Econometric Theory, **17**, 2001, 327–356. Printed in the United States of America.

STATISTICAL INFERENCE IN COINTEGRATED VECTOR AUTOREGRESSIVE MODELS WITH NONLINEAR TIME TRENDS IN COINTEGRATING RELATIONS

Pentti Saikkonen
University of Helsinki

This paper continues the work of Saikkonen (2001, *Econometric Theory* 17, 296–326) and develops an asymptotic theory of statistical inference in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations and general nonlinear parameter restrictions. Inference on parameters in cointegrating relations and short-run dynamics is studied separately. It is shown that Gaussian maximum likelihood estimators of parameters in cointegrating relations have mixed normal limiting distributions and that related Wald, Lagrange multiplier, and likelihood ratio tests for general nonlinear hypotheses have usual asymptotic chi-square distributions. These results are shown to hold even if parameters in the short-run dynamics are not identified. In that case suitable estimators of the information matrix have to be used to justify the application of Wald and Lagrange multiplier tests, whereas the likelihood ratio test is free of this difficulty. Similar results are also obtained when inference on parameters in the short-run dynamics is studied, although then Gaussian maximum likelihood estimators have usual normal limiting distributions. All results of the paper are proved without assuming existence of second partial derivatives of the likelihood function, and in some cases even differentiability with respect to nuisance parameters is not required.

1. INTRODUCTION

This paper is a sequel to Saikkonen (2001), which proved the existence and consistency of a Gaussian maximum likelihood (ML) estimator in a cointegrated vector autoregressive (VAR) model with nonlinear time trends in cointegrating relations. The practical motivation of the model was also discussed in the previous paper and more thoroughly in Ripatti and Saikkonen (2001; see also Heinesen, 1997). The treatment of Saikkonen (2001) was very general and allowed for nonlinear parameter restrictions in both the cointegrating relations and

The first version of this paper was completed while the author worked at the Research Department of the Bank of Finland, whose hospitality is gratefully acknowledged. The paper has benefited from comments of two anonymous referees and the co-editor Bruce Hansen. Address correspondence to: Pentti Saikkonen, Department of Statistics, P.O. Box 54 (Unioninkatu 37), SF-00014 University of Helsinki, Finland.

 0266-4666/01 $9.50

short-run dynamics. Further generality was achieved by proving consistency results without assuming identifiability of all underlying structural parameters. The object of this sequel is to complete the previous work by developing a corresponding asymptotic theory of statistical inference. The test procedures to be considered include the three classical likelihood based tests, namely, the Wald, score, or Lagrange multiplier (LM), and likelihood ratio (LR) tests.

In the same way as Saikkonen (2001), this paper also proves asymptotic results without assuming identifiability of all nuisance parameters. This relates our work to the recent literature on partially identified models (see Phillips, 1989; Choi and Phillips, 1992). In particular, limiting distributions of ML estimators of parameters in cointegrating vectors and related tests will be derived even if parameters in the short-run dynamics are not identified and vice versa. To some extent, identifiability of the trend parameters can also be dispensed with although, as a result of the assumed nonlinearity of the trend, the situation here is more difficult than in the case of other parameters.

It will be shown in the paper that, when the parameters of interest are identified, their ML estimators have mixed normal limiting distributions that reduce to usual normal limiting distributions in the case of short-run parameter estimators. A consequence of this is that the conventional asymptotic chi-square criterion applies to related Wald, LM, and LR tests. Proofs of these results are more complicated than usual, however, because identifiability of nuisance parameters is not assumed. This particularly holds for proving the limiting distributions of the LR tests of the paper. An advantage of the LR test over corresponding Wald and LM tests is, however, that choosing an estimator of the information matrix is not required. This choice is not quite obvious when identifiability of nuisance parameters is not assumed, although reasonable choices can be given. A further theoretical feature of the paper is that commonly used assumptions of the existence of second partial derivatives of the likelihood function are not needed and in some cases even differentiability with respect to nuisance parameters can be dispensed with.

The paper is organized as follows. Section 2 briefly presents the model and required assumptions. For a more thorough treatment and motivating discussions refer to Section 2 of Saikkonen (2001). Limiting distributions of ML estimators are derived in Section 3 and those of Wald, LM, and LR tests in Section 4. Section 5 concludes. Proofs are given in the Appendix. Finally, the employed notation is fairly standard and is discussed at the end of Section 1 of Saikkonen (2001).

2. MODEL AND ASSUMPTIONS

Consider the s-dimensional time series y_t, $t = 1,\dots,T$, which is integrated of order one $(I(1))$ and cointegrated with cointegrating rank r $(0 < r < s)$. Following Saikkonen (2001) we assume that the underlying data generation process is a Gaussian VAR process of order p. When written in the error correction

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

form the coefficient matrices of this VAR process can be general nonlinear functions of underlying structural parameters, and a nonlinear time trend depending on a parameter vector is allowed in the cointegrating relations. Specifically, the error correction form of the considered model is

$$\Delta y_t = \alpha(\psi)[\beta(\phi)' y_{t-1} - g_t(\mu)] + \sum_{j=1}^{p-1} B_j(\psi)\Delta y_{t-j} + \varepsilon_t, \qquad t = 1,\dots,T, \quad \textbf{(2.1)}$$

where $\beta(\phi) = [I_r \quad -A(\phi)]'$ $(s \times r)$ is the matrix of normalized cointegrating vectors, $\alpha(\psi)$ $(s \times r)$ and $B_j(\psi)$ $(s \times s)$ are matrices of short-run parameters with $\alpha(\psi)$ of rank r, and $g_t(\mu)$ $(r \times 1)$ is a deterministic time trend. Furthermore, the initial values $y_{-p+1},\dots,y_0$ are observable, and ε_t is Gaussian white noise, that is, $\varepsilon_t \sim NID(0,\Omega)$ with $\Omega > 0$. Conditions that guarantee that (2.1) really defines an $I(1)$ process with $\beta(\phi)' y_{t-1}$ stationary cointegrating relations are well known and can be found in Johansen (1995, Theorem 4.2) and Saikkonen (2001, Assumption 1). Here these conditions are always assumed, and the acronyms $I(1)$ model and $I(1)$ process are used to indicate this. Because we are not interested in inference on the cointegrating rank, r, we shall assume that the value of this parameter is known.

Thus, we assume that the cointegrating vectors are defined by the matrix function $A(\phi)$ $(r \times (s-r))$ and the parameter vector ϕ $(k \times 1)$ whereas the short-run dynamics are determined by the matrix functions $\alpha(\psi)$ and $B_j(\psi)$ $(j = 1,\dots,p-1)$ and the parameter vector ψ $(\ell \times 1)$. Conditions needed to prove the consistency of the ML estimators of these parameters are given in Assumptions 2 and 3 of Saikkonen (2001). These conditions include continuity of the previously mentioned functions and appropriate identification conditions that are not stated explicitly here because, by the results of the previous paper, needed consistency results will be assumed. Instead, we will only give conditions required for asymptotic inference and start with the following assumption.

Assumption 1. The true parameter value ϕ_0 is an interior point of the parameter space $\Phi \subset \mathbb{R}^k$ and in some neighborhood of ϕ_0 the function $A(\phi)$ is continuously differentiable with the matrix $\partial \operatorname{vec} A(\phi_0)/\partial\phi'$ of full column rank.

Assumption 1 is standard in the theory of asymptotic inference, although we do not need the existence of second partial derivatives of $A(\phi)$, which is often assumed. It may be noted that Assumption 1 is not very far from implying Assumption 2 of Saikkonen (2001). To see this, notice that the differentiability assumption entails that the matrix $\partial \operatorname{vec} A(\phi)/\partial\phi'$ is of full column rank in some neighborhood of ϕ_0. This implies that the function $A(\phi)$ has a continuous inverse in a compact set containing $A(\phi_0)$ but not necessarily a neighborhood of $A(\phi_0)$ (see Bartle, 1964, p. 252). Thus, the identification condition of Assumption 2(b) of Saikkonen (2001), and hence Assumption 2 as a whole, holds if we further assume that the inverse of $A(\phi)$ is defined in some neighborhood of

$A(\phi_0)$ and $A(\phi_0)$ is an interior point in the relative topology of $\overline{A(\phi)}$, the closure of $A(\Phi)$.

The next assumption is concerned with the short-run parameters of the model. We denote $B(\psi) = [B_1(\psi) \dots B_{p-1}(\psi)\alpha(\psi)]$.

Assumption 2. The true parameter value ψ_0 is an interior point of the parameter space $\Psi \subset \mathbb{R}^\ell$, and in some neighborhood of ψ_0 the function $B(\psi)$ is continuously differentiable with the matrix $\partial \operatorname{vec} B(\psi_0)/\partial\psi'$ of full column rank.

Assumption 2 is entirely similar to Assumption 1 and therefore standard in the theory of asymptotic inference. It is related to Assumption 3 of Saikkonen (2001) in the same way as Assumption 1 is related to Assumption 2 of that paper.

Now consider the trend term $g_t(\mu)$. In the same way as in Saikkonen (2001) we partition the parameter vector μ as $\mu = [\nu' \quad \gamma']'$ and, for theoretical reasons, make the additional structural assumption

$$g_t(\mu) = d(\nu) f_t(\gamma), \tag{2.2}$$

where $d(\nu)$ $(r \times r_1)$ is a time invariant matrix function of the parameter vector ν $(n \times 1)$, whereas $f_t(\gamma)$ $(r_1 \times 1)$ is a time dependent vector function of the parameter vector γ $(q \times 1)$. The sequence $f_t(\gamma)$ is specified by assuming that $f_t(\gamma) = f(t/T;\gamma)$ where $f(\cdot;\gamma)$ is a suitable function defined on the interval [0,1]. Of course, the sequence $g_t(\mu)$ is then defined by the function $g(\cdot;\mu) = d(\nu)f(\cdot;\gamma)$ so that $g_t(\mu) = g(t/T;\mu)$. The dependence of the trend term, and hence the process $y_t (t = 1,\dots,T)$, on the sample size T is not made explicit because this feature has no essential effect on asymptotic derivations (cf. Saikkonen, 2001, particularly Sect. 2.3).

Now we can give conditions required for the trend model. It is convenient to formulate some of them by using the function $g(x;\mu)$ without being explicit about the structural assumption (2.2). In the following assumption we refer to Condition 1 of Saikkonen (2001), which is not repeated here. Simple sufficient conditions for it will be discussed shortly.

Assumption 3. Let $f(x;\gamma)$ be a function from $[0,1] \times \Gamma$ to $\mathbb{R}^{r_1}$ where Γ is a compact subset of $\mathbb{R}^q$. Let $\mu = [\nu' \quad \gamma']' \in M = N \times \Gamma$ where $N \subset \mathbb{R}^n$ and let $f_t(\gamma) = f(t/T,\gamma)$ and $g_t(\mu) = d(\nu)f(t/T;\gamma)$, $t = 1,\dots,T$. Assume that the following conditions hold.

(a) The true parameter value μ_0 is an interior point of M, and in some neighborhood of μ_0 the function $g(x;\mu) = d(\nu)f(x;\gamma)$ is continuously differentiable with respect to μ.
(b) The function formed from $f(x;\gamma)$ and $\partial f(x;\gamma)/\partial\gamma_i$ $(i = 1,\dots,q)$ satisfies Condition 1 of Saikkonen (2001) when the value of γ belongs to a compact neighborhood of γ_0.
(c) The matrix $\int_0^1 [\partial g(x;\mu_0)'/\partial\mu][\partial g(x;\mu_0)/\partial\mu']dx$ is of full rank.

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

Unlike with other parameter spaces, we have found it convenient to retain the compactness assumption imposed on Γ in Saikkonen (2001). The differentiability condition in Assumption 3(a) is standard, although even here we can do without assuming the existence of second partial derivatives. The conditions in Assumption 3(b) were also used in Saikkonen (2001) to obtain orders of consistency. They imply that sample moments that appear in the derivation of limiting distributions converge in an appropriate way. Assumption 3(c) is a variant of a conventional rank condition that guarantees that estimators have nonsingular limiting distributions. Note that here, and also in other similar situations, we use the notation $\partial g(x;\mu_0)'/\partial\mu = (\partial g(x;\mu_0)/\partial\mu')'$ (cf. Lütkepohl, 1996, p. 173).

Assumption 3 is related to Assumptions 4 and 5 of Saikkonen (2001). In particular, because Assumptions 3(a) and (c) imply that the matrix $\partial \operatorname{vec} d(\nu)/\partial\nu'$ is of full column rank in some neighborhood of the true parameter value ν_0 a reasoning similar to that following Assumption 1 shows that Assumption 4 of Saikkonen (2001) nearly holds. However, except for the compactness of the parameter space Γ, the conditions in Assumption 5 of Saikkonen (2001) are much stronger than required in the present Assumption 3. Because consistency of ML estimators will here be assumed we only need local assumptions on the likelihood function instead of corresponding global assumptions needed in the consistency proofs of Saikkonen (2001). The nonlinearity of the sequence $f_t(\gamma)$ in both the time index t and the parameter γ makes the global assumptions necessarily much stronger.

As already mentioned, we will not give details of Condition 1 of Saikkonen (2001). A simple sufficient condition for Assumption 3(b) to hold is that the vector function of (x,γ) formed from $f(x;\gamma)$ and $\partial f(x;\gamma)/\partial\gamma_i$ $(i = 1,\dots,q)$ is continuously differentiable. In the case of conventional dummy variables the function $f(x;\gamma)$ is discontinuous, but because it is known and hence independent of γ these cases can also be included. The most important case that is excluded is that of structural breaks with unknown break dates or dummy variables with dates of jump depending on unknown parameters.

