions,

$$\begin{aligned} &l_{1,\tau_o-1}(\Theta,\tau,\Omega) - l_{1,\tau-1}(\Theta,\tau,\Omega) \\ &= \mathrm{tr}\left(\Omega^{-1}\textstyle\sum_{t=\tau}^{\tau_o-1} \varepsilon_{1t\tau}(\Theta)\varepsilon_{1t\tau}(\Theta)'\right) + 2\mathrm{tr}\left(\Omega^{-1}\textstyle\sum_{t=\tau}^{\tau_o-1}\varepsilon_{1t\tau}(\Theta)\varepsilon_{2t}(\Phi_2)'\right) \overset{def}{=} L_3 + L_4. \end{aligned}$$

First consider L_3 and for simplicity denote $\underline{\Psi}_1 = [\Psi_1 : \underline{\gamma}]$ and $\underline{z}_{1t\tau}^{(0)} = [w_{1t}^{(0)\prime} : \underline{d}_{t\tau}']'$. Then

$$L_3 = \mathrm{tr}\left(\Omega^{-1}\underline{\Psi}_1 \sum_{t=\tau}^{\tau_o-1} \underline{z}_{1t\tau}^{(0)}\underline{z}_{1t\tau}^{(0)\prime}\underline{\Psi}_1'\right) \geq \lambda_{\min}(\Omega^{-1})\mathrm{tr}\left(\underline{\Psi}_1 D_{1T}\left(D_{1T}^{-1}\sum_{t=\tau}^{\tau_o-1}\underline{z}_{1t\tau}^{(0)}\underline{z}_{1t\tau}^{(0)\prime}D_{1T}^{-1}\right)D_{1T}\underline{\Psi}_1'\right), \tag{A.9}$$

where $D_{1T} = \text{diag}[T^{-1/2}I_{n-r+2} : I_p]$.

Next note that

$$T^{-1/2}\sum_{t=\tau}^{\tau_o-1} w_{1t}^{(0)}\Delta d_{t-i,\tau} = O_p(T^{-1/2}), \quad i = 0,\ldots,p-1, \qquad (A.9a)$$

uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o$. Because $w_{1t}^{(0)} = [1 : \frac{t}{T} : T^{-1/2}v_{t-1}^{(0)\prime}]'$ this is obvious for the first and second components of $w_{1t}^{(0)}$. For the third component the same is true because $T^{-1/2}\max_{1\leq t\leq\tau_o}\|v_{t-1}^{(0)}\| \leq T^{-1/2}\max_{1\leq t\leq T}\|\beta_{o\perp}'x_{t-1}\| = O_p(1)$, where the equality follows form the fact that $T^{-1/2}\beta_{o\perp}'x_{[Ts]}$ obeys an invariance principle. Thus, we can conclude that

$$D_{1T}^{-1}\sum_{t=\tau}^{\tau_o-1}\underline{z}_{1t\tau}^{(0)}\underline{z}_{1t\tau}^{(0)\prime}D_{1T}^{-1} = \text{diag}\left[T^{-1}\sum_{t=\tau}^{\tau_o-1}w_{1t}^{(0)}w_{1t}^{(0)\prime} : I_p\right] + o_p(1) \qquad (A.10)$$

uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o - p$.

The next step is to observe that

$$T^{-1}\sum_{t=[T\lambda]}^{[T\lambda_o]-1} w_{1t}^{(0)}w_{1t}^{(0)\prime} \Rightarrow M_{11}(\lambda_o) - M_{11}(\lambda), \quad \underline{\lambda} \leq \lambda < \lambda_o,$$

where $M_{11}(\lambda)$ is the weak limit of $T^{-1}\sum_{t=p+1}^{[T\lambda]-1}w_{1t}^{(0)}w_{1t}^{(0)\prime}$ (cf. (A.3) of Gregory & Hansen (1996, p. 118)). It is straightforward to check that the difference $M_{11}(\lambda_o) - M_{11}(\lambda)$ is positive definite and its smallest eigenvalue is bounded from below by a positive constant when $\underline{\lambda} \leq \lambda \leq \lambda_o - \epsilon$.

The above discussion implies that, with probability approaching one, the smallest eigenvalue of the matrix on the l.h.s. of (A.10) is bounded away from zero uniformly in $[T\underline{\lambda}] \leq \tau \leq [T(\lambda_o - \epsilon)]$. Thus, with probability approaching one and in the required uniform sense,

$$L_3 \geq c_2\text{tr}(\underline{\Psi}_1 D_{1T}D_{1T}\underline{\Psi}_1') = c_2\|T^{1/2}\Psi_1\|^2 + c_2\|\underline{\gamma}\|^2,$$

where $c_2 > 0$ is a (small) constant. This implies that it only remains to show that $L_4 \geq -a_{2T}\|T^{1/2}\Psi_1\| - a_{3T}\|\underline{\gamma}\|$ with a_{2T} and a_{3T} as stated in the lemma.

To show the above mentioned inequality about L_4, conclude from the definitions that

$$L_4 = -2\text{tr}\left(\Omega^{-1}\Psi_1\sum_{t=\tau}^{\tau_o-1}w_{1t}^{(0)}\varepsilon_{2t}(\Phi_2)'\right) - 2\text{tr}\left(\Omega^{-1}\underline{\gamma}\sum_{t=\tau}^{\tau_0-1}\underline{d}_{t\tau}\varepsilon_{2t}(\Phi_2)'\right) \stackrel{def}{=} L_{41} + L_{42}.$$

Arguments similar to those already used in the proof of Lemma A.3 show that

$$|L_{41}| \leq 2\|\Omega^{-1}\|\|T^{1/2}\Psi_1\|\left\|T^{-1/2}\sum_{t=\tau}^{\tau_o-1}w_{1t}^{(0)}[\Delta x_t - \Phi_2 w_{2t}^{(0)}]'\right\| \leq a_{2T}\|T^{1/2}\Psi_1\|,$$

where $a_{2T} = O_p(1)$ in the required uniform sense.

Regarding L_{42}, one similarly obtains

$$|L_{42}| \leq 2\|\Omega^{-1}\|\|\underline{\gamma}\| \left\| \sum_{t=\tau}^{\tau_o-1} \underline{d}_{t\tau}[\Delta x_t - \Phi_2 w_{2t}^{(0)}]' \right\| \leq a_{3T}\|\underline{\gamma}\|,$$

where $a_{3T} = o_p(T^\eta)$, $\frac{1}{b} < \eta < \frac{1}{4}$, in the required uniform sense. The latter inequality follows if the last norm in the preceding expression can be replaced by $o_p(T^\eta)$. To justify this, recall that Δx_t and $w_{2t}^{(0)}$ are stationary processes with finite moments of order $b > 4$ and that Φ_2 can be assumed to belong to a bounded set. Thus, it suffices to show that $\max_{1\leq t\leq T} \|\Delta x_t\| = o_p(T^\eta)$ and similarly with Δx_t replaced by $w_{2t}^{(0)}$. This, however, can be done by using an argument entirely similar to that in (A.14) of Saikkonen & Lütkepohl (2002). The inequalities obtained for $|L_{41}|$ and $|L_{42}|$ above show that L_2 has the required lower bound and the proof is complete. □

Our next result describes the contribution of $l_{1,\tau_o+p-1}(\Theta, \tau, \Omega) - l_{1,\tau_o-1}(\Theta, \tau, \Omega)$ to $l_{1T}(\Theta, \tau, \Omega)$. We introduce the notation

$$\zeta_{t\tau}^{(0)} = (d_{t\tau} - d_{t\tau_o})\delta_1^{(0)} + \underline{\gamma}\underline{d}_{t\tau} - \underline{\gamma}^{(0)}\underline{d}_{t\tau_o}.$$

In the following lemma the relevant values of $\varepsilon_{1t\tau}(\Theta)$ can then be written as

$$\varepsilon_{1t\tau}(\Theta) = -\Psi_2 w_{1t}^{(0)} - \zeta_{t\tau}^{(0)}, \quad t = \tau_o, \ldots, \tau_o + p - 1,$$

where $\Psi_2 = \Phi_1 + [\delta_1 - \delta_1^{(0)} : 0]$. Note that here the first term in the definition of $\zeta_{t\tau}^{(0)}$ vanishes but the general definition is convenient in later derivations. Now we can formulate

Lemma A.5. There exists a constant $c_3 > 0$ such that, with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o$ and $(\Theta, \tau, \Omega) \in B_1$,

$$l_{1,\tau_o+p-1}(\Theta, \tau, \Omega) - l_{1,\tau_o-1}(\Theta, \tau, \Omega) \geq c_3 \sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 - a_{4T} \left(\sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} - a_{5T},$$

where $a_{iT} \geq 0$ and $a_{iT} = O_p(1)$ $(i = 4, 5)$.

Proof: By the definitions,

$$\begin{aligned} &l_{1,\tau_o+p-1}(\Theta, \tau, \Omega) - l_{1,\tau_o-1}(\Theta, \tau, \Omega) \\ &= \text{tr}\left(\Omega^{-1} \textstyle\sum_{t=\tau_o}^{\tau_o+p-1} \varepsilon_{1t\tau}(\Theta)\varepsilon_{1t\tau}(\Theta)'\right) + 2\text{tr}\left(\Omega^{-1} \textstyle\sum_{t=\tau_o}^{\tau_o+p-1} \varepsilon_{1t\tau}(\Theta)\varepsilon_{2t}(\Phi_2)'\right) \stackrel{def}{=} L_5 + L_6. \end{aligned}$$

Assuming $(\Theta, \tau, \Omega) \in B_1$ we find that

$$\begin{aligned} L_5 &\geq \lambda_{\min}(\Omega^{-1}) \sum_{t=\tau_o}^{\tau_o+p-1} \|\varepsilon_{1t\tau}(\Theta)\|^2 \\ &\geq \bar{\omega}^{-1} \sum_{t=\tau_o}^{\tau_o+p-1} \|\Psi_2 w_{1t}^{(0)}\|^2 + \bar{\omega}^{-1} \sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 + 2\bar{\omega}^{-1} \sum_{t=\tau_o}^{\tau_o+p-1} \zeta_{t\tau}^{(0)\prime} \Psi_2 w_{1t}^{(0)}. \end{aligned}$$

Because we can here assume that Ψ_2 is bounded (see $(A.5)$), an application of the triangle inequality and the Cauchy-Schwarz inequality shows that the absolute value of the third term in the last expression can be bounded from above by

$$const. \times \left(\sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} \left(\sum_{t=\tau_o}^{\tau_o+p-1} \|w_{1t}^{(0)}\|^2 \right)^{1/2}.$$

Here the latter square root is of order $O_p(1)$ (see the argument leading to $(A.10)$). Hence, we can conclude that

$$L_5 \geq c_3 \sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 - a_{41T} \left(\sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2}, \tag{A.11}$$

where $c_3 = \bar{\omega}^{-1} > 0$ and $a_{41T} = O_p(1)$ in the required uniform sense.

Now consider L_6. Arguments similar to those used in previous derivations combined with the present definition of $\varepsilon_{1t\tau}(\Theta)$ yield

$$|L_6| \leq 2\|\Omega^{-1}\| \left\| \Psi_2 \sum_{t=\tau_o}^{\tau_o+p-1} w_{1t}^{(0)} \varepsilon_{2t}(\Phi_2)' \right\| + 2\|\Omega^{-1}\| \left\| \sum_{t=\tau_o}^{\tau_o+p-1} \zeta_{t\tau}^{(0)} \varepsilon_{2t}(\Phi_2)' \right\|.$$

It is easy to see that the first term on the r.h.s. can be used to define the term a_{5T} in the lemma. The arguments needed are similar to those used to obtain $(A.11)$ and they can also be applied to the second term so that we can write

$$|L_6| \leq a_{5T} + a_{42T} \left(\sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2}, \tag{A.12}$$

where also $a_{42T} = O_p(1)$ in the required uniform sense. The result of the lemma now follows from $(A.11)$ and $(A.12)$ be defining $a_{4T} = a_{41T} + a_{42T}$. □

The next lemma is concerned with the contribution of $l_{1T}(\Theta, \tau, \Omega) - l_{1,\tau_o+p-1}(\Theta, \tau, \Omega)$ to $l_{1T}(\Theta, \tau, \Omega)$. Here $\varepsilon_{1t\tau}(\Theta)$ is given by

$$\varepsilon_{1t\tau}(\Theta) = -\Psi_2 w_{1t}^{(0)}, \quad t = \tau_o + p, \ldots, T.$$

Lemma A.6. There exists a constant $c_4 > 0$ such that, with probability approaching one and uniformly in $[T\underline{\lambda}] \le \tau \le \tau_o$ and $(\Theta, \tau, \Omega) \in B_1$,

$$l_{1T}(\Theta, \tau, \Omega) - l_{1,\tau_o+p-1}(\Theta, \tau, \Omega) \ge c_4 \|T^{1/2}\Psi_2\|^2 - a_{6T}\|T^{1/2}\Psi_2\|,$$

where $a_{6T} \ge 0$ and $a_{6T} = O_p(1)$.

Proof: The proof is similar to that of Lemma A.3 except for being simpler because the considered quantities are independent of τ and uniformity over this parameter is therefore of no concern. Details are omitted. □

Our next lemma is used as an alternative to Lemma A.4 in some of the subsequent derivations. The formulation of this lemma makes use of the notation $\zeta_{t\tau}^{(0)}$ employed in Lemma A.5.

Lemma A.7. There exists a constant $c_5 > 0$ such that with probability approaching one and uniformly in $[T\underline{\lambda}] \le \tau \le \tau_o - 1$ and $(\Theta, \tau, \Omega) \in B_1$,

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

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

Proof: By the definitions,

$$\begin{aligned}
& l_{1,\tau_o-1}(\Theta, \tau, \Omega) - l_{1,\tau-1}(\Theta, \tau, \Omega) \\
& = \mathrm{tr}\left(\Omega^{-1} \textstyle\sum_{t=\tau}^{\tau_o-1} \varepsilon_{1t\tau}(\Theta)\varepsilon_{1t\tau}(\Theta)'\right) + 2\mathrm{tr}\left(\Omega^{-1} \textstyle\sum_{t=\tau}^{\tau_o-1} \varepsilon_{1t\tau}(\Theta)\varepsilon_{2t}(\Phi_2)'\right) \stackrel{def}{=} L_7 + L_8.
\end{aligned}$$

Recall that $\Psi_1 = \Phi_1 + [\delta_1 : 0]$ and $\Psi_2 = \Phi_1 + [\delta_1 - \delta_1^{(0)} : 0]$. For $t = \tau, \ldots, \tau_o - 1$, we thus have $\varepsilon_{1t\tau}(\Theta) = -\Psi_1 w_{1t}^{(0)} - \underline{\gamma} \underline{d}_{t\tau} = -\Psi_2 w_{1t}^{(0)} - \zeta_{t\tau}^{(0)}$. Hence,

$$\begin{aligned}
L_7 \;\; = \;\; & \mathrm{tr}\left(\Omega^{-1}\Psi_2 \textstyle\sum_{t=\tau}^{\tau_o-1} w_{1t}^{(0)} w_{1t}^{(0)\prime} \Psi_2'\right) + \mathrm{tr}\left(\Omega^{-1} \textstyle\sum_{t=\tau}^{\tau_o-1} \zeta_{t\tau}^{(0)} \zeta_{t\tau}^{(0)\prime}\right) \\
& +2\mathrm{tr}\left(\Omega^{-1} \textstyle\sum_{t=\tau}^{\tau_o-1} \zeta_{t\tau}^{(0)} w_{1t}^{(0)\prime} \Psi_2'\right) \stackrel{def}{=} L_{71} + L_{72} + L_{73}.
\end{aligned}$$

Assume that $(\Theta,\tau,\Omega)\in B_1$. An application of the Cauchy-Schwarz inequality, the norm inequality and the triangle inequality yields

$$\begin{aligned}|L_{73}| &\leq 2\|\Omega^{-1}\|\|T^{1/2}\Psi_2\|\left\|T^{-1/2}\sum_{t=\tau}^{\tau_o-1}\zeta_{t\tau}^{(0)}w_{1t}^{(0)\prime}\right\| \\ &\leq 2\|\Omega^{-1}\|\left(\frac{\tau_o-\tau}{T}\right)^{1/2}\|T^{1/2}\Psi_2\|\left((\tau_o-\tau)^{-1}\sum_{t=\tau}^{\tau_o-1}\|w_{1t}^{(0)}\|^2\right)^{1/2}\left(\sum_{t=\tau}^{\tau_0-1}\|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2}.\end{aligned}$$

Because $\max_{1\leq t\leq T}\|w_{1t}^{(0)}\|=O_p(1)$ (see the arguments leading to $(A.10)$), the second square root in the last expression is of order $O_p(1)$ uniformly in $[T\underline{\lambda}]\leq\tau<\tau_o$. Hence,

$$|L_{73}|\leq a_{8T}\left(\frac{\tau_o-\tau}{T}\right)^{1/2}\|T^{1/2}\Psi_2\|\left(\sum_{t=\tau}^{\tau_o-1}\|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2}, \qquad (A.13)$$

where $a_{8T}=O_p(1)$ in the required uniform sense.

Next note that $L_{71}\geq 0$ and $\lambda_{\min}(\Omega^{-1})\geq\bar{\omega}^{-1}$ for $(\Theta,\tau,\Omega)\in B_1$. Consequently,

$$L_{71}+L_{72}\geq\bar{\omega}^{-1}\sum_{t=\tau}^{\tau_o-1}\|\zeta_{t\tau}^{(0)}\|^2. \qquad (A.14)$$

Now consider L_8 for which we have

$$L_8=-2\mathrm{tr}\left(\Omega^{-1}\Psi_2\sum_{t=\tau}^{\tau_o-1}w_{1t}^{(0)}\varepsilon_{2t}(\Phi_2)'\right)-2\mathrm{tr}\left(\Omega^{-1}\sum_{t=\tau}^{\tau_o-1}\zeta_{t\tau}^{(0)}\varepsilon_{2t}(\Phi_2)'\right)\stackrel{def}{=}L_{81}+L_{82}.$$

Arguments similar to those used for L_{73} show that

$$|L_{81}|\leq 2\|\Omega^{-1}\|\|T^{1/2}\Psi_2\|\left\|T^{-1/2}\sum_{t=\tau}^{\tau_o-1}w_{1t}^{(0)}\varepsilon_{2t}(\Phi_2)'\right\|\leq a_{9T}\|T^{1/2}\Psi_2\|, \qquad (A.15)$$

where $a_{9T}=O_p(1)$ in the required uniform sense. The latter inequality is obtained because, for $(\Theta,\tau,\Omega)\in B_1$, the last norm in the second expression can be replaced by $O_p(1)$ by an analog of (A.4) of Gregory & Hansen (1996, p. 118).

As for L_{82}, assume first that $\tau<\tau_o-p$ and use the Cauchy-Schwarz inequality to conclude that

$$\begin{aligned}|L_{82}| &\leq 2\|\Omega^{-1}\|\left\|\sum_{t=\tau}^{\tau_o-1}\zeta_{t\tau}^{(0)}\varepsilon_{2t}(\Phi_2)'\right\| \\ &\leq 2\|\Omega^{-1}\|\left\|\sum_{t=\tau}^{\tau+p-1}(\delta_1^{(0)}+\underline{\gamma}\underline{d}_{t\tau})\varepsilon_{2t}(\Phi_2)'\right\|+2\|\Omega^{-1}\|\left\|\delta_1^{(0)}\sum_{t=\tau+p}^{\tau_0-1}\varepsilon_{2t}(\Phi_2)'\right\| \\ &\leq 2\|\Omega^{-1}\|\left(\sum_{t=\tau}^{\tau+p-1}\|\delta_1^{(0)}+\underline{\gamma}\underline{d}_{t\tau}\|^2\right)^{1/2}\left(\sum_{t=\tau}^{\tau+p-1}\|\varepsilon_{2t}(\Phi_2)\|^2\right)^{1/2} \\ &\quad +2\|\Omega^{-1}\|\|\delta_1^{(0)}\|\left\|\sum_{t=\tau+p}^{\tau_o-1}\varepsilon_{2t}(\Phi_2)\right\|.\end{aligned}$$

Here the second inequality is based on the definitions and the triangle inequality whereas the third one also makes use of the Cauchy-Schwarz inequality and the norm inequality.

