

BREAK DATE ESTIMATION FOR VAR PROCESSES WITH LEVEL SHIFT WITH AN APPLICATION TO COINTEGRATION TESTING

PENTTI SAIKKONEN
University of Helsinki

HELMUT LÜTKEPOHL
European University Institute, Florence
and
Humboldt University Berlin

CARSTEN TRENKLER
Humboldt University Berlin

In testing for the cointegrating rank of a vector autoregressive 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 modeled as a simple shift in the level of the process. Two 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 estimators are obtained under the assumption of an increasing or decreasing size of the shift when the sample size grows. These results are used to explore the implications for testing the cointegrating rank of the process. A previously proposed likelihood ratio type test for the cointegrating rank and a modified version are considered, and their asymptotic properties are derived. It is shown that their asymptotic null distributions are unaffected by the level shift under the assumptions made for the shift size. The performance of the shift date estimators and the cointegrating rank tests in small samples is investigated by simulations.

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

We thank two referees for helpful comments, and we are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract 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. An extended version of this paper is available as an EUI discussion paper under the title “Break Date Estimation and Cointegration Testing in VAR Processes with Level Shift,” ECO 2004/21. Address correspondence to Pentti Saikkonen, e-mail: saikkone@valt.helsinki.fi.

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; Lütkepohl, Saikkonen, and Trenkler, 2004, for cointegrating rank testing). In the latter article it is assumed that a level shift has occurred in a system of time series variables at an unknown time. Lütkepohl et al. 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 test 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 addition, 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 Lütkepohl et al. (2004) in several directions. First of all we also consider another possible break date estimator. Second, we derive asymptotic properties of two break date 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 for both increasing and decreasing shift sizes when the sample size goes to infinity. These results provide interesting new insights into the properties of the estimators and explain simulation results of Lütkepohl et al. that 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 Lütkepohl et al. 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 the break date estimators and the resulting cointegrating rank tests by extending the simulation design of Lütkepohl et al.

Estimating the break date in a system of $I(1)$ variables has also been considered by Bai, Lumsdaine, and Stock (1998). These authors consider the asymptotic distribution of a pseudo maximum likelihood (ML) estimator of the break date. Although we use a similar estimator, 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 cointegrating rank tests that are based on a vector autoregressive (VAR) model with estimated break date. Although Bai et al. (1998) also discuss shift sizes that

depend on the sample size, our results go beyond their analysis because we consider increasing in addition to decreasing shift sizes.

The study is structured as follows. In Section 2, the modeling framework of Lütkepohl et al. (2004) 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 whereas it is still non-stationary after differencing just $d - 1$ times. Convergence in distribution is signified by \xrightarrow{d} , and i.i.d. stands for independently, identically distributed. The symbols for boundedness and convergence in probability are as usual $O_p(\cdot)$ and $o_p(\cdot)$, respectively. Moreover, $\|\cdot\|$ denotes the euclidean norm. The trace, 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 , $\text{diag}[A_1 : \dots : A_s]$ is the block-diagonal matrix with A_1, \dots, A_s on the diagonal. LS, RR, and VECM are used to abbreviate 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 Lütkepohl et al. (2004). 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

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

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 et al. (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 to 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

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

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

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

where $\Pi = -(I_n - A_1 - \dots - A_p)$ and $\Gamma_j = -(A_{j+1} + \dots + A_p)$ ($j = 1, \dots, p - 1$) 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 $\Psi = I_n - \Gamma_1 - \dots - \Gamma_{p-1} = I_n + \sum_{j=1}^{p-1} jA_{j+1}$ and $C = \beta_{\perp}(\alpha'_{\perp}\Psi\beta_{\perp})^{-1}\alpha'_{\perp}$, we have

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

where ξ_t is an $I(0)$ process. These properties follow from Granger’s representation theorem. Further details including a precise expression of ξ_t are given in Johansen (1995, pp. 49–52).

Multiplying (2.1) by $A(L) = I_n - A_1L - \dots - A_pL^p = I_n\Delta - \Pi L - \Gamma_1\Delta L - \dots - \Gamma_{p-1}\Delta L^{p-1}$ yields

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

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 $j = 1, \dots, p-1$. 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 two different estimators of the break date in detail, and then we will consider cointegration tests based on a model with estimated break date in Section 4.

3. SHIFT DATE ESTIMATION

In the following discussion we consider two different estimators of the shift date τ . The first one is based on estimating an unrestricted VAR model in which the cointegrating rank and 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. At the end of this section we briefly mention a third possible estimator and some of its properties. 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

As discussed previously, our first estimator of τ is based on the model

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

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 that makes it desirable to limit the search to a specific part of the sample period.

Instead of using the determinant of the residual covariance matrix as a criterion function for estimating the break date, one could consider other criteria such as the trace. We have chosen the determinant because it is in line with the Gaussian ML setup (for unknown cointegration rank), which can be viewed as the motivation for the LS estimator of the other parameters. Note, however, that we do not assume y_t to be Gaussian.

We assume 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_* , $\delta = \delta_T = T^a \delta_*$, $a \leq \frac{1}{2}$.

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 et al., 1998). For our purposes a lower bound for a is not needed because for a small shift size the break has no effect on the cointegration tests that will be considered later, even though the break date may be more difficult to estimate in that situation. An increasing shift size is treated here for completeness, and it turns out that it provides interesting insights into the actual behavior 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 $a = \frac{1}{2}$ 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 $\hat{\tau}$ that generalize results presented in Lütkepohl et al. (2004).

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 $a > 1/b$ and $\delta \neq 0$,*

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

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 $1/b < \eta < 1/4$. 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 $a = 0$, Lütkepohl et al. (2004) 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 that 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 this case, 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 $j = 0, \dots, p - 1$ can only occur if $\delta_1 \neq 0$ because $\delta \neq 0$ and $\gamma_0 = \delta - \delta_1$.

The preceding discussion implies that an overspecification of the VAR order will always make the break date estimator $\hat{\tau}$ inconsistent. This observation explains some of the small-sample results of Lütkepohl et al. (2004). 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, because γ_0 is always nonzero

when $\delta \neq 0$ (and $p \geq 1$) reasoning similar to that used previously explains why the break date will asymptotically not be estimated larger than the true one.

The second part of 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 that 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 $a > \eta - \frac{1}{2}$, only require $a > -\frac{1}{2}$ (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 $a > \eta - \frac{1}{2}$ 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 $O_p(T)$.

As mentioned in the introduction, Bai et al. (1998) considered the asymptotic distribution of the break date and found that the resulting interval estimator for the break date depends on the dimension of the system under consideration. Such dependence on the dimension of the model is not obtained with our approach, which provides orders of convergence only.

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

$$\begin{aligned} \Delta y_t = & \nu_0 + \nu_1 t + \delta_1 d_{t-1,\tau} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j,\tau} + \Pi y_{t-1} \\ & + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \quad t = p + 1, \dots, T, \end{aligned} \tag{3.3}$$

where $\delta_1 = -\Pi\delta$, as before, and the γ_j^* are as in (2.7). Thus, we can write (3.3) as

$$\begin{aligned} \Delta y_t = & \nu_0 + \nu_1 t + \left(I_n \Delta d_{t,\tau} - \sum_{j=1}^{p-1} \Gamma_j \Delta d_{t-j,\tau} - \Pi d_{t-1,\tau} \right) \delta + \Pi y_{t-1} \\ & + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \varepsilon_t, \quad t = p + 1, \dots, T. \end{aligned} \tag{3.4}$$

Unlike in the unrestricted model (3.1), the impulse dummies do not appear separately anymore in the representation (3.4) but are included in the term that also involves the shift dummy. Thus, only a single parameter vector δ is associated with all the dummy variables. Consequently, 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 (3.4) by nonlinear LS. The estimator of the break date is then obtained by minimizing an analog of (3.2) with $\hat{\varepsilon}_{i\tau}$ 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 $1/b < \eta < \frac{1}{4}$.*

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.

As a final remark on our two break date estimators we mention that, if the DGP is known to have no deterministic linear trend ($\mu_1 = 0$), the corresponding terms in (3.1), (3.3), and (3.4) may be dropped without changing the convergence rates of our break date estimators.

3.3. Ignoring Dummies in Estimating τ

Lütkepohl et al. (2004) also considered estimating the break date based on the VAR model (3.1) without including the impulse dummies. Thus the resulting break date estimator, say, $\tilde{\tau}$, is actually based on a misspecified model. In the present model framework, where the shift size depends on the sample size, it can in fact be shown that the estimator $\tilde{\tau}$ works well, provided $\delta_1 \neq 0$. More precisely, for $0 < a \leq \frac{1}{2}$, $\tilde{\tau} - \tau = o_p(1)$, and for $a \leq 0$, $\tilde{\tau} - \tau = O_p(T^{-2a/(1-2\eta)})$, where $\eta > 0$ (for details see Saikkonen, Lütkepohl, and Trenkler, 2004). Thus, although $\tilde{\tau}$ is based on a misspecified model, its convergence rate is equally as good as that of the other two estimators, provided $\delta_1 \neq 0$. 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. In the next section we con-

sider 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 Lütkepohl et al. (2004). In the following discussion it is assumed that the break date estimator is either $\hat{\tau}$ or $\hat{\tau}_R$. The objective is to test a pair of hypotheses

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

Lütkepohl et al. propose using the tests suggested by Saikkonen and Lütkepohl (2000a). 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, Lütkepohl et al. 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 LR type cointegration tests can be formed in the usual way by solving the related generalized eigenvalue problem based on the adjusted series (for details see Johansen, 1995, Thm. 6.3). The resulting test statistic will be denoted by $LR^{GLS}(r_0)$ in the following discussion.

The levels parameter μ_0 is not identified in the direction of β_\perp in our model setup, and its estimator is partly determined by the initial values in the procedure underlying the LR^{GLS} test. In fact, the dependence of the LR^{GLS} test on initial values was sometimes found to be relevant in preliminary simulations. A detailed theoretical analysis of the impact of initial values on related unit root tests is provided by Müller and Elliott (2003). Given the dependence of the LR^{GLS} tests on initial values, one may hope to improve the performance of the tests by avoiding the estimation of μ_0 . Therefore we shall also consider another 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.