3. LIMITING DISTRIBUTIONS OF ML ESTIMATORS

Consider model (2.1) with $g_t(\mu)$ as in (2.2). Conditioning on the initial values $y_{-p+1},\dots,y_0$ minus two times the logarithm of the likelihood function can be written as

$$\ell_T = T\log\det(\Omega) + \operatorname{tr}\left\{\Omega^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t'\right\}, \tag{3.1}$$

where ε_t is as in (2.1) and is interpreted as a function of the unknown parameters therein. When ML estimators exist they minimize the function ℓ_T over permissible values of the parameters. In the same way as in Saikkonen (2001) we consider a minimization problem and refer to ℓ_T as the likelihood function. It is obvious from (2.1) and (2.2) that the likelihood function only depends on the structural parameters ϕ, ψ, and ν through the corresponding reduced form parameters $A(\phi)$, $B(\psi)$, and $d(\nu)$. It will occasionally be convenient to express the likelihood function in terms of some of the reduced form parameters, and, when this is done, appropriate arguments will be added to ℓ_T and ε_t. On the other hand, we will usually drop the argument from the reduced form parameters $A(\phi)$, $B(\psi)$, and $d(\nu)$ when their dependence on the underlying structural parameters is irrelevant. A similar notational convention applies to estimators.

Theorem 3.1 and Corollary 3.1(c) of Saikkonen (2001) show that, under appropriate regularity conditions, ML estimators of the parameters A, B, d, and γ exist with probability approaching one and are consistent. Then a consistent ML estimator of the error covariance matrix Ω also exists and is defined by

$$\hat{\Omega}_T = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'. \tag{3.2}$$

Here $\hat{\varepsilon}_t$ denotes ε_t evaluated at $\hat{A}_T$, $\hat{B}_T$, $\hat{d}_T$, and $\hat{\gamma}_T$, ML estimators of the parameters A, B, d, and γ, respectively. After having established the consistency of these ML estimators Saikkonen (2001) deduced the consistency of the related structural parameter estimators $\hat{\phi}_T, \hat{\psi}_T$, and $\hat{\nu}_T$ by using suitable identification conditions. However, these identification conditions were not needed to prove the consistency of the estimators $\hat{A}_T$, $\hat{B}_T$, $\hat{d}_T$, and $\hat{\gamma}_T$. Some of the subsequent results can also be proved without assuming identifiability of all structural parameters.

Now define the parameter vector $\vartheta = [\phi' \quad \nu' \quad \gamma' \quad \psi']'$ $((k + n + q + \ell) \times 1)$ and set $\vartheta_1 = [\phi' \quad \mu']'$ where $\mu = [\nu' \quad \gamma']'$ as before. We shall first derive the limiting distribution of the ML estimator of ϑ_1. For this result we only need the consistency of the reduced form parameter estimator $\hat{B}_T$, whereas consistent estimation of the corresponding structural parameter ψ is not required. In asymptotic analyses it is therefore relevant to consider the likelihood function as a function of the parameters ϑ_1, B, and Ω and use the notation $\ell_T = \ell_T(\vartheta_1, B, \Omega)$ and $\varepsilon_t = \varepsilon_t(\vartheta_1, B)$. Note that $B \in B(\Psi)$ and that we explicitly have

$$\varepsilon_t(\vartheta_1, B) = \Delta y_t - \alpha[y_{1,t-1} - A(\phi)y_{2,t-1} - g_t(\mu)] - \sum_{j=1}^{p-1} B_j \Delta y_{t-j}, \tag{3.3}$$

where $y_t = [y_{1t}' \quad y_{2t}']'$ is partitioned into $r \times 1$ and $(s-r) \times 1$ subvectors. We also need further notation. Let $W_0(x)$ be a Brownian motion with covariance matrix Ω_0 and define $W_2(x) = e_2' \Lambda_0 W_0(x)$ where $e_2' = [0 \quad I_{s-r}]$ and Λ_0 signifies the true value of the parameter matrix $\Lambda = \beta_\perp[\alpha_\perp'(I_s - \sum_{j=1}^{p-1} \Gamma_j)\beta_\perp]^{-1}\alpha_\perp'$. Note that Λ is the moving average impact matrix that appears in Johansen's (1995, p. 49) version of Granger's representation theorem and its counterpart given in Saikkonen (2001, Sect. 2.3). If we finally define

$$F_1(x;\vartheta_1) = -\begin{bmatrix} (\partial \operatorname{vec} A(\phi)'/\partial\phi)(W_2(x) \otimes I_r) \\ \partial g(x;\mu)'/\partial\mu \end{bmatrix}$$

we can state the following theorem. Unless otherwise stated, all limits in this paper assume that $T \to \infty$.

THEOREM 3.1. *Let $y_1, \ldots, y_T$ be generated by the $I(1)$ process* (2.1) *where $g_t(\mu)$ is as in* (2.2). *Let Assumptions* 1 *and* 3 *hold and suppose that ML estimators of the parameters ϕ, μ, B, and Ω exist with probability approaching one and are consistent. Then,*

$$\begin{bmatrix} T(\hat{\phi}_T - \phi_0) \\ T^{1/2}(\hat{\mu}_T - \mu_0) \end{bmatrix} \Rightarrow \left(\int_0^1 F_1(\chi;\vartheta_{10})\alpha_0'\Omega_0^{-1}\alpha_0 F_1(\chi;\vartheta_{10})'d\chi \right)^{-1} \times \int_0^1 F_1(\chi;\vartheta_{10})\alpha_0'\Omega_0^{-1}dW_0(x).$$

Remark 3.1. Sufficient conditions for the existence and consistency of the ML estimators of the parameters ϕ, μ, B, and Ω are given in Theorem 3.1 and Corollary 3.1(a) and (c) of Saikkonen (2001). These conditions are explicitly stated in Assumptions 1, 2(a) and (b), 3(a), 4(a) and (b), and 5(a)–(d) of that paper.

An interesting feature in Theorem 3.1 is that it only assumes that the ML estimator of the reduced form parameter $B = B(\psi)$ is consistent but consistent estimation of the corresponding structural parameter ψ is not assumed. In fact, the sufficient conditions mentioned in Remark 3.1 only include continuity of the function $B(\psi)$, but identifiability of the structural parameter ψ is not assumed, and neither is differentiability of the function $B(\psi)$. Thus, we have obtained the limiting distribution of the ML estimators $\hat{\phi}_T$ and $\hat{\mu}_T$ even if the short-run parameter ψ is not identified and, therefore, cannot be consistently estimated. An intuitive explanation behind this result is that the information matrix between the parameters ϑ_1 and ψ is block diagonal so that these parameters can be estimated independently of each other. Therefore, difficulties with the estimation of ψ do not interfere with the estimation of ϑ_1. On the other hand, in the case of the parameters ϕ and μ the block diagonality of the information matrix does not hold, so that it is difficult to weaken the identification conditions implicitly assumed in Theorem 3.1 and, for example, derive the limiting distribution of $\hat{\phi}_T$ when consistency of the estimator $\hat{\mu}_T$ fails. It is easy to check that the Brownian motions $W_2(x)$ and $\alpha_0'\Omega_0^{-1}W_0(x)$ are uncorrelated and therefore independent. This implies the independence of the processes $F_1(x;\vartheta_{10})$ and $\alpha_0'\Omega_0^{-1}W_0(x)$ so that the limiting distribution in Theorem 3.1 is mixed normal. As is well known, this fact has useful implications on hypothesis testing.

We shall next obtain the limiting distribution of the short-run parameter estimator $\hat{\psi}_T$. In this context consistent estimation of the structural parameters ϕ

and ν is not required, so that it is relevant to replace them by the corresponding reduced form parameters $A \in A(\Phi)$ and $d \in d(N)$ in the definition of the likelihood function. Thus, we use the notation $\ell_T(\underline{\vartheta}_1, \psi, \Omega)$ and $\varepsilon_t = \varepsilon_t(\underline{\vartheta}_1, \psi)$ where, for simplicity, $\underline{\vartheta}_1 = (A, d, \gamma)$. An explicit expression of $\varepsilon_t(\underline{\vartheta}_1, \psi)$ is given by

$$\varepsilon_t(\underline{\vartheta}_1, \psi) = \Delta y_t - \alpha(\psi) u_{t-1}(\underline{\vartheta}_1) - \sum_{j=1}^{p-1} B_j(\psi) \Delta y_{t-j}, \tag{3.4}$$

where $u_{t-1}(\underline{\vartheta}_1) = y_{1,t-1} - Ay_{2,t-1} - df_t(\gamma)$. We also define $z_t(\underline{\vartheta}_1) = [\Delta y'_{t-1} \dots \Delta y'_{t-p+1} u_{t-1}(\underline{\vartheta}_1)']'$ and write $z_t = z_t(\underline{\vartheta}_{10})$ with $\underline{\vartheta}_{10}$ the true value of $\underline{\vartheta}_1$. Now we can state the following theorem.

THEOREM 3.2. *Let $y_1, \dots, y_T$ be generated by the $I(1)$ process* (2.1) *where $g_t(\mu)$ is as in* (2.2). *Let Assumptions* 2 *and* 3(a) *and* (b) *hold and suppose that* ML *estimators of the parameters ψ, A, d, γ, and Ω exist with probability approaching one and are consistent with $\hat{A}_T = A_0 + O_p(T^{-1})$, $\hat{d}_T = d_0 + O_p(T^{-1/2})$, and $\hat{\gamma}_T = \gamma_0 + O_p(T^{-1/2})$. Then,*

$$T^{1/2}(\hat{\psi}_T - \psi_0) \Rightarrow N(0, \Sigma_\psi(\psi_0, \Omega_0)),$$

where

$$\Sigma_\psi(\psi, \Omega) = \left(\partial \operatorname{vec} B(\psi)'/\partial\psi \left(\operatorname{plim}_T T^{-1} \sum_{t=1}^{T} z_t z_t' \otimes \Omega^{-1} \right) \partial \operatorname{vec} B(\psi)/\partial\psi' \right)^{-1}$$

and the weak limit is independent of the weak limit in Theorem 3.1.

Remark 3.2. Sufficient conditions for the existence of the ML estimators of the parameters ψ, A, d, γ, and Ω and also for all the consistency results assumed in Theorem 3.2 are given in Theorem 3.1, Corollary 3.1(b) and (c), and Proposition 3.2 of Saikkonen (2001). These conditions are explicitly stated in Assumptions 1, 2(a), 3(a) and (b), 4(a), 5(a)–(d), and 6 of that paper. The probability limit in the definition of the matrix $\Sigma_\psi(\psi, \Omega)$ exists by Lemma A.1(a) of Saikkonen (2001).

Thus, Theorem 3.2 only assumes that the ML estimators of the reduced form parameters $A = A(\phi)$ and $d = d(\nu)$ are consistent, but consistency of the corresponding structural parameter estimators $\hat{\phi}_T$ and $\hat{\nu}_T$ is not assumed. The sufficient conditions given in Remark 3.2 only include continuity of the functions $A(\phi)$ and $d(\nu)$, but identifiability of the structural parameters ϕ and ν is not assumed and neither is differentiability of the functions $A(\phi)$ and $d(\nu)$. Thus, the result of Theorem 3.2 holds even if the structural parameters ϕ and ν are not identified and, therefore, cannot be consistently estimated. Also, because the consistency of the estimator $\hat{d}_T$ does not require differentiability of the function $d(\nu)$ the differentiability requirement of Assumption 3(a) could actually be weakened in Theorem 3.2 to concern the function $f(x;\gamma)$ only. For ease of exposition we have ignored this generalization. As a whole, the treatment of nuisance parameters in Theorem 3.2 requires stronger assumptions than needed in Theorem 3.1, and any relaxations seem difficult to achieve. The fact that there is no finite dimensional reduced form counterpart of the parameter vector γ may explain why doing without any identification condition is more difficult in the case of this parameter than in the case of other parameters. Note, however, that in the special case where the sequence $f_t(\gamma)$ is known the limiting distribution in Theorem 3.2 applies without any identification conditions on nuisance parameters. This particularly holds in the standard model where the components of $f_t(\gamma)$ consist of a constant and conventional seasonal dummies. It may also be noted that if identifiability of all structural parameters is assumed the proof of Theorem 3.2 can be simplified and the orders of consistency assumed of the nuisance parameter estimators can then be obtained from Theorem 3.1.

When all structural parameters are identified the estimators in Theorem 3.1 and 3.2 are asymptotically independent of each other. It is straightforward to show that then they are also asymptotically independent of the error covariance matrix estimator $\hat{\Omega}_T$, which has the same limiting distribution as in the case where the values of the other parameters are a priori known and not estimated. Moreover, it can be shown that the likelihood ratio of our model belongs to the locally asymptotically mixed normal (LAMN) family so that ML estimators are asymptotically efficient (see, e.g., Basawa and Scott, 1983; Phillips, 1991; Jeganathan, 1995).

We close this section by noting that it is straightforward to extend the results of Theorems 3.1 and 3.2 to allow for smooth parameter constraints to be considered in the next section. Specifically, consider restrictions of the form

$$h(\vartheta) = [h_\phi(\phi)'\ h_\nu(\nu)'\ h_\gamma(\gamma)'\ h_\psi(\psi)']' = 0, \tag{3.5}$$

where $h(\vartheta)$ is a continuously differentiable vector function of dimension $\not\ell \leq k + n + q + \ell$. Thus, we assume that each component of ϑ is restricted separately. Although more general restrictions might be considered we shall restrict ourselves to this special case because potential generalizations are likely to be of minor practical interest. When the restrictions in (3.5) are assumed all of our previous results are easily seen to hold. To demonstrate this, note first that in the consistency results of Saikkonen (2001) no particular assumptions were made of the parameter spaces Φ, N, and Ψ, so that we can simply assume that they are defined in such a way that the restrictions implied by (3.5) hold. The parameter space Γ was assumed compact, but this causes no problems because we only need to consider the redefined parameter space $\Gamma \cap \{h_\gamma(\gamma) = 0\}$, which is compact. Thus, it follows that with appropriate interpretations of the parameter spaces all the consistency results proved in Saikkonen (2001) and used in Theorems 3.1 and 3.2 still hold. This implies that the limiting distributions of these ML estimators can be derived by using a standard Lagrange multiplier method and arguments similar to those in the proofs of Theorems 3.1 and 3.2 (cf., e.g., Kohn, 1978).