In the last expression

$$(\tau_o - \tau - p)^{-2\eta} \sum_{t=\tau}^{\tau+p-1} \|\varepsilon_{2t}(\Phi_2)\|^2 = O_p(1), \quad \frac{1}{b} < \eta < \frac{1}{4},$$

and

$$(\tau_o - \tau - p)^{-1/2} \left\| \sum_{t=\tau+p}^{\tau_o-1} \varepsilon_{2t}(\Phi_2) \right\| = O_p(1)$$

uniformly in $[T\underline{\lambda}] \leq \tau < \tau_o - p$ and $(\Theta, \tau, \Omega) \in B_1$. Here the latter result can be concluded from the Hájek-Rényi inequality given in Proposition 1 of Bai (1994). The former can be obtained by an argument similar to that used to prove (A.14) of Saikkonen & Lütkepohl (2002).

Combining the above discussion on L_{82} shows that

$$\begin{aligned} |L_{82}| &\leq a_{71T} \left[\left(\sum_{t=\tau}^{\tau+p-1} \|\delta_1^{(0)} + \underline{\gamma} \underline{d}_{t\tau})\|^2 \right)^{1/2} + (\tau_o - \tau - p)^{1/2} \|\delta_1^{(0)}\| \right] \\ &= a_{71T} \left[\left(\sum_{t=\tau}^{\tau+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} + \left(\sum_{t=\tau+p}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} \right], \end{aligned}$$

where $a_{71T} = O_p((\tau_o - \tau)^\eta)$ in the required uniform sense and the equality follows from definitions. Since for any real numbers $a \geq 0$ and $b \geq 0$ we have $a + b \leq \sqrt{2}(a^2 + b^2)^{1/2}$ it follows that

$$|L_{82}| \leq \sqrt{2} a_{71T} \left(\sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2}. \qquad (A.16)$$

In the proof of this result it was assumed that $\tau < \tau_o - p$ but it also holds for $\tau_o - p \leq \tau < \tau_o$. In that case arguments similar to those used for L_{73} give

$$|L_{82}| \leq 2\|\Omega^{-1}\| \left(\sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} \left(\sum_{t=\tau}^{\tau_o-1} \|\varepsilon_{2t}(\Phi_2)\|^2 \right)^{1/2}$$

and $(A.16)$ holds with $a_{71T} = O_p(1)$. The result of the lemma is obtained from the definitions of L_7 and L_8 in conjunction with $(A.13)-(A.16)$ by defining $c_5 = \bar{\omega}^{-1}$, $a_{7T} = \sqrt{2} a_{71T} / (\tau_o - \tau)^\eta$ and a_{8T} and a_{9T} as done in $(A.13)$ and $(A.15)$, respectively. □

In the proof of the next lemma as well as in subsequent proofs, frequent use will be made of the elementary inequality

$$a_2x^2 - a_1x - a_0 \geq -\frac{a_1^2}{4a_2} - a_0, \quad x \geq 0, \tag{A.17}$$

which holds for $a_0, a_1 \geq 0$ and $a_2 > 0$.

Lemma A.8. Let $\epsilon > 0$ and $B_2 = \{(\Theta, \tau, \Omega) : \|T^{1/2-\eta}\Phi_1\|^2 + \|T^{1/2-\eta}\Psi_2\|^2 \leq \epsilon^2\}$, where $\frac{1}{b} < \eta < \frac{1}{4}$ is the same as in Lemma A.7. Then,

$$\inf_{(\Theta,\tau,\Omega)\in B_2^c} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) > 0$$

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

Proof: By the definitions and Lemma A.2,

$$\begin{aligned} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) &= l_{1T}(\Theta, \tau, \Omega) + l_{2T}(\Phi_2, \Omega) - l_{2T}(\Phi_{2o}, \Omega_o) \\ &\geq l_{1T}(\Theta, \tau, \Omega) + \inf_{(\Phi_2,\Omega)} l_{2T}(\Phi_2, \Omega) - l_{2T}(\Phi_{2o}, \Omega_o) \quad (A.18) \\ &= l_{1T}(\Theta, \tau, \Omega) + O_p(1) \end{aligned}$$

Thus, it suffices to show that, for some $\epsilon_* > 0$,

$$\inf_{(\Theta,\tau,\Omega)\in B_2^c} T^{-2\eta} l_{1T}(\Theta, \tau, \Omega) \geq \epsilon_* \tag{A.19}$$

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

From Lemma A.1 it follows that we only need to prove $(A.19)$ with the set B_2^c replaced by $B_1 \cap B_2^c$. Let $0 < \epsilon_1 \leq \lambda_o - \underline{\lambda}$ and define the sets

$$B_{21} = B_1 \cap B_2^c \cap \{(\Theta, \tau, \Omega) : [T\underline{\lambda}] \leq \tau \leq [T(\lambda_o - \epsilon_1)]\}$$

and

$$B_{22} = B_1 \cap B_2^c \cap \{(\Theta, \tau, \Omega) : [T(\lambda_o - \epsilon_1)] < \tau \leq \tau_o\}.$$

According to what was said above, it suffices to establish $(A.19)$ separately with B_2^c replaced by B_{21} and B_{22}. Here we are free to choose the value of ϵ_1. Whatever our choice, Lemma A.4 can be applied on the set B_{21} on which we shall first concentrate.

From Lemmas A.4, A.5 and $(A.17)$ we first find that, uniformly in B_{21},

$$T^{-2\eta}l_{1,\tau_o-1}(\Theta,\tau,\Omega)-T^{-2\eta}l_{1,\tau-1}(\Theta,\tau,\Omega)\geq-(a_{2T}^2+a_{3T}^2)/4c_2T^{2\eta}=o_p(1)$$

and

$$T^{-2\eta}l_{1,\tau_o+p-1}(\Theta,\tau,\Omega)-l_{1,\tau_o-1}(\Theta,\tau,\Omega)\geq-\frac{a_{4T}^2}{4c_3T^{2\eta}}-\frac{a_{5T}}{T^{2\eta}}=o_p(1).$$

Combining these inequalities with those obtained from Lemmas A.3 and A.6 shows that, uniformly in B_{21},

$$\begin{aligned}T^{-2\eta}l_{1T}(\Theta,\tau,\Omega) \;\geq\;& c_1\|T^{1/2-\eta}\Phi_1\|^2-T^{-\eta}a_{1T}\|T^{1/2-\eta}\Phi_1\|\\ &+c_4\|T^{1/2-\eta}\Psi_2\|^2-T^{-\eta}a_{6T}\|T^{1/2-\eta}\Psi_2\|+o_p(1).\end{aligned}\qquad (A.20)$$

Denote $c^*=\min(c_1,c_4)$ and $a_T^*=\sqrt{2}\max(a_{1T},a_{6T})$. Then the preceding inequality implies that, uniformly in B_{21},

$$\begin{aligned}T^{-2\eta}l_{1T}(\Theta,\tau,\Omega) \;\geq\;& c^*(\|T^{1/2-\eta}\Phi_1\|^2+\|T^{1/2-\eta}\Psi_2\|^2)\\ &-\frac{1}{\sqrt{2}}T^{-\eta}a_T^*(\|T^{1/2-\eta}\Phi_1\|+\|T^{1/2-\eta}\Psi_2\|)+o_p(1).\end{aligned}$$

For simplicity, denote $\varphi_T^2=\|T^{1/2-\eta}\Phi_1\|^2+\|T^{1/2-\eta}\Psi_2\|^2$ and note that the sum of the two norms in the last expression above is at most $\sqrt{2}\varphi_T$. Thus, uniformly in B_{21},

$$T^{-2\eta}l_{1T}(\Theta,\tau,\Omega)\geq c^*\varphi_T^2-T^{-\eta}a_T^*\varphi_T+o_p(1)=c^*\varphi_T^2\left(1-\frac{a_T^*}{c^*T^\eta\varphi_T}\right)+o_p(1).\qquad (A.21)$$

Because $\varphi_T>\epsilon$ on B_{21} and $a_T^*=O_p(1)$ uniformly in B_{21}, this shows that $(A.19)$ holds with B_2^c replaced by B_{21}.

Now consider proving $(A.19)$ with B_2^c replaced by B_{22}. Here we can use Lemmas A.3, A.5, A.6 and A.7 to conclude that, with probability approaching one and uniformly in B_{22},

$$\begin{aligned}T^{-2\eta}l_{1,T}(\Theta,\tau,\Omega) \;\geq\;& c_1\|T^{1/2-\eta}\Phi_1\|^2-T^{-\eta}a_{1T}\|T^{1/2-\eta}\Phi_1\|\\ &+c_4\|T^{1/2-\eta}\Psi_2\|^2-T^{-\eta}(a_{6T}+a_{9T})\|T^{1/2-\eta}\Psi_2\|\\ &+c_3T^{-2\eta}\textstyle\sum_{t=\tau_o}^{\tau_o+p+1}\|\zeta_{t\tau}^{(0)}\|^2-T^{-\eta}a_{4T}\left(T^{-2\eta}\sum_{t=\tau_o}^{\tau_o+p+1}\|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2}-T^{-2\eta}a_{5T}\\ &+c_5T^{-2\eta}\textstyle\sum_{t=\tau}^{\tau_o-1}\|\zeta_{t\tau}^{(0)}\|^2\\ &-\left[a_{7T}\left(\frac{\tau_o-\tau}{T}\right)^\eta+a_{8T}\left(\frac{\tau_o-\tau}{T}\right)^{1/2}\|T^{1/2-\eta}\Psi_2\|\right]\left(T^{-2\eta}\textstyle\sum_{t=\tau}^{\tau_o-1}\|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2}.\end{aligned}\qquad (A.22)$$

Here it is understood that a_{9T} and the last two terms on the r.h.s. are deleted if $\tau=\tau_o$ because then Lemma A.7 becomes redundant. By $(A.17)$ the sum of the fifth, sixth and seventh terms on the r.h.s. is of order $o_p(1)$ uniformly in B_{22} and the sum of the last two terms can be bounded from below by $-\frac{1}{4c_5}\left[a_{7T}\left(\frac{\tau_o-\tau}{T}\right)^{\eta}+a_{8T}\left(\frac{\tau_o-\tau}{T}\right)^{1/2}\|T^{1/2-\eta}\Psi_2\|\right]^2$. Thus, expanding the square and inserting the result to the r.h.s. of the preceding inequality yields, uniformly in B_{22},

$$\begin{aligned} T^{-2\eta}l_{1T}(\Theta,\tau,\Omega) \;\geq\; & c_1\|T^{1/2-\eta}\Phi_1\|^2-T^{-\eta}a_{1T}\|T^{1/2-\eta}\Phi_1\| \\ & +c_{4T}(\tau)\|T^{1/2-\eta}\Psi_2\|^2-a_{10T}(\tau)\|T^{1/2-\eta}\Psi_2\|-a_{11T}(\tau)+o_p(1), \end{aligned} \tag{A.23}$$

where

$$c_{4T}(\tau)=c_4-\frac{a_{8T}^2}{4c_5}\left(\frac{\tau_0-\tau}{T}\right),$$

$$a_{10T}(\tau)=T^{-\eta}a_{6T}+T^{-\eta}a_{9T}+\frac{a_{7T}a_{8T}}{2c_5}\left(\frac{\tau_o-\tau}{T}\right)^{1/2+\eta}$$

and

$$a_{11T}(\tau)=\frac{a_{7T}^2}{4c_5}\left(\frac{\tau_o-\tau}{T}\right)^{2\eta}.$$

Note that here $a_{6T},\ldots,a_{9T}$ are of order $O_p(1)$ uniformly in B_{22} and that, on B_{22}, $(\tau_o-\tau)/T\leq 2\epsilon_1$, say. Since we are here free to choose the value of ϵ_1 we can choose it so small that the following two conditions hold with probability approaching one and uniformly in B_{22}: (i) $c_{4T}(\tau)\geq c_4/2$ and (ii) $a_{10T}(\tau)$ and $a_{11T}(\tau)$ become smaller than any preassigned positive number. Taking these facts into account and comparing the inequality $(A.23)$ with $(A.20)$ shows that there are only two points which make the previous proof based on inequality $(A.20)$ directly inapplicable in the present context. These points are that instead of the terms $T^{-\eta}a_{6T}=o_p(1)$ and $o_p(1)$ we have in $(A.23)$ $a_{10T}(\tau)$ and $a_{11T}(\tau)+o_p(1)$, respectively, which are not of order $o_p(1)$ but can only be replaced by an arbitrarily small positive number independent of parameters. However, this is sufficient for the application of essentially the same proof as previously. Indeed, we can conclude that, uniformly in B_{22}, an analog of $(A.21)$ holds except that in the last expression T^{η} is replaced by a fixed positive number which can be assumed as large as we wish and $o_p(1)$ is replaced by a fixed negative number which, in absolute value, can be assumed as small as we wish. In particular, we can assume that T^{η} and $o_p(1)$ in $(A.21)$ are replaced by M/ϵ and $-\epsilon/M$, respectively, where M can be chosen arbitrarily large. This shows that we can make the r.h.s. of the present version of $(A.21)$ larger than some $\epsilon_*>0$ with probability approaching one. Thus, there is a choice of

ϵ_1 such that $(A.19)$ holds with B_2^c replaced by B_{21} and B_{22}. This completes the proof. □

The next lemma is similar to Lemma A.8 except that it deals with the short-run parameter Φ_2.

Lemma A.9. Let $\epsilon > 0$ and $B_3 = \{(\Theta, \tau, \Omega) : \|T^{1/2-\eta}(\Phi_2 - \Phi_{2o})\| \le \epsilon\}$, where $\frac{1}{b} < \eta < \frac{1}{4}$ is the same as in Lemma A.7. Then,

$$\inf_{(\Theta,\tau,\Omega)\in B_3^c} l_T(\Theta, \tau, \Omega) - l_T(\Theta_o, \tau_o, \Omega_o) > 0$$

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

Proof: By Lemma A.1 it suffices to prove the result with B_3^c replaced by $B_1 \cap B_3^c$. First consider the break dates $[T\underline{\lambda}] \le \tau \le [T(\lambda_o - \epsilon_1)]$ and note that the derivation of the inequality in $(A.21)$ is valid for these break dates and for all $(\Theta, \tau, \Omega) \in B_1 \cap B_3^c$. It is also valid for every $\epsilon_1 > 0$. Thus, an application of $(A.17)$ shows that in this part of the parameter space $T^{-2\eta} l_{1T}(\Theta, \tau, \Omega) \ge o_p(1)$ holds uniformly. Next note that the inequality $(A.23)$ is valid for $[T(\lambda_o - \epsilon)] < \tau \le \tau_o$ and for all $(\Theta, \tau, \Omega) \in B_1 \cap B_3^c$. Moreover, as the discussion after that inequality reveals, we can, with a suitable (small) choice of ϵ_1, use $(A.17)$ to obtain an analog of $(A.21)$ from which we conclude that, with probability approaching one and uniformly in the considered part of the parameter space, $T^{-2\eta} l_{1T}(\Theta, \tau, \Omega) \ge -\epsilon_2$, where $\epsilon_2 > 0$ can be chosen arbitrarily small. From the above discussion and the first equality in $(A.18)$ it thus follows that we need to show that, for some $\epsilon_* > 0$,

$$\inf_{(\Theta,\tau,\Omega)\in B_3^c} T^{-2\eta} l_{2T}(\Phi_2, \Omega) - T^{-2\eta} l_{2T}(\Phi_2, \Omega) \ge \epsilon_*$$

with probability approaching one. Arguments needed to show this are similar to those used in previous proofs and also very similar to those used to prove the consistency of the LS estimators of the parameters Φ_2 and Ω in the standard regression model $\Delta x_t = \Phi_2 w_t^{(0)} + \varepsilon_t$. Details are straightforward and are omitted. □

The next lemma again makes use of the notation $\zeta_{t\tau}^{(0)}$ introduced for Lemma A.5.

Lemma A.10. Let $B_4 = \{(\Theta, \tau, \Omega) : (\tau_o - \tau)^{-2\eta} \sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \leq M^2\}$, where $\tau < \tau_o$ and $\frac{1}{b} < \eta < \frac{1}{4}$ is the same as in Lemma A.7. Then, there exists a real number $M_0 > 0$ such that, for all $M \geq M_0$,

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

with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o - 1$. If the quantity $(\tau_o - \tau)^{-2\eta}$ in the definition of the set B_4 is replaced by $T^{-2\eta}$ the same conclusion holds.

Proof: From $(A.18)$ it follows that it suffices to show that there exists a real number $M_0 > 0$ such that, for all $M \geq M_0$ and any $M_1 > 0$,

$$\inf_{(\Theta,\tau,\Omega)\in B_4^c} l_{1T}(\Theta, \tau, \Omega) > M_1 \qquad (A.24)$$

with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o - 1$. From Lemmas A.1, A.8 and A.9 it further follows that here the set B_4^c can be replaced by $B_1 \cap B_2 \cap B_3 \cap B_4^c$. From $(A.19)$ it can be seen that the value of ϵ in the definition of B_2 can be chosen arbitrarily small.

We wish to apply Lemmas A.3, A.5, A.6 and A.7 to obtain a lower bound for $l_{1T}(\Theta, \tau, \Omega)$. This lower bound can be obtained by multiplying both sides of the inequality $(A.22)$ by $T^{2\eta}$. By $(A.17)$ the contribution of the first four terms to the r.h.s. of the resulting inequality can be replaced by $O_p(1)$. This is also the case for the seventh term. Hence, we can write

$$\begin{aligned} l_{1T}(\Theta, \tau, \Omega) \ \geq\ & c_3 \sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 - a_{4T} \left(\sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} + c_5 \sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \\ & - \left(a_{7T}(\tau_o - \tau)^\eta + a_{8T}(\tau_o - \tau)^{1/2}\|\Psi_2\| \right) \left(\sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} + O_p(1). \end{aligned}$$

This holds uniformly in $B_1 \cap B_2 \cap B_3 \cap B_4^c$ and $[T\underline{\lambda}] \leq \tau \leq \tau_o - 1$. In this part of the parameter space we also have

$$(\tau_o - \tau)^{1/2}\|\Psi_2\| = (\tau_o - \tau)^\eta \left(\frac{\tau_o - \tau}{T} \right)^{1/2-\eta} \|T^{1/2-\eta}\Psi_2\| \leq \epsilon(\tau_o - \tau)^\eta$$

and $a_{4T} \leq a_{4T}(\tau_o - \tau)^\eta$ (see Lemma A.8). Denote $c^* = \min(c_3, c_5)$, $a_T^* = \max(a_{4T},\, a_{7T} + \epsilon a_{8T})$ and for simplicity, $\xi_\tau^2 = \sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2$. From the lower bound obtained for $l_{1T}(\Theta, \tau, \Omega)$ above we can then further obtain

$$l_{1T}(\Theta, \tau, \Omega) \geq c^*\xi_\tau^2 - a_T^*(\tau_o - \tau)^\eta \xi_\tau + O_p(1) = c^*\xi_\tau^2 \left(1 - \frac{a_T^*(\tau_o - \tau)^\eta}{c^*\xi_\tau} \right) + O_p(1). \qquad (A.25)$$

Again, this holds uniformly in $B_1 \cap B_2 \cap B_3 \cap B_4^c$ and $[T\underline{\lambda}] \leq \tau \leq \tau_o - 1$. Now, on B_4^c, $\xi_\tau > M(\tau_o - \tau)^\eta$ so that, for all M large enough and with probability approaching one, we can make the r.h.s. of $(A.25)$ larger than any preassigned number $M_1 > 0$. Thus, we have established $(A.24)$ and thereby the first assertion of the lemma. The second assertion is obvious by $(A.25)$ and the discussion thereafter. $\square$

Before proceeding to new proofs we discuss how Lemmas A.3-A.10 are formulated when $\tau \geq \tau_o$.

