Econometric Theory, **17**, 2001, 296–326. Printed in the United States of America.

CONSISTENT ESTIMATION IN COINTEGRATED VECTOR AUTOREGRESSIVE MODELS WITH NONLINEAR TIME TRENDS IN COINTEGRATING RELATIONS

PENTTI SAIKKONEN
University of Helsinki

This paper studies the consistency of the Gaussian maximum likelihood estimator in a cointegrated vector autoregressive model with nonlinear time trends in cointegrating relations. The results are proved in a coordinate free framework that readily allows for general nonlinear parameter restrictions and makes it possible to show the consistency of reduced form parameter estimators without assuming identifiability of underlying structural parameters. Various consistency results for structural parameter estimators can then be obtained by imposing suitable identification conditions for the parameters of interest but not necessarily for nuisance parameters. Orders of consistency are also obtained because they are needed to develop a related asymptotic theory of statistical inference.

1. INTRODUCTION

The statistical theory of cointegrated systems has experienced considerable development since the seminal paper by Engle and Granger (1987). As far as parameter estimation is concerned, most of the research has been focused on problems that are linear or can be handled by methods applicable in linear problems. Papers dealing with theoretical aspects of genuinely nonlinear estimation problems are still rather few. Reference can be made to Phillips (1991), Saikkonen (1993a, 1995), Wooldridge (1994), Johansen (1997), Pesaran and Shin (1997), and Park and Phillips (1998). Except for Wooldridge (1994) and Park and Phillips (1998), this previous work has been restricted to models where nonlinearity only appears in parameters and not in variables. Both of these two papers study nonlinear least squares estimation in single equation cointegrating regressions. Wooldridge (1994) only allows for nonlinear functions of stationary variables, whereas Park and Phillips (1998) concentrate on a non-

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 useful 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.

 © 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

linear regression model with a scalar regressor integrated of order one ($I(1)$). The purpose of this paper and its sequel (Saikkonen, 2001) is to develop a fairly complete asymptotic estimation theory for Gaussian vector autoregressive (VAR) models with general parameter restrictions and nonlinear deterministic time trends in cointegrating relations. This paper formulates the considered model and proves the consistency of the maximum likelihood (ML) estimator, whereas the sequel deals with asymptotic inference. It is hoped that the methods of proof used in these two papers can also be useful in other contexts involving nonlinearities and integrated processes.

A major motivation of including nonlinear time trends in cointegrating relations is to allow for the effect of structural changes such as level shifts. Such structural changes may be sudden or gradual, and they should ideally have a clear interpretation. If they are erroneously neglected serious misinterpretations of the estimated model can result. In particular, the cointegrating rank of the model may be incorrectly estimated. Ripatti and Saikkonen (2001; see also Heinesen, 1997) provide a more thorough discussion of these aspects in the context of a VAR model and illustrate them by an empirical application to interest rate data. Similar ideas have also been used in the related univariate problem of testing for an autoregressive unit root. The recent paper by Leybourne, Newbold, and Vougas (1998) is pertinent to our work and contains a good discussion of the potential usefulness of models with gradual structural changes. The wide literature on sudden structural breaks is also worth mentioning, although the case of an unknown break date is excluded from our analysis (for some recent work on structural breaks in cointegrated systems see Quintos, 1997; Seo, 1998; Inoue, 1999; and the references therein).

The rest of the paper is organized as follows. Section 2 formulates the model to be considered in subsequent sections and also in Saikkonen (2001). The existence and consistency of the ML estimator are studied in Section 3. Unlike in some previous work (see, e.g., Saikkonen, 1993a; Wooldridge, 1994) our results are obtained for a global maximizer of the likelihood function instead of a local maximizer. Our consistency results are proved in a coordinate free framework similar to that used by Johansen (1997) in a model without nonlinear time trends. A convenient feature of this approach is that it makes it possible to obtain consistency results without requiring identifiability of all structural parameters of the model. This relates our work to the recent research on the estimation of partially identified models studied by Phillips (1989) and Choi and Phillips (1992). Finally, Section 4 provides a discussion of the results with potential extensions, and an Appendix contains a number of useful auxiliary results in addition to proofs of the theorems.