4. HYPOTHESIS TESTING

We consider testing general nonlinear restrictions on the parameters ϑ_1 and ψ. Following the development of the previous section our treatment will be divided into two parts. First we obtain tests for the null hypothesis

$$\mathcal{H}_1 : h_1(\vartheta_{10}) = [h_\phi(\phi_0)' \; h_\nu(\nu_0)' \; h_\gamma(\gamma_0)']' = 0, \tag{4.1}$$

where $h_1(\vartheta_1)$ is a continuously differentiable vector function of dimension $f_1 \leq k + n + q$. As a result of the assumed structure of the function $h_1(\vartheta)$, the matrix $H_1(\vartheta_1) = \partial h_1(\vartheta_1)/\partial\vartheta_1'$ is of the form $H_1(\vartheta_1) = \text{diag}[\partial h_1(\phi)/\partial\phi' \; \partial h_1(\nu)/\partial\nu' \; \partial h_1(\gamma)/\partial\gamma']$, and, as usual, it is supposed to be of full row rank at $\vartheta_1 = \vartheta_{10}$. If some components of ϑ_1 are not involved in the null hypothesis the definitions of $h_1(\vartheta_1)$ and $H_1(\vartheta_1)$ are modified in an obvious way.

The Wald test for the preceding null hypothesis tests whether $h_1(\hat{\vartheta}_{1T})$ is significantly different from zero. A general form of the test statistic is

$$\mathcal{W}_1 = h_1(\hat{\vartheta}_{1T})'[H_1(\hat{\vartheta}_{1T})\hat{M}_{1T}^{-1}H_1(\hat{\vartheta}_{1T})']^{-1}h_1(\hat{\vartheta}_{1T}). \tag{4.2}$$

Here $\hat{M}_{1T}$ is any (nonsingular) matrix with the property

$$\Upsilon_T^{-1}\hat{M}_{1T}\Upsilon_T^{-1} = \Upsilon_T^{-1}\sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})'\Upsilon_T^{-1} + o_p(1), \tag{4.3}$$

where $\Upsilon_T = \text{diag}[TI_k \quad T^{1/2}I_{n+q}]$ and

$$F_{1t}(\vartheta_1) = -\begin{bmatrix} (\partial \operatorname{vec} A(\phi)'/\partial\phi)(y_{2,t-1} \otimes I_r) \\ \partial g_t(\mu)'/\partial\mu \end{bmatrix}.$$

Clearly, $F_{1t}(\vartheta_1)$ is an empirical counterpart of the function $F_1(x;\vartheta_1)$, so that the first term on the right hand side of (4.3) is an obvious sample analog of the matrix that is inverted in the weak limit of Theorem 3.1. Instead of only considering this particular choice of the matrix $\hat{M}_{1T}$ we wish to be more general and also allow for alternative possibilities. In particular, because the preceding choice makes use of the asymptotic orthogonality of the parameters ϑ_1 and ψ it may not work well in small or moderate samples. A conventional way to allow for the effects of the short-run parameter ψ is to choose $\hat{M}_{1t}$ as

$$\begin{aligned} \hat{M}_{1T} = {} & \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})' - \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}F_{2t}(\hat{\vartheta}_T)' \\ & \times \left(\sum_{t=1}^{T} F_{2t}(\hat{\vartheta}_T)\hat{\Omega}_T^{-1}F_{2t}(\hat{\vartheta}_T)'\right)^{-1} \sum_{t=1}^{T} F_{2t}(\hat{\vartheta}_T)\hat{\Omega}_T^{-1}\hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})', \end{aligned} \tag{4.4}$$

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

where $F_{2t}(\vartheta) = -(\partial \operatorname{vec} B(\psi)'/\partial\psi)(z_t(\vartheta_1) \otimes I_s)$ and $z_t(\vartheta_1) = [\Delta y'_{t-1} \ldots \Delta y'_{t-p+1} u_{t-1}(\vartheta_1)']'$ with $u_{t-1}(\vartheta_1) = y_{1,t-1} - A(\phi)y_{2,t-1} - g_t(\mu)$. This choice is based on the outer product form of the observed information matrix of the parameters ϑ_1 and ψ. Therefore we can only justify its use when the nuisance parameter ψ is identified and the function $B(\psi)$ is continuously differentiable. Because the limiting distribution of test statistic $\mathcal{W}_1$ can be derived without assuming identifiability of the parameter ψ it is reasonable to consider a modification of (4.4). A natural possibility is to replace $F_{2t}(\vartheta)$ by $-z_t(\vartheta_1) \otimes I_s$, which means that in the test statistic the parameter B is treated as if it were unconstrained, although in the estimation its values are restricted to the set $B(\Psi)$. Using Lemma A.1 in Saikkonen (2001) and arguments similar to those in the Appendix it can be readily seen that the preceding choices of the matrix $\hat{M}_{1T}$ satisfy condition (4.3).

To develop LM and LR tests for the null hypothesis $\mathcal{H}_1$, constrained ML estimators of the relevant parameters are needed. As discussed in the previous section, the existence and consistency of these constrained estimators obtain under the same conditions as in the case of the corresponding unconstrained estimators. The constrained ML estimator of the parameter ϑ_1 is denoted by $\tilde{\vartheta}_{1T}$, and a similar notation is used for constrained ML estimators of other parameters.

For the LM test statistic it is reasonable to consider the likelihood function explicitly as a function of the parameters ϑ_1 and $B \in B(\Psi)$. Thus, in (3.1) we use $\varepsilon_t = \varepsilon_t(\vartheta_1, B)$ (see (3.3)) and denote $\ell_T = \ell_T(\vartheta_1, B, \Omega)$. Then, by straightforward differentiation,

$$\partial\ell_T(\vartheta_1, B, \Omega)/\partial\vartheta_1 = -2\sum_{t=1}^{T} F_{1t}(\vartheta_1)\alpha'\Omega^{-1}\varepsilon_t(\vartheta_1, B), \tag{4.5}$$

where $F_{1t}(\vartheta_1)$ is as in (4.3) and satisfies $F_{1t}(\vartheta_1)\alpha' = -\partial\varepsilon_t(\vartheta_1, B)'/\partial\vartheta_1$. Now we can introduce our LM test statistic

$$\begin{aligned}\mathcal{LM}_1 = {} & \tfrac{1}{4}(\partial\ell_T(\tilde{\vartheta}_{1T}, \tilde{B}_T, \tilde{\Omega}_T)/\partial\vartheta_1')\widetilde{M}_{1T}^{-1} H_1(\tilde{\vartheta}_{1T})'[H_1(\tilde{\vartheta}_{1T})\widetilde{M}_{1T}^{-1} H_1(\tilde{\vartheta}_{1T})']^{-1} \\ & \times H_1(\tilde{\vartheta}_{1T})\widetilde{M}_{1T}^{-1}(\partial\ell_T(\tilde{\vartheta}_{1T}, \tilde{B}_T, \tilde{\Omega}_T)/\partial\vartheta_1),\end{aligned} \tag{4.6}$$

where $\widetilde{M}_{1T}$ is any matrix such that $\Upsilon_T^{-1}\widetilde{M}_{1T}\Upsilon_T^{-1}$ satisfies (4.3) with the estimators on the right hand side replaced by their constrained counterparts. The division by four in the test statistic is needed because we are working with minus two times the logarithm of the likelihood function. Of course, the role of the matrix $\widetilde{M}_{1T}$ in test statistic $\mathcal{LM}_1$ is entirely similar to that of $\hat{M}_{1T}$ in test statistic $\mathcal{W}_1$. Thus a possible choice for $\widetilde{M}_{1T}$ is given by the constrained version of (4.4) (if the parameter ψ is identified) or its modification discussed following (4.4). That these choices satisfy the required analog of (4.3) can be seen in the same way as in the case of $\hat{M}_{1T}$.

Finally, consider the LR test statistic that is defined by

$$\mathcal{LR}_1 = \ell_T(\tilde{\vartheta}_{1T}, \tilde{B}_T, \tilde{\Omega}_T) - \ell_T(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T). \tag{4.7}$$

A convenient feature of the LR test statistic is that no choice of matrices such as $\hat{M}_{1T}$ or $\widetilde{M}_{1T}$ is needed because this choice is automatically built into the test statistic. This appears useful when the identification of the nuisance parameter ψ is not assumed because then the use of the most natural choices (4.4) and its constrained version is not justified by our results.

The following theorem shows that a standard chi-square criterion applies to all three test statistics introduced previously. In the proof of this theorem it is convenient to make explicit use of the continuity of the function $B(\psi)$, which in Theorem 3.1 was only an implicit assumption behind the consistency of the reduced form ML estimator $\hat{B}_T$.

THEOREM 4.1. *Suppose that the assumptions of Theorem* 3.1 *hold and that the null hypothesis* $\mathcal{H}_1$ *is true. Suppose further that the following assumptions hold.*

(a) *The assumptions made of the existence and consistency of the unconstrained* ML *estimators in Theorem* 3.1 *also hold for the corresponding constrained* ML *estimators.*
(b) *The estimators* $\hat{B}_T$ *and* $\tilde{B}_T$ *satisfy* $\hat{B}_T = B_0 + O_p(T^{-1/2})$ *and* $\tilde{B}_T = B_0 + O_p(T^{-1/2})$.
(c) *The function* $B(\psi)$ *in Assumption* 2 *is continuous.*
(d) *The function* $h_1(\vartheta_1)$ $(f_1 \times 1)$ *in* (4.1) *is continuously differentiable with the matrix* $H_1(\vartheta_{10}) = \partial h_1(\vartheta_{10})/\partial\vartheta_1'$ $(f_1 \times (k+n+q))$ *of full row rank.*

Then, test statistics $\mathcal{W}_1$, $\mathcal{LM}_1$, *and* $\mathcal{LR}_1$ *have an asymptotic chi-square distribution with* f_1 *degrees of freedom.*

Remark 4.1. Assumption (a) of Theorem 4.1 holds under the same conditions as the corresponding assumption in Theorem 3.1 (see Remark 3.1 and the discussion at the end of Sect. 3). Assumption (b) holds under the conditions of Proposition 3.2 of Saikkonen (2001). Explicit sufficient conditions for all the consistency assumptions needed in Theorem 4.1 are given in Assumptions 1, 2(a) and (b), 3(a), 4(a) and (b), 5(a)–(d), and 6 of Saikkonen (2001).

The result of Theorem 4.1 is hardly surprising because similar results have also been obtained previously (see, e.g., Phillips, 1991; Johansen, 1991, 1995). However, the required assumptions make our result different from most of its previous counterparts. Compared with the assumptions of Theorem 3.1, we have been forced to strengthen the consistency assumption made for the estimator $\hat{B}_T$ by assuming an order of consistency, and a similar assumption is also needed for the constrained counterpart of $\hat{B}_T$. These additional assumptions are not restrictive, however, and the same can be said about the other assumptions of Theorem 4.1. In fact, the discussion in Remark 4.1 implies that sufficient conditions for Theorem 4.1 to hold are fairly standard except for the novel feature

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

that identifiability of the short-run parameter ψ is not required and neither is differentiability of the likelihood function with respect to ψ. This particularly means that the result of Theorem 4.2 holds even if the parameter ψ cannot be consistently estimated.

It can be shown that test statistics $\mathcal{W}_1$, $\mathcal{LM}_1$, and $\mathcal{LR}_1$ are consistent, and, although this is not studied in this paper, we would expect that they also have the same limiting distribution under conventional local alternatives. It should be noted, however, that in general this limiting distribution is not noncentral chi square (cf. Basawa and Scott, 1983, Ch. 3; Saikkonen, 1993, Theorem 5.3). Thus, as far as asymptotic properties are concerned, it is not possible to choose between these three tests. If identification of the nuisance parameter ψ is an issue the LR test seems convenient because, unlike the Wald and LM tests, it does not require choosing a matrix such as $\hat{M}_{1T}$ or $\tilde{M}_{1T}$ that may be somewhat problematic in small or moderate samples. It is probably this feature that also makes the derivation of the limiting distribution of test statistic $\mathcal{LR}_1$ more difficult than that of test statistics $\mathcal{W}_1$ and $\mathcal{LM}_1$ (see the Appendix). Of course, in finite samples the behavior of the test statistics $\mathcal{W}_1$, $\mathcal{LM}_1$, and $\mathcal{LR}_1$ may be different so that simulation studies of their small sample properties and, particularly, of the effects of different choices of the matrices $\hat{M}_{1T}$ and $\tilde{M}_{1T}$ would be of great interest. Such simulations are outside the scope of this paper, however.

Now consider testing restrictions on the short-run parameter ψ. The null hypothesis is

$$\mathcal{H}_2 : h_2(\psi_0) = 0, \tag{4.8}$$

where $h_2(\psi)$ is a continuously differentiable vector function of dimension $f_2 \leq \ell$ and the matrix $H_2(\psi) = \partial h_2(\psi)/\partial\psi'$ is of full row rank at $\psi = \psi_0$. The tests to be developed for this null hypothesis do not require identifiability of the parameters ϕ and ν.