The counterpart of Lemma A.3 is concerned with the time points $t = p+1, \ldots, \tau_o - 1$ and break dates $\tau_o \leq \tau \leq [T\bar{\lambda}]$ but is otherwise similar to Lemma A.3 and its proof is similar to the proof of Lemma A.6 in that uniformity in τ is of no concern.

The next time points of interest are now $t = \tau_o, \ldots, \tau_o + p - 1$ so that we need to consider a counterpart of Lemma A.5. Here we write

$$\begin{aligned} \varepsilon_{1t\tau}(\Theta) &= -\Phi_1 w_{1t}^{(0)} - d_{t\tau_o}(\delta_1 - \delta_1^{(0)}) - (d_{t\tau} - d_{t\tau_o})\delta_1 - \underline{\gamma}\underline{d}_{t\tau} + \underline{\gamma}^{(0)}\underline{d}_{t\tau_o} \\ &= -\Psi_2 w_{1t}^{(0)} - \zeta_{t\tau}, \qquad t = \tau_o, \ldots, \tau_o + p - 1, \end{aligned}$$

where $\Psi_2 = \Phi_1 + [\delta_1 - \delta_1^{(0)} : 0]$ as before and $\zeta_{t\tau} = (d_{t\tau} - d_{t\tau_o})\delta_1 + \underline{\gamma}\underline{d}_{t\tau} - \underline{\gamma}^{(0)}\underline{d}_{t\tau_o}$. In other words, in place of $\zeta_{t\tau}^{(0)}$ we now use an analogous variable defined by using the parameter δ_1 instead of $\delta_1^{(0)}$. However, replacing $\zeta_{t\tau}^{(0)}$ in Lemma A.5 by $\zeta_{t\tau}$ is clearly possible, as can be seen from the given proof.

Instead of the time points $t = \tau_o + p, \ldots, \tau - 1$ it is next reasonable to consider the time points $t = \tau_o + p, \ldots, \tau + p - 1$. Then the number of time points is the same as in Lemmas A.4 and A.7. Changes in parameters have to be made, though. Now

$$\varepsilon_{1t\tau}(\Theta) = -\Phi_1 w_{1t}^{(0)} + d_{t\tau_o}\delta_1^{(0)} - d_{t\tau}\delta_1 - \underline{\gamma}\underline{d}_{t\tau} = -\Psi_1^{(0)} w_{1t}^{(0)} - (d_{t\tau}\delta_1 + \underline{\gamma}\underline{d}_{t\tau}), \quad t = \tau_o + p, \ldots, \tau + p - 1,$$

where $\Psi_1^{(0)} = \Phi_1 - [\delta_1^{(0)} : 0]$. Thus, we now have the matrix $\Psi_1^{(0)}$ in place of Ψ_1 used in Lemma A.4 and, as above, the former is defined by using $\delta_1^{(0)}$ instead of δ_1 in Ψ_1. The parameter $\underline{\gamma}$ used in Lemma A.4 is also changed by adding δ_1 to its columns. With these replacements the counterpart of Lemma A.4 applies with $[T(\lambda_o + \epsilon)] \leq \tau \leq [T\bar{\lambda}]$.

Next consider the counterpart of Lemma A.7 which is also concerned with time points $t = \tau_o + p, \ldots, \tau_o + p - 1$. Here the preceding expression of $\varepsilon_{1t\tau}(\Theta)$ is modified to the form

$$\varepsilon_{1t\tau}(\Theta) = -\Psi_2 w_{1t}^{(0)} - \zeta_{t\tau}, \quad t = \tau_o + p, \ldots, \tau + p - 1,$$

where Ψ_2 is as defined in the proof of Lemma A.7. In the counterpart of Lemma A.7 we then have $\zeta_{t\tau}$ in place of $\zeta_{t\tau}^{(0)}$ and $\tau_o + 1 \leq \tau \leq [T\bar{\lambda}]$. The proof can again be basically obtained by following the previous proof.

The counterpart of Lemma A.6 is straightforward. The relevant time points are $t = \tau, \ldots, T$ and the obtained lower bound is as before except for the obvious change in the values of τ which become $\tau_o \leq \tau \leq [T\bar{\lambda}]$. The proof is also changed and becomes similar to the proof of Lemma A.3.

It is not difficult to check that the modified versions of Lemmas A.3 - A.7 can be used to show that the results of Lemmas A.8 and A.9 also apply for $\tau_o \leq \tau \leq [T\bar{\lambda}]$. Regarding Lemma A.10, when $\tau_o + 1 \leq \tau \leq [T\bar{\lambda}]$, the set B_4 is defined as

$$B_4 = \left\{ (\Theta, \tau, \Omega) : (\tau_o - \tau)^{-2\eta} \sum_{t=\tau_o}^{\tau+p-1} \|\zeta_{t\tau}\|^2 \leq M^2 \right\}$$

but otherwise the same result obtains.

Now we can turn to our next lemma which is central in studying asymptotic properties of the break date estimator. Recall that $\delta_{1o} = -\Pi_o \delta_o = -\alpha_o \beta_o' \delta_o$, where $\delta_o = T^a \delta_*$. Thus, $\delta_{1o} \neq 0$ if and only if ${\beta_o}' \delta_o \neq 0$. Note also that we shall use the convention that the infimum over an empty set is ∞.

Lemma A.11. Let $M > 0$. Assume that $\delta_{1o} \neq 0$ and define $B_5 = \{(\Theta, \tau, \Omega) : (|\tau_o - \tau| - p)\|\delta_{1o}\|^{2/(1-2\eta)} \leq M\}$, where $\frac{1}{b} < \eta < \frac{1}{4}$ is the same as in Lemma A.7 or its counterpart when $\tau > \tau_o$. Then there exists a real number $M_0 > 0$ such that, for all $M \geq M_0$,

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

with probability approaching one. If $\delta_{1o} = 0$ the same result holds with the set B_5 replaced by $B_{50} = \{(\Theta, \tau, \Omega) : T^{-2\eta} \sum_{t=p+1}^{T} \|\underline{\gamma} \underline{d}_{t\tau} - \underline{\gamma}_o \underline{d}_{t\tau_o}\|^2 \leq M\}$.

Proof: Assume first that $\tau < \tau_o - p$ and $\delta_{1o} \neq 0$. From Lemmas A.1, A.8 and A.9 it follows that we can replace the set B_5^c by $B_1 \cap B_2 \cap B_3 \cap B_5^c$.

By the definitions, $\delta_1^{(0)} = -\Pi \delta_o = -\alpha^{(0)} \beta_o' \delta_o - \rho^{(0)} \beta_{o\perp}' \delta_o$, where $\beta_o' \delta_o \neq 0$. On B_3, $\|\alpha^{(0)} - \alpha_o\| \leq \epsilon T^{\eta - 1/2}$ and, on B_2, $\|\rho^{(0)}\| \leq \epsilon T^{\eta-1}$ (see Lemmas A.8 and A.9). Thus, since

$\delta_{1o} = -\alpha_o \beta_o' \delta_o$ and $\delta_o = T^a \delta_*$,

$$\begin{aligned}
\|\delta_1^{(0)} - \delta_{1o}\| &\leq \|\alpha^{(0)} - \alpha_o\| \|\beta_o'\delta_o\| + \|\rho^{(0)}\| \|\beta_{o\perp}'\delta_o\| \\
&\leq T^a \|\delta_*\| \epsilon (\|\beta_o\| T^{\eta-1/2} + \|\beta_{o\perp}\| T^{\eta-1}) \\
&\leq cT^{\eta+a-1/2}\epsilon
\end{aligned}$$

for some positive and finite constant c. Hence, because $\zeta_{t\tau}^{(0)} = (d_{t\tau} - d_{t\tau_o})\delta_1^{(0)} = \delta_1^{(0)}$ for $t = \tau + p, \ldots, \tau_o - 1$, we have on $B_1 \cap B_2 \cap B_3 \cap B_5^c$,

$$\begin{aligned}
\left((\tau_o - \tau)^{-2\eta} \sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2} &\geq \left((\tau_o - \tau)^{-2\eta} \sum_{t=\tau+p}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2} \\
&= (\tau_o - \tau)^{-\eta} (\tau_o - \tau - p)^{1/2} \|\delta_1^{(0)}\| \\
&= \left(1 - \tfrac{p}{\tau_o - \tau}\right)^{\eta} (\tau_o - \tau - p)^{1/2-\eta} \|\delta_1^{(0)}\| \\
&\geq \left(\tfrac{1}{p+1}\right)^{\eta} \left((\tau_o - \tau - p)\|\delta_{1o}\|^{2/(1-2\eta)}\right)^{1/2-\eta} \left(1 - \tfrac{\|\delta_1^{(0)} - \delta_{1o}\|}{\|\delta_{1o}\|}\right) \\
&\geq \tfrac{M^{1/2-\eta}}{(p+1)^{\eta}} \left(1 - \tfrac{c\epsilon T^{\eta-1/2}}{\|\alpha_o \beta_o' \delta_*\|}\right).
\end{aligned} \tag{A.26}$$

Here the fourth relation makes use of the triangle inequality. For all T and M large enough the last expression can be made larger than the real number M_0 in Lemma A.10. Thus, the stated result follows from Lemma A.10.

Now consider the case $\tau > \tau_o + p$ but maintain the assumption $\delta_{1o} \neq 0$. Then, using the counterparts of Lemmas A.8 and A.9 we can proceed in the same way as in the case $\tau < \tau_o - p$ until the relations $(A.26)$ which start now as

$$\left((\tau - \tau_o)^{-2\eta} \sum_{t=\tau_o}^{\tau+p-1} \|\zeta_{t\tau}\|^2\right)^{1/2} \geq \left((\tau - \tau_o)^{-2\eta} \sum_{t=\tau_o+p}^{\tau-1} \|\zeta_{t\tau}\|^2\right)^{1/2} = (\tau - \tau_o)^{-\eta} (\tau - \tau_o - p)^{1/2} \|\delta_1\|.$$

Thus, in place of $\delta_1^{(0)}$ we have now δ_1. However, from the counterpart of Lemma A.8 we find that, on B_2, $\|\delta_1 - \delta_1^{(0)}\| \leq \epsilon T^{\eta-1/2}$ and a straightforward modification of the arguments in the latter part of $(A.26)$ combined with the present version of Lemma A.10 give the desired result.

Next assume that $\delta_{1o} = 0$ and $\tau \leq \tau_o$. In this case we use the inequality

$$T^{-2\eta} \sum_{t=\tau}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 \geq T^{-2\eta} \sum_{t=\tau}^{*\ \tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2, \tag{A.27}$$

where the summation on the r.h.s. is over the values of t for which $\Delta d_{t\tau_o} \neq 0$ or $\Delta d_{t\tau} \neq 0$. Clearly the number of such time points is at most $2p$.

From the definitions it follows that

$$\begin{aligned}\zeta_{t\tau}^{(0)} &= (d_{t\tau}-d_{t\tau_o})\delta_1^{(0)}+\Delta d_{t\tau_o}\delta_1^{(0)}+\sum_{j=1}^{p-1}\Delta d_{t-j,\tau_o}(\Gamma_j-\Gamma_{jo})\delta_o\\ &\quad+\sum_{j=0}^{p-1}\Delta d_{t-j,\tau}\gamma_j-\Delta d_{t\tau_o}\delta_o+\sum_{j=1}^{p-1}\Delta d_{t-j,\tau_o}\Gamma_{jo}\delta_o.\end{aligned}$$

Notice that here $\Gamma_{jo}\delta_o=-\gamma_{jo}$ ($j=1,\dots,p-1$) and, since now $\delta_{1o}=0$, $\delta_o=\gamma_{0o}$. Thus, the sum of the last three terms equals $\underline{\gamma}\underline{d}_{t\tau}-\underline{\gamma}_o\underline{d}_{t\tau_o}$ and we wish to show that the contribution of the first three terms to the r.h.s. of (A.27) can be ignored. To this end, note that now $\delta_1^{(0)}=-\rho^{(0)}\beta_{o\perp}'\delta_o$ so that, on B_2, $\|\delta_1^{(0)}\|\le c\epsilon T^{\eta+a-1}$ for some $0\le c<\infty$ (see Lemma A.8). Furthermore, on B_3, $\|(\Gamma_j-\Gamma_{jo})\delta_o\|\le\|\Gamma_j-\Gamma_{jo}\|\|\delta_o\|\le\epsilon T^{\eta+a-1/2}\|\delta_*\|$ ($j=1,\dots,p-1$) (see Lemma A.9). Using these facts and the triangle inequality we find that

$$\left(T^{-2\eta}{\sum_{t=\tau}^{\tau_o+p-1}}^{*}\|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2}\ge\left(T^{-2\eta}{\sum_{t=\tau}^{\tau_o+p-1}}^{*}\|\underline{\gamma}\underline{d}_{t\tau}-\underline{\gamma}_o\underline{d}_{t\tau_o}\|^2\right)^{1/2}-const\times T^{a-1/2}\epsilon.$$

On the r.h.s. the summation can be extended to all $t=p+1,\dots,T$. This means that on $B_1\cap B_2\cap B_3\cap B_{50}^c$ the last expression becomes larger than the real number M_0 in Lemma A.10 for all T and M large enough. Thus, the stated result follows from the latter part of Lemma A.10.

Finally, assume that $\delta_{1o}=0$ and $\tau>\tau_o$. In place of (A.27) we then have a similar inequality with $t=\tau_o,\dots,\tau+p-1$ and $\zeta_{t\tau}^{(0)}$ replaced by $\zeta_{t\tau}$. However, using the fact that $\|\delta_1-\delta_1^{(0)}\|\le\epsilon T^{1/2-\eta}$ on B_2 it is straightforward to show that the proof can be reduced to a form entirely similar to that in the case $\tau\le\tau_o$. This completes the proof of the lemma. □

Now we can prove Theorem 3.1. As discussed earlier, the estimator $\hat\tau$ can also be obtained by minimizing -2 times the Gaussian log-likelihood function $l_T(\Theta,\tau,\Omega)$. First consider the case $a>0$ and $\delta_{1o}\neq 0$. By Lemma A.11 we can then concentrate on the break dates $\tau_o-p\le\tau\le\tau_o+p$. First consider the case $\tau_o-p\le\tau\le\tau_o$. If $\gamma_{jo}=0$ for all $j=0,\dots,p-1$, Lemma A.11 shows that, asymptotically, $\tau_o-p\le\hat\tau\le\tau_o$, as required. Next suppose that $\gamma_{j0,o}\neq 0$ and consider the break dates $\tau_o-p\le\tau\le\tau_o-p+j_0$. For any of these break dates we have

$$\zeta_{t\tau}^{(0)}=-\sum_{j=j_0}^{p-1}\gamma_j^{(0)}\Delta d_{t-j,\tau_o},\quad t=\tau_o+j_0,\dots,\tau_o+p-1.$$

Suppose first that $j_0 > 0$. Then, since $\gamma_{j_0}^{(0)} - \gamma_{j_0,o} = -(\Gamma_{j_0} - \Gamma_{j_0,o})\delta_o$, we have for $(\Theta, \tau, \Omega) \in B_3$,

$$\begin{aligned}\left((\tau_o - \tau)^{-2\eta} \sum_{t=\tau}^{\tau_o - p+1} \|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2} &\geq (\tau_o - \tau)^{-\eta}\|\gamma_{j_0}^{(0)}\| \\ &\geq \|\gamma_{j_0,o}\| - \epsilon T^{\eta-1/2}\|\delta_o\| \\ &= T^a(\|\Gamma_{j_0,o}\delta_*\| - \epsilon T^{\eta-1/2}\|\delta_*\|).\end{aligned}$$

Because $\gamma_{j_0,o} = -T^a\Gamma_{j_0,o}\delta_* \neq 0$, the last quantity tends to infinity as $T \to \infty$. Hence, we can conclude from Lemmas A.9 and A.10 that asymptotically the function $l_T(\Theta, \tau, \Omega)$ is not minimized for $\tau \leq \tau_o - p + j_0$. Now consider the case $j_0 = 0$. From the definitions it follows that $\gamma_o^{(0)} - \gamma_{0o} = \delta_{1o} - \delta_1^{(0)}$, where $\|\delta_{1o} - \delta_1^{(0)}\| \leq cT^{\eta+a-1/2}\epsilon$ on $B_2 \cap B_3$ (see the beginning of the proof of Lemma A.11). Hence, since $\gamma_{0o} = T^a(\delta_* + \alpha_o\beta_o'\delta_*) \neq 0$, the proof given in the case $j_0 > 0$ applies with obvious changes and shows that asymptotically $\hat{\tau} \leq \tau_o - p$ cannot occur.

To complete the proof of the first assertion, consider the case $\tau_o + 1 \leq \tau \leq \tau_o + p$. By the definitions we then have $\zeta_{\tau_o\tau} = -\delta_1 - \gamma_0^{(0)} = -\delta_o + (\delta_1^{(0)} - \delta_1)$, where $\delta_o \neq 0$ and $\|\delta_1^{(0)} - \delta_1\| \leq \epsilon T^{\eta-1/2}$ for $(\Theta, \tau, \Omega,) \in B_2 \cap B_3$ (see Lemma A.8 and the definition of Ψ_2 given before Lemma A.5). In the same way as in the preceding case we can thus conclude from Lemmas A.8, A.9 and A.10 that asymptotically $\hat{\tau} > \tau_o$ cannot occur. This completes the proof of the first assertion in the case $a > 0$ and $\delta_{1o} \neq 0$.

Next assume that $a > \eta > 1/b$ and $\delta_{1o} = 0$. Then, if $\tau \leq \tau_o - p + j_0$ and $j_0 > 0$,

$$T^{-2\eta} \sum_{t=p+1}^{T} \|\underline{\gamma}\underline{d}_{t\tau} - \underline{\gamma}_o\underline{d}_{t\tau_o}\|^2 \geq T^{-2\eta}\|\gamma_{j_0,o}\|^2 = T^{2a-2\eta}\|\Gamma_{j_0,o}\delta_*\|^2. \tag{A.28}$$

Because the last quantity tends to infinity as $T \to \infty$ it follows from the latter part of Lemma A.11 that asymptotically $\hat{\tau} \leq \tau_o - p + j_0$ cannot occur. If $j_0 = 0$, we have $\gamma_{0o} = \delta_o - \delta_{1o} = \delta_o \neq 0$ and $(A.28)$ holds with $\Gamma_{j_0,o}\delta_*$ replaced by δ_*. Hence the same conclusion also obtains for $j_0 = 0$.

If $\tau > \tau_o$ the l.h.s. of $(A.28)$ can be bounded from below by $T^{-2\eta}\|\gamma_{0o}\|^2 = T^{-2\eta}\|\delta_o\|^2 = T^{2a-2\eta}\|\delta_*\|^2$, and the situation is similar to the case $j_0 = 0$ above.

Finally, the second part of the theorem follows directly from the first part of Lemma A.11. This completes the proof of Theorem 3.1.