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 α'_\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 $[\alpha'_\perp : \dots : \alpha'_\perp]$ 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_* . The result is $\theta_* = \beta'_\perp C (\sum_{j=0}^{p-1} \gamma_j^* - \Psi_\beta \theta)$, 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

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

The test will be based on the series

$$\tilde{y}_t^{(0)} = y_t - \tilde{\mu}_1 t - \tilde{\delta} d_{t\hat{\tau}} = \mu_0 + x_t - (\tilde{\mu}_1 - \mu_1) t - \tilde{\delta} d_{t\hat{\tau}} + \delta d_{t\hat{\tau}}, \tag{4.1}$$

which are adjusted for the deterministic trend and the shift term. 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

$$\Delta \tilde{y}_t^{(0)} = \Pi^+ \tilde{y}_{t-1}^{(+)} + \sum_{j=1}^{p-1} \Gamma_j \Delta \tilde{y}_{t-j}^{(0)} + e_{t\tilde{r}}, \tag{4.2}$$

where $\tilde{y}_{t-1}^{(+)} = [\tilde{y}_{t-1}^{(0)'}, 1]'$ 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 by solving the related generalized eigenvalue problem, as before. We will denote the LR statistic for the null hypothesis $\text{rk}(\Pi) = r_0$ by $LR^{PAR}(r_0)$ in the following discussion because only a partial set of parameters associated with the deterministic part is estimated in the first step. Its limiting distribution differs from that of $LR^{GLS}(r_0)$ and also from the one given in Theorem 6.3 of Johansen (1995) for the corresponding LR test statistic. We have the following result for the case where the shift occurs in the cointegrating relations ($\delta_1 \neq 0$) and the shift size increases with the sample size. A proof is also given in the Appendix.

THEOREM 4.1. *Suppose Assumptions 1–3 hold. If $\delta_1 \neq 0$, $0 < a < \frac{1}{2}$, and $H_0(r_0)$ is true,*

$$LR^{GLS}(r_0) \xrightarrow{d} \text{tr} \left\{ \left(\int_0^1 B_*(s) dB_*(s)' \right)' \left(\int_0^1 B_*(s) B_*(s)' ds \right)^{-1} \times \left(\int_0^1 B_*(s) dB_*(s)' \right) \right\}$$

and

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

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 $\int_0^1 B_+(s) dB(s)' - \int_0^1 B_+(s) dsB(1)'$, for example.

Several remarks are worth making regarding this theorem. First, a similar result for their break date estimators and cointegrating rank test was obtained by Lütkepohl et al. (2004) under more restrictive assumptions regarding the break size. The limiting distribution of $LR^{GLS}(r_0)$ in Theorem 4.1 is the same as its earlier counterpart in Lütkepohl et al., whereas the limiting distribution of $LR^{PAR}(r_0)$ differs in that the process $B_+(s)$ appears in place of the Brownian

bridge $B_*(s)$. The reason is of course that an intercept term is included in the auxiliary model on which $LR^{PAR}(r_0)$ is based. On the other hand, the limiting distribution of $LR^{PAR}(r_0)$ 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 our Theorem 4.1. Notice that the term $\int_0^1 B_+(s) dB_*(s)'$ consists of two components. The first one is

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

and the second one is

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

Second, in the case without trend in the model, that is, $\mu_1 = 0$ a priori and hence $\tilde{\mu}_1 = 0$, the processes $B_+(s)$ and $B_*(s)$ in the limiting distributions in Theorem 4.1 can be replaced by $[B(s)', 1]'$ and $B(s)$, respectively. Then the limiting distribution of the test statistic $LR^{PAR}(r_0)$ is the same as the limiting distribution of the corresponding LR test statistic in Theorem 6.3 of Johansen (1995). This result can be proved by making appropriate modifications to the proof of Theorem 4.1 in the Appendix. Moreover, in this case the limiting distribution of $LR^{GLS}(r_0)$ is the same as that of an LR test based on a model without any deterministic terms.

Third, from the proof of Theorem 4.1 it is apparent that the same limiting distributions are 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_0 and μ_1 are estimated in the first step leading to $LR^{GLS}(r_0)$, whereas only μ_1 is estimated in the first step of the $LR^{PAR}(r_0)$ 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 discussion has no effect on the limiting distributions of the cointegration tests. The same result was obtained by Lütkepohl et al. (2004) for $LR^{GLS}(r_0)$ in a more limited model framework. It may be worth emphasizing that such a result will not be obtained if instead of our estimation procedures for the deterministic parameters, the Johansen (1995) ML approach is applied to a model with estimated shift date (see also Johansen, Mosconi, and Nielsen, 2000, for a discussion of the case when the break date is known).

Extensions of our results in different directions are conceivable. In particular, limiting results as in Theorem 4.1 can also be obtained under other assumptions for the shift size. For example, if $\delta_1 \neq 0$, the theorem holds more generally for $a < \frac{1}{2}$. In particular, it holds for $a = 0$, where the shift size does not depend

on the sample size, and for $a < 0$, where the shift size decreases with increasing sample size. In fact, $a = \frac{1}{2}$ is the only case where a different result for the limiting distributions of the cointegration tests may be obtained. To get the same distributions as in Theorem 4.1, we then need the additional assumption that the break date estimator is consistent. This condition is satisfied for $\hat{\tau}_R$ but requires further assumptions for $\hat{\tau}$ (see Theorem 3.1(i)). Proofs for other assumptions regarding the shift are not given here because they require a separate treatment of different cases, which complicates the presentation. For details see, however, the discussion paper version of this article (Saikkonen et al., 2004). We have treated the case where the shift actually occurs in the cointegrating relations and the shift size may be large because in this case our theory can help to explain some simulation results of Lütkepohl et al. (2004), as we will see in Section 5.

It also 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 and Lütkepohl (2000a). 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.

An extension of our framework to the case where a break occurs not only in the levels of the series but also in the trend slopes may be desirable for applied work. However, such an extension is not straightforward, and the limiting distribution of the cointegrating rank tests is likely to be affected by the break date in this case.

To apply the cointegrating rank tests we need critical values for the second limiting distribution in Theorem 4.1. The limiting distribution of $LR^{GLS}(r_0)$ is the same as in Lütkepohl et al. (2004), and critical values are available in Lütkepohl and Saikkonen (2000, Table 1). The second limiting distribution in Theorem 4.1 is simulated numerically by approximating the standard Brownian motions with T -step random walks of dimension $n - r_0$, as in Johansen (1995, Sect. 15.1). The percentiles in Table 1 are based on sample lengths of $T = 1,000$ using independent standard normal variates for the error terms and 100,000 replications of the simulation experiment. The computations are done with GAUSS V5.

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 cointegra-

TABLE 1. Percentiles of limiting distribution of $LR^{PAR}(r_0)$

$n - r_0$	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

tion test procedures. 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, and Saikkonen, 2001):

$$x_t = A_1 x_{t-1} + \varepsilon_t = \begin{bmatrix} \psi & 0 \\ 0 & I_{n-r} \end{bmatrix} x_{t-1} + \varepsilon_t,$$

$$\varepsilon_t \sim \text{i.i.d. } N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{n-r} \end{bmatrix} \right), \tag{5.1}$$

where $\psi = \text{diag}(\psi_1, \dots, \psi_r)$ and Θ are $(r \times r)$ and $(r \times (n - r))$ matrices, respectively. As shown by Toda, this type of process 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 that leave such tests invariant. Obviously, if $|\psi_i| < 1$ ($i = 1, \dots, r$) 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 $\Pi = -(I_n - A_1) = \text{diag}(\psi - I_r, 0)$ and, thus, $\delta_1 = -\Pi\delta$ can only be nonzero if level shifts occur in stationary components of the DGP.

Samples are simulated by starting with initial values of zero and discarding the first 50 observations. We have considered a sample size of $T = 100$. The number of replications is 1,000. Thus, the standard error of an estimator of a true rejection probability P is $s_P = \sqrt{P(1-P)/1,000}$, for example, $s_{0.05} = 0.007$. Moreover, we use different VAR orders p , although the true order is $p = 1$, to explore the impact of this quantity on the estimation and testing results. In all simulations the search procedures are applied to all possible break points τ from the fifth up to and including the 96th observation. This choice corresponds to the situation where no prior knowledge on the break date is available. Therefore a search is performed over the full sample period except for some observations at the beginning and at the end. Recall that our theoretical results exclude the possibility of a break at the very beginning or at the very end of the sample. Leaving out 4% of the observations at both ends is to some extent arbitrary. Because we will consider VAR orders up to $p = 3$, a break closer than three periods to the end of the sample results in one or more impulse dummies in (3.1) being zero throughout the sample period and therefore can not be handled in our setup. We decided to stop the search close to the end of the feasible period and treat the beginning and the end of the sample symmetrically in this respect. In practice, some prior knowledge on the break date may be available that can be used to narrow the period where a search is necessary. In that case it may be easier for an estimator to find the true break date, and, hence, the results for the break date estimation and cointegration testing may improve relative to those obtained in our simulations.

To compute the estimator $\hat{\tau}_R$ we use a nonlinear LS estimation method by applying the Gauss–Newton algorithm to minimize the sum of squared residuals corresponding to (3.4). The iterations of the algorithm stop if the change in $D_{\hat{\tau}_R} = \det[(T - p)^{-1} \sum_{t=p+1}^T \hat{\varepsilon}_{t\tau}^R \hat{\varepsilon}_{t\tau}^{R'}]$ from iteration i to $i + 1$ is less than $(T - p)^{-n}$, where $\hat{\varepsilon}_{t\tau}^R$ ($t = p + 1, \dots, T$) are the residual vectors from the nonlinear estimation of (3.4). 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.