Our general Wald test statistic for the preceding null hypothesis is

$$\mathcal{W}_2 = h_2(\hat{\psi}_T)'[H_2(\hat{\psi}_T)\hat{M}_{2T}^{-1}H_2(\hat{\psi}_T)']^{-1}h_2(\hat{\psi}_T), \tag{4.9}$$

where $\hat{M}_{2T}$ is any (nonsingular) matrix with the property

$$T^{-1}\hat{M}_{2T} = T^{-1}\sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T)\hat{\Omega}_T^{-1}F_{2t}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T)' + o_p(1). \tag{4.10}$$

Here $F_{2t}(\underline{\vartheta}_1, \psi) = -(\partial \operatorname{vec} B(\psi)'/\partial\psi)(z_t(\underline{\vartheta}_1) \otimes I_s)$ with $z_t(\underline{\vartheta}_1)$ defined following (3.4). Note that we have $F_{2t}(\underline{\vartheta}_1, \psi) = \partial\varepsilon_t(\underline{\vartheta}_1, \psi)'/\partial\psi$ and that $F_{2t}(\underline{\vartheta}_1, \psi)$ equals $F_{2t}(\vartheta)$ (see (4.4)) except that it is expressed as a function of the reduced form parameters A and d instead of the corresponding structural parameters ϕ and ν. The role of the matrix $\hat{M}_{2T}$ is similar to that of $\hat{M}_{1T}$ in test statistic $\mathcal{W}_1$. If all the parameters are identified and the functions $A(\phi)$ and $g_t(\mu)$ are continuously differentiable one can proceed as in (4.4) and choose

$$\hat{M}_{2T} = \sum_{t=1}^{T} F_{2t}(\hat{\vartheta}_T)\hat{\Omega}_T^{-1}F_{2t}(\hat{\vartheta}_T)' - \sum_{t=1}^{T} F_{2t}(\hat{\vartheta}_T)\hat{\Omega}_T^{-1}\hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})'$$
$$\times \left(\sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})'\right)^{-1} \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}F_{2t}(\hat{\vartheta}_T)'. \quad \textbf{(4.11)}$$

This choice can be modified to yield a test statistic that does not require identifiability of the nuisance parameters ϕ and ν. Of course, then $F_{2t}(\hat{\vartheta}_T)$ is replaced by $F_{2t}(\hat{\underline{\vartheta}}_{1T},\hat{\psi}_T)$, whereas a natural modification of $F_{1t}(\hat{\vartheta}_{1T})$ is given by $[-(y_{2,t-1}\otimes I_r)' \quad -(\partial f_t(\hat{\gamma}_T)/\partial\gamma')'\hat{d}_T' \quad -f_t(\hat{\gamma}_T)'\otimes I_{n+q}]'$. This means that in the test statistic the parameters A and d are treated as if they were unconstrained, although in the estimation their values are restricted to the sets $A(\Phi)$ and $d(N)$, respectively. That these choices of the matrix $\hat{M}_{2T}$ satisfy condition (4.10) can again be seen by using Lemma A.1 of Saikkonen (2001) and arguments similar to those in the Appendix.

We can also construct LM and LR tests for the null hypothesis $\mathcal{H}_2$. In this context it is relevant to consider the likelihood function as a function of the parameters $\underline{\vartheta}_1$, ψ, and Ω. Thus, in (3.1) we use $\varepsilon_t = \varepsilon_t(\underline{\vartheta}_1,\psi)$ see ((3.4)) and denote $\ell_T = \ell_T(\underline{\vartheta}_1,\psi,\Omega)$. Straightforward differentiation and the expression of $F_{2t}(\underline{\vartheta}_1,\psi)$ given following (4.10) show that

$$\partial\ell_T(\underline{\vartheta}_1,\psi,\Omega)/\partial\psi = 2\sum_{t=1}^{T} F_{2t}(\underline{\vartheta}_1,\psi)\Omega^{-1}\varepsilon_t(\underline{\vartheta}_1,\psi). \quad \textbf{(4.12)}$$

This expression is used to define our general LM test statistic

$$\mathcal{LM}_2 = \tfrac{1}{4}(\partial\ell_T(\tilde{\underline{\vartheta}}_{1T},\tilde{\psi}_T,\tilde{\Omega}_T)/\partial\psi')\widetilde{M}_{2T}^{-1}H_2(\tilde{\psi}_T)'[H_2(\tilde{\psi}_T)\widetilde{M}_{2T}^{-1}H_2(\tilde{\psi}_T)']^{-1}$$
$$\times H_2(\tilde{\psi}_T)\widetilde{M}_{2T}^{-1}(\partial\ell_T(\tilde{\underline{\vartheta}}_{1T},\tilde{\psi}_T,\tilde{\Omega}_T)/\partial\psi), \quad \textbf{(4.13)}$$

where $\widetilde{M}_{2T}$ is any matrix such that $T^{-1}\widetilde{M}_{2T}$ satisfies (4.10) with the estimators on the right hand side replaced by their constrained counterparts. Of course, the matrix $\widetilde{M}_{2T}$ can be modified in the same way as the matrix $\hat{M}_{2T}$ in test statistic $\mathcal{W}_2$, and these modifications can be justified in the same way as in the case of $\hat{M}_{2T}$. For the LR test no choice of matrices like these is needed. The test statistic is defined by

$$\mathcal{LR}_2 = \ell_T(\tilde{\underline{\vartheta}}_{1T},\tilde{\psi}_T,\tilde{\Omega}_T) - \ell_T(\hat{\underline{\vartheta}}_{1T},\hat{\psi}_T,\hat{\Omega}_T). \quad \textbf{(4.14)}$$

The following theorem shows that a conventional chi-square criterion applies to all three test statistics. The proof of this theorem assumes consistency of the infeasible ML estimator of the parameter $\underline{\vartheta}_1$ obtained by assuming that the values of the parameters ψ and Ω are a priori known. This estimator, denoted by $\dot{\underline{\vartheta}}_{1T}$, is obtained by minimizing the function $\ell_T(\underline{\vartheta}_1,\psi_0,\Omega_0)$ so that technically it can be interpreted as a constrained ML estimator of the parameter

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

$\underline{\vartheta}_1$. It will also be assumed that the infeasible ML estimators of the parameters ψ and Ω obtained by assuming that the values of the parameters ϕ and μ (or $\underline{\vartheta}_1 = (A, d, \gamma)$) are a priori known are consistent. These ML estimators are denoted by $\dot{\psi}_T$ and $\dot{\Omega}_T$, respectively. Consistency of the constrained versions of all these infeasible ML estimators is also assumed.

THEOREM 4.2. *Suppose that the assumptions of Theorem* 3.2 *hold and that the null hypothesis $\mathcal{H}_2$ is true. Suppose further that the following assumptions hold.*

(a) *The assumptions made of the existence and consistency of the unconstrained* ML *estimators in Theorem* 3.2 *also hold for the corresponding constrained* ML *estimators and for the infeasible* ML *estimators $\dot{\vartheta}_{1T}$, $\dot{\psi}_T$, and $\dot{\Omega}_T$ and the constrained versions of $\dot{\psi}_T$ and $\dot{\Omega}_T$.*
(b) *The function $h_2(\psi)(f_2 \times 1)$ in* (4.8) *is continuously differentiable with the matrix $H_2(\psi_0) = \partial h_2(\psi_0)/\partial\psi'$ $(f_2 \times \ell)$ of full row rank.*

Then, test statistics $\mathcal{W}_2$, $\mathcal{LM}_2$, and $\mathcal{LR}_2$ have an asymptotic chi-square distribution with f_2 degrees of freedom.

Remark 4.2. Although Saikkonen (2001) does not explicitly consider the case where the value of the error covariance matrix is known it is not difficult to see that similar consistency results can also be proved in this case. Thus, the consistency properties of the infeasible ML estimators discussed in assumption (a) of Theorem 4.2 are the same as those of the other ML estimators therein. Consequently, this assumption as a whole holds under the same conditions as the corresponding assumptions in Theorem 3.2 (see Remark 3.2 and the discussion at the end of Sect. 3). Explicit sufficient conditions for all the consistency assumptions needed in Theorem 4.2 can be found in Assumptions 1, 2(a), 3(a) and (b), 4(a), 5(a)–(d), and 6 of Saikkonen (2001).

Compared with its previous counterparts the result of Theorem 4.2 is again expected. However, as the discussion in Remark 4.2 implies, a novel feature of our result is that it has been obtained without assuming identifiability of the nuisance parameters ϕ and ν or even that the likelihood function is differentiable with respect to these parameters. Thus, the application of test statistics $\mathcal{W}_2$, $\mathcal{LM}_2$, and $\mathcal{LR}_2$ is justified even if the parameters ϕ and ν cannot be consistently estimated. Of course, in the same way as in Theorem 3.2 appropriate identifiability conditions are needed to guarantee consistent estimation of the parameters d and γ.

Unlike in Theorem 3.2 we have needed additional consistency assumptions about infeasible ML estimators of the parameters $\underline{\vartheta}_1$, ψ, and Ω. This is somewhat unusual and may be an artifact of the employed method of proof, which, as a result of the potential nonidentifiability of the parameters ϕ and ν, is not quite standard. We wish to emphasize, however, that these additional assumptions can be justified by the assumptions mentioned in Remark 4.2. It may also be noted that similar consistency results about infeasible ML estimators of the parameters B and Ω are also used in the proof of Theorem 4.1 (see the proof of Lemma A.1 in the Appendix) but there they can be deduced from the likelihood function $\ell_T(\vartheta_{10}, B, \Omega)$ by using the assumed continuity of the function $B(\psi)$ and arguments in Saikkonen (2001). A simplifying fact in that case is that $\ell_T(\vartheta_{10}, B, \Omega)$ is the likelihood function of a standard linear regression model with asymptotically stationary regressors and Gaussian white noise errors.

Finally, note that our previous discussion given for test statistics $\mathcal{W}_1$, $\mathcal{LM}_1$, and $\mathcal{LR}_1$ also applies to test statistics $\mathcal{W}_2$, $\mathcal{LM}_2$, and $\mathcal{LR}_2$ except that under conventional local alternatives a standard noncentral chi-square limiting distribution can be expected because now the limit theory is based on an ordinary normal distribution and not on a mixed normal distribution.

5. CONCLUSION

This paper has completed the work initiated in Saikkonen (2001) by developing an asymptotic theory of statistical inference in cointegrated VAR processes with nonlinear time trends in cointegrating relations. We have shown that ML estimators have normal or mixed normal limiting distributions and that Wald, LM, and LR tests with usual asymptotic chi-square distributions can be applied even if some nuisance parameters of the model are not identified. In the case of Wald and LM tests potential lack of identification means, however, that care is needed in choosing an estimator for the information matrix, whereas LR tests are free of this difficulty. Although the given theory appears fairly complete several possible extensions can be considered. Some of them are discussed in Saikkonen (2001, Sect. 4) and will hopefully be studied in the future.

REFERENCES

Bartle, R.G. (1964) *The Elements of Real Analysis.* New York: Wiley.

Basawa, I.V. & D.J. Scott (1983) *Asymptotic Optimal Inference for Nonergodic Models.* New York: Springer-Verlag.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural models. *Journal of Econometrics* 51, 113–150.

Gallant, A.R. (1987) *Nonlinear Statistical Models.* New York: Wiley.

Hansen, B.E. (1992) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Heinesen, E. (1997) Structural breaks in error correction models. *Oxford Bulletin of Economics and Statistics* 59, 187–192.

Jeganathan, P. (1995) Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–887.

Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1581.

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

Kohn, R. (1978) Asymptotic properties of time domain Gaussian estimators. *Advances in Applied Probability* 10, 339–359.

Lütkepohl, H. (1996) *Handbook of Matrices.* Chichester: Wiley.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.

Ripatti, A. & P. Saikkonen (2001) Vector autoregressive processes with nonlinear time trends in cointegrating relations. *Macroeconomic Dynamics*, forthcoming.

Saikkonen, P. (1993) Continuous weak convergence and stochastic equicontinuity results for integrated processes with an application to the estimation of a regression model. *Econometric Theory* 9, 155–188.

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

APPENDIX

A.1. Proofs for Section 3.

A.1.1. Proof of Theorem 3.1. Note first that when the estimators $\hat{\vartheta}_{1T}$, $\hat{B}_T$, and $\hat{\Omega}_T$ exist they minimize the function $\ell_T(\vartheta_1, B, \Omega)$. From the consistency of the estimator $\hat{\vartheta}_{1T}$ and Assumptions 1 and 3(a) it follows that $\hat{\vartheta}_{1T}$ is an interior point of $\Phi \times M$ with probability approaching one. Thus, because $\hat{\vartheta}_{1T}$ can also be obtained by minimizing the function $\ell_T(\vartheta_1, \hat{B}_T, \hat{\Omega}_T)$, we have $\partial \ell_T(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T)/\partial \vartheta_1 = 0$. Using this identity and the expression of $\partial \ell_T(\vartheta_1, B, \Omega)/\partial \vartheta_1$ in (4.5) we can therefore write

$$
\begin{aligned}
&2\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\vartheta_{10}) \hat{\alpha}_T' \hat{\Omega}_T^{-1} \varepsilon_t(\vartheta_{10}, \hat{B}_T) \\
&\quad = 2\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T}) \hat{\alpha}_T' \hat{\Omega}_T^{-1} [\varepsilon_t(\vartheta_{10}, \hat{B}_T) - \varepsilon_t(\hat{\vartheta}_{1T}, \hat{B}_T)] \\
&\qquad + 2\Upsilon_T^{-1} \sum_{t=1}^{T} [F_{1t}(\vartheta_{10}) - F_{1t}(\hat{\vartheta}_{1T})] \hat{\alpha}_T' \hat{\Omega}_T^{-1} \varepsilon_t(\vartheta_{10}, \hat{B}_T),
\end{aligned}
\tag{A.1}
$$

where $F_{1t}(\vartheta_1)$ and Υ_T are defined following (4.3) and an explicit definition of $\varepsilon_t(\vartheta_1, B)$ can be found in (3.3). We shall show next that the second term in the last expression is of order $o_p(1)$. This term consists of two components corresponding to the partition of $F_{1t}(\vartheta_1)$ into $-(\partial \operatorname{vec} A(\phi)'/\partial\phi)(y_{2,t-1} \otimes I_r)$ and $-\partial g_t(\mu)'/\partial\mu$. To analyze the first one, notice that $\varepsilon_t(\vartheta_{10}, \hat{B}_T) = \Delta y_t - \hat{B}_T z_t$ where $z_t = z_t(\vartheta_{10}) = [\Delta y_{t-1}' \ldots \Delta y_{t-p+1}'\; u_{t-1}(\vartheta_{10})']'$ with $u_{t-1}(\vartheta_{10}) = y_{1,t-1} - A(\phi_0) y_{2,t-1} - g_t(\mu_0)$. As the discussion at the end of Section 2.3 of Saikkonen (2001) shows, z_t is an asymptotically stationary process. Thus, it can be shown that the first component in the last expression of (A.1) is

$$
o_p(1) T^{-1} \sum_{t=1}^{T} (y_{2,t-1} \otimes I_r) \hat{\alpha}_T' \hat{\Omega}_T^{-1} [\Delta y_t - \hat{B}_T z_t] = o_p(1),
$$