A.2 Proof of Theorem 3.2

The break date estimator $\hat{\tau}_R$ can also be obtained by minimizing the objective function $l_T(\Theta, \tau, \Omega)$ over the relevant restricted part of the parameter space. Compared to the previous unrestricted estimation the parameters δ_1 and $\underline{\gamma}$ in $(A.2)$ are no more freely varying but (smooth) functions of the parameters δ, $\rho^{(0)}$, $\alpha^{(0)}$ and $\Gamma_1, \ldots, \Gamma_{p-1}$. Specifically, $\delta_1 = -\Pi\delta = -\alpha^{(0)}\beta_o'\delta - \rho^{(0)}\beta_{o\perp}'\delta$, $\gamma_0 = \delta - \delta_1$ and $\gamma_j = -\Gamma_j\delta$ $(j = 1, \ldots, p-1)$. Unlike with the unconstrained estimation it is not quite obvious that these restricted estimators exist. This fact will therefore be justified first. After that the proof follows straightforwardly from the results used to prove Theorem 3.1.

Define

$$y_t^{(\tau)} = x_t - (d_{t\tau} - d_{t\tau_o})\delta_o. \tag{A.29}$$

Using $y_t^{(\tau)}$ in place of x_t we can obtain an analog of $(A.2)$ in which $d_{t\tau_o}$ and $\underline{d}_{t\tau_o}$ are replaced by $d_{t\tau}$ and $\underline{d}_{t\tau}$, respectively, and $u_{t-1}^{(0)}$ and $v_{t-1}^{(0)}$ are replaced by analogs defined in terms of $y_t^{(\tau)}$ instead of x_t. In other words, in place of $u_{t-1}^{(0)}$ and $v_{t-1}^{(0)}$ we use $u_{t-1}^{(\tau)} = \beta_o' y_{t-1}^{(\tau)}$ and $v_{t-1}^{(\tau)} = \beta_{o\perp}' y_{t-1}^{(\tau)}$, respectively. In place of $(A.3)$ we then have

$$\Delta y_t^{(\tau)} = \Phi w_t^{(\tau)} + (\Xi - \Xi^{(0)})q_{t\tau} + \varepsilon_t, \quad t = p+1, p+2, \ldots,$$

where $w_t^{(\tau)}$ is an obvious modification of $w_t^{(0)}$.

Clearly, we can express the vector $\varepsilon_{t\tau}(\Theta)$ as

$$\varepsilon_{t\tau}(\Theta) = \Delta y_t^{(\tau)} - \Phi w_t^{(\tau)} - (\Xi - \Xi^{(0)})q_{t\tau}$$

and use this expression in the previous definition of $l_T(\Theta, \tau, \Omega)$. To demonstrate the existence of a minimizer of the objective function $l_T(\Theta, \tau, \Omega)$ it also appears convenient to use the reparameterization $\Theta \to \Theta^{(0)} = [\Phi : \Xi - \Xi^{(0)}]$. Thus, if for simplicity we denote $\underline{z}_t^{(\tau)} = [\Delta y_t^{(\tau)\prime} : w_t^{(\tau)\prime} : q_{t\tau}']'$ and $R(\Theta^{(0)}) = [I_n : -\Phi : \Xi - \Xi^{(0)}]$ we can write the relevant objective function as

$$l_T(\Theta^{(0)}, \tau, \Omega) = (T-p)\log\det\Omega + \operatorname{tr}\left(\Omega^{-1}R(\Theta^{(0)})\sum_{t=p+1}^{T}\underline{z}_t^{(\tau)}\underline{z}_t^{(\tau)\prime}R(\Theta^{(0)})'\right). \tag{A.30}$$

Note that in the present context the parameter Θ has the same meaning as before except that it is treated as a (smooth) function of the parameters $\nu_0^{(0)}$, $\nu_1^{(0)}$, δ, $\rho^{(0)}$, $\alpha^{(0)}$ and $\Gamma_1, \ldots, \Gamma_{p-1}$. Because the parameter $\Xi^{(0)}$ is also a (smooth) function of (some of) these parameters the same is true for the parameter $\Theta^{(0)}$. All these parameter restrictions are taken into account when the minimization of the objective function $l_T(\Theta^{(0)}, \tau, \Omega)$ is considered. Notice that, because the objective function is expressed as a function of the "reduced form" parameter $\Theta^{(0)}$, the role of the parameter restrictions is to define the permissible space of $\Theta^{(0)}$. A similar idea, of course, applies to the previous parameterization of the objective function, that is, to $l_T(\Theta, \tau, \Omega)$ (cf. Saikkonen (2001) and the references therein for a similar approach).

A useful consequence of the fact that we can still interpret the objective function $l_T(\Theta, \tau, \Omega,)$ as a function of the "reduced form" parameter Θ and only restrict its permissible space is that results obtained to prove Theorem 3.1 can be applied straightforwardly even here. In particular, we wish to apply Lemma A.11 to conclude that, when the existence of a minimizer of the objective function $l_T(\Theta, \tau, \Omega)$ is studied in the present setup, values of the break date parameter τ can be restricted as implied by this lemma. Of course, this conclusion also holds when the objective function is parameterized as $l_T(\Theta^{(0)}, \tau, \Omega)$.

To justify the application of Lemma A.11, we first discuss how Lemmas A.1 - A.10 have to be modified to match the present setup. Notice that the existence of a minimizer of the objective function $l_T(\Theta, \tau, \Omega)$ is not needed to prove Lemmas A.1 - A.11 and the same is also true for their modified versions to be discussed below.

First note that Lemma A.2 is still used in its previous form and, because it is concerned with unrestricted values of Φ_2 and Ω, it obviously applies in the present context. Lemma A.1 is simply modified by replacing B_1 by the intersection of the restricted parameter space of (Θ, τ, Ω) and values for which the inequality constraints in $(A.4)$ and $(A.5)$ hold. This restricted version of the parameter space B_1 is then used to replace B_1 in Lemmas A.3 - A.7. It is straightforward to check that the previous proofs of these Lemmas apply in essence despite the differences in parameter spaces.

Next consider Lemmas A.8 - A.10, where, in addition to B_1, also the parameter spaces B_2, B_3 and B_4 are redefined to allow for the employed restrictions. Again, it is not difficult to check that the previous proofs carry over. It is also easy to see that the modifications needed for Lemmas A.3 - A.10 can be done in the case $\tau \geq \tau_o$.

Because analogs of Lemmas A.1 - A.10 hold in the present context, it is further straightforward to show that the result of Lemma A.11 also holds with the parameter space B_5 redefined to account for the employed restrictions. Thus, we can conclude that when searching for a minimizer of the objective function $l_T(\Theta^{(0)}, \tau, \Omega)$, the value of the break date parameter τ can be restricted as implied by Lemma A.11. Specifically, if $\delta_{1o} \neq 0$, Lemma A.11 directly shows that $\tau_o - p \leq \tau \leq \tau_o + p$ can be assumed. If $\delta_{1o} = 0$ and $a > b$, we can even assume $\tau_o - p + 1 \leq \tau \leq \tau_o + p - 1$, as the argument used to prove the corresponding case of Theorem 3.1(i) readily shows.

We shall now show that the function $l_T(\Theta^{(0)}, \tau, \Omega)$ and hence $l_T(\Theta, \tau, \Omega)$ has a minimizer with probability approaching one. In what follows, reference to Lemmas A.1 - A.11 will be understood to mean the present restricted setup. We first show the following intermediate result, where the matrix $D_T = \text{diag}[T^{-1/2}I : I_p]$ is used. Its dimension equals the dimension of the vector $\underline{z}_t^{(\tau)}$.

Lemma A.12. There exists an $\epsilon_* > 0$ such that

$$\lambda_{\min}\left(D_T^{-1}\sum_{t=p+1}^{T}\underline{z}_t^{(\tau)}\underline{z}_t^{(\tau)\prime}D_T^{-1}\right) \geq \epsilon_* \qquad (A.31)$$

with probability approaching one and uniformly in τ, when the value of the break date parameter τ can be restricted as implied by Lemma A.11.

Proof: The values of τ can be restricted depending on the value of a and whether $\delta_{1o} = 0$ or not. Different cases will therefore be discussed separately.

Case (i): $a > 0$ **and** $\delta_{1o} \neq 0$ **or** $a > \eta > 1/b$

From Lemma A.11 we can then conclude that, if a minimizer of $l_T(\Theta^{(0)}, \tau, \Omega)$ exists, in large samples it must be such that the corresponding τ is in the interval $[\tau_o - p, \tau_o + p]$. If $\delta_{1o} \neq 0$ this follows directly from the first part of Lemma A.11. If $\delta_{1o} = 0$ (and $a > \eta > 1/b$) the same conclusion can be drawn from the second part of the lemma by the argument used in the proof of Theorem 3.1 to obtain $(A.28)$.

To justify $(A.31)$, assume first that $\underline{a < \frac{1}{2}}$. Then the moment matrix in $(A.31)$ behaves asymptotically in the same way as in the proof of Theorem 3.1 in that the vectors $\Delta y_t^{(\tau)}$ and $w_t^{(\tau)}$ in the definition of $\underline{z}_t^{(\tau)}$ can be replaced by analogs defined in terms of x_t. This follows by observing that, when $|\tau_o - \tau| \leq p$, the latter term on the r.h.s. of $(A.29)$ satisfies

$$\left\|T^{-1}\sum_{t=p+1}^{T}(d_{t\tau} - d_{t\tau_o})^2\delta_o\delta_o'\right\| \leq const \times T^{2a-1}\sum_{t=p+1}^{T}|d_{t\tau} - d_{t\tau_o}| \leq const \times T^{2a-1}. \qquad (A.32)$$

When $a < 1/2$ the last quantity converges to zero and the desired conclusion is readily obtained.

If $a = \frac{1}{2}$ the latter term on the r.h.s. of $(A.29)$ has an impact but $(A.31)$ still obtains. To see this, suppose first that $\delta_{1o} \neq 0$. Then, as $|\tau - \tau_o| \leq p$, the latter term on the r.h.s. of $(A.29)$ behaves like an impulse dummy. Because now $\delta_o = T^{1/2}\delta_*$ this term affects the asymptotic behavior of the moment matrix in $(A.31)$ but, as can be readily seen, it only affects the diagonal and off-diagonal elements related to $u_{t-1}^{(\tau)}$ and $\Delta y_{t-j}^{(\tau)}$ $(j = 0, \ldots, p-1)$. Moreover, the impact is such that asymptotically the moment matrix in $(A.31)$ only differs from that obtained in the previous case by an additive positive semidefinite matrix. Thus, from this fact and the result of the previous case one again obtains $(A.31)$.

Next assume that $\delta_{1o} = 0$ and $a = \frac{1}{2}$. Here the situation is similar to the preceding case except for being simpler because now $u_{t-1}^{(\tau)} = \beta_o' x_{t-1} = u_{t-1}^{(0)}$. Thus, we again get $(A.31)$ and, thus, we have justified $(A.31)$ in the case of the first part of the theorem. It remains to consider the second part for which the following assumption is made.

Case (ii): $a \leq 0$ **and** $\delta_{1o} \neq 0$

If $a = 0$ it follows from the first part of Lemma A.11 that we can assume $|\tau - \tau_o|$ to be bounded and arguments similar to those in the case $0 < a < 1/2$ and $\delta_{1o} \neq 0$ show $(A.31)$. If $a < 0$ we cannot restrict the values of τ. However, from $(A.32)$ it can be seen that the vectors $\Delta y_t^{(\tau)}$ and $w_t^{(\tau)}$ in the definition of $\underline{z}_t^{(\tau)}$ can be replaced by analogs defined in terms of x_t. Arguments similar to those used in the proof of Theorem 3.1 then show that $(A.31)$ also holds in the present case. (In particular, analogs of (A.3) and (A.4) of Gregory & Hansen (1996) and (A.14) of Saikkonen & Lütkepohl (2002) can be used to handle sums of cross products between $[\Delta x_t' : w_t^{(0)'}]'$ and $q_{t\tau}$.) □

We have now shown that when searching for a minimizer of the function $l_T(\Theta^{(0)}, \tau, \Omega)$ we can in both parts of Theorem 3.2 restrict the values of the break date τ in such a way that $(A.31)$ holds with probability approaching one and uniformly in τ.

Using Lemma A.12 we can analyze the function $l_T(\Theta^{(0)}, \tau, \Omega)$ in the same way as in the proof of Proposition 3.1 of Saikkonen (2001, pp. 320-321) and conclude that it suffices to search for a minimizer of $l_T(\Theta^{(0)}, \tau, \Omega)$ in that part of the parameter space where, in addition to the restrictions on τ, we also have $0 < \underline{\omega} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq \bar{\omega} < \infty$ and $\|\Theta^{(0)}\| \leq \bar{M} < \infty$.

We shall demonstrate that the parameter space defined by all these restrictions is compact. To this end, note first that the restrictions imposed on $\Theta^{(0)}$ are of the form $h(\Theta^{(0)}) = 0$, where $h(\cdot)$ is a continuous function. Thus, because the unrestricted parameter space of $\Theta^{(0)}$ is the whole Euclidean space, it follows that the restricted space is closed and its intersection with parameter values restricted by $0 < \underline{\omega} \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq \bar{\omega} < \infty$ and $\|\Theta^{(0)}\| \leq \bar{M} < \infty$ is compact. The continuity of the function $l_T(\cdot, \tau, \cdot)$ therefore ensures that, for every relevant value of τ, a minimizer exists with probability approaching one. This proves the (asymptotic) existence of the nonlinear LS estimators of $\Theta^{(0)}$, τ, Ω and hence also that of Θ.

To prove part (i) of Theorem 3.2, first consider the case $\delta_{1o} \neq 0$ and assume that $\tau \leq \tau_o - 1$. As noticed above, we can also assume that $\tau_o - p \leq \tau$. Using the definitions we can express the vector $\zeta_{t\tau}^{(0)}$ as

$$\begin{aligned}
\zeta_{t\tau}^{(0)} &= \delta_1^{(0)} d_{t\tau} + \delta \Delta d_{t\tau} - \delta_1 \Delta d_{t\tau} - \sum_{j=1}^{p-1} \Gamma_j \delta \Delta d_{t-j,\tau} - \delta_1^{(0)} d_{t\tau_o} - \delta_o \Delta d_{t\tau_o} \\
&\quad + \delta_1^{(0)} \Delta d_{t\tau_o} + \sum_{j=1}^{p-1} \Gamma_j \delta_o \Delta d_{t-j,\tau_o}.
\end{aligned}$$

Taking the assumed restrictions into account we can write this further as

$$\begin{aligned}
\zeta_{t\tau}^{(0)} &= -(\delta_1 - \delta_1^{(0)}) d_{t\tau} + \left(\Delta d_{t\tau} - \textstyle\sum_{j=1}^{p-1} \Gamma_j \Delta d_{t-j,\tau} - (\alpha^{(0)} \beta_o' + \rho^{(0)} \beta_{o\perp}') d_{t-1,\tau} \right) \delta \\
&\quad - \left(\Delta d_{t\tau_o} - \textstyle\sum_{j=1}^{p-1} \Gamma_j \Delta d_{t-j,\tau_o} - (\alpha^{(0)} \beta_o' + \rho^{(0)} \beta_{o\perp}') d_{t-1,\tau_o} \right) \delta_o
\end{aligned} \qquad (A.33)$$

Here we have also made use of the facts that $\delta_1 = -\Pi\delta$ and $\delta_1^{(0)} = -\Pi\delta_o$ with $\Pi = \alpha^{(0)}\beta_o' + \rho^{(0)}\beta_{o\perp}'$.

To show that asymptotically the function $l_T(\Theta, \tau, \Omega)$ cannot be minimized for $\tau_o - p \leq \tau \leq \tau_o - 1$, we consider two cases separately. In the first case it is assumed that $\delta \geq T^a \epsilon_*$, where $\epsilon_* > 0$ is arbitrary. The second case will then assume that $\delta < T^a \epsilon_*$.

Now consider parameter values for which $\tau_o - p \leq \tau \leq \tau_o - 1$ and $\delta \geq T^a \epsilon_*$ hold for some $\epsilon_* > 0$. By Lemma A.8 we can also assume that $\|\delta_1 - \delta_1^{(0)}\| \leq \epsilon T^{\eta - 1/2}$. Using this, $(A.33)$

and the above mentioned parameter restrictions, we find that

$$\begin{aligned}\left((\tau_o-\tau)^{-2\eta}\sum_{t=\tau}^{\tau_o+p-1}\|\zeta_{t\tau}^{(0)}\|^2\right)^{1/2} &\geq p^{-\eta}\|\zeta_{\tau\tau}^{(0)}\| \\ &= p^{-\eta}\|\delta-(\delta_1-\delta_1^{(0)})\| \\ &\geq p^{-\eta}(\|\delta\|-\|\delta_1-\delta_1^{(0)}\|) \\ &\geq p^{-\eta}T^a\|\epsilon_*\|-\epsilon p^{-\eta}T^{\eta-1/2}.\end{aligned}$$

Because the last quantity tends to infinity with T, it follows from Lemma A.10 that asymptotically $\tau_o-p\leq\hat{\tau}_R\leq\tau_o-1$ cannot occur.

For parameter values $\tau_o-p\leq\tau\leq\tau_o-1$ and $\delta<T^a\epsilon_*$ we can also use $(A.33)$ and Lemma A.10. First note that, by Lemma A.8, the norm of the first term on the r.h.s. of (A.33) can be bounded by $\epsilon T^{\eta-1/2}$. Next, from Lemmas A.8 and A.9 it follows that the term in front of δ in the second term on the r.h.s. of $(A.33)$ can be assumed bounded and so the norm of the whole term can be bounded by a quantity of the form $c_1\epsilon_*T^a$, where $0<c_1<\infty$. Similar arguments can also be used to show that, at least for $t=\tau_o$, the norm of the third term on the r.h.s. of $(A.33)$ can be bounded from below by a quantity of the form $c_2\|\delta_*\|T^a$, where $0<c_2<\infty$ and $\|\delta_*\|\neq 0$. Thus, since ϵ_* can be chosen arbitrarily small, the asymptotic behavior of $(\tau_o-\tau)^{-2\eta}\sum_{t=\tau}^{\tau_o+p-1}\|\zeta_{t\tau}^{(0)}\|^2\geq p^{-2\eta}\sum_{t=\tau}^{\tau_o+p-1}\|\zeta_{t\tau}^{(0)}\|^2$ is dominated by the third term on the r.h.s. of $(A.33)$ and the preceding discussion implies that this sum tends to infinity with T. From this and Lemma A.10 we can conclude that asymptotically $\tau_o-p\leq\hat{\tau}_R\leq\tau_o-1$ cannot occur.

Thus, we have shown that, when $\delta_{1o}\neq 0$, asymptotically $\hat{\tau}_R<\tau_0$ cannot occur. A similar argument with $\zeta_{t\tau}^{(0)}$ replaced by $\zeta_{t\tau}$ and with Lemma A.10 replaced by its corresponding counterpart shows that asymptotically $\hat{\tau}_R>\tau_o$ cannot occur either.

Now suppose that $\delta_{1o}=0$ and consider the break dates $\tau_o-p\leq\tau\leq\tau_o-1$. Instead of $(A.33)$ we use a slightly different representation of $\zeta_{t\tau}^{(0)}$ given by

$$\begin{aligned}\zeta_{t\tau}^{(0)} &= -\Delta d_{t\tau}(\delta_1-\delta_1^{(0)})-d_{t-1,\tau_o}\delta_1^{(0)}+d_{t-1,\tau}\delta_1^{(0)}+\textstyle\sum_{j=1}^{p-1}\Delta d_{t-j,\tau_o}(\Gamma_j-\Gamma_{jo})\delta_o \\ &\quad+\left(\Delta d_{t\tau}-\textstyle\sum_{j=1}^{p-1}\Gamma_j\Delta d_{t-j,\tau}\right)\delta-\left(\Delta d_{t\tau_o}-\textstyle\sum_{j=1}^{p-1}\Gamma_{jo}\Delta d_{t-j,\tau_o}\right)\delta_o.\end{aligned} \qquad (A.33a)$$

This representation can be obtained from the definitions (cf. the similar representation used in the proof of Lemma A.11). As with the case $\delta_{1o}\neq 0$, our treatment will be divided into two separate cases.