The interpretation of the simulation results is done in two steps. First, we analyze the ability of the shift date estimators to locate the true break point. Second, we discuss the small-sample properties of the corresponding cointegration tests based on these estimators. This discussion includes a comparison of the LR^{PAR} and LR^{GLS} test procedures.

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 $\tau = 50$. Afterward, we comment briefly on the importance of the value of τ , the innovation correlation, and the results of a four-dimensional DGP with two cointe-

gration relations without presenting detailed results. The latter DGP has been considered to study the properties of the procedures in the case of more complicated processes. Finally, we examine situations where $\delta_1 = -\Pi\delta = 0$ and, hence, according to Theorems 3.1(i) and 3.2(i), consistent estimation requires larger shift sizes.

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 Figure 1a. 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 before, we have $\delta_1 = \Pi\delta = \alpha\beta'\delta \neq 0$ in (3.1) and, hence, $\theta = \beta'\delta \neq 0$ in (2.7).

The performance of the estimators $\hat{\tau}_R$ and $\hat{\tau}$ is very similar, although the former estimator is more successful in finding the correct break date for small shift magnitudes. Only if $\delta_{(1)} = 3$ and $\delta_{(1)} = 5$ does $\hat{\tau}$ perform slightly better. For large values of $\delta_{(1)}$ both estimators perform identically. In fact, the cases of $\delta_{(1)} = 3$ and $\delta_{(1)} = 5$ represent the few exceptions in all our simulation experiments where $\hat{\tau}$ outperforms $\hat{\tau}_R$. These observations also hold if one considers the small band $[\tau - 2; \tau + 2]$ instead of the single true value of τ only to evaluate the break date estimator. The frequency of break date estimates $\hat{\tau}$ and $\hat{\tau}_R$ in the interval $[\tau - 2; \tau + 2]$ is denoted by $\hat{\tau}(\text{band})$ and $\hat{\tau}_R(\text{band})$, respectively, in Figure 1. Obviously, the frequency of finding the true τ increases for

TABLE 2. Break date estimates for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, VAR order $p = 1$, true break point $\tau = 50$, sample size $T = 100$, $\delta_{(2)} = \delta_{(3)} = 0$

Break date estimates	$\delta_{(1)} = 1$	$\delta_{(1)} = 2$	$\delta_{(1)} = 3$	$\delta_{(1)} = 5$	$\delta_{(1)} = 7$	$\delta_{(1)} = 10$
$\hat{\tau}$ (Ignoring nonlinear restrictions)						
<48	0.425	0.205	0.011	0.000	0.000	0.000
=48/49	0.031	0.030	0.003	0.000	0.000	0.000
=50	0.092	0.572	0.967	1.000	1.000	1.000
=51/52	0.015	0.003	0.000	0.000	0.000	0.000
>52	0.437	0.190	0.019	0.000	0.000	0.000
$\hat{\tau}_R$ (Considering nonlinear restrictions)						
<48	0.417	0.172	0.023	0.000	0.000	0.000
=48/49	0.023	0.010	0.001	0.000	0.000	0.000
=50	0.109	0.641	0.947	0.999	1.000	1.000
=51/52	0.016	0.007	0.003	0.000	0.000	0.000
>52	0.435	0.170	0.026	0.001	0.000	0.000

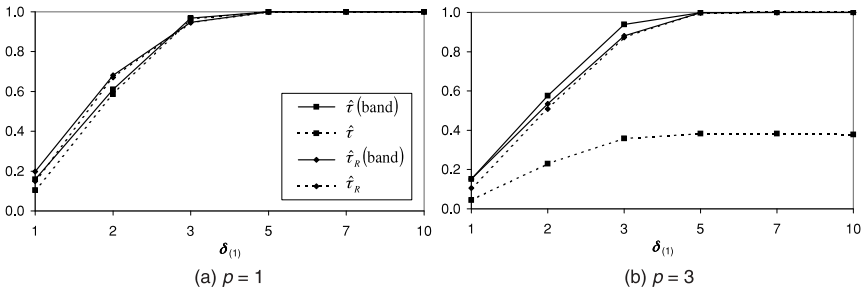


FIGURE 1. Relative frequency of true break point estimates ($\hat{\tau}$, $\hat{\tau}_R$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}$ (band), $\hat{\tau}_R$ (band)) for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, sample size $T = 100$, true break point $\tau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

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 1 to 10 may be interpreted as changing a or δ_* in Assumption 3 accordingly.

Next, we have fitted a VAR(3) model although the true DGP has only an order $p = 1$, and we give the results in Table 3 and Figure 1b. In this case $j_0 = 0$ in Theorem 3.1(i) because $\gamma_1 = \gamma_2 = 0$. Thus, under the conditions of that theorem, $Pr\{48 \leq \hat{\tau} \leq 50\} \rightarrow 1$. In line with this result, $\hat{\tau}$ does not find the true break point $\tau = 50$ even if $\delta_{(1)}$ is large. In fact, we observe in Table 3 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, focusing on the band $[\tau - 2; \tau + 2]$, $\hat{\tau}$ is slightly more successful than $\hat{\tau}_R$ for $\delta_{(1)} = 2$ and 3 (see Figure 1b).

To analyze possible effects of the location of τ we also studied break points $\tau = 10, 25, 75$, and 90 using the same three-dimensional DGP as before. We found that it is only slightly more difficult to detect the more extreme break points if small shift magnitudes are considered. In the case of large shift magnitudes the location of the break date becomes even less important for the estimation results. These observations were made for both break date estimators and both VAR orders $p = 1$ and 3.

Next, we 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. The absence of instantaneous error term correlation made it more difficult for both estimators to locate the true break point regardless of the lag order. This outcome can be explained by the fact that we considered a shift only in one of the three components so that a weaker link of the components owing to $\Theta = (0, 0)$ complicated the break date search. In this situation, $\hat{\tau}_R$ was always

TABLE 3. Break date estimates for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, VAR order $p = 3$, true break point $\tau = 50$, sample size $T = 100$, $\delta_{(2)} = \delta_{(3)} = 0$

Break date estimates	$\delta_{(1)} = 1$	$\delta_{(1)} = 2$	$\delta_{(1)} = 3$	$\delta_{(1)} = 5$	$\delta_{(1)} = 7$	$\delta_{(1)} = 10$
$\hat{\tau}$ (Ignoring nonlinear restrictions)						
<48	0.436	0.226	0.028	0.000	0.000	0.000
=48/49	0.083	0.353	0.595	0.629	0.640	0.626
=50	0.042	0.194	0.350	0.371	0.360	0.374
=51/52	0.014	0.002	0.000	0.000	0.000	0.000
>52	0.425	0.225	0.027	0.000	0.000	0.000
$\hat{\tau}_R$ (Considering nonlinear restrictions)						
<48	0.400	0.231	0.060	0.004	0.000	0.000
=48/49	0.023	0.010	0.005	0.000	0.000	0.000
=50	0.091	0.510	0.866	0.993	1.000	1.000
=51/52	0.015	0.014	0.006	0.000	0.000	0.000
>52	0.471	0.235	0.063	0.003	0.000	0.000

the most successful procedure, and usually its advantage was even more pronounced than in the case of $\Theta = (0.4, 0.8)$.

Our results with respect to the more complicated four-dimensional DGP with a cointegrating rank $r = 2$ ($\psi_1 = \psi_2 = 0.7$) and $\Theta = ([0.4 \ 0.4]': [0.4 \ 0.4]')$ clearly indicated that the performance of the break date estimators deteriorates in the case of smaller level shifts. We chose a shift vector of the form $\delta = (\delta_{(1)}, \delta_{(2)}, \delta_{(3)}, \delta_{(4)})'$ with $\delta_{(1)}$ ranging again from 1 to 10 and $\delta_{(2)} = \delta_{(3)} = \delta_{(4)} = 0$. By adding a further dimension to the process the importance of the break in only one component is weakened, which may explain the lower number of correct break date estimates. The relative rankings of the two estimation procedures did not change, however.

Finally, we examined 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. First, we considered a three-dimensional process as in the base case but with $\delta_{(3)}$ ranging from 1 to 10 and $\delta_{(1)} = \delta_{(2)} = 0$. Because 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 simulated a case of co-breaking. Second, we used a three-dimensional process with $\psi_1 = 1$ so that the cointegrating rank is $r = 0$. In the 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 Figure 2. Clearly, it was more difficult for both procedures to locate τ in this situation. In particular, the case of no cointegration is much more difficult to deal with than co-breaking (compare Figures 2c and 2d with Figures 2a and 2b). The poor performance of the shift date estimation 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 requires large shift magnitudes if $\delta_1 = 0$.

To sum up, the constrained estimator $\hat{\tau}_R$ is usually at least as good as $\hat{\tau}$ and often superior in finding the true break date. The small-sample results for $\hat{\tau}$ are in line with our asymptotic derivations, which say that this procedure may estimate the break date too early when the VAR order is overspecified.

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 differ very little even in those cases where the break date estimators perform differently. Therefore, we only discuss the outcomes for our three-dimensional base DGP for which we have $\delta_1 = \Pi\delta \neq 0$, the case treated in Theorem 4.1.