The following notation will be used throughout the paper. The vec operator stacks the columns of a matrix one below another, and the symbol $\otimes$ indicates the Kronecker product for matrices. For an $n \times m$ matrix A of full column rank, $A_\perp$ is an $n \times (n - m)$ matrix of full column rank and such that $A'A_\perp = 0$ $(0 < m < n)$. If A is a square matrix its smallest and largest eigenvalues are

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

denoted by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$, respectively, whereas $\det(A)$ and $\mathrm{tr}(A)$ denote the determinant and trace of A, respectively. When applied to matrices, the inequality $A > B$ ($A \geq B$) means that the difference $A - B$ is positive definite (semidefinite). By $\|A\|$ we denote the Euclidean matrix norm of the matrix A, that is, $\|A\| = \{\mathrm{tr}(A'A)\}^{1/2}$. In the context of parameters and parameter spaces the subscript 0 is used to emphasize "true" parameter values, and $\bar{S}$ is used for the closure of the set S. The standard notation $D[0,1]^n$ signifies the product space of n copies of $D[0,1]$, the space of functions on $[0,1]$ that are right continuous and have left hand limits. The symbols $\Rightarrow$ and $\xrightarrow{p}$ signify weak convergence and convergence in probability, respectively, and $[\cdot]$ denotes the integer part of the indicated real number. The usual lag operator is denoted by L whereas $\Delta = 1 - L$ is the related difference operator.

2. MODEL AND ASSUMPTIONS

2.1. Specification of the Stochastic Part of the Model

Let $y_t, t = 1,\ldots,T$ be an s-dimensional time series generated by a VAR process of order p. The error correction form of the process reads as

$$\Delta y_t = d_t + \Pi y_{t-1} + \sum_{j=1}^{p-1} B_j \Delta y_{t-j} + \varepsilon_t, \qquad t = 1,2,\ldots, \tag{2.1}$$

where Π ($s \times s$) and B_j ($s \times s$) are unknown parameters and d_t ($s \times 1$) is a deterministic sequence to be discussed in the next section. Furthermore, the initial values $y_{-p+1},\ldots,y_0$ are observable and ε_t is Gaussian white noise, that is, $\varepsilon_t \sim NID(0,\Omega)$ with $\Omega > 0$. This last assumption is only made for simplicity and could be replaced by a more general martingale difference assumption as in Sims, Stock, and Watson (1990). Because we are interested in time series whose stochastic components are integrated of order one and cointegrated we make the following assumption, which implies that the version of Granger's representation theorem in Johansen (1995, p. 49) applies.

Assumption 1.

(a) The matrix Π in (2.1) is of rank r ($0 < r < s$) so that it can be written as

$$\Pi = \alpha\beta', \tag{2.2}$$

where α and β are $s \times r$ matrices of full column rank.

(b) If $\det[(I_s - \sum_{j=1}^{p-1} B_j z^j)(1 - z) - \Pi z] = 0$ then $|z| > 1$ or $z = 1$.

(c) The matrix $\alpha_\perp'(I_s - \sum_{j=1}^{p-1} B_j)\beta_\perp$ is of full rank $s - r$.

When Assumption 1 holds, the columns of the matrix β are cointegrating vectors so that, with a suitable specification of initial values, both Δy_t and $\beta' y_t$ are stationary around deterministic trends, although the process y_t itself contains a nonstationary $I(1)$ component (see Johansen, 1995, p. 49). In this paper

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

the cointegrating rank, r, is supposed to be correctly specified, possibly with the help of some test procedure.

We also need additional conditions to guarantee the uniqueness of cointegrating vectors. This is part of the following assumption.

Assumption 2.

(a) The matrix of cointegrating vectors is of the form $\beta = \beta(\phi) = [I_r \ -A(\phi)]'$ where $A(\phi)$ is a continuous function from a set $\Phi \subset \mathbb{R}^k$ to $\mathbb{R}^{r\times(s-r)}$, $\underline{k \leq r(s-r)}$.
(b) $A(\phi_0)$ is an interior point of $A(\Phi)$ in the relative topology of $\overline{A(\Phi)}$, and in some neighborhood of ϕ_0 the function $A(\phi)$ is bijective and has a continuous inverse in some neighborhood of $A(\phi_0)$.