In the first one the parameter δ is restricted as $\delta \geq T^a \epsilon_*$, where $\epsilon_* > 0$ is arbitrary and $a > \eta > 1/b$. From the preceding representation of $\zeta_{t\tau}^{(0)}$ it then follows that

$$\begin{aligned}
\left(T^{-2\eta} \sum_{t=p+1}^{T} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} &\geq T^{-\eta} \|\zeta_{\tau\tau}^{(0)}\| \\
&= T^{-\eta} \|\delta - (\delta_1 - \delta_1^{(0)})\| \\
&\geq T^{-\eta} (\|\delta\| - \|\delta_1 - \delta_1^{(0)}\|) \\
&\geq T^{a-\eta} \|\epsilon_*\| - \epsilon T^{-1/2}.
\end{aligned}$$

Here the last inequality makes use of the fact that $\|\delta_1 - \delta_1^{(0)}\| \leq \epsilon T^{\eta - 1/2}$ can be assumed by Lemma A.8. Because the last quantity tends to infinity with T, it follows from the latter result of Lemma A.10 that asymptotically $\tau_o - p \leq \hat{\tau}_R \leq \tau_o - 1$ cannot occur.

When $\delta < T^a \epsilon_*$ $(a > \eta > 1/b)$ is assumed, $(A.33a)$ and Lemma A.11 give the desired result much in the same way as in the case $\delta_{1o} \neq 0$, where $(A.33)$ was used instead of $(A.33a)$. First note that the norm of the first four terms on the r.h.s. of $(A.33a)$ can be bounded by a quantity of the form $\epsilon c T^{\eta + a - 1/2}$, where $0 < c < \infty$. This follows from Lemma A.8 and arguments used to prove Lemma A.11 for $\delta_{1o} = 0$. Next, in the same way as in the case $\delta_{1o} \neq 0$ one can show that the term in front of δ in the fifth term on the r.h.s. of $(A.33a)$ can be assumed bounded and, hence, the norm of the whole term can be bounded by a quantity of the form $c_1 \epsilon_* T^a$, where $0 < c_1 < \infty$. By similar arguments we finally find that, at least for $t = \tau_o$, the norm of the last term on the r.h.s. of $(A.33a)$ can be bounded below by a quantity of the form $c_2 \|\delta_*\| T^a$, where $0 < c_2 < \infty$ and $\|\delta_*\| \neq 0$. Thus, since ϵ_* can be chosen arbitrarily small, the asymptotic behavior of $\left(T^{-2\eta} \sum_{t=p+1}^{T} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2}$ is dominated by the last term on the r.h.s. of $(A.33a)$ and it follows from the latter result of Lemma A.10 that asymptotically $\tau_o - p \leq \hat{\tau}_R \leq \tau_o - 1$ cannot occur.

Thus, we have shown that, when $\delta_{1o} = 0$, we asymptotically cannot have $\hat{\tau}_R < \tau_o$. Again a similar proof with $\zeta_{t\tau}^{(0)}$ replaced by $\zeta_{t\tau}$ and Lemma A.10 replaced by its corresponding counterpart shows that asymptotically $\hat{\tau}_R > \tau_o$ cannot occur either. This completes the proof of part (i) of the theorem in the case $\delta_{1o} = 0$. Part (ii) is a consequence of the (asymptotic) existence of $\hat{\tau}_R$ and Lemma A.11. This completes the proof of Theorem 3.2.

A.3 Proof of Theorem 3.3

The estimator $\tilde{\tau}$ can be obtained by minimizing the Gaussian likelihood function $l_T(\Theta, \tau, \Omega)$ subject to the restriction $\underline{\gamma} = 0$. Thus, we can consider minimizing the objective function

$$l_T^*(\Theta_1, \tau, \Omega) = (T-p)\log\det\Omega + \operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{t\tau}^*(\Theta_1)\varepsilon_{t\tau}^*(\Theta_1)'\right),$$

where $\Theta_1 = [\Phi : \delta_1]$ and $\varepsilon_{t\tau}^*(\Theta_1) = \Delta x_t - \Phi w_t^{(0)} - \delta_1 d_{t\tau} + \delta_1^{(0)} d_{t\tau_o} + \underline{\gamma}^{(0)}\underline{d}_{t\tau_o}$. We can express $\varepsilon_{t\tau}^*(\Theta_1)$ as

$$\varepsilon_{t\tau}^*(\Theta_1) = \varepsilon_{1t\tau}^*(\Theta_1) + \varepsilon_{2t}(\Phi_2),$$

where $\varepsilon_{1t\tau}^*(\Theta_1) = -\Phi_1 w_{1t}^{(0)} - \delta_1 d_{t\tau} + \delta_1^{(0)} d_{t\tau_o} + \underline{\gamma}^{(0)}\underline{d}_{t\tau_o}$ and $\varepsilon_{2t}(\Phi_2)$ is as defined in Section A.1. Analogously to $l_T(\Theta, \tau, \Omega)$ we can decompose $l_T^*(\Theta_1, \tau, \Omega)$ as

$$l_T^*(\Theta_1, \tau, \Omega) = l_{1T}^*(\Theta_1, \tau, \Omega) + l_{2T}(\Phi_2, \Omega),$$

where

$$l_{1T}^*(\Theta_1, \tau, \Omega) = \operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{1t\tau}^*(\Theta_1)\varepsilon_{1t\tau}^*(\Theta_1)'\right) + 2\operatorname{tr}\left(\Omega^{-1}\sum_{t=p+1}^{T}\varepsilon_{1t\tau}^*(\Theta_1)\varepsilon_{2t}(\Phi_2)'\right)$$

and, as before,

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

We shall now discuss modifications of Lemmas A.1 - A.10 based on the objective function $l_T^*(\Theta_1, \tau, \Omega)$. Some of them are minor and are therefore only briefly mentioned. For Lemmas A.5 and A.7 - A.10, new statements of the results will be provided which will be furnished with a superscript '*' to indicate the correspondence to the previous results. We will refer to all the modified versions of the lemmas by attaching a superscript '*' to the number even if no explicit statement of the result is presented.

The result of Lemma A.1 holds with $l_T(\Theta, \tau, \Omega)$ replaced by $l_T^*(\Theta_1, \tau, \Omega)$ and the set B_1 redefined by replacing the parameter Θ by Θ_1. In subsequent discussions this redefinition of B_1 will be denoted by B_1^*. To see that this modification of Lemma A.1 holds, notice that from the proof of that lemma it can be seen that we first need to show that $T^{-1}l_T^*(\Theta_{1o}, \tau_o, \Omega_o) = O_p(1)$. However, because $T^{-1}l_T^*(\Theta_{1o}, \tau_o, \Omega_o)$ differs from the expression in the middle of $(A.6)$ only in that ε_t is replaced by $\varepsilon_t^* = \varepsilon_t + \underline{\gamma}_o\underline{d}_{t\tau_o}$, this follows by straightforward application of the weak law of large numbers and the assumption $\underline{\gamma}_o = O(T^a)$, $a \leq 1/2$. The proof can be completed by repeating the latter part of the proof of Lemma A.1 because therein only time points are involved for which $\varepsilon_{t\tau}(\Theta) = \varepsilon^*_{t\tau}(\Theta_1)$ holds.

Because Lemma A.2 is concerned with $l_{2T}(\Phi_2, \Omega)$, it can be used as before, whereas the result of Lemma A.3 holds with $l_{1,\tau-1}(\Theta, \tau, \Omega)$ replaced by $l^*_{1,\tau-1}(\Theta_1, \tau, \Omega)$ and the set B_1 replaced by B^*_1. This latter fact is obvious because $l_{1,\tau-1}(\Theta, \tau, \Omega) = l^*_{1,\tau-1}(\Theta_1, \tau, \Omega)$ when $\tau \leq \tau_o$.

The result of Lemma A.4 holds with the inequality replaced by

$$l^*_{1,\tau_o-1}(\Theta_1, \tau, \Omega) - l^*_{1,\tau-1}(\Theta_1, \tau, \Omega) \geq c_2 \|T^{1/2}\Psi_1\|^2 - a_{2T}\|T^{1/2}\Psi_1\|$$

and the set B_1 replaced by B^*_1. (Here c_2 and a_{2T} have the same properties as in Lemma A.4.) A proof of this result is obtained by following the proof of Lemma A.4 with the restriction $\underline{\gamma} = 0$ imposed. Because this means that only the upper left hand corner of the matrix on the l.h.s. of $(A.10)$ needs to be analyzed and the term L_{42} can be ignored, the desired result readily follows.

The result of Lemma A.6 applies with $l_{1T}(\Theta, \tau, \Omega) - l_{1,\tau_o+p-1}(\Theta, \tau, \Omega)$ replaced with $l^*_{1T}(\Theta_1, \tau, \Omega) - l^*_{1,\tau_o+p-1}(\Theta_1, \tau, \Omega)$ and the set B_1 replaced by B^*_1. This is obvious because the two differences have identical values.

Because more substantial modifications are required for the remaining lemmas, we formulate new versions of them. For an analog of Lemma A.5 we introduce the notation

$$\zeta^{(*)}_{t\tau} = (d_{t\tau} - d_{t\tau_o})\delta^{(0)}_1 - \underline{\gamma}^{(0)}\underline{d}_{t\tau_o}.$$

Because $\Psi_2 = \Phi_1 + [\delta_1 - \delta^{(0)}_1 : 0]$, we have

$$\varepsilon^*_{1t\tau}(\Theta_1) = -\Psi_2 w^{(0)}_{1t} - \zeta^{(*)}_{t\tau}, \quad t = \tau_o, \ldots, \tau_o + p - 1.$$

Lemma A.5*. There exists a constant $c_3 > 0$ such that, with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o$ and $(\Theta_1, \tau, \Omega) \in B^*_1$,

$$l^*_{1,\tau_o+p-1}(\Theta_1, \tau, \Omega) - l^*_{1,\tau_o-1}(\Theta_1, \tau, \Omega) \geq c_3 \sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta^{(*)}_{t\tau}\|^2 - a_{4T}\left(\sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta^{(*)}_{t\tau}\|^2\right)^{1/2} - a_{5T},$$

where a_{4T} and a_{5T} are as in Lemma A.5.

Proof: In the same way as in the proof of Lemma A.5 we can write

$$l^*_{1,\tau_o+p-1}(\Theta_1,\tau,\Omega) - l^*_{1,\tau_o-1}(\Theta_1,\tau,\Omega) = L^*_5 + L^*_6,$$

where L^*_5 and L^*_6 are as L_5 and L_6, respectively, except for being defined by using $\varepsilon^*_{1t\tau}(\Theta_1)$ instead of $\varepsilon_{1t\tau}(\Theta)$. The stated result follows because the analysis given for L_5 and L_6 in the proof of Lemma A.5 applies also here with $\zeta^{(0)}_{t\tau}$ replaced by $\zeta^{(*)}_{t\tau}$. □

Lemma A.7*. There exists a constant $c_5 > 0$ such that, with probability approaching one and uniformly in $[T\underline{\lambda}] \le \tau \le \tau_o - 1$ and $(\Theta_1,\tau,\Omega) \in B^*_1$,

$$\begin{aligned}
&l^*_{1,\tau_o-1}(\Theta_1,\tau,\Omega) - l^*_{1,\tau-1}(\Theta_1,\tau,\Omega) \\
&\ge c_5 \textstyle\sum_{t=\tau}^{\tau_o-1} \|\zeta^{(*)}_{t\tau}\|^2 - \left(a_{7T} + a_{8T}\left(\frac{\tau_o-\tau}{T}\right)^{1/2} \|T^{1/2}\Psi_2\|\right)\left(\sum_{t=\tau}^{\tau_o-1} \|\zeta^{(*)}_{t\tau}\|^2\right)^{1/2} \\
&\quad - a_{9T}\|T^{1/2}\Psi_2\|,
\end{aligned}$$

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

Proof: In the same way as in the proof of Lemma A.7 we can write,

$$l^*_{1,\tau_o-1}(\Theta_1,\tau,\Omega) - l^*_{1,\tau-1}(\Theta_1,\tau,\Omega) = L^*_7 + L^*_8,$$

where L^*_7 and L^*_8 differ from L_7 and L_8, respectively, only in that $\varepsilon^*_{1t\tau}(\Theta_1)$ is used in place of $\varepsilon_{1t\tau}(\Theta)$. By the definitions we can also write $\varepsilon^*_{1t\tau}(\Theta_1) = -\Psi_2 w^{(0)}_{1t} - \zeta^{(*)}_{t\tau}$, $t = \tau, \ldots, \tau_o - 1$, (cf. the corresponding representation of $\varepsilon_{1t\tau}(\Theta)$). An inspection of the proof of Lemma A.7 reveals that the analysis given for L_7 applies to L^*_7 if only the quantity $\zeta^{(0)}_{t\tau}$ is replaced by $\zeta^{(*)}_{t\tau}$. Thus, we can write $L^*_7 = L^*_{71} + L^*_{72} + L^*_{73}$, where $|L^*_{73}|$ and $L^*_{71} + L^*_{72}$ satisfy inequality $(A.13)$ and $(A.14)$, respectively, except that $\zeta^{(0)}_{t\tau}$ on the r.h.s. is replaced by $\zeta^{(*)}_{t\tau}$.

Regarding L^*_8 we can write $L^*_8 = L_{81} + L^*_{82}$, where L_{81} satisfies $(A.15)$. Thus, we can conclude from the preceding discussion that it only remains to show that

$$|L^*_{82}| \le a_{7T}\left(\sum_{t=\tau}^{\tau_o-1} \|\zeta^{(*)}_{t\tau}\|^2\right)^{1/2},$$

where a_{7T} is as stated in Lemma A.7*. To see this notice that $\zeta_{t\tau}^{(*)} = \delta_1^{(0)}$, $t = \tau, \ldots, \tau_o - 1$, and therefore (cf. the analysis given for L_{82})

$$
\begin{aligned}
|L_{82}^*| &\leq 2\|\Omega^{-1}\| \left\| \sum_{t=\tau}^{\tau_o-1} \zeta_{t\tau}^{(*)} \varepsilon_{2t}(\Phi_2)' \right\| \\
&\leq 2\|\Omega^{-1}\| \|\delta_1^{(0)}\| \left\| \sum_{t=\tau}^{\tau_o-1} \varepsilon_{2t}(\Phi_2) \right\| \\
&= 2\|\Omega^{-1}\| \left(\sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(*)}\|^2 \right)^{1/2} \left\| (\tau_o - \tau)^{-1/2} \sum_{t=\tau}^{\tau_o-1} \varepsilon_{2t}(\Phi_2) \right\|.
\end{aligned}
$$

This gives the desired result because the last norm is of order $O_p(1)$ uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o - 1$, as discussed in the proof of Lemma A.7. □

Because the quantity η does not appear in the results of Lemmas A.4* and A.7*, the subsequent modifications of Lemmas A.8, A.9 and A.10 can be generalized accordingly.

Lemma A.8*. Let $\epsilon > 0$ and $B_2^* = \{(\Theta_1, \tau, \Omega) : \|T^{(1/2)-\eta}\Phi_1\|^2 + \|T^{(1/2)-\eta}\Psi_2\|^2 \leq \epsilon^2\}$, where $0 < \eta < \frac{1}{2}$. Then,

$$
\inf_{(\Theta_1,\tau,\Omega)\in B_2^{*c}} l_T^*(\Theta_1, \tau, \Omega) - l_T^*(\Theta_{1o}, \tau_o, \Omega_o) > 0
$$

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

Proof: We can follow the proof of Lemma A.8. First note that in place of $(A.18)$ we now have

$$
\begin{aligned}
l_T^*(\Theta_1, \tau, \Omega) - l_T^*(\Theta_{1o}, \tau_o, \Omega_o) &= l_{1T}^*(\Theta_1, \tau, \Omega) - l_{1T}^*(\Theta_{1o}, \tau_o, \Omega_o) + l_{2T}(\Phi_2, \Omega) - l_2(\Phi_{2o}, \Omega_o) \\
&\geq l_{1T}^*(\Theta_1, \tau, \Omega) - l_{1T}^*(\Theta_{1o}, \tau_o, \Omega_o) + O_p(1),
\end{aligned}
$$

where the inequality is obtained from Lemma A.2 in the same way as in $(A.18)$. Here $l_{1T}^*(\Theta_{1o}, \tau_o, \Omega_o)$ is not zero. By the definition of $\varepsilon_{1t\tau_o}^*(\Theta_1)$ we have $\varepsilon_{1t\tau_o}^*(\Theta_{1o}) = \underline{\gamma}_o \underline{d}_{t\tau_o}$ and, hence,

$$
\begin{aligned}
l_{1T}^*(\Theta_{1o}, \tau_o, \Omega_o) &= \mathrm{tr}\left(\Omega_o^{-1} \underline{\gamma}_o \sum_{t=\tau_o}^{\tau_o+p-1} \underline{d}_{t\tau_o} \underline{d}_{t\tau_o}' \underline{\gamma}_o' \right) + 2\mathrm{tr}\left(\Omega_o^{-1} \underline{\gamma}_o \sum_{t=\tau_o}^{\tau_o+p-1} \varepsilon_t' \right) \\
&= \mathrm{tr}\left(\Omega_o^{-1} \underline{\gamma}_o \underline{\gamma}_o' \right) + 2\mathrm{tr}\left(\Omega_o^{-1} \underline{\gamma}_o \sum_{t=\tau_o}^{\tau_o+p-1} \varepsilon_t' \right) \\
&\geq \bar{\omega}^{-1} \|\underline{\gamma}_o\|^2 - a_{*T} \|\underline{\gamma}_o\|,
\end{aligned}
$$

where $\bar{\omega}^{-1} > 0$ and $a_{*T} = O_p(1)$. Here the inequality can be justified by arguments similar to those used to analyze L_5 and L_6 in the proof of Lemma A.5. By $(A.17)$ the last quantity can be bounded from below by $-a_{*T}^2\bar{\omega}/4 = O_p(1)$ and it follows that

$$l_T^*(\Theta_1, \tau, \Omega) - l_T^*(\Theta_{1o}, \tau_o, \Omega_o) \geq l_{1T}^*(\Theta_1, \tau, \Omega) + O_p(1).$$

Thus, we can conclude that it suffices to show that, for some $\epsilon_* > 0$,

$$\inf_{(\Theta_1,\tau,\Omega)\in B_2^{*c}} T^{-2\eta} l_{1T}^*(\Theta_1, \tau, \Omega) \geq \epsilon_* \tag{A.34}$$

with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o$. From Lemma A.1* it then follows that $(A.34)$ can be established with B_2^{*c} replaced by $B_1^* \cap B_2^{*c}$. This in turn can be done separately for the sets B_{21}^* and B_{22}^* defined by modifying the definitions of B_{21} and B_{22}, respectively, in an obvious way.

Consider the l.h.s. of $(A.34)$ with B_2^{*c} replaced by B_{21}^*. From Lemmas A.3*, A.4*, A.5*, A.6* and the result in $(A.17)$ we then find that $T^{-2\eta}l_{1T}^*(\Theta_1, \tau, \Omega)$ has a lower bound exactly of the same form as obtained for $T^{-2\eta}l_{1T}(\Theta, \tau, \Omega)$ in $(A.20)$. Then an analog of $(A.21)$ also holds and we can proceed in the same way as in the proof of Lemma A.8 and show that $(A.34)$ holds when B_2^{*c} is replaced by B_{21}^*.