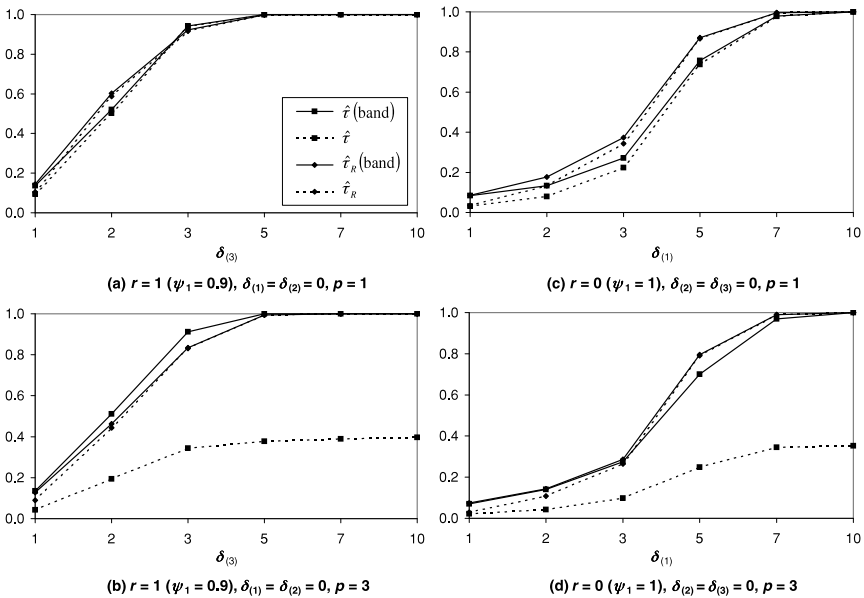


FIGURE 2. Relative frequency of true break point estimates ($\hat{\tau}$, $\hat{\tau}_R$) or of estimates in interval $\tau \pm 2$ ($\hat{\tau}$ (band), $\hat{\tau}_R$ (band)) for three-dimensional DGPs with $\Theta = (0.4, 0.8)$ (a and b), $\Theta = (0, 0)$ (c and d), sample size $T = 100$, true break point $\tau = 50$.

The results are given in Tables 4 and 5. To be precise, we present the rejection frequencies for the null hypothesis $H_0: r = r_0$ when the LR^{PAR} and LR^{GLS} tests are applied to a process with known or estimated shift date. 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 discussion when we refer to this case.

In Table 4 we see for fitted VAR(1) models that in particular the sizes of LR^{PAR} are clearly higher than the nominal 5% level in cases of small shift magnitudes, for which we obtain many incorrect break date locations. 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 . The small-sample powers (not adjusted for the variations in the sizes) are roughly constant if $\hat{\tau}$ is used whereas an increase in power for small values of $\delta_{(1)}$ can be observed for $\hat{\tau}_R$.

TABLE 4. Relative rejection frequencies of cointegration rank tests for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, VAR order $p = 1$, true break point $\tau = 50$, sample size $T = 100$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$

	Known break date	Estimated break date					
		$\delta_{(1)} = 1$	$\delta_{(1)} = 2$	$\delta_{(1)} = 3$	$\delta_{(1)} = 5$	$\delta_{(1)} = 7$	$\delta_{(1)} = 10$
<i>LR^{PAR}</i> cointegration rank test based on $\hat{\tau}$							
$r_0 = 0$	0.596	0.610	0.597	0.598	0.596	0.596	0.596
$r_0 = 1$	0.065	0.116	0.095	0.067	0.065	0.065	0.065
$r_0 = 2$	0.010	0.018	0.009	0.010	0.010	0.010	0.010
<i>LR^{GLS}</i> cointegration rank test based on $\hat{\tau}$							
$r_0 = 0$	0.544	0.546	0.544	0.543	0.544	0.544	0.544
$r_0 = 1$	0.026	0.058	0.046	0.028	0.026	0.026	0.026
$r_0 = 2$	0.003	0.003	0.001	0.003	0.003	0.003	0.003
<i>LR^{PAR}</i> cointegration rank test based on $\hat{\tau}_R$							
$r_0 = 0$	0.596	0.672	0.628	0.603	0.596	0.596	0.596
$r_0 = 1$	0.065	0.104	0.085	0.070	0.065	0.065	0.065
$r_0 = 2$	0.010	0.015	0.011	0.010	0.010	0.010	0.010
<i>LR^{GLS}</i> cointegration rank test based on $\hat{\tau}_R$							
$r_0 = 0$	0.544	0.618	0.579	0.549	0.545	0.544	0.544
$r_0 = 1$	0.026	0.058	0.045	0.027	0.026	0.026	0.026
$r_0 = 2$	0.003	0.003	0.002	0.003	0.003	0.003	0.003

TABLE 5. Relative rejection frequencies of cointegration rank tests for three-dimensional DGP with $r = 1$ ($\psi_1 = 0.9$), $\Theta = (0.4, 0.8)$, VAR order $p = 3$, true break point $\tau = 50$, sample size $T = 100$, nominal significance level 0.05, $\delta_{(2)} = \delta_{(3)} = 0$

	Known break date	Estimated break date					
		$\delta_{(1)} = 1$	$\delta_{(1)} = 2$	$\delta_{(1)} = 3$	$\delta_{(1)} = 5$	$\delta_{(1)} = 7$	$\delta_{(1)} = 10$
<i>LR^{PAR}</i> cointegration rank test based on $\hat{\tau}$							
$r_0 = 0$	0.412	0.450	0.441	0.409	0.379	0.366	0.375
$r_0 = 1$	0.062	0.099	0.084	0.078	0.073	0.072	0.067
$r_0 = 2$	0.017	0.017	0.018	0.019	0.021	0.020	0.017
<i>LR^{GLS}</i> cointegration rank test based on $\hat{\tau}$							
$r_0 = 0$	0.337	0.339	0.323	0.287	0.253	0.241	0.228
$r_0 = 1$	0.030	0.038	0.041	0.028	0.024	0.019	0.015
$r_0 = 2$	0.005	0.005	0.004	0.004	0.005	0.006	0.004
<i>LR^{PAR}</i> cointegration rank test based on $\hat{\tau}_R$							
$r_0 = 0$	0.412	0.484	0.448	0.421	0.412	0.412	0.412
$r_0 = 1$	0.062	0.093	0.069	0.062	0.062	0.062	0.062
$r_0 = 2$	0.017	0.016	0.012	0.016	0.017	0.017	0.017
<i>LR^{GLS}</i> cointegration rank test based on $\hat{\tau}_R$							
$r_0 = 0$	0.337	0.419	0.381	0.348	0.337	0.337	0.337
$r_0 = 1$	0.030	0.038	0.043	0.032	0.030	0.030	0.030
$r_0 = 2$	0.005	0.004	0.003	0.004	0.005	0.005	0.005

The impact of the shift magnitude on the small-sample properties of the cointegration tests is similar if VAR(3) models are fitted and $\hat{\tau}_R$ is used (see Table 5). Note, however, that the tests' small-sample power is clearly lower than in the VAR(1) case even if the true shift date is known. Regarding $\hat{\tau}$ we observe some differences if VAR(3) models are applied (compare Table 5). Here, the sizes of *LR^{GLS}* and powers of both tests fall below the values for a known shift date when $\delta_{(1)}$ is large. Obviously, the effect of the wrong locations on the small-sample properties becomes important if the shift magnitude is large.

Three important observations can be made with respect to the relative small-sample properties of the two cointegration tests *LR^{PAR}* and *LR^{GLS}*. First, the new test *LR^{PAR}* rejects somewhat too often if the null hypothesis is true ($r = 1$), even if the shift date is known. These higher rejection frequencies were also found for other DGPs. Thus, it may be worth exploring small-sample cor-

rections for the new tests in future work. Second, for increasing shift magnitudes the relative performance of LR^{PAR} and LR^{GLS} based on the estimator $\hat{\tau}_R$ is in general similar. Third, using $\hat{\tau}$ the drop in small-sample power in the case of large shift magnitudes when the VAR order is overspecified is more pronounced for LR^{GLS} than for LR^{PAR} . Thus, our new test proposal appears to be less affected by the incorrect break date estimates.

The overall conclusion from our simulations is that the estimator $\hat{\tau}_R$ is generally preferable to $\hat{\tau}$. Taking account of the nonlinear restrictions is beneficial to both locating the shift date and 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 $\hat{\tau}_R$ is used. Given the size distortions of LR^{PAR} even if the shift date is known, size correcting procedures may be worth exploring for this test in the future.

6. CONCLUSIONS

We have analyzed the asymptotic properties of two estimators for the shift date in a cointegrated VAR process with level shift. The shift is modeled by a simple shift dummy variable. The first estimator is based on an unrestricted VAR model, and the second one is obtained by taking into account the relation between the parameters of the stochastic and deterministic parts of the model. Asymptotic properties of both 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 Lütkepohl et al. (2004), who also consider the first estimator 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 cointegrating rank tests. These tests differ from those considered by Lütkepohl et al. 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 the adjusted series. The asymptotic distributions of the tests are 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 the most successful one in locating the true shift date. Moreover, a superior break date estimator tends to improve the small-sample properties of subsequent cointegration tests. 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 Lütkepohl et al. (2004) and the new tests of the present paper shows, however, that the latter tend to reject a true

null hypothesis more often than the Lütkepohl et al. 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.

Another direction for extending our results may be to develop a joint procedure for determining the break date and the cointegrating rank. Our two-step procedure, where the break date is estimated in the first stage and then tests for the cointegrating rank are performed, is easy to use in practice and therefore has some appeal in applied work. An alternative approach may be to determine the break date and cointegrating rank jointly, for instance by minimizing a model selection criterion. We leave such extensions for future research.

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.
- Gregory, A.W. & B.E. Hansen (1996) Residual-based tests for cointegration in models with regime shifts. *Journal of Econometrics* 70, 99–126.
- Hubrich, K., H. Lütkepohl, & P. Saikkonen (2001) A review of systems cointegration tests. *Econometric Reviews* 20, 247–318.
- Johansen, S. (1995) *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. 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.
- Lütkepohl, H. & P. Saikkonen (2000) Testing for the cointegrating rank of a VAR process with a time trend. *Journal of Econometrics* 95, 177–198.
- Lütkepohl, H., P. Saikkonen, & C. Trenkler (2004) Testing for the cointegrating rank of a VAR process with level shift at unknown time. *Econometrica* 72, 647–662.
- Müller, U.K. & G. Elliott (2003) Tests for unit roots and the initial condition. *Econometrica* 71, 1269–1286.
- 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 non-linear time trends in cointegrating relations. *Econometric Theory* 17, 296–326.
- Saikkonen, P. & H. Lütkepohl (2000a) Testing for the cointegrating rank of a VAR process with structural shifts. *Journal of Business & Economic Statistics* 18, 451–464.
- Saikkonen, P. & H. Lütkepohl (2000b) Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process. *Journal of Time Series Analysis* 21, 435–456.
- Saikkonen, P. & H. Lütkepohl (2002) Testing for a unit root in a time series with a level shift at unknown time. *Econometric Theory* 18, 313–348.
- Saikkonen, P., H. Lütkepohl, & C. Trenkler (2004) Break Date Estimation and Cointegration Testing in VAR Processes with Level Shift. European University Institute, Florence, Discussion paper ECO 2004/21.
- 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.