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

where the term $o_p(1)$ on the left hand side just replaces the difference $\partial \operatorname{vec} A(\hat{\phi}_T)'/\partial\phi - \partial \operatorname{vec} A(\phi_0)'/\partial\phi$ and the equality follows from Lemma A.1(b) of Saikkonen (2001). The second component in the last expression of (A.1) is

$$T^{1/2}\sum_{t=1}^{T}[\partial g_t(\hat{\mu}_T)'/\partial\mu - \partial g_t(\mu_0)'/\partial\mu]\hat{\alpha}_T'\hat{\Omega}_T^{-1}[\Delta y_t - \hat{B}_T z_t] = o_p(1).$$

To justify the equality, use Theorem 2.1 of Saikkonen (2001) to conclude that $\max_{1\le t\le T} T^{-1/2}\|\sum_{j=1}^{t} z_j\| = O_p(1)$ and similarly with z_j replaced by Δy_j. The desired result can then be deduced from Lemma A.3(b) of Saikkonen (2001) and the fact that, by Assumption 3(b), $\partial g(x;\mu)/\partial\mu$ satisfies Condition 1 of the same paper in some compact neighborhood of μ_0. Thus, using the definition of $\varepsilon_t(\vartheta_1, B)$ (see (3.3)) we can write (A.1) as

$$\begin{aligned}
&2\Upsilon_T^{-1}\sum_{t=1}^{T}F_{1t}(\vartheta_{10})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\varepsilon_t(\vartheta_{10},\hat{B}_T)\\
&\quad = 2\Upsilon_T^{-1}\sum_{t=1}^{T}F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}[\varepsilon_t(\vartheta_{10},\hat{B}_T) - \varepsilon_t(\hat{\vartheta}_{1T},\hat{B}_T)] + o_p(1)\\
&\quad = -2\Upsilon_T^{-1}\sum_{t=1}^{T}F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T[(\hat{A}_T - A_0)y_{2,t-1} + (g_t(\hat{\mu}_T) - g_t(\mu_0))] + o_p(1)\\
&\quad = 2\Upsilon_T^{-1}\sum_{t=1}^{T}F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T F_{1t}(\bar{\vartheta}_{1T})'(\hat{\vartheta}_{1T} - \vartheta_{10}) + o_p(1), \qquad \textbf{(A.2)}
\end{aligned}$$

where the last equality is based on a mean value expansion and $F_{1t}(\bar{\vartheta}_{1T})$ signifies a matrix whose ith row equals the ith row of the matrix $F_{1t}(\bar{\vartheta}_{1T}^{(i)})$ with $\bar{\vartheta}_{1T}^{(i)} = a_i\vartheta_{10} + (1-a_i)\hat{\vartheta}_{1T}$, $0 \le a_i \le 1$, $i = 1,\dots,k+n+q$.

Now consider the first expression in (A.2) and notice that $\varepsilon_t(\vartheta_{10},\hat{B}_T) = \Delta y_t - \hat{B}_T z_t$ and $\varepsilon_t(\vartheta_{10}, B_0) = \varepsilon_t$. Because $\partial g(x;\mu)/\partial\mu$ satisfies Condition 1 of Saikkonen (2001) in some compact neighborhood of μ_0 we find from the definition of $F_{1t}(\vartheta_1)$, the consistency of the involved estimators, and Lemma A.1(b) and (c) of Saikkonen (2001) that

$$\begin{aligned}
2\Upsilon_T^{-1}\sum_{t=1}^{T}F_{1t}(\vartheta_{10})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\varepsilon_t(\vartheta_{10},\hat{B}_T) &= 2\Upsilon_T^{-1}\sum_{t=1}^{T}F_{1t}(\vartheta_{10})\alpha'\Omega^{-1}\varepsilon_t + o_p(1)\\
&\Rightarrow 2\int_0^1 F_1(x;\vartheta_{10})\alpha_0'\Omega_0^{-1}dW_0(x). \qquad \textbf{(A.3)}
\end{aligned}$$

To justify the latter relation, note first that, by Theorem 2.1 of Saikkonen (2001), $y_t = x_t - K_0(L)\alpha_0 g_t(\mu_0) + P_{\beta_\perp}y_0$ where $x_t = \Lambda_0\sum_{j=1}^{t}\varepsilon_j + K_0(L)\varepsilon_t$, $P_{\beta_\perp} = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'$ is evaluated at the true parameter value, $K_0(L) = \sum_{j=1}^{\infty}K_{j0}L^j$ is a rational matrix function of the lag operator L such that its power series representation converges absolutely in an open disk containing the unit disk, and $g_t(\mu_0) = 0$, $t < 0$. Let $x_{2,t-1}$ contain the last $s - r$ components of x_t and note that, in the same way as in the proof of Lemma A.1(f) of Saikkonen (2001), the replacement of $y_{2,t-1}$ in the definition of $F_{1t}(\vartheta_{10})$ by

$x_{2,t-1}$ and further by $\Lambda_0 \sum_{j=1}^{t} \varepsilon_j$ causes an error of order $o_p(1)$. Thus, arguments used in the proof of Lemma A.1(f) of Saikkonen (2001) give

$$\left(T^{-1/2} y_{2,[Tx]}, T^{-1/2} \Lambda_0 \sum_{j=1}^{[Tx]} \varepsilon_j \right) \Rightarrow (W_2(x), \Lambda_0 W_0(x))$$

and

$$\left(\partial g_{[Tx]}(\mu_0)/\partial \mu, T^{-1/2} \Lambda_0 \sum_{j=1}^{[Tx]} \varepsilon_j \right) \Rightarrow (\partial g(x;\mu_0)/\partial \mu, \Lambda_0 W_0(x))$$

jointly in the Skorohod topology. Hence, (A.3) follows from Theorem 2.1 of Hansen (1992).

Next consider the first term in the last expression of (A.2) and note that

$$\begin{aligned} &\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T}) \hat{\alpha}_T' \hat{\Omega}_T^{-1} \hat{\alpha}_T F_{1t}(\bar{\vartheta}_{1T})' \Upsilon_T^{-1} \\ &\quad = \Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\vartheta_{10}) \alpha_T' \Omega_T^{-1} \alpha_T F_{1t}(\vartheta_{10})' \Upsilon_T^{-1} + o_p(1) \\ &\quad \Rightarrow \int_0^1 F_1(x;\vartheta_{10}) \alpha_0' \Omega_0^{-1} \alpha_0 F_1(x;\vartheta_{10})' dx. \end{aligned} \qquad \textbf{(A.4)}$$

The equality is a straightforward consequence of the consistency of the involved estimators, Lemmas A.1(e) and A.3(a) of Saikkonen (2001), and the facts that, by Assumption 3(b), $\max_{1\le t\le T} T^{-1/2} \| y_{2,t-1} \| = \max_{1\le t\le T} T^{-1/2} \| x_{2,t-1} \| + o_p(1) = O_p(1)$ and $\max_{1\le t\le T} \| \partial g_t(\mu)/\partial \mu \| = O(1)$ uniformly over μ in some compact neighborhood of μ_0. The weak convergence in (A.4) is obtained from Lemma A.1(d)–(f) of Saikkonen (2001). Using Assumptions 1 and 3(c) it is straightforward to check that the limit in (A.4) is positive definite (a.s.). From Theorem 2.1 of Hansen (1992) it further follows that the weak convergences in (A.3) and (A.4) hold jointly so that the result of the theorem is obtained from (A.2), (A.3), and (A.4). ■

A.1.2. Proof of Theorem 3.2. Here it is relevant to define the likelihood function in terms of the parameter $\underline{\vartheta}_1$ and use the notation $\ell_T(\underline{\vartheta}_1, \psi, \Omega)$. Assumption 2 and the assumed consistency of the estimator $\hat{\psi}_T$ imply that $\hat{\psi}_T$ is an interior point of Ψ with probability approaching one and, because it minimizes the function $\ell_T(\hat{\underline{\vartheta}}_{1T}, \psi, \hat{\Omega}_T)$, we have $\partial \ell_T(\hat{\underline{\vartheta}}, \hat{\psi}_T, \hat{\Omega}_T)/\partial \psi = 0$. This in conjunction with the expression of $\partial \ell_T(\underline{\vartheta}_1, \psi, \Omega)/\partial \psi$ in (4.12) and the fact that $\varepsilon_t(\underline{\vartheta}_{10}, \psi_0) = \varepsilon_t$ yields

$$\begin{aligned} &2T^{-1/2} \sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T}, \psi_0) \hat{\Omega}_T^{-1} \varepsilon_t(\hat{\underline{\vartheta}}_{1T}, \psi_0) \\ &\quad = 2T^{-1/2} \sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T) \hat{\Omega}_T^{-1} [\varepsilon_t(\hat{\underline{\vartheta}}_{1T}, \psi_0) - \varepsilon_t(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T)] \\ &\qquad + 2T^{-1/2} \sum_{t=1}^{T} [F_{2t}(\hat{\underline{\vartheta}}_{1T}, \psi_0) - F_{2t}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T)] \hat{\Omega}_T^{-1} \varepsilon_t \\ &\qquad + 2T^{-1/2} \sum_{t=1}^{T} [F_{2t}(\hat{\underline{\vartheta}}_{1T}, \psi_0) - F_{2t}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T)] \hat{\Omega}_T^{-1} [\varepsilon_t(\hat{\underline{\vartheta}}_{1T}, \psi_0) - \varepsilon_t], \end{aligned} \qquad \textbf{(A.5)}$$

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

where $F_{2t}(\underline{\vartheta}_1, \psi)$ is defined following (4.10) and $\varepsilon_t(\underline{\vartheta}_1, \psi)$ in (3.4). We wish to show that the second and third terms on the right hand side are of order $o_p(1)$. Using the definition of $z_t(\underline{\vartheta}_1)$ following (3.4) and the fact that $z_t(\underline{\vartheta}_{10}) = z_t$ we find that the second term on the right hand side of (A.5) can be written as

$$\begin{aligned} & o_p(1)T^{-1/2}\sum_{t=1}^{T}[z_t(\hat{\underline{\vartheta}}_{1T}) \otimes I_s]\hat{\Omega}_T^{-1}\varepsilon_t \\ &\quad = o_p(1)T^{-1/2}\sum_{t=1}^{T}(z_t \otimes I_s)\hat{\Omega}_T^{-1}\varepsilon_t \\ &\qquad + o_p(1)T^{-1/2}\sum_{t=1}^{T}[(z_t(\hat{\underline{\vartheta}}_{1T}) - z_t) \otimes I_s]\hat{\Omega}_T^{-1}\varepsilon_t, \end{aligned} \tag{A.6}$$

where the term $o_p(1)$ just replaces the difference $\partial \operatorname{vec} B(\hat{\psi}_T)'/\partial\psi - \partial \operatorname{vec} B(\psi)'/\partial\psi$. That the first term on the right hand side is of order $o_p(1)$ is a simple consequence of the facts that $\hat{\Omega}_T^{-1} = O_p(1)$ and that $(z_t \otimes I_s)\varepsilon_t$ is a square integrable martingale difference sequence. To show the same for the second one, notice that the only nonzero elements of $z_t(\hat{\underline{\vartheta}}_{1T}) - z_t$ are given by $u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10}) = -(\hat{A}_T - A_0)y_{2,t-1} - (\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0))$. Thus, we may consider

$$\begin{aligned} & o_p(1)T^{-1/2}\sum_{t=1}^{T}[(\hat{A}_T - A_0)y_{2,t-1} \otimes I_s]\hat{\Omega}_T^{-1}\varepsilon_t \\ &\quad + o_p(1)T^{-1/2}\sum_{t=1}^{T}[(\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)) \otimes I_s]\hat{\Omega}_T^{-1}\varepsilon_t = o_p(1). \end{aligned}$$

To justify the equality, one can first use Lemma A.1(c) of Saikkonen (2001) to conclude that the second term on the left hand side is of order $o_p(1)$. For the first term the same conclusion follows from Lemma A.1(b) of the same paper and the assumption $\hat{A}_T - A_0 = O_p(T^{-1})$. Thus, we have shown that (A.6) or the second term on the right hand side of (A.5) is of order $o_p(1)$.

To show that the third term on the right hand side of (A.5) is of order $o_p(1)$, conclude from (3.4) that $\varepsilon_t(\hat{\underline{\vartheta}}_{1T}, \psi_0) - \varepsilon_t = -\alpha_0(u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10}))$. Thus, in the same way as in (A.6) we may consider

$$\begin{aligned} & o_p(1)T^{-1/2}\sum_{t=1}^{T}[z_t(\hat{\underline{\vartheta}}_{1T}) \otimes I_s]\hat{\Omega}_T^{-1}\alpha_0[u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})] \\ &\quad = o_p(1)T^{-1/2}\sum_{t=1}^{T}[z_t \otimes I_s]\hat{\Omega}_T^{-1}\alpha_0[u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})] \\ &\qquad + o_p(1)T^{-1/2}\sum_{t=1}^{T}[(z_t(\hat{\underline{\vartheta}}_{1T}) - z_t) \otimes I_s]\hat{\Omega}_T^{-1}\alpha_0[u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})]. \end{aligned} \tag{A.7}$$

Here we can further write

$$\begin{aligned} u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10}) &= -(\hat{A}_T - A_0)y_{2,t-1} - (\hat{d}_T - d_0)f_t(\hat{\gamma}_T) \\ &\quad - d_0(f_t(\hat{\gamma}_T) - f_t(\gamma_0)). \end{aligned} \tag{A.8}$$

Recall the assumptions $\hat{A}_T - A_0 = O_p(T^{-1})$, $\hat{d}_T - d_0 = O_p(T^{-1/2})$ and $\hat{\gamma}_T - \gamma_0 = O_p(T^{-1/2})$ and also that $\max_{1\le t\le T}\| f_t(\gamma)\|$ and $\max_{1\le t\le T}\|\partial f_t(\gamma)/\partial\gamma\|$ are bounded in some neighborhood of γ_0 by the assumed version of Condition 1 of Saikkonen (2001). Thus, because $\max_{1\le t\le T} T^{-1/2}\|y_{2,t-1}\| = O_p(1)$, it follows from (A.8) after a mean value expansion of $f_t(\hat{\gamma}_T) - f_t(\gamma_0)$ that $\max_{1\le t\le T} \|u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})\| = O_p(T^{-1/2})$. Because it is also easy to check that $\max_{1\le t\le T} E\|z_t\| = O(1)$ we find that the first term on the right hand side of (A.7) is of order $o_p(1)$. That the same is true for the second one can be readily seen by recalling that the only nonzero elements of $z_t(\hat{\underline{\vartheta}}_{1T}) - z_t$ are given by $u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})$. Thus, we have shown that (A.7) is of order $o_p(1)$ and thereby that the third term on the right hand side of (A.5) is of order $o_p(1)$.