Now consider proving $(A.34)$ with B_2^{*c} replaced by B_{22}^*. Here we can obtain an analog of $(A.22)$ by using Lemmas A.3*, A.5*, A.6* and A.7*. This yields a lower bound for $T^{-2\eta}l_{1T}^*(\Theta_1, \tau, \Omega)$ which differs from that obtained for $T^{-2\eta}l_{1T}(\Theta, \tau, \Omega)$ in $(A.22)$ only in that (i) $\zeta_{t\tau}^{(0)}$ is replaced by $\zeta_{t\tau}^{(*)}$ and (ii) $a_{7T}\left(\frac{\tau_o-\tau}{T}\right)^\eta$ is replaced by $a_{7T}T^{-\eta}$. This implies that we can proceed in the same way as after $(A.22)$ and obtain the following analog of $(A.23)$:

$$\begin{aligned} T^{-2\eta}l_{1T}^*(\Theta_1, \tau, \Omega) \quad \geq \quad & c_1\|T^{(1/2)-\eta}\Phi_1\|^2 - T^{-\eta}a_{1T}\|T^{(1/2)-\eta}\Phi_1\| \\ & + c_{4T}(\tau)\|T^{(1/2)-\eta}\Psi_2\|^2 - a_{10T}^*(\tau)\|T^{(1/2)-\eta}\Psi_2\| - a_{11T}^* + o_p(1), \end{aligned}$$

where $c_{4T}(\tau)$ is as in $(A.23)$,

$$a_{10T}^*(\tau) = T^{-\eta}a_{6T} + T^{-\eta}a_{9T} + T^{-\eta}\frac{a_{7T}a_{8T}}{2c_5}\left(\frac{\tau_o - \tau}{T}\right)^{1/2}$$

and

$$a_{11T}^* = \frac{a_{7T}^2}{T^{-2\eta}4c_5}.$$

Thus, it follows that we have, with probability approaching one and uniformly in B_{22}^*, $c_{4T}(\tau) \geq c_4/2$, $a_{10T}^*(\tau) = o_p(1)$ and $a_{11T}^* = o_p(1)$, and the situation becomes exactly the same as in the case of the set B^*_{21}. This completes the proof of Lemma A.8*. □

Lemma A.9*. Let $\epsilon > 0$ and $B^*_3 = \{(\Theta_1, \tau, \Omega) : \|T^{(1/2)-\eta}(\Phi_2 - \Phi_{2o})\| \leq \epsilon\}$, where η is the same as in Lemma A.8*. Then,

$$\inf_{(\Theta_1,\tau,\Omega)\in B^{*c}_3} l^*_T(\Theta_1, \tau, \Omega) - l^*_T(\Theta_{1o}, \tau_o, \Omega_o) > 0$$

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

Proof: The proof can be obtained in the same way as that of Lemma A.9. Instead of using results obtained in the proof of Lemma A.8 we, of course, use corresponding results discussed in the proof of Lemma A.8*. Details are omitted. □

Lemma A.10*. (i) Assume that $0 < a \leq 1/2$, let $\epsilon > 0$ and $B^*_4 = \{(\Theta_1, \tau, \Omega) : (\tau_o - \tau)^{(1/2)-\eta}\|\delta^{(0)}_1\| \leq T^a\epsilon\}$, where $\tau < \tau_o$ and η is the same as in Lemma A.8*. Then,

$$\inf_{(\Theta_1,\tau,\Omega)\in B^{*c}_4} l^*_T(\Theta_1, \tau, \Omega) - l^*_T(\Theta_{1o}, \tau_o, \Omega_o) > 0$$

with probability approaching one and uniformly in $[T\underline{\lambda}] \leq \tau \leq \tau_o - 1$.
(ii) Assume that $a \leq 0$ and let $B^*_{40} = \{(\Theta_1, \tau, \Omega) : (\tau_o - \tau)\|\delta_{1o}\|^{2/(1-2\eta)} \leq M\}$, where $\tau < \tau_o$ and η is the same as in Lemma A.8*. Then, there exists a real number M_0 such that, for all $M \geq M_0$,

$$\inf_{(\Theta_1,\tau,\Omega)\in B^{*c}_{40}} l^*_T(\Theta_1, \tau, \Omega) - l^*_T(\Theta_{1o}, \tau_o, \Omega_o) > 0$$

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

Proof: (i) In the same way as in the proof of Lemma A.10 it can first be seen that it suffices to show that $l^*_{1T}(\Theta_1, \tau, \Omega)$ satisfies an analog of $(A.24)$ (instead of $(A.18)$ we use its analog discussed in the proof of Lemma A.8*). This in turn can be done by using Lemmas A.3*, A.5*, A.6* and A.7* to obtain a lower bound for $l^*_{1T}(\Theta_1, \tau, \Omega)$. This lower bound divided by $T^{2\eta}$ was discussed in the proof of Lemma A.8*. From that discussion it can be concluded that

$$l_{1T}^*(\Theta_1,\tau,\Omega) \;\geq\; c_3\sum_{t=\tau_o}^{\tau_o+p-1}\|\zeta_{t\tau}^{(*)}\|^2 - a_{4T}\left(\sum_{t=\tau_o}^{\tau_o+p-1}\|\zeta_{t\tau}^{(*)}\|^2\right)^{1/2} + c_5\sum_{t=\tau}^{\tau_o-1}\|\zeta_{t\tau}^{(*)}\|^2$$
$$-\left(a_{7T}+a_{8T}(\tau_o-\tau)^{1/2}\|\Psi_2\|\right)\left(\sum_{t=\tau}^{\tau_o-1}\|\zeta_{t\tau}^{(*)}\|^2\right)^{1/2} + O_p(1).$$

This holds uniformly in $B_1^*\cap B_2^*\cap B_3^*\cap B_4^{*c}$ and $[T\underline{\lambda}]\leq\tau\leq\tau_o-1$. By $(A.17)$, the first two terms on the r.h.s. can be bounded from below by $-a_{4T}^2/4c_3$ and in what follows they will be absorbed in the term $O_p(1)$. The term $(\tau_o-\tau)^{1/2}\|\Psi_2\|$ can be bounded from above by $\epsilon_1(\tau_o-\tau)^{\eta}$, where $\epsilon_1>0$ (cf. the proof of Lemma A.10). Taking these facts into account and observing that $\zeta_{t\tau}^{(*)}=\delta_1^{(0)}$ for $t=\tau,\ldots,\tau_o-1$, we can write the preceding inequality as

$$\begin{aligned} l_{1T}^*(\Theta_1,\tau,\Omega) &\geq c_5(\tau_o-\tau)\|\delta_1^{(0)}\|^2 - a_T^*(\tau_o-\tau)^{\eta+1/2}\|\delta_1^{(0)}\| + O_p(1) \\ &= (\tau_o-\tau)\|\delta_1^{(0)}\|^2\left(1-\frac{a_T^*(\tau_o-\tau)^{\eta-1/2}}{c_5\|\delta_1^{(0)}\|}\right)+O_p(1). \end{aligned} \tag{A.35}$$

Here $a_T^*=O_p(1)$, which as well as the whole result, holds uniformly in $B_1^*\cap B_2^*\cap B_3^*\cap B_4^{*c}$ (cf. $(A.25)$). On B_4^{*c} we have $\|\delta_1^{(0)}\|>(\tau_o-\tau)^{\eta-1/2}T^a\epsilon$. Thus, with probability approaching one, the last expression in $(A.35)$ exceeds any preassigned real number. This proves the desired result.

(ii) Here the proof is obtained by following the previous proof of $(A.35)$. The rest of the proof is then similar to that of Lemma A.10. □

The results of the adjusted Lemmas can be modified in a straightforward way to the case $\tau\geq\tau_o$. Using Lemma A.10* and its modification we can prove Theorem 3.3.

Suppose that $0<a\leq 1/2$ and $\tau<\tau_o$. Then, in the same way as in $(A.26)$, we find that

$$\begin{aligned} (\tau_o-\tau)^{(1/2)-\eta}\|\delta_1^{(0)}\| &\geq (\tau_o-\tau)^{(1/2)-\eta}\|\delta_{1o}\|\left(1-\frac{\|\delta_1^{(0)}-\delta_{1o}\|}{\|\delta_{1o}\|}\right) \\ &\geq (\tau_o-\tau)^{(1/2)-\eta}T^a\|\alpha_o\beta_o'\delta_*\|\left(1-\frac{\epsilon c T^{\eta-1/2}}{\|\alpha_o\beta_o'\delta_*\|}\right), \end{aligned} \tag{A.36}$$

where $0<c<\infty$. As $\tau_o-\tau\geq 1$ the last expression tends to infinity and Lemma A.10*(i) and the fact that $\tilde{\tau}$ exists (for all T larger than some constant) implies that asymptotically $\tilde{\tau}<\tau_o$ cannot occur. Because a similar analysis can be carried out in the case $\tau<\tau_o$ the proof of the first assertion of Theorem 3.3 is complete.

To prove the second assertion, write $(\tau_o-\tau)^{(1/2)-\eta}\|\delta_{1o}\|$ in the second expression of $(A.36)$ as $\big((\tau_o-\tau)\|\delta_{1o}\|^{2/(1-2\eta)}\big)^{(1/2)-\eta}$ and proceed in the same way as in $(A.26)$. The desired result is then obtained from Lemma A.10*(ii). $\square$

A.4 Proof of Lemma 4.1

Because we do not need values from $\mathcal{T}$ other than the estimated and the true value in the following, we denote the latter by τ and hence drop the subscript 'o' for convenience. For simplicity we will denote the break date estimator by $\hat{\tau}$. This estimator can be any one of the three estimators considered in Section 3 unless explicit distinctions are made. Moreover, we will make reference to Saikkonen & Lütkepohl (2000) several times and therefore abbreviate that article by S&L.

Properties of RR estimators

We shall first show that the RR estimators of the parameters based on equation (2.7) with the unknown break date τ replaced by the estimator $\hat{\tau}$ satisfy appropriate consistency properties. This replacement changes the VECM (2.7) to

$$\Delta y_t = \nu + \alpha(\beta' y_{t-1} - \phi(t-1) - \theta d_{t-1,\hat{\tau}}) + \sum_{j=1}^{p-1}\Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1}\gamma_j^* \Delta d_{t-j,\hat{\tau}} + \varepsilon_{t\hat{\tau}}, \qquad (A.37)$$

where

$$\varepsilon_{t\hat{\tau}} = \varepsilon_t + \alpha_o\beta_o'\delta_o(d_{t-1,\hat{\tau}} - d_{t-1,\tau}) - \sum_{j=0}^{p-1}\gamma_{jo}^*(\Delta d_{t-j,\hat{\tau}} - \Delta d_{t-j,\tau}). \qquad (A.38)$$

Write

$$y_t = \mu_{0o} + \mu_{1o}t + \delta_o d_{t\hat{\tau}} + y_{t\hat{\tau}}, \qquad (A.39)$$

where $y_{t\hat{\tau}} = x_t - \delta_o(d_{t\hat{\tau}} - d_{t\tau})$. Using the transformation $y_t \to \mu_{0o} + \mu_{1o}t + \delta_o d_{t\hat{\tau}} + y_{t\hat{\tau}}$ we can transform the preceding VECM to the form

$$\Delta y_{t\hat{\tau}} = \nu^{(0)} + \alpha(\beta' y_{t-1,\hat{\tau}} - \phi^{(0)}(t-1) - \theta^{(0)} d_{t-1,\hat{\tau}}) + \sum_{j=1}^{p-1}\Gamma_j \Delta y_{t-j,\hat{\tau}} + \sum_{j=0}^{p-1}\gamma_j^{*(0)} \Delta d_{t-j,\hat{\tau}} + \varepsilon_{t\hat{\tau}}, \quad (A.40)$$

where

$$\begin{aligned}
\nu^{(0)} &= \nu + \alpha\beta'\mu_{0o} - \Psi\mu_{1o} \\
\phi^{(0)} &= \phi - \beta'\mu_{1o} \\
\theta^{(0)} &= \theta - \beta'\delta_o \\
\gamma_0^{*(0)} &= \delta - \delta_o
\end{aligned}$$

and

$$\gamma_j^{*(0)} = \gamma_j^* + \Gamma_j\delta_o \qquad (j = 1, \ldots, p-1).$$

Note that the true values of these parameters are zero. RR estimators of the parameters in $(A.40)$ are obtained by transforming the RR estimators based on $(A.37)$ in the same way as the corresponding parameters (e.g., $\tilde{\phi}^{(0)} = \tilde{\phi} - \tilde{\beta}'\mu_{1o}$). Asymptotic properties of these transformed estimators are derived below. The needed derivations make use of the following general results.

Lemma A.13. Let J_{tT} be a (possibly) random vector such that $\max_{1\le t\le T}\|J_{tT}\| = O_p(1)$ and J_t a vector valued stochastic process satisfying $\sup_t E\|J_t\|^2 < \infty$. Then, if $\hat{\tau} - \tau = O_p(T^{-2a/(1-2\eta)})$, $\eta - \frac{1}{2} \le a \le 0$ and $\frac{1}{b} < \eta < \frac{1}{4}$,

(i) $T^{-\eta-1/2}\sum_{t=p+1}^{T} J_{tT}\delta_o'(d_{t\hat{\tau}} - d_{t\tau}) = O_p(T^{a-\eta-\frac{1}{2}-\frac{2a}{(1-2\eta)}}) = O_p(1)$

(ii) $T^{-\eta-1/2}\sum_{t=p+1}^{T} J_t\delta_o'(d_{t\hat{\tau}} - d_{t\tau}) = O_p(T^{a-\eta-a/(1-2\eta)}) = O_p(1)$.

Proof: To prove (i), use the triangle inequality to conclude that the norm of the considered sum is dominated by $T^{-\eta-1/2}\max_{1\le t\le T}\|J_{tT}\|\|\delta_o\||\hat{\tau} - \tau|$. The stated orders in probability follow from this and the assumptions.

For (ii) we can use the triangle inequality and the Cauchy-Schwarz inequality to show that the norm of the considered sum is dominated by

$$T^{-\eta}\|\delta_o\|\left(T^{-1}\sum_{t=p+1}^{T}\|J_t\|^2\right)^{1/2}\left(\sum_{t=p+1}^{T}|d_{t\hat{\tau}} - d_{t\tau}|^2\right)^{1/2} = O_p\left(T^{a-\eta}\right)|\hat{\tau} - \tau|^{1/2}.$$

Here the equality is due to the assumptions which also readily show that the last expression is of the stated order in probability. □

Note that the assumption $a \geq \eta - 1/2$ is equivalent to $-2a/(1-2\eta) \leq 1$ and is not restrictive because $\hat{\tau} - \tau$ is necessarily $O_p(T)$, as already mentioned in Section 3.

Now we can prove the asymptotic properties of the RR estimators discussed above. We denote by $\tilde{\alpha}_0$ and $\tilde{\beta}_0$ normalized versions of the estimators $\tilde{\alpha}$ and $\tilde{\beta}$, respectively, such that $\tilde{\beta}_0 = \tilde{\beta}((\beta_o'\beta_o)^{-1}\beta_o'\tilde{\beta})^{-1}$.

Lemma A.14. Under the conditions of Lemma 4.1,

$$\tilde{\beta}_0 - \beta_o = O_p(T^{\eta-1}) \qquad (A.41)$$

$$\tilde{\phi}^{(0)} = O_p(T^{\eta-3/2}) \qquad (A.42)$$

$$\tilde{\theta}^{(0)} = O_p(T^{\eta-1/2}) \qquad (A.43)$$

$$\tilde{\alpha}_0 - \alpha_o = O_p(T^{\eta-1/2}) \qquad (A.44)$$

$$\tilde{\Gamma}_j - \Gamma_{jo} = O_p(T^{\eta-1/2}) \qquad (A.45)$$

$$\tilde{\nu}^{(0)} = O_p(T^{\eta-1/2}) \qquad (A.46)$$

$$\tilde{\gamma}_j^{*(0)} = o_p(T^{\eta}), \quad j = 0, \ldots, p-1 \qquad (A.47)$$

$$\tilde{\Omega} - \Omega = O_p(T^{\eta-1/2}) \qquad (A.48)$$

where $1/b < \eta < 1/4$.

Proof: We shall prove Lemma A.14 for the following different cases: (a) $\delta_{1o} \neq 0$ and $a \leq 0$; (b) $\delta_{1o} = 0$ and $a \leq 1/b$; (c) all remaining cases. In Cases (a) and (b) all three break date estimators $\hat{\tau}$, $\hat{\tau}_R$ and $\tilde{\tau}$ are relevant. In Case (c), also all three break date estimators are relevant except when $\delta_{1o} = 0$ and $a > 1/b$. In the latter situation, $\tilde{\tau}$ is not considered because this case is not covered by Theorem 3.3. We start by introducing some notation.

Define $\mathbf{x}_{t\hat{\tau}} = [y_{t-1,\hat{\tau}}' : (t-1) : d_{t-1,\hat{\tau}}]'$ and $\mathbf{p}_{1t\hat{\tau}} = [1 : \Delta y_{t-1,\hat{\tau}}' : \cdots : \Delta y_{t-p+1,\hat{\tau}}']'$. Then equation $(A.40)$ can be expressed as

$$\Delta y_{t\hat{\tau}} = \alpha\psi'\mathbf{x}_{t\hat{\tau}} + \Lambda\mathbf{p}_{t\hat{\tau}} + \varepsilon_{t\hat{\tau}}, \quad t = p+1, p+2, \ldots, \qquad (A.49)$$

where $\mathbf{p}_{t\hat{\tau}} = [\mathbf{p}'_{1t\hat{\tau}} : \underline{d}'_{t\hat{\tau}}]'$, $\psi' = [\beta' : -\phi^{(0)} : -\theta^{(0)}]$ and $\Lambda = [\Lambda_1 : \gamma_0^{*(0)}, \ldots, \gamma_{p-1}^{*(0)}]$ with $\Lambda_1 = [\nu^{(0)} : \Gamma_1 : \cdots : \Gamma_{p-1}]$. Here $\underline{d}_{t\hat{\tau}} = [\Delta d_{t\hat{\tau}}, \ldots, \Delta d_{t-p+1,\hat{\tau}}]'$. The RR estimators of α, ψ and Ω can be obtained as follows. Define

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

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

and

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

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

$$\det(\lambda S_{11\hat{\tau}} - S_{10\hat{\tau}} S_{00\hat{\tau}}^{-1} S_{01\hat{\tau}}) = 0. \qquad (A.50)$$

When the RR estimator of ψ is available, those of the other parameters can be obtained by replacing ψ by its estimator in $(A.49)$ and using usual LS formulas for the obtained auxiliary regression model.

Case(a): $\delta_{1o} \neq 0$ **and** $a \leq 0$

We start by studying the RR estimator of ψ. In the same way as in S&L we follow the proof of Lemma 13.1 of Johansen (1995) and transform equation $(A.50)$ to

$$\det(\lambda A'_T S_{11\hat{\tau}} A_T - A'_T S_{10\hat{\tau}} S_{00\hat{\tau}}^{-1} S_{01\hat{\tau}} A_T) = 0, \qquad (A.51)$$

where

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

and, consequently,

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

Recall the definition $w_t^{(0)} = [1 : \frac{t}{T} : T^{-1/2}v_{t-1}^{(0)\prime} : u_{t-1}^{(0)\prime} : \Delta x_{t-1}' : \cdots : \Delta x_{t-p+1}']'$ (see $(A.3)$) and note that for all $a \leq 1/2$,

$$T^{-1/2} \sum_{t=\tau}^{[T\bar{\lambda}]-1} w_t^{(0)} \delta_o' \Delta d_{t-i,\tau} = o_p\left(T^{\eta+a-1/2}\right), \quad i = 0, \ldots, p-1, \tag{$A.52$}$$

uniformly in $[T\underline{\lambda}] \leq \tau < [T\bar{\lambda}]$. For the first component of $w_t^{(0)}$, that is, for $w_{1t}^{(0)}$, this is essentially justified in the proof of Lemma A.4 because $\|\delta_o\| = T^a \|\delta_*\|$ (see Equation $(A.9a)$). Regarding $w_{2t}^{(0)}$, consider its first component $u_{t-1}^{(0)}$. Since $u_{t-1}^{(0)} = \beta_o' x_{t-1}$ is stationary we can use an argument similar to that for (A.14) of Saikkonen & Lütkepohl (2002) and conclude that $\max_{1 \leq t \leq T} \left\| u_{t-1}^{(0)} \right\| = o_p\left(T^{\eta}\right)$, which gives the desired result. A similar reasoning applies to the remaining components of $w_{2t}^{(0)}$, that is, to Δx_{t-j} $(j = 1, \ldots, p-1)$, and, hence, we have established $(A.52)$.