APPENDIX: Proofs

Some parts of the proofs are similar to those of the corresponding results stated in Lütkepohl et al. (2004) 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 $T \rightarrow \infty$. When obtaining weak convergences in a product space of $D([\underline{\lambda}, \bar{\lambda}])$ we frequently make use of results given in Appendix A.1 of Gregory and 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 that 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., $\mu_{0o}, \mu_{1o}, \tau_o$, 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

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

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

$$\begin{aligned} \Delta x_t &= \nu_0^{(0)} + \nu_1^{(0)}t + \delta_1 d_{tr} + \gamma d_{tr} - \delta_1^{(0)} d_{tr_o} - \gamma^{(0)} d_{tr_o} + \Pi x_{t-1} \\ &\quad + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \quad t = p + 1, p + 2, \dots \end{aligned} \tag{A.1}$$

Here

$$\nu_0^{(0)} = \nu_0 + \Pi\mu_{0o} - \Psi\mu_{1o} - \Pi\mu_{1o},$$

$$\nu_1^{(0)} = \nu_1 + \Pi\mu_{1o},$$

$$\gamma = [\gamma_0 : \dots : \gamma_{p-1}],$$

$$d_{tr} = [\Delta d_{tr} : \dots : \Delta d_{t-p+1, \tau}]',$$

$$\delta_1^{(0)} = -\Pi\delta_o,$$

and

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

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

It will also be convenient to use the transformation $\Pi x_{t-1} = \alpha^{(0)} u_{t-1}^{(0)} + \rho^{(0)} v_{t-1}^{(0)}$, where $u_{t-1}^{(0)} = \beta_o' x_{t-1}$, $v_{t-1}^{(0)} = \beta_{o\perp}' x_{t-1}$, $\alpha^{(0)} = \alpha \beta_o' \beta_o (\beta_o' \beta_o)^{-1}$, and $\rho^{(0)} = \alpha \beta_o' \beta_{o\perp} (\beta_{o\perp}' \beta_{o\perp})^{-1}$. 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

$$\Delta x_t = \nu_0^{(0)} + \nu_1^{(0)} t + \delta_1 d_{tr} + \underline{\gamma} d_{tr} - \delta_1^{(0)} d_{tr_o} - \underline{\gamma}^{(0)} d_{tr_o} + \alpha^{(0)} u_{t-1}^{(0)} + \rho^{(0)} v_{t-1}^{(0)} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \quad t = p + 1, p + 2, \dots \tag{A.2}$$

Denote $q_{tr} = [d_{tr} : d_{tr}']'$ and

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

With this notation (A.2) becomes

$$\Delta x_t = \Phi w_t^{(0)} + \Xi q_{tr} - \Xi^{(0)} q_{tr_o} + \varepsilon_t, \quad t = p + 1, p + 2, \dots, \tag{A.3}$$

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

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

Then

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

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,

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

The definition of $\varepsilon_{tr}(\Theta)$ (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 previously. Before turning to this issue we note that the preced-

ing 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

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

and

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

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

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

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

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

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

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

and

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

Hence,

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

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

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

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 $t = [T\bar{\lambda}] + p, \dots, T$. 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

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

Then $w_t^{(0)} = [w_{1t}^{(0)'} : w_{2t}^{(0)'}]'$, 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 $\Phi_2 = [\alpha^{(0)} : \Gamma_1 : \dots : \Gamma_{p-1}]$. With these definitions,

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

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

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

where

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

and

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

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 $\tau \geq \tau_o$ (cf. Bai et al., 1998).

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

LEMMA A.3. *There exists a constant $c_1 > 0$ such that, with probability approaching one and uniformly in $[T\lambda] \leq \tau \leq \tau_o$ 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 $t \leq \tau - 1$, $\varepsilon_{1t}(\Theta) = -\Phi_1 w_{1t}^{(0)}$ and, consequently,

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

For L_1 we have

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

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 $[T\lambda] \leq \tau \leq \tau_o$ for the second eigenvalue can be seen by using an analog of (A.3) of Gregory and 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 and Hansen (1996, p. 118), the norm in the middle of the last expression is of order $O_p(1)$ uniformly in $[T\lambda] \leq \tau \leq \tau_0$ 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

$$\varepsilon_{1t\tau}(\Theta) = -\Psi_1 w_{1t}^{(0)} - \gamma d_{t\tau}, \quad t = \tau, \dots, \tau_o - 1,$$

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

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

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

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

Proof. By the definitions,

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

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

$$\begin{aligned} L_3 &= \text{tr} \left(\Omega^{-1} \Psi_1 \sum_{t=\tau}^{\tau_o-1} \underline{z}_{1t\tau}^{(0)} \underline{z}_{1t\tau}^{(0)'} \Psi_1' \right) \\ &\geq \lambda_{\min}(\Omega^{-1}) \text{tr} \left(\Psi_1 D_{1T} \left(D_{1T}^{-1} \sum_{t=\tau}^{\tau_o-1} \underline{z}_{1t\tau}^{(0)} \underline{z}_{1t\tau}^{(0)'} D_{1T}^{-1} \right) D_{1T} \Psi_1' \right), \end{aligned} \tag{A.9}$$

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

Next note that

$$T^{-1/2} \sum_{t=\tau}^{\tau_o-1} w_{1t}^{(0)} \Delta d_{t-i,\tau} = O_p(T^{-1/2}), \quad i = 0, \dots, p - 1,$$

uniformly in $[T\lambda] \leq \tau < \tau_o$. Because $w_{1t}^{(0)} = [1 : t/T : T^{-1/2}v_{t-1}^{(0)}]'$ 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 from the fact that $T^{-1/2}\beta'_{o\perp} x_{[Ts]}$ obeys an invariance principle. Thus, we can conclude that

$$D_{1T}^{-1} \sum_{t=\tau}^{\tau_o-1} \underline{z}_{1T}^{(0)} \underline{z}_{1T}^{(0)'} D_{1T}^{-1} = \text{diag} \left[T^{-1} \sum_{t=\tau}^{\tau_o-1} w_{1t}^{(0)} w_{1t}^{(0)'} : I_p \right] + o_p(1) \tag{A.10}$$

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

The next step is to observe that

$$T^{-1} \sum_{t=[T\underline{\lambda}]}^{[T\underline{\lambda}_o]-1} w_{1t}^{(0)} w_{1t}^{(0)'} \Rightarrow M_{11}(\lambda_o) - M_{11}(\underline{\lambda}), \quad \underline{\lambda} \leq \lambda < \lambda_o,$$

where $M_{11}(\lambda)$ is the weak limit of $T^{-1} \sum_{t=p+1}^{[T\underline{\lambda}]-1} w_{1t}^{(0)} w_{1t}^{(0)'}$ (cf. (A.3) of Gregory and 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 preceding 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,

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

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 previously mentioned inequality about L_4 , conclude from the definitions that

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

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

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

where $a_{3T} = o_p(T^\eta)$, $1/b < \eta < \frac{1}{4}$, 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 and Lütkepohl (2002). The inequalities obtained for $|L_{41}|$ and $|L_{42}|$ 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

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

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

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

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 the following lemma.

LEMMA A.5. *There exists a constant $c_3 > 0$ such that, with probability approaching one and uniformly in $[T\lambda] \leq \tau \leq \tau_o$ 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,

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

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

$$\begin{aligned} L_5 &\geq \lambda_{\min}(\Omega^{-1}) \sum_{t=\tau_o}^{\tau_o+p-1} \|\varepsilon_{1t\tau}(\Theta)\|^2 \\ &\geq \bar{\omega}^{-1} \sum_{t=\tau_o}^{\tau_o+p-1} \|\Psi_2 w_{1t}^{(0)}\|^2 + \bar{\omega}^{-1} \sum_{t=\tau_o}^{\tau_o+p-1} \|\zeta_{t\tau}^{(0)}\|^2 + 2\bar{\omega}^{-1} \sum_{t=\tau_o}^{\tau_o+p-1} \zeta_{t\tau}^{(0)'} \Psi_2 w_{1t}^{(0)}. \end{aligned}$$

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

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

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) by 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

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

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

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

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

Proof. The proof is similar to that of Lemma A.3. ■

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 $[T\lambda] \leq \tau \leq \tau_o - 1$ and $(\Theta, \tau, \Omega) \in B_1$,*

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

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

Proof. By the definitions,

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

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

$$L_7 = \text{tr} \left(\Omega^{-1} \Psi_2 \sum_{t=\tau}^{\tau_o-1} w_{1t}^{(0)} w_{1t}^{(0)'} \Psi_2' \right) + \text{tr} \left(\Omega^{-1} \sum_{t=\tau}^{\tau_o-1} \zeta_{t\tau}^{(0)} \zeta_{t\tau}^{(0)'} \right) + 2\text{tr} \left(\Omega^{-1} \sum_{t=\tau}^{\tau_o-1} \zeta_{t\tau}^{(0)} w_{1t}^{(0)'} \Psi_2' \right) \stackrel{\text{def}}{=} L_{71} + L_{72} + L_{73}.$$

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

$$|L_{73}| \leq 2 \|\Omega^{-1}\| \|T^{1/2} \Psi_2\| \left\| T^{-1/2} \sum_{t=\tau}^{\tau_o-1} \zeta_{t\tau}^{(0)} w_{1t}^{(0)'} \right\| \leq 2 \|\Omega^{-1}\| \left(\frac{\tau_o - \tau}{T} \right)^{1/2} \|T^{1/2} \Psi_2\| \left((\tau_o - \tau)^{-1} \sum_{t=\tau}^{\tau_o-1} \|w_{1t}^{(0)}\|^2 \right)^{1/2} \left(\sum_{t=\tau}^{\tau_o-1} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2}.$$

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 $[T\lambda] \leq \tau < \tau_o$. Hence,

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

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,

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

Now consider L_8 , for which we have

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

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

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

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 and Hansen (1996, p. 118).