The preceding discussion and the identity $\varepsilon_t(\hat{\underline{\vartheta}}_{1T},\psi) = \Delta y_t - B(\psi)z_t(\hat{\underline{\vartheta}}_{1T})$ imply that we can write (A.5) as

$$
\begin{aligned}
&2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T},\psi_0)\hat{\Omega}_T^{-1}\varepsilon_t(\hat{\underline{\vartheta}}_{1T},\psi_0)\\
&\quad = 2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T},\hat{\psi}_T)\hat{\Omega}_T^{-1}[B(\hat{\psi}_T) - B(\psi_0)]z_t(\hat{\underline{\vartheta}}_{1T}) + o_p(1)\\
&\quad = -2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T},\hat{\psi}_T)\hat{\Omega}_T^{-1}F_{2t}(\hat{\underline{\vartheta}}_{1T},\bar{\psi}_T)'(\hat{\psi}_T - \psi_0) + o_p(1), \qquad \textbf{(A.9)}
\end{aligned}
$$

where the second equality is based on a mean value expansion with the notation $\bar{\psi}_T$ defined in the same way as its analog in the proof of Theorem 3.1. Now consider the extreme expressions in (A.9) and first note that

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T}\psi_0)\hat{\Omega}_T^{-1}\varepsilon_t(\hat{\underline{\vartheta}}_{1T},\psi_0) &= T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\vartheta}_{10},\psi_0)\Omega_0^{-1}\varepsilon_t + o_p(1)\\
&\Rightarrow N(0,\Sigma_{\psi}(\psi_0,\Omega_0)^{-1}), \qquad \textbf{(A.10)}
\end{aligned}
$$

where the matrix $\Sigma_{\psi}(\psi_0,\Omega_0)^{-1}$ is well defined by Lemma A.1 of Saikkonen (2001). The stated weak convergence is obtained from a standard martingale central limit theorem and the fact that in the definition of $F_{2t}(\underline{\vartheta}_{10},\psi_0)$ the vector $z_t(\underline{\vartheta}_{10}) = z_t$ can be replaced by $\zeta_t = [\Delta x'_{t-1}\dots\Delta x'_{t-p+1} \quad x'_{t-1}\beta_0]'$ (cf. the proof of Lemma A.1 of Saikkonen, 2001). To justify the equality in (A.10), observe that

$$
\varepsilon_t(\hat{\underline{\vartheta}}_{1T},\psi_0) - \varepsilon_t = \alpha_0(\hat{A}_T - A_0)y_{2,t-1} + \alpha_0(\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0))
$$

and

$$
F_{2t}(\hat{\underline{\vartheta}}_{1T},\psi_0) - F_{2t}(\underline{\vartheta}_{10},\psi_0) = -(\partial\,\mathrm{vec}\,B(\psi_0)'/\partial\psi)[(z_t(\hat{\underline{\vartheta}}_{1T}) - z_t)\otimes I_s],
$$

where the only nonzero elements of $z_t(\hat{\underline{\vartheta}}_{1T}) - z_t$ are given by $u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})$. An explicit expression of $u_{t-1}(\hat{\underline{\vartheta}}_{1T}) - u_{t-1}(\underline{\vartheta}_{10})$ can be found in (A.8) where $\hat{A}_T - A_0 = O_p(T^{-1})$, $\hat{d}_T - d_0 = O_p(T^{-1/2})$, and $\hat{\gamma}_T - \gamma_0 = O_p(T^{-1/2})$ by assumption. The equality in (A.10) can now be established by using these facts, the consistency of the estimator $\hat{\Omega}_T$, and a straightforward though somewhat tedious application of Lemmas

A.1 and A.3 of Saikkonen (2001) combined with a (possibly termwise) mean value expansion of $d_0(f_t(\hat{\gamma}_T) - f_t(\gamma_0))$ when needed. Thus, we have established (A.10) as a whole.

Next consider

$$T^{-1}\sum_{t=1}^{T} F_{2t}(\hat{\underline{\vartheta}}_{1T},\hat{\psi}_T)\hat{\Omega}_T^{-1}F_{2t}(\hat{\underline{\vartheta}}_{1T},\bar{\psi}_T) = T^{-1}\sum_{t=1}^{T} F_{2t}(\underline{\vartheta}_{10},\psi_0)\Omega_0^{-1}F_{2t}(\underline{\vartheta}_{10},\psi_0)' + o_p(1)$$

$$= \Sigma_{\psi}(\psi_0,\Omega_0)^{-1} + o_p(1). \qquad \textbf{(A.11)}$$

Here the justification of the first equality is similar to that in (A.10) except that the situation is now simpler because the standardization is by T and not by $T^{1/2}$. The second equality follows from the definitions of $F_{2t}(\underline{\vartheta}_1,\psi)$ and $\Sigma_{\psi}(\psi_0,\Omega_0)$ and Lemma A.1(a) of Saikkonen (2001). Combining (A.9)–(A.11) gives the stated limiting distribution. To show its independence of the limiting distribution in Theorem 3.1, note that these limiting distributions are determined by the joint weak limit of partial sum processes formed from ε_t and $(\zeta_t \otimes I_s)\Omega_0^{-1}\varepsilon_t$. Because we have $E(\zeta_t \otimes I_s)\Omega_0^{-1}\varepsilon_t\varepsilon_{t+j}' = 0$ for all j the weak limits of these partial sum processes are uncorrelated Brownian motions and therefore independent. This completes the proof. ■

Note that the proof of Theorem 3.2 explicitly makes use of the assumed orders of consistency. In particular, the proofs given for the equalities in (A.7) and (A.10) are not possible if mere consistency is assumed, and even the result $\hat{A}_T = A_0 + o_p(T^{-1/2})$ proved in Theorem 3.1 of Saikkonen (2001) does not suffice.

A.2. Proofs for Section 4.

A.2.1. Intermediate Results. Before starting to prove Theorems 4.1 and 4.2 two auxiliary results needed to obtain the limiting distributions of the LR tests will be given. These results make use of a decomposition of the likelihood function introduced in equation (3.6) of Saikkonen (2001). First decompose the vector $\varepsilon_t(\vartheta_1,B)$ in (3.3) as $\varepsilon_t(\vartheta_1,B) = \varepsilon_{1t}(\vartheta_1,B) + \varepsilon_{2t}(B)$ where

$$\varepsilon_{1t}(\vartheta_1,B) = \alpha(A - A_0)y_{2,t-1} + \alpha[df_t(\gamma) - d_0 f_t(\gamma_0)]$$

and

$$\varepsilon_{2t}(B) = \Delta y_t - Bz_t$$

with $z_t = z_t(\underline{\vartheta}_{10})$ as before. Using these definitions we can write

$$\ell_T(\vartheta_1,B,\Omega) = \ell_{1T}(\vartheta_1,B,\Omega) + \ell_{2T}(B,\Omega),$$

where

$$\ell_{1T}(\vartheta_1,B,\Omega) = 2\,\mathrm{tr}\left(\Omega^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\vartheta_1,B)\varepsilon_{2t}(B)'\right) + \mathrm{tr}\left(\Omega^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\vartheta_1,B)\varepsilon_{1t}(\vartheta_1,B)'\right)$$

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

and

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

The notations $\ell_T(\underline{\vartheta}_1, \psi, \Omega)$, $\ell_{1T}(\underline{\vartheta}_1, \psi, \Omega)$, and $\ell_{2T}(\psi, \Omega)$ are used when the parameters $\underline{\vartheta}_1$ and ψ are used instead of ϑ_1 and B, respectively. Now we can prove the following.

LEMMA A.1. *Suppose the conditions of Theorem* 4.1 *hold and let* $\bar{\Omega}_T$ *be any random matrix with the property* $\bar{\Omega}_T = \Omega_0 + o_p(1)$. *Then,*

(a) $\ell_{2T}(\hat{B}_T, \bar{\Omega}_T) - \ell_{2T}(\tilde{B}_T, \bar{\Omega}_T) = o_p(1)$,
(b) $\ell_{1T}(\hat{\vartheta}_{1T}, \hat{B}_T, \bar{\Omega}_T) - \ell_{1T}(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T) = o_p(1)$,
(c) $\ell_{1T}(\tilde{\vartheta}_{1T}, \tilde{B}_T, \bar{\Omega}_T) - \ell_{1T}(\tilde{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T) = o_p(1)$.

Proof. Let $\dot{B}_T$ and $\dot{\Omega}_T$ be the (infeasible) ML estimators of B and Ω, respectively, obtained by minimizing $\ell_{2T}(B,\Omega)$ over $\{(B,\Omega): B \in \overline{\mathcal{B}(\Psi)}, \Omega > 0\}$. Specializing Theorem 3.1 and Proposition 3.2(d) of Saikkonen (2001) to the case where ϑ_1 is restricted by $\vartheta_1 = \vartheta_{10}$ one can see that these estimators exist with probability approaching one and satisfy $\dot{B}_T = B_0 + O_p(T^{-1/2})$ and $\dot{\Omega}_T = \Omega_0 + o_p(1)$. (Note that this only requires continuity of the function $B(\psi)$.) To prove (a), it suffices to show that

$$\ell_{2T}(\hat{B}_T, \bar{\Omega}_T) - \ell_{2T}(\dot{B}_T, \bar{\Omega}_T) = o_p(1) \tag{A.12}$$

and that the same result holds with $\hat{B}_T$ replaced by $\tilde{B}_T$. First observe that, by the definition of the estimators $\hat{\vartheta}_{1T}$, $\hat{B}_T$, and $\hat{\Omega}_T$,

$$\begin{aligned} 0 &\geq \ell_T(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T) - \ell_T(\hat{\vartheta}_{1T}, \dot{B}_T, \hat{\Omega}_T) \\ &= \ell_{1T}(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T) - \ell_{1T}(\hat{\vartheta}_{1T}, \dot{B}_T, \hat{\Omega}_T) + \ell_{2T}(\hat{B}_T, \hat{\Omega}_T) - \ell_{2T}(\dot{B}_T, \hat{\Omega}_T). \end{aligned} \tag{A.13}$$

We shall show later that

$$\ell_{1T}(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T) - \ell_{1T}(\hat{\vartheta}_{1T}, \dot{B}_T, \hat{\Omega}_T) = o_p(1). \tag{A.14}$$

Assuming that this holds one obtains from (A.13)

$$\begin{aligned} 0 &\geq \ell_{2T}(\hat{B}_T, \hat{\Omega}_T) - \ell_{2T}(\dot{B}_T, \hat{\Omega}_T) + o_p(1) \\ &= \operatorname{tr}\left\{\dot{\Omega}_T^{-1}\left(\sum_{t=1}^{T} \varepsilon_{2t}(\hat{B}_T)\varepsilon_{2t}(\hat{B}_T)' - \sum_{t=1}^{T} \varepsilon_{2t}(\dot{B}_T)\varepsilon_{2t}(\dot{B}_T)'\right)\right\} \\ &\quad + \operatorname{tr}\left\{(\hat{\Omega}_T^{-1} - \dot{\Omega}_T^{-1})\left(\sum_{t=1}^{T} \varepsilon_{2t}(\hat{B}_T)\varepsilon_{2t}(\hat{B}_T)' - \sum_{t=1}^{T} \varepsilon_{2t}(\dot{B}_T)\varepsilon_{2t}(\dot{B}_T)'\right)\right\} \\ &\quad + o_p(1). \end{aligned} \tag{A.15}$$

Here the equality is a direct consequence of the definition of $\ell_{2T}(B,\Omega)$, which also shows that the first term in the last expression equals $\ell_{2T}(\hat{B}_T, \dot{\Omega}_T) - \ell_{2T}(\dot{B}_T, \dot{\Omega}_T)$. On the other hand, the second term in the last expression is of order $o_p(1)$ because both $\hat{\Omega}_T$ and $\dot{\Omega}_T$ are consistent estimators of Ω and because the difference between the two sums is of order $O_p(1)$. Because $\hat{B}_T = B_0 + O_p(T^{-1/2})$ and $\dot{B}_T = B_0 + O_p(T^{-1/2})$ this last fact can

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

be established in a straightforward manner by using Lemma A.1(a) and (b) of Saikkonen (2001) and the identities

$$\varepsilon_{2t}(\dot{B}_T) = \varepsilon_{2t}(\hat{B}_T) + (\hat{B}_T - \dot{B}_T)z_t$$
$$= \varepsilon_t - (\hat{B}_T - B_0)z_t + (\hat{B}_T - \dot{B}_T)z_t. \quad \textbf{(A.16)}$$

Thus, we can conclude from (A.15) that

$$0 \geq \ell_{2T}(\hat{B}_T, \dot{\Omega}_T) - \ell_{2T}(\dot{B}_T, \dot{\Omega}_T) + o_p(1).$$

However, by the definition of the estimators $\dot{B}_T$ and $\dot{\Omega}_T$, the difference on the right hand side is nonnegative so that it must be of order $o_p(1)$, and we have established (A.12) in the special case $\bar{\Omega}_T = \dot{\Omega}_T$. To prove the same result in the general case, just observe that the arguments used for the equality in (A.15) also show that $\ell_{2T}(\hat{B}_T, \bar{\Omega}_T) - \ell_{2T}(\dot{B}_T, \bar{\Omega}_T) = \ell_{2T}(\hat{B}_T, \dot{\Omega}_T) - \ell_{2T}(\dot{B}_T, \dot{\Omega}_T) + o_p(1)$. Thus, to complete the proof of (A.12), we have to justify (A.14).