Now, using $(A.52)$ and Lemma A.13 in conjunction with the definition of $y_{t\hat{\tau}}$ (see $(A.39)$) and the assumption $a \leq 0$ we can show that

$$S_{00\hat{\tau}} = S_{00\tau} + o_p(1), \tag{$A.53$}$$

$$S_{01\hat{\tau}} A_T = S_{01\tau} A_T + o_p(1) \tag{$A.54$}$$

and

$$A_T' S_{11\hat{\tau}} A_T = A_T' S_{11\tau} A_T + o_p(1). \tag{$A.55$}$$

Details are straightforward but somewhat tedious and will be omitted.

From $(A.53) - (A.55)$ and the proof of Lemma 2.1 of S&L we can conclude that the estimators $\tilde{\beta}_0$, $\tilde{\phi}^{(0)}$ and $\tilde{\theta}^{(0)}$ are consistent of orders $o_p(T^{-1/2})$, $o_p(T^{-1})$ and $o_p(1)$, respectively. After this, the consistency properties of the other estimators can be studied by using the LS estimation of the auxiliary regression model

$$\Delta y_{t\hat{\tau}} = \alpha(\tilde{\psi}' \mathbf{x}_{t\hat{\tau}}) + \Lambda \mathbf{p}_{t\hat{\tau}} + \text{error}, \tag{$A.56$}$$

where $\tilde{\psi}$ is the RR estimator of ψ and the error has the representation $\varepsilon_{t\hat{\tau}} - \alpha_o(\tilde{\psi} - \psi_o)' \mathbf{x}_{t\hat{\tau}}$ (see $(A.17)$ and $(A.31)$). Using the consistency results obtained for $\tilde{\psi}$ in conjunction with Lemma A.13 and $(A.52)$ it can be shown that the replacement of $\hat{\tau}$ by the true break date in the appropriately standardized moment matrix of this LS estimation causes an error of order $o_p(1)$. Here the needed arguments are similar to those already used and the employed standardization is by the matrix $[T^{-1/2}I : I_p]$ where the latter identity matrix is used for the impulse dummies in $\mathbf{p}_{t\hat{\tau}}$. It is further straightforward to see that this standardized moment matrix is asymptotically positive definite and block diagonal between the impulse dummies and other regressors. The consistency results obtained for $\tilde{\psi}$ and $(A.52)$ also readily show that the off diagonal blocks are of order $o_p(T^{\eta-1/2})$.

The next step is to study the sums of cross products between the regressors and errors in $(A.56)$. Here we can also use a standardization of the form $[T^{-1/2}I : I_p]$ and show that these sums of cross products are of order $[o_p(T^{1/2}) : o_p(T^{\eta})]$, where the latter order is related to the impulse dummies and the former to the other regressors. From this and what was said about the asymptotic behavior of the standardized moment matrix we obtain $(A.47)$ and that $T^{1/2}(\tilde{\alpha}_0 - \alpha_o) = o_p(T^{1/2})$, $T^{1/2}(\tilde{\Gamma}_j - \Gamma_{jo}) = o_p(T^{1/2})$ and $T^{1/2}\tilde{\nu}^{(0)} = o_p(T^{1/2})$ or consistency of the estimators $\tilde{\alpha}_0$, $\tilde{\Gamma}_j$ $(j = 1, \ldots, p-1)$ and $\tilde{\nu}^{(0)}$. After this, consistency of $\tilde{\Omega}$ is also straightforward to obtain by using the auxiliary regression model $(A.56)$.

These results can be proved by using again the consistency results obtained for $\tilde{\psi}$ in conjunction with Lemma A.13 and $(A.52)$. Some of the involved details will be illustrated later when orders of consistency are obtained. Here we only note that the order $o_p(T^{\eta})$ for the sums of cross products between the impulse dummies and errors of the auxiliary regression model $(A.56)$ results because the error term contains the component $\varepsilon_{t\hat{\tau}}$ which in turn contains the component ε_t (see $(A.38)$). Hence, we have to find out the order of $\sum_{t=p+1}^{T} \Delta d_{t-j,\hat{\tau}}\varepsilon_t$ and $(A.52)$ gives the result $o_p(T^{\eta})$.

To obtain the orders of consistency stated in $(A.41)$ – $(A.43)$, we follow LST and note that the first order conditions for $\tilde{\psi}$, the RR estimator of ψ, can be expressed as

$$0 = \tilde{\alpha}_0'\tilde{\Omega}^{-1}\left(T^{-\eta}S_{\varepsilon 1\hat{\tau}}B_T - \tilde{\alpha}_0 T^{-\eta}U_T'[T^{-1}B_T'S_{11\hat{\tau}}B_T] - T^{-\eta}(\tilde{\alpha}_0 - \alpha_o)\psi_o'S_{11\hat{\tau}}B_T\right), \quad (A.57)$$

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

$$U_T = \begin{bmatrix} T\beta_{o\perp}'\tilde{\beta}_0 \\ T^{3/2}\tilde{\phi}^{(0)\prime} \\ (T-\tau)^{1/2}\tilde{\theta}^{(0)\prime} \end{bmatrix}$$

and

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

Now, notice that B_T is formed by the last $(n-r+2)$ columns of $T^{1/2}A_T$ and that $\psi_o'\mathbf{x}_{t\hat{\tau}} = \beta_o' y_{t-1,\hat{\tau}}$. Using these facts, the definition of $y_{t\hat{\tau}}$ (see $(A.39)$), $(A.52)$, and Lemma A.13, it is straightforward to proceed in the same way as in the above consistency proof and show that in $(A.57)$

$$T^{-\eta}S_{\varepsilon 1\hat{\tau}}B_T = O_p(1),$$

$$T^{-1}B_T'S_{11\hat{\tau}}B_T = T^{-1}B_T'S_{11\tau}B_T + o_p(1)$$

and

$$T^{-\eta}(\tilde{\alpha}_0 - \alpha_o)\psi_o'S_{11\hat{\tau}}B_T = o_p(1).$$

In the same way as in the proof of Lemma 2.1 of S&L we can then conclude from $(A.57)$ that $T^{-\eta}U_T = O_p(1)$ and furthermore $(A.41) - (A.43)$ hold. To illustrate the needed arguments, we consider one detail and demonstrate that

$$T^{-1-\eta}\sum_{t=p+1}^{T}\beta_o' y_{t-1,\hat{\tau}}y_{t-1,\hat{\tau}}'\beta_{o\perp} = T^{-1-\eta}\sum_{t=p+1}^{T}\beta_o' x_{t-1}x_{t-1}'\beta_{o\perp} + O_p(1) = O_p(1).$$

The l.h.s. is contained in $T^{-\eta}\psi_o'S_{11\hat{\tau}}B_T$ and hence also in $T^{-\eta}S_{\varepsilon 1\hat{\tau}}B_T$. Because the sum in the second expression is of order $O_p(T)$ by well-known properties of stationary and $I(1)$ processes, this result follows from

$$T^{-1-\eta}\sum_{t=p+1}^{T}\beta_o' x_{t-1}\delta_o'(d_{t-1,\hat{\tau}} - d_{t-1,\tau}) = o_p(1),$$

$$T^{-1-\eta}\sum_{t=p+1}^{T}\beta_{o\perp}' x_{t-1}\delta_o'\beta_o(d_{t-1,\hat{\tau}} - d_{t-1,\tau}) = O_p(1)$$

and

$$T^{-1-\eta}\sum_{t=p+1}^{T}\beta_o'\delta_o\delta_o'\beta_o(d_{t-1,\hat{\tau}} - d_{t-1,\tau}) = o_p(1).$$

Because $\beta_o' x_{t-1}$ is second order stationary the first of these results follows from Lemma A.13(ii). The second one can be justified by Lemma A.13(i) because $\max_{1\leq t\leq T} T^{-1/2}\|x_t\| = O_p(1)$ by well-known properties of $I(1)$ processes. The third result can also be obtained from Lemma A.13(i) by choosing $J_{tT} = \beta'\delta_o$.

We still have to justify the orders of consistency for the estimators $\tilde{\alpha}_0$, $\tilde{\Gamma}_j$ $(j = 1, \ldots, p-1)$, $\tilde{\nu}^{(0)}$ and $\tilde{\Omega}$. Here we can proceed in the same way as in the corresponding consistency proof but use the improved consistency result obtained for the estimator $\tilde{\psi}$ to improve the convergence rate of the sums of cross products between the regressors other than the impulse dummies and the error in the auxiliary regression model $(A.56)$. Specifically, we can show that this improved convergence rate is $O_p(T^{\eta-1/2})$ which combined with arguments used in the consistency proof implies $(A.44)-(A.46)$. After this, $(A.48)$ can be proved by using the auxiliary regression model $(A.56)$ and orders of consistency obtained for other parameters.

To illustrate some of the details in the proof of the above mentioned improved convergence rate between the regressors and errors of the auxiliary regression model, we consider the regressors $\tilde{\psi}'\mathbf{x}_{t\hat{\tau}}$. We need to consider

$$\begin{aligned} &T^{-\eta-1/2}\textstyle\sum_{t=p+1}^{T}\tilde{\psi}'\mathbf{x}_{t\hat{\tau}}(\varepsilon_{t\hat{\tau}}-\alpha_o(\tilde{\psi}-\psi_o)\mathbf{x}_{t\hat{\tau}})' \\ &=T^{-\eta-1/2}\textstyle\sum_{t=p+1}^{T}\tilde{\psi}'\mathbf{x}_{t\hat{\tau}}\varepsilon_t'-T^{-\eta-1/2}\sum_{t=p+1}^{T}\tilde{\psi}'\mathbf{x}_{t\hat{\tau}}\mathbf{x}_{t\hat{\tau}}'(\tilde{\psi}-\psi_o)\alpha_o'+o_p(1), \end{aligned} \tag{A.58}$$

where the order term is straightforward to justify by using the definition of $\varepsilon_{t\hat{\tau}}$ (see $(A.38)$), Lemma A.13 and $(A.52)$. The vector $\tilde{\psi}'\mathbf{x}_{t\hat{\tau}}$ contains three components of which the first one is $\tilde{\beta}_0'y_{t-1,\hat{\tau}}$. In the first term on the r.h.s. of $(A.58)$ we concentrate on this term and conclude from the definition of $y_{t\hat{\tau}}$ (see $(A.39)$) that

$$T^{-\eta-1/2}\sum_{t=p+1}^{T}\tilde{\beta}_0'y_{t-1,\hat{\tau}}\varepsilon_t'=T^{-\eta-1/2}\sum_{t=p+1}^{T}\tilde{\beta}_0'x_{t-1}\varepsilon_t'-T^{-\eta-1/2}\sum_{t=p+1}^{T}\tilde{\beta}_0'\delta_o(d_{t-1,\hat{\tau}}-d_{t-1,\tau})\varepsilon_t'=O_p(1),$$

where the latter equality follows from the consistency result obtained for $\tilde{\beta}_0$, Lemma A.13(ii) and well-known properties of stationary and $I(1)$ processes. A similar result holds for the two other components of $\tilde{\psi}'\mathbf{x}_{t\hat{\tau}}$. Here the weaker orders of consistency $\tilde{\beta}_0=\beta_o+o_p(T^{-1/2})$, $\tilde{\phi}^{(0)}=o_p(T^{-1})$ and $\tilde{\theta}^{(0)}=o_p(1)$ would actually suffice. However they do not suffice for the second term on the r.h.s. of $(A.58)$. To see this, note that here we, for instance, need to consider the matrix $T^{-\eta-1/2}(\tilde{\beta}_0-\beta_o)'\sum_{t=p+1}^{T}y_{t-1,\hat{\tau}}y_{t-1,\hat{\tau}}'(\tilde{\beta}_0-\beta_o)$ which explodes if only $\tilde{\beta}_0-\beta_o=o_p(T^{-1/2})$ holds. However, with the results given by $(A.41)-(A.43)$ even the second term on the r.h.s. of $(A.58)$ can be handled and it follows that the l.h.s. of $(A.58)$ is of order $O_p(1)$. This completes the proof in the case $\delta_{1o}\neq 0$ and $a\leq 0$.

Case (b): $a\leq 1/b$ **and** $\delta_{1o}=0$

In this case nothing can be said about the asymptotic behavior of the break date estimator so that $\hat{\tau}$ can take any value between $[T\underline{\lambda}]$ and $[T\bar{\lambda}]$. When $a=0$ this case was considered in LST and the arguments used therein can be modified for all $a\leq 1/b$.

In this case we cannot use Lemma A.13 but we still have $(A.52)$ and $(A.53)$, as we shall now demonstrate. The former is simple because by the definition of $y_{t\hat{\tau}}$ (see $(A.39)$) 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 $\Delta d_{t-j,\tau}$ $(j = 0, \ldots, p-1)$ but not on the corresponding step dummies. Thus, because $(A.52)$ still holds and $a + \eta < 1/2$, we can proceed in the same way as in the proof of Lemma 2.1 of S&L and show first 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 $o_p(1)$. The employed arguments are similar to those used in the case $\delta_{1o} \neq 0$ and $a \leq 0$. Because in $S_{00\tau}$ the vector $\mathbf{p}_{t\tau}$ can similarly be replaced by $\mathbf{p}_{1t\tau}$ (see S&L), $(A.52)$ follows.

Now consider $(A.54)$. Using $(A.52)$ and the definition of $y_{t\hat{\tau}}$ 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 $o_p(1)$ (cf. the justification of $(A.61)$ below). Next, an application of $(A.52)$ and the definition of $y_{t\hat{\tau}}$ gives

$$T^{-1} \sum_{t=p+1}^{T} \mathbf{p}_{1t\hat{\tau}} \mathbf{p}'_{1t\hat{\tau}} = T^{-1} \sum_{t=p+1}^{T} \mathbf{p}_{1t\tau} \mathbf{p}'_{1t\tau} + o_p(1) \qquad (A.59)$$

and

$$T^{-1} \sum_{t=p+1}^{T} \Delta y_{t\hat{\tau}} \mathbf{p}'_{1t\hat{\tau}} = T^{-1} \sum_{t=p+1}^{T} \Delta y_{t\tau} \mathbf{p}'_{1t\tau} + o_p(1). \qquad (A.60)$$

In addition to these results, we have

$$T^{-1} \sum_{t=p+1}^{T} \Delta y_{t-j,\hat{\tau}} \mathbf{x}'_{t\hat{\tau}} A_T = T^{-1} \sum_{t=p+1}^{T} \Delta y_{t-j,\tau} \mathbf{x}'_{t\tau} A_T + o_p(1) \quad (j = 0, \ldots, p-1). \qquad (A.61)$$

To justify this relation, notice that now $\beta_o' y_{t-1,\hat{\tau}} = \beta_o' x_{t-1}$. Thus, an application of $(A.52)$ readily shows that $(A.61)$ 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 $T^{3/2}$. The same can be said about the $(n+1)$th column. That the stated result holds for the last column can be seen from

$$\begin{aligned} T^{-1} \textstyle\sum_{t=p+1}^{T} \Delta y_{t-j,\hat{\tau}} d_{t-1,\tau} &= T^{-1} \textstyle\sum_{t=\tau+1}^{T} \Delta x_{t-j} + o_p(1) \\ &= o_p(1) \qquad (j = 0, \ldots, p-1) \end{aligned} \qquad (A.62)$$

uniformly in $[T\underline{\lambda}] \leq \tau \leq [T\bar{\lambda}]$. Here the former equality is an immediate consequence of the definition of $y_{t\hat{\tau}}$ and the assumption $a \leq 1/b$ whereas the latter follows because an invariance principle applies to partial sums of Δx_t.

In addition to $(A.59)$ - $(A.61)$, we also need to consider the matrix $T^{-1} \sum_{t=p+1}^{T} \mathbf{p}_{1t\hat{\tau}} \mathbf{x}'_{t\hat{\tau}} A_T$. 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

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

However, the contribution of this element to the matrix on the l.h.s. of $(A.54)$ 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 $(A.59)$ 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 $(A.60)$ is of order $o_p(1)$. Both of these results follow from the definition of $y_{t\hat{\tau}}$ and the fact that Δx_t obeys a weak law of large numbers. Taking these results together, we can thus conclude that $(A.54)$ also holds when $\delta_1 = 0$.

Since $(A.53)$ and $(A.54)$ hold we can write

$$A_T' S_{10\hat{\tau}} S_{00\hat{\tau}}^{-1} S_{01\hat{\tau}} A_T = A_T' S_{10\tau} S_{00\tau}^{-1} S_{01\tau} A_T + o_p(1) = \begin{bmatrix} \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & 0 \end{bmatrix} + o_p(1), \qquad (A.63)$$

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 $(A.55)$ which, given $(A.63)$, would be sufficient for the consistency of the RR estimator of ψ. However, $(A.55)$ 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 $\beta_o' y_{t-1,\hat{\tau}} = \beta_o' x_{t-1}$ by the definition of $y_{t\hat{\tau}}$, 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 $(A.52)$ and the definition of $y_{t\hat{\tau}}$ in the same way as in the case of $(A.53)$, $(A.54)$ and $(A.62)$. Thus, because we also have $\bar{S}_{11\hat{\tau}}^{21} = o_p(1)$, the above discussion and $(A.63)$ show that equation $(A.51)$ is to order $o_p(1)$ identical to

$$\det(\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta}) \det(\lambda \bar{S}_{11\hat{\tau}}^{22}) = 0. \qquad (A.64)$$

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 arguments similar to those used below $(A.10)$ show that $\bar{S}^{22}_{11[T\lambda]}$ converges weakly in $D([\underline{\lambda}, \bar{\lambda}])$ to a (a.s.) positive definite limit. In fact, arguments used to arrive at $(A.64)$ show that $\bar{S}^{22}_{11[T\lambda]}$ 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 $\tilde{\beta}_0$, $\tilde{\phi}^{(0)}$ and $\tilde{\theta}^{(0)}$ are consistent of orders $o_p(T^{-1/2})$, $o_p(T^{-1})$ and $o_p(1)$, respectively. After this, the asymptotic behavior of the other RR estimators can be studied by using the auxiliary regression model $(A.56)$ in the same way as in the case $\delta_{1o} \neq 0$ and $a \leq 0$. In addition to the obtained consistency properties of the estimators $\tilde{\beta}_0$, $\tilde{\phi}^{(0)}$ and $\tilde{\theta}^{(0)}$, the employed arguments make use of the facts that now $\psi_o'\mathbf{x}_{t\tau} = \beta_o' x_{t-1}$ and that in the error of the auxiliary regression model $(A.56)$ the component $\varepsilon_{t\hat{\tau}}$ only depends on ε_t and impulse dummies but not on step dummies (see $(A.38)$). Thus, Lemma A.13 is not needed and it suffices to use $(A.52)$.