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

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

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 $[T\lambda] \leq \tau < \tau_o - p$ 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 and Lütkepohl (2002).

Combining the preceding discussion of L_{82} shows that

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

where $a_{71T} = O_p((\tau_o - \tau)^\eta)$ in the required uniform sense and the equality follows from definitions. Because 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

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

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 and also 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 $1/b < \eta < \frac{1}{4}$ 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 $[T\lambda] \leq \tau \leq \tau_o$.

Proof. By the definitions and Lemma A.2,

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

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 $[T\lambda] \leq \tau \leq \tau_o$.

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 previously, 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 and 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} ,

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

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} ,

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

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 is at most $\sqrt{2}\varphi_T$. Thus, uniformly in B_{21} ,

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

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} ,

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

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 $-(1/4c_5)[a_{7T}((\tau_o - \tau)/T)^\eta + a_{8T}((\tau_o - \tau)/T)^{1/2} \|T^{1/2-\eta}\Psi_2\|]^2$. Thus, expanding the square and inserting the result on the r.h.s. of the preceding inequality yields, uniformly in B_{22} ,

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

where

$$\begin{aligned}
 c_{4T}(\tau) &= c_4 - \frac{a_{8T}^2}{4c_5} \left(\frac{\tau_o - \tau}{T} \right), \\
 a_{10T}(\tau) &= T^{-\eta}a_{6T} + T^{-\eta}a_{9T} + \frac{a_{7T}a_{8T}}{2c_5} \left(\frac{\tau_o - \tau}{T} \right)^{1/2+\eta},
 \end{aligned}$$

and

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

Note that here a_{6T}, \dots, 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. Because we are free here 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 that make the previ-

ous 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 that can be assumed as large as we wish and $o_p(1)$ is replaced by a fixed negative number that, 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 $B_3 = \{(\Theta, \tau, \Omega) : \|T^{1/2-\eta}(\Phi_2 - \Phi_{2o})\| \leq \epsilon\}$, where $1/b < \eta < \frac{1}{4}$ 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 $[T\lambda] \leq \tau \leq \tau_o$.

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 $[T\lambda] \leq \tau \leq [T(\lambda_o - \epsilon_1)]$ 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 $T^{-2\eta}l_{1T}(\Theta, \tau, \Omega) \geq o_p(1)$ holds uniformly. Next note that the inequality (A.23) is valid for $[T(\lambda_o - \epsilon)] < \tau \leq \tau_o$ 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, $T^{-2\eta}l_{1T}(\Theta, \tau, \Omega) \geq -\epsilon_2$, where $\epsilon_2 > 0$ can be chosen arbitrarily small. From the preceding discussion and the first equality in (A.18) it thus follows that we need to show that, for some $\epsilon_* > 0$,

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

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 $\zeta_{tr}^{(0)}$ introduced for Lemma A.5.

LEMMA A.10. Let $B_4 = \{(\Theta, \tau, \Omega) : (\tau_o - \tau)^{-2\eta} \sum_{i=\tau}^{\tau_o+p-1} \|\zeta_{ir}^{(0)}\|^2 \leq M^2\}$, where $\tau < \tau_o$ and $1/b < \eta < \frac{1}{4}$ 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 $M_1 > 0$,

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

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

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

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, $\xi_\tau^2 = \sum_{i=\tau}^{\tau_o+p-1} \|\zeta_{ir}^{(0)}\|^2$. From the lower bound obtained for $l_{1T}(\Theta, \tau, \Omega)$ previously we can then further obtain

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

Again, this holds uniformly in $B_1 \cap B_2 \cap B_3 \cap B_4^c$ and $[T\lambda] \leq \tau \leq \tau_o - 1$. 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. ■

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 $t = p + 1, \dots, \tau_o - 1$ and break dates $\tau_o \leq \tau \leq [T\bar{\lambda}]$ but is otherwise similar to Lemma A.3.

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

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

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

Instead of the time points $t = \tau_o + p, \dots, \tau - 1$ it is next reasonable to consider the time points $t = \tau_o + p, \dots, \tau + p - 1$. 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

$$\begin{aligned} \varepsilon_{1\tau}(\Theta) &= -\Phi_1 w_{1t}^{(0)} + d_{\tau_o} \delta_1^{(0)} - d_{\tau} \delta_1 - \underline{\gamma}d_{\tau} \\ &= -\Psi_1^{(0)} w_{1t}^{(0)} - (d_{\tau} \delta_1 + \underline{\gamma}d_{\tau}), \quad t = \tau_o + p, \dots, \tau + p - 1, \end{aligned}$$

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 before, the former is defined by using $\delta_1^{(0)}$ instead of δ_1 in Ψ_1 . The parameter γ 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 $t = \tau_o + p, \dots, \tau_o + p - 1$. Here the preceding expression of $\varepsilon_{1\tau}(\Theta)$ is modified to the form

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

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

The counterpart of Lemma A.6 is straightforward. The relevant time points are $t = \tau, \dots, T$, 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 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 $1/b < \eta < \frac{1}{4}$ 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 $B_{50} = \{(\Theta, \tau, \Omega) : T^{-2\eta} \sum_{t=p+1}^T \|\underline{\gamma}_{d_{t\tau}} - \underline{\gamma}_o \underline{d}_{t\tau_o}\|^2 \leq M\}$.

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

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

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 $t = \tau + p, \dots, \tau_o - 1$, we have on $B_1 \cap B_2 \cap B_3 \cap B_5^c$,

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

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

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

Thus, in place of $\delta_1^{(0)}$ we now have 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

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

where $\mathcal{I}_{\tau, \tau_o} = \{t \mid t \in \{\tau, \dots, \tau_o + p - 1\} \text{ and } (\Delta d_{t\tau_o} \neq 0 \text{ or } \Delta d_{t\tau} \neq 0)\}$, that is, 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

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

Notice that here $\Gamma_{j_o}\delta_o = -\gamma_{j_o}$ ($j = 1, \dots, p - 1$) and, because now $\delta_{1o} = 0$, $\delta_o = \gamma_{0o}$. Thus, the sum of the last three terms equals $\gamma \underline{d}_{t\tau} - \gamma_o \underline{d}_{t\tau_o}$, 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 $\delta_1^{(0)} = -\rho^{(0)}\beta'_{o1}\delta_o$ so that, on B_2 , $\|\delta_1^{(0)}\| \leq c\epsilon T^{\eta+a-1}$ for some $0 \leq c < \infty$ (see Lemma A.8). Furthermore, on B_3 , $\|(\Gamma_j - \Gamma_{j_o})\delta_o\| \leq \|\Gamma_j - \Gamma_{j_o}\| \|\delta_o\| \leq \epsilon T^{\eta+a-1/2} \|\delta_s\|$ ($j = 1, \dots, p - 1$) (see Lemma A.9). Using these facts and the triangle inequality we find that

$$\left(T^{-2\eta} \sum_{t \in \mathcal{I}_{\tau, \tau_o}} \|\zeta_{t\tau}^{(0)}\|^2 \right)^{1/2} \geq \left(T^{-2\eta} \sum_{t \in \mathcal{I}_{\tau, \tau_o}} \|\gamma \underline{d}_{t\tau} - \gamma_o \underline{d}_{t\tau_o}\|^2 \right)^{1/2} - const \times T^{a-1/2}\epsilon.$$

On the r.h.s. the summation can be extended to all $t = p + 1, \dots, T$. This means that on $B_1 \cap B_2 \cap B_3 \cap B_5^c$ 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 $\|\delta_1 - \delta_1^{(0)}\| \leq \epsilon T^{1/2-\eta}$ on B_2 it is straightforward to show that the proof can be reduced to a form entirely similar to that in the case $\tau \leq \tau_o$. 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 $\tau_o - p \leq \tau \leq \tau_o + p$. First consider the case $\tau_o - p \leq \tau \leq \tau_o$. If $\gamma_{j_o} = 0$ for all $j = 0, \dots, p - 1$, Lemma A.11 shows that, asymptotically, $\tau_o - p \leq \hat{\tau} \leq \tau_o$, as required. Next suppose that $\gamma_{j_o, o} \neq 0$ and consider the break dates $\tau_o - p \leq \tau \leq \tau_o - p + j_o$. For any of these break dates we have

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

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

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

Because $\gamma_{j_o, o} = -T^a \Gamma_{j_o, o} \delta_* \neq 0$, the last quantity tends to infinity as $T \rightarrow \infty$. Hence, we can conclude from Lemmas A.9 and A.10 that asymptotically the function $l_T(\Theta, \tau, \Omega)$ is not minimized for $\tau \leq \tau_o - p + j_o$. Now consider the case $j_o = 0$. 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, because $\gamma_{0o} = T^a(\delta_* + \alpha_o \beta'_o \delta_*) \neq 0$, the proof given in the case $j_o > 0$ 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 $\zeta_{\tau_o\tau} = -\delta_1 - \gamma_o^{(0)} = -\delta_o + (\delta_1^{(0)} - \delta_1)$, 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.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 $\tau \leq \tau_o - p + j_o$ and $j_o > 0$,

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

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

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 $j_o = 0$.

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 γ in (A.2) are no more freely varying but (smooth) functions of the parameters δ , $\rho^{(0)}$, $\alpha^{(0)}$, and $\Gamma_1, \dots, \Gamma_{p-1}$. Specifically, $\delta_1 = -\Pi\delta = -\alpha^{(0)}\beta'_o\delta - \rho^{(0)}\beta'_{o\perp}\delta$, $\gamma_0 = \delta - \delta_1$, and $\gamma_j = -\Gamma_j\delta$ ($j = 1, \dots, p - 1$). 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

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

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 $\Theta \rightarrow \Theta^{(0)} = [\Phi : \Xi - \Xi^{(0)}]$. Thus, if for simplicity we denote $\underline{z}_t^{(\tau)} = [\Delta y_t^{(\tau)'} : w_t^{(\tau)'} : q_{t\tau}']'$ and $R(\Theta^{(0)}) = [I_n : -\Phi : \Xi - \Xi^{(0)}]$ we can write the relevant objective function as

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

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 $\Gamma_1, \dots, \Gamma_{p-1}$. 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 permis-

sible 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 subsequently.