From the definition of $\ell_{1T}(\vartheta_1, B, \Omega)$ it follows that

$$\ell_{1T}(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T) - \ell_{1T}(\hat{\vartheta}_{1T}, \dot{B}_T, \hat{\Omega}_T)$$
$$= 2\,\mathrm{tr}\left\{\hat{\Omega}_T^{-1}\left(\sum_{t=1}^{T} \varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T)\varepsilon_{2t}(\hat{B}_T)' - \sum_{t=1}^{T} \varepsilon_{1t}(\hat{\vartheta}_{1T}, \dot{B}_T)\varepsilon_{2t}(\dot{B}_T)'\right)\right\}$$
$$+ \mathrm{tr}\left\{\hat{\Omega}_T^{-1}\left(\sum_{t=1}^{T} \varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T)\varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T)' - \sum_{t=1}^{T} \varepsilon_{1t}(\hat{\vartheta}_{1T}, \dot{B}_T)\varepsilon_{1t}(\hat{\vartheta}_{1T}, \dot{B}_T)'\right)\right\}. \quad \textbf{(A.17)}$$

From the definitions one also obtains

$$\varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T) = \hat{\alpha}_T(\hat{A}_T - A_0)y_{2,t-1} + \hat{\alpha}_T[\hat{d}_T f(\hat{\gamma}_T) - d_0 f_t(\gamma_0)] \quad \textbf{(A.18)}$$

and

$$\varepsilon_{1t}(\hat{\vartheta}_{1T}, \dot{B}_T) = \varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T) - (\hat{\alpha}_T - \dot{\alpha}_T)(\hat{A}_T - A_0)y_{2,t-1}$$
$$- (\hat{\alpha}_T - \dot{\alpha}_T)[\hat{d}_T f(\hat{\gamma}_T) - d_0 f_t(\gamma_0)]. \quad \textbf{(A.19)}$$

Because the assumptions of Theorem 3.1 are supposed to hold we can conclude that $\hat{A}_T = A_0 + O_p(T^{-1})$, $\hat{d}_T = d_0 + O_p(T^{-1/2})$, and $\hat{\gamma}_T = \gamma_0 + O_p(T^{-1/2})$, whereas the function $f_t(\gamma)$ is assumed to satisfy Condition 1 of Saikkonen (2001) in some compact neighborhood of γ_0. That the first term on the right hand side of (A.17) is of order $o_p(1)$ can be established by using these facts, the identities (A.16), (A.18), and (A.19), the consistency properties of the involved estimators, and Lemmas A.1 and A.3 of Saikkonen (2001). The proof for the second term on the right hand side of (A.17) is entirely similar except that in the quantity containing cross products of the last terms on the right hand sides of (A.18) and (A.19) a (termwise) mean value expansion is first employed. (If desired, a mean value expansion can also be used in several other cases.) Details of these derivations are straightforward but somewhat tedious and will therefore be omitted. For the proofs of parts (b) and (c) it is useful to note that the preceding arguments can also be used to show that all sums on the right hand side of (A.17) are of order $O_p(1)$.

Thus, we have proved (A.14) and thereby (A.12). To prove (A.12) with $\hat{B}_T$ replaced by $\tilde{B}_T$, first notice that the considered restrictions only concern the parameter ϑ_1 and not B. This means that the inequality in (A.13) holds even if $\hat{\vartheta}_{1T}$, $\hat{B}_T$, and $\hat{\Omega}_T$ are replaced by their constrained counterparts that have the same consistency properties as $\hat{\vartheta}_{1T}$, $\hat{B}_T$, and $\hat{\Omega}_T$ (see the discussion at the end of Sect. 3). Thus, the previous proof of (A.12) also applies in the case of constrained estimators implying the first assertion of the lemma. The arguments used to show that (A.17) is of order $o_p(1)$ readily prove the second and third assertions. ■

Lemma A.1 is used to obtain the limiting distribution of test statistic $\mathcal{LR}_1$. For test statistic $\mathcal{LR}_2$ we use the following lemma.

LEMMA A.2. *Suppose the assumptions of Theorem* 4.2 *hold and let* $\bar{\Omega}_T$ *be any random matrix with the property* $\bar{\Omega}_T = \Omega_0 + o_p(1)$. *Then,*

$$\ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T, \bar{\Omega}_T) - \ell_{1T}(\tilde{\underline{\vartheta}}_{1T}, \tilde{\psi}_T, \bar{\Omega}_T) = o_p(1).$$

Proof. First note that

$$\ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \tilde{\psi}_T, \bar{\Omega}_T) - \ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) = o_p(1). \tag{A.20}$$

Because $\ell_{1T}(\underline{\vartheta}_1, \psi, \Omega)$ depends on ψ only through $B(\psi)$ and because $\hat{B}_T$ and the other involved estimators satisfy the same consistency properties as in (A.17) this can be seen by arguments entirely similar to those used to show that the right hand side of (A.17) is of order $o_p(1)$ and also in the proofs of parts (b) and (c) of Lemma A.1. Further, because the same arguments apply with $\hat{\underline{\vartheta}}_{1T}$ and $\hat{\psi}_T$ replaced by $\tilde{\underline{\vartheta}}_{1T}$ and $\tilde{\psi}_T$, respectively, the result of the lemma follows if we show that

$$\ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) - \ell_{1T}(\tilde{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) = o_p(1). \tag{A.21}$$

Denote $\dot{\underline{\vartheta}}_{1T} = (\dot{A}_T, \dot{d}_T, \dot{\gamma}_T)$. To prove (A.21), it suffices to show that

$$\ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) - \ell_{1T}(\dot{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) = o_p(1) \tag{A.22}$$

and similarly with $\hat{\underline{\vartheta}}_{1T}$ replaced by $\tilde{\underline{\vartheta}}_{1T}$. To this end, notice that

$$\ell_{1T}(\dot{\underline{\vartheta}}_{1T}, \hat{\psi}_T, \hat{\Omega}_T) - \ell_{1T}(\dot{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) = o_p(1)$$

again by arguments similar to those used to show that the right hand side of (A.17) is of order $o_p(1)$ and also to prove parts (b) and (c) of Lemma A.1. Using this result and (A.20) one obtains

$$\begin{aligned} 0 &\geq \ell_T(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T, \hat{\Omega}_T) - \ell_T(\dot{\underline{\vartheta}}_{1T}, \hat{\psi}_T, \hat{\Omega}_T) \\ &= \ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \hat{\psi}_T, \hat{\Omega}_T) - \ell_{1T}(\dot{\underline{\vartheta}}_{1T}, \hat{\psi}_T, \hat{\Omega}_T) \\ &= \ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) - \ell_T(\dot{\underline{\vartheta}}_{1T}, \psi_0, \Omega_0) + o_p(1). \end{aligned}$$

By the definition of $\dot{\underline{\vartheta}}_{1T}$ the difference in the last expression is nonnegative so that (A.22) follows. Because the same arguments apply with $\hat{\underline{\vartheta}}_{1T}$ replaced by $\tilde{\underline{\vartheta}}_{1T}$ the proof of the lemma is complete. ■

A.2.2. Proof of Theorem 4.1. First consider the Wald test and note that we have $\Upsilon_T(\hat{\vartheta}_{1T} - \vartheta_{10}) = O_p(1)$, $\hat{B}_T = B_0 + O_p(T^{-1/2})$, and $\hat{\Omega}_T = \Omega_0 + o_p(1)$ by assumption. Thus, because the matrices Υ_T and $H_1(\vartheta_1)$ commute and $h_1(\vartheta_{10}) = 0$, a standard mean value expansion yields

$$\Upsilon_T h_1(\hat{\vartheta}_{1T}) = H_1(\vartheta_{10})\Upsilon_T(\hat{\vartheta}_{1T} - \vartheta_{10}) + o_p(1),$$

whereas (A.4) shows that

$$\Upsilon_T^{-1}\hat{M}_{1T}\Upsilon_T^{-1} \Rightarrow \int_0^1 F_1(x;\vartheta_{10})\alpha_0'\Omega_0^{-1}\alpha_0 F_1(x;\vartheta_{10})'dx \overset{\text{def}}{=} M_1. \quad \textbf{(A.23)}$$

Hence, because $H(\hat{\vartheta}_{1T}) = H(\vartheta_{10}) + o_p(1)$ we can conclude from the preceding discussion and Theorem 3.1 that

$$\mathcal{W}_1 \Rightarrow \left(H_1(\vartheta_{10})\int_0^1 F_1(x;\vartheta_{10})\alpha_0'\Omega_0^{-1}dW_0(x)\right)'\left(H_1(\vartheta_{10})'M_1^{-1}H_1(\vartheta_{10})\right)^{-1}$$
$$\times\left(H_1(\vartheta_{10})\int_0^1 F_1(x;\vartheta_{10})\alpha_0'\Omega_0^{-1}dW_0(x)\right).$$

The stated limiting distribution is obtained by observing that conditional on $F_1(x;\vartheta_{10})$ the $f_1 \times 1$ vector in the parentheses is normally distributed with zero mean and covariance matrix $H_1(\vartheta_{10})'M_1^{-1}H(\vartheta_{10})$ (cf. Theorem 3.1 and the discussion following it).

Next consider the LM test and note that $\Upsilon_T(\tilde{\vartheta}_{1T} - \vartheta_{10}) = O_p(1)$, $\tilde{B}_T = B_0 + O_p(T^{-1/2})$, and $\tilde{\Omega}_T = \Omega_0 + o_p(1)$ by assumption. Analogously to (A.1) we have

$$\begin{aligned}
&2\Upsilon_T^{-1}\sum_{t=1}^T F_{1t}(\vartheta_{10})\hat{\alpha}_T'\tilde{\Omega}_T^{-1}\varepsilon_t(\vartheta_{10},\tilde{B}_T)\\
&\quad = 2\Upsilon_T^{-1}\sum_{t=1}^T F_{1t}(\tilde{\vartheta}_{1T})\tilde{\alpha}_T'\tilde{\Omega}_T^{-1}\varepsilon_t(\tilde{\vartheta}_{1T},\tilde{B}_T)\\
&\qquad + 2\Upsilon_T^{-1}\sum_{t=1}^T F_{1t}(\tilde{\vartheta}_{1T})\tilde{\alpha}_T'\tilde{\Omega}_T^{-1}[\varepsilon_t(\vartheta_{10},\tilde{B}_T) - \varepsilon_t(\tilde{\vartheta}_{1T},\tilde{B}_T)]\\
&\qquad + 2\Upsilon_T^{-1}\sum_{t=1}^T [F_{1t}(\vartheta_{10}) - F_{1t}(\tilde{\vartheta}_{1T})]\tilde{\alpha}_T'\tilde{\Omega}_T^{-1}\varepsilon_t(\vartheta_{10},\tilde{B}_T).
\end{aligned} \quad \textbf{(A.24)}$$

Because the consistency properties of the involved estimators are the same as in (A.1) the last term on the right hand side of (A.24) is of order $o_p(1)$ by the arguments given for the corresponding term in (A.1). Further, in the same way as in (A.2) we can use a mean value expansion to express the second term on the right hand side of (A.24) as

$$2\Upsilon_T^{-1}\sum_{t=1}^T F_{1t}(\tilde{\vartheta}_{1T})\tilde{\alpha}_T'\tilde{\Omega}_T^{-1}\tilde{\alpha}_T F_{1t}(\bar{\vartheta}_{1T})'(\tilde{\vartheta}_{1T} - \vartheta_{10}) = 2\widetilde{M}_T\Upsilon_T(\tilde{\vartheta}_{1T} - \vartheta_{10}) + o_p(1),$$

where $\bar{\vartheta}_{1T}$ is defined in the same way as before. To justify the equality, notice that it results from replacing $\bar{\vartheta}_{1T}$ by $\tilde{\vartheta}_{1T}$ and this can be justified in the same way as the equality in (A.4). Thus, observing that (A.3) also holds with $\hat{B}_T$ and $\hat{\Omega}_T$ replaced by $\tilde{B}_T$ and $\tilde{\Omega}_T$, respectively, we find from the preceding discussion that (A.24) can be written as

$$2\Upsilon_T^{-1}\sum_{t=1}^{T} F_{1t}(\vartheta_{10})\alpha'\Omega^{-1}\varepsilon_t$$

$$= 2\Upsilon_T^{-1}\sum_{t=1}^{T} F_{1t}(\tilde{\vartheta}_{1T})\tilde{\alpha}_T'\tilde{\Omega}_T^{-1}\varepsilon_t(\tilde{\vartheta}_{1T},\tilde{B}_T) + 2(\Upsilon_T^{-1}\widetilde{M}_{1T}\Upsilon_T^{-1})\Upsilon_T(\tilde{\vartheta}_{1T}-\vartheta_{10}) + o_p(1). \tag{A.25}$$

Here the first term on the right hand side equals $-\Upsilon_T^{-1}\partial\ell_T(\tilde{\vartheta}_{1T},\tilde{B}_T,\tilde{\Omega}_T)/\partial\vartheta_1$ (see (4.5)). As for the second term, use the identity $h_1(\tilde{\vartheta}_{1T}) = 0$ and a standard mean value expansion to conclude that $H_1(\tilde{\vartheta}_{1t})\Upsilon_T(\tilde{\vartheta}_{1T}-\vartheta_{10}) = o_p(1)$. Thus, one obtains from (A.25) that

$$H_1(\tilde{\vartheta}_{1T})(\Upsilon_T^{-1}\widetilde{M}_{1T}\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\partial\ell_T(\tilde{\vartheta}_{1T},\tilde{B}_T,\tilde{\Omega}_T)/\partial\vartheta_1$$

$$= -2H_1(\tilde{\vartheta}_{1T})(\Upsilon_T^{-1}\widetilde{M}_{1T}\Upsilon_T^{-1})^{-1}\Upsilon_T^{-1}\sum_{t=1}^{T} F_{1t}(\vartheta_{10})\alpha'\Omega^{-1}\varepsilon_t + o_p(1). \tag{A.26}$$

To complete the proof, observe that (A.23) also holds with $\hat{M}_{1T}$ replaced by $\widetilde{M}_{1T}$ so that the limiting distribution of test statistic $\mathcal{LM}_1$ can be obtained from (A.26) by using (A.3) and arguments similar to those used for $\mathcal{W}_1$.