Without going into details we note that the standardized moment matrix of the auxiliary regression model $(A.56)$ is again asymptotically positive definite and block diagonal between the impulse dummies and other regressors (the standardization is again by the matrix $[T^{-1/2}I : I_p]$). Moreover, it is not difficult to check that the off diagonal blocks are of order $o_p(T^{\eta-1/2})$. This can be seen by using $(A.52)$ and observing that here the vector δ_o is absent in most of the cases. As in the previous case $\delta_{1o} \neq 0$ and $a \leq 0$, we can also here show that the correspondingly standardized sums of cross products between the regressors and error in the auxiliary regression model $(A.56)$ is of order $[o_p(T^{1/2}) : o_p(T^{\eta})]$ with the partition as before. Thus, $(A.47)$ and the consistency of the estimators $\tilde{\alpha}_0$, $\tilde{\Gamma}_j$ $(j = 1, \ldots, p-1)$, $\tilde{\nu}^{(0)}$ and $\tilde{\Omega}$ follows in the same way as in the previous case.

To obtain the orders of consistency, consider the estimator $\tilde{\psi}$ and the first order conditions $(A.57)$. Recall that now $\psi_o'\mathbf{x}_{t\hat{\tau}} = \beta_o' x_{t-1}$. Using this fact it can be seen that $T^{-\eta}\psi_o' S_{11\hat{\tau}} B_T = O_p(1)$ and $T^{-\eta} S_{\varepsilon 1\hat{\tau}} B_T = O_p(1)$. The employed arguments are similar to those used in the above consistency proof and in the proof of Theorem 3.1. To illustrate, note that in order to prove $T^{-\eta}\psi_o' S_{11\hat{\tau}} B_T = O_p(1)$ we need to show that

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

and

$$T^{-1-\eta} \sum_{t=p+1}^{T} \beta_o' x_{t-1} x_{t-1}' \beta_{o\perp} - T^{-1-\eta} \sum_{t=p+1}^{T} \beta_o' x_{t-1} (d_{t\hat{\tau}} - d_{t\tau}) \delta_o' \beta_{o\perp} = O_p(1).$$

The former of these results follows from the Hájek-Rényi inequality given in Proposition 1 of Bai (1994) and the order is actually $o_p(1)$. Regarding the latter, the first term on the l.h.s. is of order $o_p(1)$ by well-known properties of stationary and $I(1)$ processes. That the same is true for the second term can be seen by using the fact that $\|\delta_o'\beta_{o\perp}\| = O(T^a) = O(T^\eta)$ in conjunction with the argument used to show $(A.62)$. 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 $(A.57)$ and what has been said above we now find that $T^{-\eta}U_T = O_p(1)$. In the same way as in the case $\delta_{1o} \neq 0$ this implies $(A.41)$ - $(A.43)$.

We still have to obtain the stated orders of consistency for the estimators $\tilde{\alpha}_0$, $\tilde{\Gamma}_j$ $(j = 1, \ldots, p-1)$, $\tilde{\nu}^{(0)}$ and $\tilde{\Omega}$. In the same way as in the case $\delta_{1o} \neq 0$ and $a \leq 0$ it suffices to use the improved consistency result obtained for the estimator $\tilde{\psi}$ to improve the earlier convergence rate of sums of cross products between the regressors other than impulse dummies and the error in the auxiliary regression model $(A.56)$. We omit details, which are similar to those used earlier, and only note that instead of $o_p(T^{1/2})$ the convergence rate obtained for the above mentioned sums of cross products is now $O_p(T^{\eta-1/2})$. Given this improvement, the previous consistency results obtained for the estimators $\tilde{\alpha}_0$, $\tilde{\Gamma}_j$ $(j = 1, \ldots, p-1)$, $\tilde{\nu}^{(0)}$ and $\tilde{\Omega}$ can be improved to the stated form. This completes the proof in the case $\delta_{1o} = 0$ and $a \leq 1/b$.

Case (c)

Because in the preceding proof the estimator $\hat{\tau}$ could take any values between $[T\underline{\lambda}]$ and $[T\bar{\lambda}]$ it is straightforward to use the same arguments and show that the results of Lemma A.14 also hold when $\delta_o = 0$. This is actually fairly obvious because when $\delta_o = 0$ the vectors $\varepsilon_{t\hat{\tau}}$ and $y_{t\hat{\tau}}$ simplify to ε_t and x_t, respectively, and the proof will be simplified in many places. Details are omitted.

To prove the remaining cases we first note that the result of Lemma A.14 holds also when the break date is assumed known. In this case we can even replace the quantity η by zero except in $(A.47)$ where $O_p(1)$ is obtained. A formal proof of this can be obtained from S&L by observing that the omission of some impulse dummies from the model considered by S&L is of no significance and that the same is true for the dependence of the parameters δ on the sample size. The latter fact is clear because the results of Lemma A.14 are formulated for the transformed parameters used in model $(A.40)$ and the true values of these transformed parameters are zero.

Because the result of Lemma A.14 holds when the break date is assumed known it also holds when the break date can be consistently estimated, that is, when $\hat{\tau}-\tau = o_p(1)$. Indeed, then all the analysis can be restricted to that part of the sample space where $\hat{\tau} = \tau$ holds and the probability of this can be made arbitrarily close to unity for all T large enough. This means that we have established the results of the lemma for the constrained estimator $\hat{\tau}_R$ when $\delta_{1o} \neq 0$ and $a > 0$ or $\delta_{1o} = 0$ and $a > 1/b$ (Theorem 3.2(i)) and for the estimator $\tilde{\tau}$ when $\delta_{1o} \neq 0$ and $a > 0$ (Theorem 3.3(i)).

If $j_0 = p - 1$ in Theorem 3.1(i) the preceding argument also applies to the estimator $\hat{\tau}$ when $\delta_{1o} \neq 0$ and $a > 0$ or $\delta_{1o} = 0$ and $a > 1/b$. For other values of j_0 further arguments are needed. By Theorem 3.1(i) it suffices to consider any value of the break date such that $\tau_o - p + 1 + j_0 \leq \tau \leq \tau_o$, where we have included the subscript 'o' to signify the true break date. For simplicity, consider the case $j_0 = p - 2$ and $\tau = \tau_o - 1$. It is easy to see that even though the break date is misspecified by one we can still consider (2.7) as a correctly specified model if we only redefine the parameters $\gamma_0^*, \ldots, \gamma_{p-1}^*$ as $\gamma_0^* = \alpha\beta'\delta$, $\gamma_1^* = \delta$, and $\gamma_j^* = -\Gamma_{j-1}\delta$, $j = 2, \ldots, p-1$. By assumption we then have $\gamma_{p-1}^* \neq 0$ and $\Gamma_{p-1}\delta = 0$. With these new definitions the error term of model (2.7) is still ε_t and the analysis given in the case of a known break date can be used. Because the other parameters of the model are not affected by the redefinition of the parameters γ_j^* $(j = 0, \ldots, p-1)$ the obtained consistency results will be the same as in the case where the true break date is known. The same argument can clearly be extended to other values of j_0. This completes the proof of Lemma A.14. □

Now we can prove Lemma 4.1 and we start with the results (4.1) and (4.2). Recall the definitions $\nu = -\alpha\beta'\mu_0 + \Psi\mu_1$, $\nu^{(0)} = \nu + \alpha\beta'\mu_{0o} - \Psi\mu_{1o}$ and $\phi^{(0)} = \phi - \beta'\mu_{1o}$ which imply that

$$\nu^{(0)} = -\alpha\beta'(\mu_0 - \mu_{0o}) + \Psi(\mu_1 - \mu_{1o}) = -\alpha\beta'(\mu_0 - \mu_{0o}) + \Psi_\beta\phi^{(0)} + \Psi_{\beta_\perp}\beta_\perp'(\mu_1 - \mu_{1o}).$$

Here the latter equality is obtained by arguments similar to those used to define the estimator $\tilde{\phi}_*$. These arguments further show that $\beta_\perp'(\mu_1 - \mu_{1o}) = \beta_\perp' C(\nu^{(0)} - \Psi_\beta\phi^{(0)})$ and the same relation applies to estimators. Thus, we have

$$\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_1) = \tilde{\beta}'_\perp \tilde{C}(\tilde{\nu}^{(0)} - \tilde{\Psi}_\beta \tilde{\phi}^{(0)}).$$

Here and in what follows the subscripts 'o' and '0' are omitted from true parameter values and the estimators of α and β, respectively, to simplify notation. By Lemma A.14, one obtains from the previous equality

$$\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_1) = \tilde{\beta}'_\perp \tilde{C}\tilde{\nu}^{(0)} + o_p(T^{-1/2}).$$

Now recall that in the auxiliary regression model $(A.56)$, Λ contains $\Lambda_1 = [\nu^{(0)} : \Gamma_1 : \cdots : \Gamma_{p-1}]$ and that $\tilde{\nu}^{(0)}$ can be viewed as the LS estimator of $\nu^{(0)}$. Hence, $\tilde{\beta}'_\perp \tilde{C}\tilde{\nu}^{(0)}$ can be obtained by LS from the auxiliary regression model

$$\tilde{\beta}'_\perp \tilde{C}\Delta y_{t\hat{\tau}} = \Lambda^* \mathbf{p}_{t\hat{\tau}} + \text{error}, \tag{A.65}$$

where Λ^* is a conformable coefficient matrix and the error has the representation $\tilde{\beta}'_\perp \tilde{C}\varepsilon_{t\hat{\tau}} - \tilde{\beta}'_\perp \tilde{C}\alpha(\tilde{\psi} - \psi)\mathbf{x}_{t\hat{\tau}}$ and, by the definition of C and Lemma A.14, $\tilde{\beta}'_\perp \tilde{C}\alpha = O_p(T^{\eta - 1/2})$. Using this fact, Lemma A.14 and arguments similar to those used in its proof, it is straightforward to show that the asymptotic properties of the LS estimator of the parameter Λ in the auxiliary regression model $(A.65)$ can be obtained by assuming that the error equals $\tilde{\beta}'_\perp \tilde{C}\varepsilon_{t\hat{\tau}}$. The same arguments and the definition of $\varepsilon_{t\hat{\tau}}$ (see $(A.38)$) further show that the error can be assumed to be $\tilde{\beta}'_\perp \tilde{C}\varepsilon_t$ or even $\beta'_\perp C\varepsilon_t$. Since it is also straightforward to show that the estimation of the intercept term in $(A.65)$ is asymptotically independent of the estimation of the other regression coefficients we can conclude that

$$\begin{aligned} T^{1/2}\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_1) &= \tilde{\beta}'_\perp \tilde{C} T^{-1/2} \sum_{t=p+1}^{T} \varepsilon_t + o_p(1) \\ &= \beta'_\perp C T^{-1/2} \sum_{t=p+1}^{T} \varepsilon_t + o_p(1), \end{aligned}$$

where the latter equality is again due to Lemma A.14. This and a standard central limit theorem yield

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

To obtain (4.2) we need to show that $\tilde{\beta}_\perp$ on the l.h.s. can be replaced by $\beta_\perp$. To see this, write

$$(\tilde{\beta}_\perp - \beta_\perp)'(\tilde{\mu}_1 - \mu_1) = (\tilde{\beta}_\perp - \beta_\perp)'\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\beta}'(\tilde{\mu}_1 - \mu_1) + (\tilde{\beta}_\perp - \beta_\perp)'\tilde{\beta}_\perp(\tilde{\beta}'_\perp\tilde{\beta}_\perp)^{-1}\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_1). \tag{A.66}$$

By the consistency of the estimator $\tilde{\beta}$ and the result just obtained the latter term on the r.h.s. is of order $o_p(T^{-1/2})$ and the same is true for the former because $\tilde{\beta}'(\tilde{\mu}_1 - \mu_1) = \tilde{\phi}^{(0)} = O_p(T^{\eta-3/2})$ by Lemma A.14. From this last result one can obtain (4.1) because $\tilde{\beta}$ can be replaced by β by an argument similar to that used in $(A.66)$.

Now consider the estimator $\tilde{\delta}$. From its derivation we get the identity $\sum_{j=0}^{p-1} \gamma_j^* - \Psi\delta_o = \Psi(\delta - \delta_o)$. By the definitions, this is equivalent to

$$\sum_{j=0}^{p-1} \gamma_j^{*(0)} = \Psi_\beta \beta'(\delta - \delta_o) + \Psi_{\beta_\perp} \beta_\perp'(\delta - \delta_o).$$

Because the same relation applies to estimators, arguments similar to those used to define the estimator $\tilde{\delta}$ yield

$$\tilde{\beta}_\perp'(\tilde{\delta} - \delta_o) = \tilde{\beta}_\perp' \tilde{C} \left(\sum_{j=0}^{p-1} \tilde{\gamma}_j^{*(0)} - \tilde{\Psi}_\beta \tilde{\theta}^{(0)} \right).$$

Lemma A.14 implies that the r.h.s. of this equality is of order $o_p(T^\eta)$. The same result holds even if $\tilde{\beta}_\perp$ on the l.h.s. is replaced by $\beta_\perp$, as can be seen by proceeding in the same way as in $(A.66)$. Thus, we have established (4.4). To obtain (4.3), notice that $\tilde{\beta}'(\tilde{\delta} - \delta) = \tilde{\theta}^{(0)} = O_p(T^{\eta-1/2})$ and the stated result follows by using an argument similar to that in $(A.66)$. This completes the proof of Lemma 4.1.

A.5 Proof of Theorem 4.1

The structure of our proof of Theorem 4.1 is similar to that of Theorem 11.1 of Johansen (1995). Therefore we just outline the arguments in the following.

First note that the limiting distribution of the test statistic can be derived by assuming that the true value of the parameter μ_0 is zero. Thus, we can write equation (4.5) as

$$\tilde{y}_t^{(0)} = x_t - (\tilde{\mu}_1 - \mu_1)t - (\tilde{\delta} - \delta)d_{t\tau} - (\tilde{\delta} - \delta)(d_{t\hat{\tau}} - d_{t\tau}) - \delta(d_{t\hat{\tau}} - d_{t\tau}). \qquad (A.67)$$

Using this representation and arguments similar to those used to derive the asymptotic properties of the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ (see the proof of Lemma 4.1) we can now mimic the proof given in Johansen (1995, pp. 158-160) and see that all the quantities which therein converge in probability to constants will here converge in probability to the same constants. However, quantities which in Johansen (1995, pp. 158-160) converge weakly to functionals of a Brownian motion will here converge weakly to different functionals of a Brownian motion.

Here these weak limits are determined by the weak limit of $T^{-1/2}\beta_\perp'\tilde{y}_{[Ts]}^{(0)}$. We have

$$\begin{aligned} &T^{-1/2}\beta_\perp'\tilde{y}_{[Ts]}^{(0)} = T^{-1/2}\beta_\perp' x_{[Ts]} - T^{-1/2}\beta_\perp'(\tilde{\mu}_1 - \mu_1)t + o_p(1) \\ &\xrightarrow{d} \beta_\perp' C(W(s) - sW(1)) \stackrel{def}{=} \beta_\perp' C W_+(s), \end{aligned} \tag{A.68}$$

where $W(s)$ is an $(n - r_0)$- dimensional Brownian motion with covariance matrix Ω and hence the limit is a linear transformation of the Brownian bridge $W_+(s) = W(s) - sW(1)$. The error term in the equality is understood to hold in the Skorohod topology.

To justify $(A.68)$, first consider the equality. Because $\beta_\perp'(\tilde{\delta} - \delta) = o_p(T^\eta)$ with $1/b < \eta < 1/4$ by (4.4) of Lemma 4.1 it is clear that the contribution of the third and fourth terms on the r.h.s. of $(A.67)$ to $T^{-1/2}\beta_\perp'\tilde{y}_{[Ts]}^{(0)}$ is asymptotically negligible. The same argument also applies to the fifth term on the r.h.s. of $(A.67)$ when $a < 1/2$. For $a = 1/2$ we have $T^{-1/2}\delta = O(1)$ but $\hat{\lambda} - \lambda = o_p(1)$ which implies that the contribution of the fifth term is of order $o_p(1)$ (in the Skorohod topology) and the desired conclusion follows. As for the weak convergence in $(A.68)$, it can be justified by a standard functional central limit theorem and (4.2) of Lemma 4.1 by observing that the limit of the second expression is determined by the process ε_t (see Johansen (1995, Eq. (B.24))) and the proof of (4.2).

Note that to obtain $(A.68)$ in the case $a = 1/2$, the assumption of consistent break date estimation included in the formulation of the theorem is not needed. However, this is not the case for some of the convergence statements referred to after $(A.67)$. For instance, to mimic the proof in Johansen (1995, pp. 158-160) we have to consider the matrix $(T - p)^{-1}\sum_{t=p+1}^{T}\Delta\tilde{y}_t^{(0)}\Delta\tilde{y}_t^{(0)\prime}$ which contains the additive component

$$\delta\delta'(T-p)^{-1}\sum_{t=p+1}^{T}(\Delta d_{t\hat{\tau}} - \Delta d_{t\tau})^2 = \delta_*\delta_*'\frac{T^{2a}}{T-p}\sum_{t=p+1}^{T}(\Delta d_{t\hat{\tau}} - \Delta d_{t\tau})^2. \tag{A.69}$$

Now consider the unconstrained estimator of Section 3.1 and suppose that $a = 1/2$ and $j_0 < p - 1$. From Theorem 3.1 we can only conclude that $Pr\{\tau - p + 1 + j_0 \le \hat{\tau} \le \tau\} \to 1$ and it does not follow that the quantity in $(A.69)$ is asymptotically negligible.

To continue the proof, so far we have demonstrated that the limiting distribution of the test statistic $LR(r_0)$ can be derived by ignoring the last two terms on the r.h.s. of $(A.67)$. As for the first three terms on the r.h.s. of $(A.67)$, their asymptotic behavior is not affected by the fact that the size of the break is allowed to depend on the sample size (cf. Lemma 4.1). This means that we have reduced the problem to that of a known break date studied by S&L. From Lemma 4.1 and the proof of Theorem 3.1 of S&L it can be seen that, when $\mu_0 = 0$

is assumed, the test statistic in the theorem is asymptotically equivalent to a similar test statistic based on an analog of (4.6) defined by replacing $\tilde{y}_{t-1}^{(+)}$ by $\tilde{y}_{t-1}^{(0)}$. It is straightforward to show that the use of $\tilde{y}_{t-1}^{(+)}$ instead of $\tilde{y}_{t-1}^{(0)}$ changes the limiting distribution of the test statistic as stated in the theorem. In other words, since the vector $\tilde{y}_{t-1}^{(+)}$ is obtained from $\tilde{y}_{t-1}^{(0)}$ by augmenting with unity, the same augmentation results in one of the two Brownian bridges in the limiting distribution obtained in Theorem 3.1 of S&L. Technical details, which are similar to the corresponding two cases in Johansen (1995, Section 11.2) are straightforward and will be omitted.

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.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press.

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*, 20, 247 - 318.

Inoue, A. (1999), Tests of cointegrating rank with a trend-break, *Journal of Econometrics*, 90, 215 - 237.

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

Johansen, S., R. Mosconi & 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.

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 & C. Trenkler (2004), Testing for the cointegrating rank of a VAR process with level shift at unknown time, *Econometrica*, 72, 647 - 662.

Marsaglia, G. (2000), The Monster, a Random Number Generator with Period over 10^{2587} As Long As the Previously Touted Longest Period One, *mimeo*, Department of Statistics, Florida State University.

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

Saikkonen, P. (2001), Consistent estimation in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations, *Econometric Theory*, 17, 296-326.

Saikkonen, P. & H. Lütkepohl (2000), 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 (2002), Testing for a unit root in a time series with a level shift at unknown time, *Econometric Theory*, 18, 313 - 348.

Seo, B. (1998), Tests for structural change in cointegrated systems, *Econometric Theory*, 14, 222 - 259.

Toda, H.Y. (1994), Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present, *Review of Economics and Statistics*, 76, 66 - 79.

Toda, H.Y. (1995), Finite sample performance of likelihood ratio tests for cointegration ranks in vector autoregressions, *Econometric Theory*, 11, 1015-1032.