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 $l_T(\Theta, \tau, \Omega)$ have 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 $z_t^{(\tau)}$.

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

$$\lambda_{\min} \left(D_T^{-1} \sum_{t=p+1}^T z_t^{(\tau)} z_t^{(\tau)'} D_T^{-1} \right) \geq \epsilon_* \tag{A.31}$$

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 $a < \frac{1}{2}$. 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

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

When $a < \frac{1}{2}$ 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 $\Delta y_{t-j}^{(\tau)}$ ($j = 0, \dots, p - 1$). 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 semi-definite 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 $0 < a < \frac{1}{2}$ 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 and Hansen, 1996, and (A.14) of Saikkonen and 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 , and 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 noted previously, we can also assume that $\tau_o - p \leq \tau$. Using the definitions we can express the vector $\zeta_{\tau\tau}^{(0)}$ as

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

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

$$\begin{aligned} \zeta_{\tau\tau}^{(0)} &= -(\delta_1 - \delta_1^{(0)})d_{\tau\tau} + \left(\Delta d_{\tau\tau} - \sum_{j=1}^{p-1} \Gamma_j \Delta d_{\tau-j,\tau} - (\alpha^{(0)}\beta'_o + \rho^{(0)}\beta'_{o\perp})d_{\tau-1,\tau} \right) \delta \\ &\quad - \left(\Delta d_{\tau\tau_o} - \sum_{j=1}^{p-1} \Gamma_j \Delta d_{\tau-j,\tau_o} - (\alpha^{(0)}\beta'_o + \rho^{(0)}\beta'_{o\perp})d_{\tau-1,\tau_o} \right) \delta_o. \end{aligned} \tag{A.33}$$

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 previously mentioned parameter restrictions, we find that

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

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, because 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 $\hat{\tau}_R < \tau_o$ 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

$$\begin{aligned} \zeta_{t\tau}^{(0)} &= -\Delta d_{t\tau}(\delta_1 - \delta_1^{(0)}) - d_{t-1,\tau_o} \delta_1^{(0)} + d_{t-1,\tau} \delta_1^{(0)} + \sum_{j=1}^{p-1} \Delta d_{t-j,\tau_o} (\Gamma_j - \Gamma_{j_o}) \delta_o \\ &+ \left(\Delta d_{t\tau} - \sum_{j=1}^{p-1} \Gamma_j \Delta d_{t-j,\tau} \right) \delta - \left(\Delta d_{t\tau_o} - \sum_{j=1}^{p-1} \Gamma_{j_o} \Delta d_{t-j,\tau_o} \right) \delta_o. \end{aligned} \tag{A.34}$$

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

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

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.34) 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.34). First note that the norm of the first four terms on the r.h.s. of (A.34) 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.34) 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.34) 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, because ϵ_* can be chosen arbitrarily small, the asymptotic behavior of $(T^{-2\eta} \sum_{t=p+1}^T \|\zeta_{t\tau}^{(0)}\|^2)^{1/2}$ is dominated by the last term on the r.h.s. of (A.34), 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. Hence, the proof of Theorem 3.2 is complete.

A.3. Proof of Theorem 4.1. For simplicity we will denote the break date estimator by $\hat{\tau}$. This estimator can be either of the two estimators considered in Section 3 unless explicit distinctions are made. From the assumptions $\delta_1 \neq 0$ and $0 < a < \frac{1}{2}$ and Theorems 3.1 and 3.2 it follows that asymptotically $\tau_o - p \leq \hat{\tau} \leq \tau_o$ can be assumed. This fact will be used in several arguments of the proof without explicit notice.

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_o replaced by the estimator $\hat{\tau}$ satisfy appropriate consistency properties. This replacement changes the VECM (2.7) to

$$\Delta y_t = \nu + \alpha(\beta' y_{t-1} - \phi(t-1) - \theta d_{t-1, \hat{\tau}}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j^* \Delta d_{t-j, \hat{\tau}} + \varepsilon_{t\hat{\tau}}, \tag{A.35}$$

where

$$\varepsilon_{t\hat{\tau}} = \varepsilon_t + \alpha_o \beta_o' \delta_o (d_{t-1, \hat{\tau}} - d_{t-1, \tau_o}) - \sum_{j=0}^{p-1} \gamma_{jo}^* (\Delta d_{t-j, \hat{\tau}} - \Delta d_{t-j, \tau_o}). \tag{A.36}$$

Write

$$y_t = \mu_{0o} + \mu_{1o} t + \delta_o d_{t\hat{\tau}} + y_t^{(\hat{\tau})}, \tag{A.37}$$

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

$$\begin{aligned} \Delta y_t^{(\hat{\tau})} &= \nu^{(0)} + \alpha(\beta' y_{t-1}^{(\hat{\tau})} - \phi^{(0)}(t-1) - \theta^{(0)} d_{t-1, \hat{\tau}}) \\ &+ \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j}^{(\hat{\tau})} + \sum_{j=0}^{p-1} \gamma_j^{*(0)} \Delta d_{t-j, \hat{\tau}} + \varepsilon_{t\hat{\tau}}, \end{aligned} \tag{A.38}$$

where $\nu^{(0)} = \nu + \alpha \beta' \mu_{0o} - \Psi \mu_{1o}$, $\phi^{(0)} = \phi - \beta' \mu_{1o}$, $\theta^{(0)} = \theta - \beta' \delta_o$, $\gamma_0^{*(0)} = \delta - \delta_o$, and $\gamma_j^{*(0)} = \gamma_j^* + \Gamma_j \delta_o$ ($j = 1, \dots, p-1$). Note that the true values of these parameters are zero. RR estimators of the parameters in (A.38) are obtained by transforming the RR estimators based on (A.35) 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

subsequently. 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.13. *Under the conditions of Theorem 4.1, $\tilde{\beta}_0 - \beta_o = O_p(T^{-1})$, $\tilde{\Phi}^{(0)} = O_p(T^{-3/2})$, $\tilde{\theta}^{(0)} = O_p(T^{-1/2})$, $\tilde{\alpha}_0 - \alpha_o = O_p(T^{-1/2})$, $\tilde{\Gamma}_j - \Gamma_{j_o} = O_p(T^{-1/2})$, $\tilde{\nu}^{(0)} = O_p(T^{-1/2})$, $\tilde{\gamma}_j^{*(0)} = O_p(1)$, $j = 0, \dots, p - 1$, $\tilde{\Omega} - \Omega = O_p(T^{-1/2})$.*

Proof. We first note that the result of Lemma A.13 also holds when the break date is assumed known. A formal proof of this can be obtained by following the proof of Lemma 2.1 of Saikkonen and Lütkepohl (2000a) and observing that the omission of some impulse dummies from the model considered by Saikkonen and Lütkepohl is of no significance and that the same is true for the dependence of the parameter δ on the sample size. The latter fact is clear because the results of Lemma A.13 are formulated by using the transformed model (A.38) in which the true values of the deterministic parameters are zero.

Because the result of Lemma A.13 holds when the break date is assumed known it also holds when the break date can be consistently estimated, that is, when $\hat{\tau} - \tau_o = o_p(1)$. Indeed, then the analysis can be restricted to that part of the sample space where $\hat{\tau} = \tau_o$ holds and the probability of this can be made arbitrarily close to unity for all T large enough. This proves the results of the lemma for the constrained estimator $\hat{\tau}_R$.

If $j_0 = p - 1$ in Theorem 3.1(i) the preceding argument also applies to the unconstrained estimator $\hat{\tau}$. 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 $\tau_o - p + 1 + j_0 \leq \tau \leq \tau_o$. 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) a correctly specified model if we only redefine the parameters $\gamma_0^*, \dots, \gamma_{p-1}^*$ as $\gamma_0^* = \alpha\beta'\delta$, $\gamma_1^* = \delta$, and $\gamma_j^* = -\Gamma_{j-1}\delta$, $j = 2, \dots, p - 1$. By assumption we then have $\gamma_{p-1}^* \neq 0$ whereas $\Gamma_{p-1}\delta = 0$. 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^* ($j = 0, \dots, p - 1$) 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. ■

Properties of the new estimators of the deterministic parameters. 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 one of the estimators $\hat{\tau}$ or $\hat{\tau}_R$.