Finally, consider the LR test and conclude from a mean value expansion that

$$\mathcal{LR}_1 = \ell_T(\tilde{\vartheta}_{1T},\tilde{B}_T,\tilde{\Omega}_T) - \ell_T(\hat{\vartheta}_{1T},\hat{B}_T,\hat{\Omega}_T)$$

$$= \operatorname{tr}\left\{\bar{\Omega}_T^{-1}\left(\sum_{t=1}^{T}\varepsilon_t(\tilde{\vartheta}_{1T},\tilde{B}_T)\varepsilon_t(\tilde{\vartheta}_{1T},\tilde{B}_T)' - \sum_{t=1}^{T}\varepsilon_t(\hat{\vartheta}_{1T},\hat{B}_T)\varepsilon_t(\hat{\vartheta}_{1T},\hat{B}_T)'\right)\right\},$$

where $\bar{\Omega}_T$ is defined in the same way as other similar notations and has the property $\bar{\Omega}_T = \Omega_0 + o_p(1)$. Using the identity $\varepsilon_t(\vartheta_1,B) = \varepsilon_{1t}(\vartheta_1,B) + \varepsilon_{2t}(B)$ in the last expression and the decomposition of the log-likelihood function given in Section A.2.1 it is straightforward to check that

$$\mathcal{LR}_1 = \ell_{1T}(\tilde{\vartheta}_{1T},\tilde{B}_T,\bar{\Omega}_T) - \ell_{1T}(\hat{\vartheta}_{1T},\hat{B}_T,\bar{\Omega}_T) + \ell_{2T}(\tilde{B}_T,\bar{\Omega}_T) - \ell_{2T}(\hat{B}_T,\bar{\Omega}_T)$$

$$= \ell_{1T}(\tilde{\vartheta}_{1T},\hat{B}_T,\hat{\Omega}_T) - \ell_{1T}(\hat{\vartheta}_{1T},\hat{B}_T,\hat{\Omega}_T) + o_p(1),$$

where the latter equality follows from Lemma A.1. Using the definition of $\ell_{1T}(\vartheta_1,B,\Omega)$ we can thus write

$$\mathcal{LR}_1 = 2\operatorname{tr}\left\{\hat{\Omega}_T^{-1}\sum_{t=1}^{T}[\varepsilon_{1t}(\tilde{\vartheta}_{1T},\hat{B}_T) - \varepsilon_{1t}(\hat{\vartheta}_{1T},\hat{B}_T)]\varepsilon_{2t}(\hat{B}_T)'\right\}$$

$$+ \operatorname{tr}\left\{\hat{\Omega}_T^{-1}\sum_{t=1}^{T}[\varepsilon_{1t}(\tilde{\vartheta}_{1T},\hat{B}_T) - \varepsilon_{1t}(\hat{\vartheta}_{1T},\hat{B}_T)]\varepsilon_{1t}(\tilde{\vartheta}_{1T},\hat{B}_T)'\right\}$$

$$+ \operatorname{tr}\left\{\hat{\Omega}_T^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\hat{\vartheta}_{1T},\hat{B}_T)[\varepsilon_{1t}(\tilde{\vartheta}_{1T},\hat{B}_T) - \varepsilon_{1t}(\hat{\vartheta}_{1T},\hat{B}_T)]'\right\} + o_p(1).$$

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

Here we can use a standard mean value expansion and further Lemmas A.1 and A.3 of Saikkonen (2001) to show that replacing $\varepsilon_{1t}(\tilde{\vartheta}_{1T}, \hat{B}_T)$ by $\varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T) - \hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})'(\tilde{\vartheta}_{1T} - \hat{\vartheta}_{1T})$ causes an error of order $o_p(1)$. Details are similar to those used for (A.17). Hence, after straightforward calculations we find that

$$\begin{aligned}\mathcal{LR}_1 &= -2(\tilde{\vartheta}_{1T} - \hat{\vartheta}_{1T})' \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1}[\varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T) + \varepsilon_{2t}(\hat{B}_T)] \\ &\quad + (\tilde{\vartheta}_{1T} - \hat{\vartheta}_{1T})' \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1} \hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})' \tilde{\vartheta}_{1t} - \hat{\vartheta}_{1T}) + o_p(1) \\ &= (\tilde{\vartheta}_{1T} - \hat{\vartheta}_{1T})' \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1} \hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})'(\tilde{\vartheta}_{1T} - \hat{\vartheta}_{1T}) + o_p(1). \qquad \textbf{(A.27)}\end{aligned}$$

Here the latter equality follows because the sum in the first term on the right hand side of the first equality equals $-2^{-1}\partial\ell_{1T}(\hat{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T)/\partial\vartheta_1 = 0$. Using this identity and arguments similar to those given previously we also find that

$$\begin{aligned}&\Upsilon_T^{-1}\partial\ell_{1T}(\tilde{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T)/\partial\vartheta_1 \\ &\quad = -2\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\tilde{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1} \varepsilon_t(\tilde{\vartheta}_{1T}, \hat{B}_T) \\ &\quad = -2\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\tilde{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1}[\varepsilon_{1t}(\hat{\vartheta}_{1T}, \hat{B}_T) + \varepsilon_{2t}(\hat{B}_T)] \\ &\qquad - 2\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\tilde{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1}[\varepsilon_t(\tilde{\vartheta}_{1T}, \hat{B}_T) - \varepsilon_t(\hat{\vartheta}_{1T}, \hat{B}_T)] \\ &\quad = -2\Upsilon_T^{-1} \sum_{t=1}^{T} F_{1t}(\hat{\vartheta}_{1T})\hat{\alpha}_T' \hat{\Omega}_T^{-1} \hat{\alpha}_T F_{1t}(\hat{\vartheta}_{1T})'(\tilde{\vartheta}_{1T} - \hat{\vartheta}_{1T}) + o_p(1).\end{aligned}$$

Using similar arguments it can further be seen that $\Upsilon_T^{-1}\partial\ell_{1T}(\tilde{\vartheta}_{1T}, \hat{B}_T, \hat{\Omega}_T)/\partial\vartheta_1 = \Upsilon_T^{-1}\partial\ell_{1T}(\tilde{\vartheta}_{1T}, \tilde{B}_T, \tilde{\Omega}_T)/\partial\vartheta_1 + o_p(1)$ and $\Upsilon_T^{-1}\widetilde{M}_{1T}\Upsilon_T^{-1} = \Upsilon_T^{-1}\hat{M}_{1T}\Upsilon_T^{-1} + o_p(1)$. Combining these facts with the expression of test statistic $\mathcal{LR}_1$ given in (A.27) and using the definition of the matrix $\widetilde{M}_{1T}$ yields

$$\begin{aligned}\mathcal{LR}_1 &= 4^{-1}(\Upsilon_T^{-1}\partial\ell_{1T}(\tilde{\vartheta}_{1T}, \tilde{B}_T, \tilde{\Omega}_T)/\partial\vartheta_1)'(\Upsilon_T^{-1}\widetilde{M}_{1T}\Upsilon_T^{-1})^{-1}(\Upsilon_T^{-1}\partial\ell_{1T}(\tilde{\vartheta}_{1T}, \tilde{B}_T, \tilde{\Omega}_T)/\partial\vartheta_1) \\ &\quad + o_p(1). \qquad \textbf{(A.28)}\end{aligned}$$

Making use of (A.26), (A.28), and a standard argument based on Lagrange multipliers it can finally be seen that $\mathcal{LR}_1 = \mathcal{LM}_1 + o_p(1)$ (cf. Gallant, 1987, pp. 229–230). Thus, because we have proved that test statistic $\mathcal{LM}_1$ has the stated limiting distribution the same is true for test statistic $\mathcal{LR}_1$. ■

A.2.3. Proof of Theorem 4.2. First consider the Wald test and note that, by Theorem 3.2 and the continuity of $H_2(\psi), H_2(\hat{\psi}_T) = H_2(\psi_0) + o_p(1)$ and $h_2(\hat{\psi}_T) = H_2(\psi_0)(\hat{\psi}_T - \psi_0) + o_p(T^{-1/2})$. Because (A.11) implies that $T^{-1}\hat{M}_{2T} = \Sigma_\psi(\psi_0, \Omega_0)^{-1} + o_p(1)$ the stated result follows from Theorem 3.2.

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press

As for the LM test, note first that $\tilde{\psi}_T = \psi_0 + O_p(T^{-1/2})$ and $\hat{\alpha}_T = \alpha_0 + O_p(T^{-1/2})$ by Theorem 3.2 and the discussion at the end of Section 3, whereas $\tilde{\Omega}_T = \Omega_0 + o_p(1)$ by assumption. Next recall that we showed that the two last terms on the right hand side of (A.5) are of order $o_p(1)$. Thus, because the consistency properties of the constrained estimators are the same as their unconstrained counterparts the arguments used for (A.5) can be repeated to show that

$$2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\psi_0)\tilde{\Omega}_T^{-1}\varepsilon_t(\underline{\tilde{\vartheta}}_{1T},\psi_0)$$

$$= 2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)\tilde{\Omega}_T^{-1}\varepsilon_t(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)$$

$$+ 2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)\tilde{\Omega}_T^{-1}[\varepsilon_t(\underline{\tilde{\vartheta}}_{1T},\psi_0) - \varepsilon_t(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)] + o_p(1).$$

Further, arguments similar to those used in (A.9) show that we can write the second term in the last expression as

$$2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)\tilde{\Omega}_T^{-1}F_{2t}(\underline{\tilde{\vartheta}}_{1T},\bar{\psi}_T)(\tilde{\psi}_T - \psi_0)$$

$$= 2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)\tilde{\Omega}_T^{-1}F_{2t}(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)'(\tilde{\psi}_T - \psi_0) + o_p(1),$$

where the notation $\bar{\psi}_T$ is defined as before and the equality can be justified in the same way as in (A.11). Thus, using the definition of $\widetilde{M}_{2T}$ we can conclude that

$$2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\psi_0)\tilde{\Omega}_T^{-1}\varepsilon_t(\underline{\tilde{\vartheta}}_{1T},\psi_0)$$

$$= 2T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T)\tilde{\Omega}_T^{-1}\varepsilon_t(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T) + 2(T^{-1}\widetilde{M}_{2T})T^{1/2}(\tilde{\psi}_T - \psi_0) + o_p(1).$$

Here the first term on the right hand side equals $T^{-1/2}\partial\ell_T(\tilde{\vartheta}_{1T},\tilde{\psi}_T,\tilde{\Omega}_T)/\partial\psi$. Thus, observing that (A.10) also holds with $\underline{\hat{\vartheta}}_{1T}$ and $\hat{\Omega}_T$ replaced by $\underline{\tilde{\vartheta}}_{1T}$ and $\tilde{\Omega}_T$, respectively, and proceeding in the same way as following (A.25) we find that $T^{1/2}H_2(\tilde{\psi}_T)(\hat{\psi}_T - \psi_0) = o_p(1)$ and, furthermore,

$$H_2(\tilde{\psi}_T)(T^{-1}\widetilde{M}_{2T})^{-1}T^{-1/2}\partial\ell_T(\underline{\tilde{\vartheta}}_{1T},\tilde{\psi}_T,\tilde{\Omega}_T)/\partial\psi$$

$$= 2H_2(\tilde{\psi}_T)(T^{-1}\widetilde{M}_{2T})^{-1}T^{-1/2}\sum_{t=1}^{T} F_{2t}(\underline{\vartheta}_{10},\psi_0)\Omega_0^{-1}\varepsilon_t + o_p(1).$$

Because $T^{-1}\widetilde{M}_{2T}$ is asymptotically equivalent to $T^{-1}\hat{M}_{2T}$ the limiting distribution of test statistic $\mathcal{LM}_2$ can be obtained from this by using (A.10) and arguments similar to those used for $\mathcal{W}_2$.

Finally, consider the LR test. By a mean value expansion, similar to that in the proof of Theorem 4.1, we have

$$\begin{aligned}\mathcal{LR}_2 &= \ell_{1T}(\underline{\tilde{\vartheta}}_{1T}, \tilde{\psi}_T, \bar{\Omega}_T) - \ell_{1T}(\underline{\hat{\vartheta}}_{1T}, \hat{\psi}_T, \bar{\Omega}_T) + \ell_{2T}(\tilde{\psi}_T, \bar{\Omega}_T) - \ell_{2T}(\hat{\psi}_T, \bar{\Omega}_T)\\ &= \ell_{2T}(\tilde{\psi}_T, \bar{\Omega}_T) - \ell_{2T}(\hat{\psi}_T, \bar{\Omega}_T) + o_p(1),\end{aligned}$$

where the latter equality follows from Lemma A.2. Now consider the infeasible estimator $\dot{\psi}_T$ and note that it is straightforward to specialize Theorem 3.2 for it and show that $\dot{\psi}_T$ has the same limiting distribution as $\hat{\psi}_T$. In fact, (A.9)–(A.11) and standard arguments applied to the estimator $\dot{\psi}_T$ show that $\dot{\psi}_t$ and $\hat{\psi}_T$ are asymptotically equivalent in the sense that $T^{1/2}(\hat{\psi}_T - \dot{\psi}_T) = o_p(1)$. The same arguments can also be used to show that $T^{1/2}(\tilde{\psi}_T - \ddot{\psi}_T) = o_p(1)$ where $\ddot{\psi}_T$ is the constrained counterpart of $\dot{\psi}_T$. Thus, because the estimators $\dot{\psi}_T$ and $\dot{\Omega}_T$ and their constrained counterparts can be obtained from the likelihood function $\ell_{2T}(\psi, \Omega)$ it is straightforward to conclude from the preceding representation of $\mathcal{LR}_2$ that

$$\mathcal{LR}_2 = \ell_{2T}(\dot{\psi}_T, \dot{\Omega}_T) - \ell_{2T}(\ddot{\psi}_T, \ddot{\Omega}_T) + o_p(1).$$

Because standard asymptotics hold for ML estimators based on the likelihood function $\ell_{2T}(\psi, \Omega)$ the standard chi-square limiting is also obtained for LR tests based on it and hence for $\mathcal{LR}_2$. ■

https://doi.org/10.1017/S0266466601172038 Published online by Cambridge University Press