LEMMA A.14. *Under the conditions of Theorem 4.1, the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ have the following properties:*

$$\beta'(\tilde{\mu}_1 - \mu_{1_o}) = O_p(T^{-3/2}), \tag{A.39}$$

$$T^{1/2}\beta'_\perp(\tilde{\mu}_1 - \mu_{1_o}) \xrightarrow{d} N(0, \beta'_\perp C\Omega C'\beta_\perp), \tag{A.40}$$

$$\beta'(\tilde{\delta} - \delta_o) = O_p(T^{-1/2}), \tag{A.41}$$

$$\beta'_\perp(\tilde{\delta} - \delta_o) = O_p(1). \tag{A.42}$$

Proof. We start with the results (A.39) and (A.40). 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

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

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

$$\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_{1o}) = \tilde{\beta}'_\perp \tilde{C}(\tilde{\nu}^{(0)} - \tilde{\Psi}_\beta\tilde{\phi}^{(0)}).$$

Here and in what follows the subscript 0 is omitted from the estimators of α and β to simplify the notation. By Lemma A.13, one obtains from the previous equality

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

Note that the estimator $\tilde{\nu}^{(0)}$ can be viewed as the LS estimator of the parameter $\nu^{(0)}$ in the auxiliary regression model obtained by replacing $\beta'y_{t-1}^{(\hat{\tau})} - \phi^{(0)}(t-1) - \theta^{(0)}d_{t-1,\hat{\tau}}$ in (A.38) by its observed analog $\beta'y_{t-1}^{(\hat{\tau})} - \tilde{\phi}^{(0)}(t-1) - \tilde{\theta}^{(0)}d_{t-1,\hat{\tau}}$. This implies that $\tilde{\beta}'_\perp \tilde{C}\tilde{\nu}^{(0)}$ can be obtained by LS from the auxiliary regression model

$$\tilde{\beta}'_\perp \tilde{C}\Delta y_t^{(\hat{\tau})} = \Lambda \mathbf{p}_{t\hat{\tau}} + \text{error}, \tag{A.43}$$

where $\mathbf{p}_{t\hat{\tau}} = [1, \Delta y_{t-1}^{(\hat{\tau})}, \dots, \Delta y_{t-p+1}^{(\hat{\tau})}, d_{t\hat{\tau}}]'$, Λ is a conformable coefficient matrix, and the error has the representation $\tilde{\beta}'_\perp \tilde{C}\varepsilon_{t\hat{\tau}} - \tilde{\beta}'_\perp \tilde{C}\alpha[(\tilde{\beta} - \beta)y_{t-1}^{(\hat{\tau})} + \tilde{\phi}^{(0)}(t-1) + \tilde{\theta}^{(0)}d_{t-1,\hat{\tau}}]$. By the definition of C and Lemma A.13, $\tilde{\beta}'_\perp \tilde{C}\alpha = O_p(T^{-1/2})$. Using this fact, Lemma A.13, and the assumptions, it is straightforward to show that the asymptotic properties of the LS estimator of the parameter Λ in the auxiliary regression model (A.43) 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.36)) further show that the error can be assumed to be $\tilde{\beta}'_\perp \tilde{C}\varepsilon_t$ or even $\beta'_\perp C\varepsilon_t$. Because it is also straightforward to show that the estimation of the intercept term in (A.43) is asymptotically independent of the estimation of the other regression coefficients we can conclude that

$$T^{1/2}\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_{1o}) = \beta'_\perp CT^{-1/2} \sum_{t=p+1}^T \varepsilon_t + o_p(1).$$

This and a standard central limit theorem yield

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

To obtain (A.40) we need to show that $\tilde{\beta}'_\perp$ on the l.h.s. can be replaced by β'_\perp . To see this, write

$$\begin{aligned} (\tilde{\beta}'_\perp - \beta'_\perp)(\tilde{\mu}_1 - \mu_{1o}) &= (\tilde{\beta}'_\perp - \beta'_\perp)\tilde{\beta}(\tilde{\beta}'\tilde{\beta})^{-1}\tilde{\beta}'(\tilde{\mu}_1 - \mu_{1o}) \\ &\quad + (\tilde{\beta}'_\perp - \beta'_\perp)\tilde{\beta}'_\perp(\tilde{\beta}'_\perp\tilde{\beta}'_\perp)^{-1}\tilde{\beta}'_\perp(\tilde{\mu}_1 - \mu_{1o}). \end{aligned} \tag{A.44}$$

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 $\tilde{\beta}'(\tilde{\mu}_1 - \mu_{1o}) = \tilde{\beta}^{(0)} = O_p(T^{-3/2})$ by Lemma A.13. From this last result one can obtain (A.39) because $\tilde{\beta}$ can be replaced by β using an argument similar to that used in (A.44).

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

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

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.13 implies that the r.h.s. of this equality is of order $O_p(1)$. Moreover, $\tilde{\beta}'(\tilde{\delta} - \delta_o) = \tilde{\theta}^{(0)} = O_p(T^{-1/2})$. Thus, (A.41) and (A.42) follow because in these results $\tilde{\beta}$ and $\tilde{\beta}_{\perp}$ can be replaced by β and β_{\perp} , respectively, by using an argument similar to that in (A.44). This completes the proof of Lemma A.14. ■

Proof of the limiting distribution of LR^{PAR} . The structure of our proof of the limiting distribution of $LR^{PAR}(r_0)$ is similar to that of Theorem 11.1 of Johansen (1995). Therefore we just outline the arguments in the following discussion.

First note that the limiting distribution of the test statistic $LR^{PAR}(r_0)$ can be derived by assuming that the true value of the parameter μ_0 is zero. Thus, we can write equation (4.1) as

$$\tilde{y}_t^{(0)} = x_t - (\tilde{\mu}_1 - \mu_{1o})t - (\tilde{\delta} - \delta_o) d_{t\tau_o} - (\tilde{\delta} - \delta_o)(d_{t\tilde{\tau}} - d_{t\tau_o}) - \delta_o(d_{t\tilde{\tau}} - d_{t\tau_o}). \tag{A.45}$$

Using this representation, the assumption $a < \frac{1}{2}$, and the asymptotic properties of the estimators $\tilde{\mu}_1$ and $\tilde{\delta}$ obtained in Lemma A.14, we can now mimic the proof given in Johansen (1995, pp. 158–160) and see that all the quantities that therein converge in probability to constants will here converge in probability to the same constants. However, quantities that 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

$$T^{-1/2}\beta'_{\perp}\tilde{y}_{[Ts]}^{(0)} = T^{-1/2}\beta'_{\perp}x_{[Ts]} - T^{-1/2}\beta'_{\perp}(\tilde{\mu}_1 - \mu_{1o})t + o_p(1) \xrightarrow{d} \beta'_{\perp}C(W(s) - sW(1)) \stackrel{def}{=} \beta'_{\perp}CW_+(s), \tag{A.46}$$

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.46), first consider the equality. Because $\beta'_{\perp}(\tilde{\delta} - \delta_o) = O_p(1)$ by (A.42) of Lemma A.14, it is clear that the contribution of the third and fourth terms on the

r.h.s. of (A.45) to $T^{-1/2}\beta'_\perp\tilde{y}_{[T\tau]}^{(0)}$ is asymptotically negligible. The same argument also applies to the fifth term on the r.h.s. of (A.45) because $a < \frac{1}{2}$. As for the weak convergence in (A.46), it can be justified by a standard functional central limit theorem and (A.40) of Lemma A.14 by observing that the limit of the second expression is determined by the process ε_t (see Johansen, 1995, eqn. (B.24), and the proof of (A.40)).

The preceding discussion implies that the limiting distribution of the test statistic $LR^{PAR}(r_0)$ can be derived by ignoring the last three terms on the r.h.s. of (A.45). This means that in the same way as in Saikkonen and Lütkepohl (2000a) we have reduced the problem to that of no break studied by Saikkonen and Lütkepohl (2000b). From Lemma A.14 and the proof of Theorem 3 of Saikkonen and Lütkepohl (2000b) it can be seen that, when $\mu_{0o} = 0$ is assumed, the trace test statistic in that theorem is asymptotically equivalent to a similar test statistic based on an analog of (4.2) 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, because 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 of Saikkonen and Lütkepohl (2000b). Technical details, which are similar to the corresponding two cases in Johansen (1995, Sect. 11.2), are straightforward and will be omitted.

Asymptotic properties of the GLS estimators of the deterministic parameters. Because the break date estimator is asymptotically between $\tau_o - p$ and τ_o it is straightforward to follow the proof of Theorem 2.1 of Saikkonen and Lütkepohl (2000a) (case $a_1 < 1$) and obtain asymptotic properties of the GLS estimators of the parameters μ_0 , μ_1 , and δ . Denoting these GLS estimators by $\tilde{\mu}_0$, $\tilde{\mu}_1$, and $\tilde{\delta}$, it can be demonstrated that (A.39)–(A.42) of Lemma A.14 hold for $\tilde{\mu}_1$ and $\tilde{\delta}$ except that in (A.42) $O_p(T^a)$ replaces $O_p(1)$. For $\tilde{\mu}_0$ it follows that $\beta'(\tilde{\mu}_0 - \mu_{0o}) = O_p(T^{-1/2})$ and $\beta'_\perp(\tilde{\mu}_0 - \mu_{0o}) = O_p(T^a)$.

Limiting distribution of LR^{GLS} . The test statistic $LR^{GLS}(r_0)$ is defined as $LR^{PAR}(r_0)$ except that now $\tilde{y}_t^{(0)} = y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t - \tilde{\delta} d_{t\tilde{\tau}}$. Instead of (A.45) we therefore have

$$\begin{aligned} \tilde{y}_t^{(0)} &= x_t - (\tilde{\mu}_0 - \mu_{0o}) - (\tilde{\mu}_1 - \mu_{1o})t - (\tilde{\delta} - \delta_o) d_{t\tilde{\tau}} \\ &\quad - (\tilde{\delta} - \delta_o)(d_{t\tilde{\tau}} - d_{t\tau_o}) - \delta_o(d_{t\tilde{\tau}} - d_{t\tau_o}), \end{aligned} \tag{A.47}$$

where the estimators on the r.h.s. satisfy the rates of convergence obtained previously. It is straightforward to check that, under the conditions of Theorem 4.1, the rates of convergence obtained for $\tilde{\delta}$ are sufficient for the fifth term on the r.h.s. of (A.47) to have no effect on the asymptotic properties of the second sample moments on which the test statistic is based, and the same is true for the last term, which is as in (A.45). Thus, the problem reduces to that of a known break date studied by Saikkonen and Lütkepohl (2000a), and the limiting distribution of the test statistic is obtained from Theorem 3.1 of that paper. Here it suffices to note the following points. First, the dependence of the break size on the sample size has no effect because the needed arguments only involve the difference $\tilde{\delta} - \delta_o$. Second, it is not difficult to check that the rate of convergence $\beta'_\perp(\tilde{\delta} - \delta_o) = o_p(T^{1/2})$ suffices instead of $\beta'_\perp(\tilde{\delta} - \delta_o) = O_p(1)$, which could be used in Saikkonen and Lütkepohl (2000a) and the same is true for $\beta'_\perp(\tilde{\mu}_0 - \mu_{0o}) = o_p(T^{1/2})$. Thus, we have demonstrated that, under the conditions of Theorem 4.1, the test statistic has the same limiting distribution as in Saikkonen and Lütkepohl (2000a).