Thus, we assume that the cointegrating vectors can be specified by a continuous, possibly nonlinear, function of a k-dimensional parameter vector ϕ. Note that we do not assume compactness of the parameter space, as is often done in nonlinear estimation problems. A price to be paid for this is that proving the existence and consistency of a ML estimator of ϕ gets somewhat complicated. In fact, we shall first prove these results for the reduced form parameter matrix $A = A(\phi)$ when its parameter space is extended from $A(\Phi)$ to $\overline{A(\Phi)}$. Assuming that $A(\phi_0)$ is an interior point of $A(\Phi)$ in the relative topology of $\overline{A(\Phi)}$ ensures that in large samples a consistent estimator of $A(\phi)$ defined in $\overline{A(\Phi)}$ still takes values in $A(\Phi)$. Readers who wish to avoid topological technicalities may instead of the first condition of Assumption 2(b) simply assume that $A(\Phi)$ is open or closed. In the latter case the true parameter value $A(\phi_0)$ may be a boundary point of $A(\Phi)$. In its general form Assumption 2(b) rules out the possibility that $A(\phi_0)$ is among certain "corner" points of $A(\Phi)$. To give a concrete though somewhat artificial example, suppose that $A(\Phi) = \{(a_1, a_2) \in \mathbb{R}^2 : -1 < a_1 < 1, -1 < a_2 \leq 1\}$. In this case the true parameter value $A(\phi_0)$ may not be $(-1,1)$ or $(1,1)$. If $A(\phi_0)$ were one of these two points a consistent estimator taking values in $\overline{A(\Phi)}$ could even in large samples estimate a_1 by 1 or -1 and would therefore not belong to $A(\Phi)$.

As a whole Assumption 2(b) implies that a ML estimator of ϕ is well defined when the sample size is large enough and, furthermore, that a consistent estimator of ϕ can be recovered from a consistent estimator of $A(\phi)$. Assumption 2(b) is actually an identification condition for the parameter ϕ based on the well-known idea discussed in Rothenberg (1971, p. 585). This identification condition is slightly stronger than the commonly used requirement $A(\phi) \neq A(\phi_0)$, $\phi \neq \phi_0$, although from a practical point of view there is probably not much difference between these two conditions. An advantage of the formulation of Assumption 2 is its generality. Indeed, because the parameter space Φ is an arbitrary subset of $\mathbb{R}^k$ our treatment also covers the case where conventional restrictions are imposed on the parameter vector ϕ. It may also be noted that, in addition to $A(\phi_0)$, Assumption 2(b) also permits ϕ_0 to be a boundary point of its parameter space.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

Assuming that the first r rows of β form an identity matrix is a convenient and frequently used normalizing restriction. It is easy to see that it is equivalent to the requirement that the last $s - r$ components of y_t are not cointegrated. Thus, the validity of this assumption is testable (see Saikkonen, 1992; Boswijk, 1996; Luukkonen, Ripatti, and Saikkonen, 1999). When the parameter matrix α is unrestricted, one can have a specification such as $\beta' = [\beta_1(\phi)\ \beta_2(\phi)]$ with $\beta_1(\phi)$ $(r \times r)$ nonsingular. This is obvious from (2.2) because $\alpha\beta'$ can be replaced by $\alpha_*[I_r\ -A(\phi)]$ with $\alpha_* = \alpha\beta_1(\phi)$ and $A(\phi) = -\beta_1(\phi)^{-1}\beta_2(\phi)$. If we have $\beta_1(\phi) = \beta_1$ and $\beta_2(\phi) = \beta_2$ with β_1 and β_2 unrestricted except for identifying zero restrictions our specification for the cointegrating relations $\beta' y_t$ becomes similar to that used in traditional simultaneous equations models. Note that in this case the matrix $A(\phi)$ is a nonlinear function of the unknown parameters in the matrix β even though this matrix itself is only restricted by zero restrictions. Saikkonen (1993b) provides a general discussion of structural cointegrating relations of this kind. It is easy to see that the identification conditions given therein are sufficient for Assumption 2. In VAR models similar structural cointegrating relations have recently been employed by Johansen and Juselius (1994), Boswijk (1995a, 1995b, 1996), and Pesaran and Shin (1997) among others. The last authors also consider the case where the cointegrating vectors are specified by a nonlinear function of an underlying structural parameter vector (for similar work, see also Johansen, 1991, 1997; Phillips, 1991).

We will also allow for the possibility that the short-run parameters α and $B_1, \ldots, B_{p-1}$ are nonlinear continuous functions of an underlying parameter vector. This is formalized in the following assumption.

Assumption 3.

(a) The parameter matrix $B = [B_1 \ldots B_{p-1}\alpha]$ is of the form $B = B(\psi)$ where $B(\psi)$ is a continuous function from a set $\Psi \subset \mathbb{R}^\ell$ to $\mathbb{R}^{s\times(p-1)s+r}$.

(b) $B(\psi_0)$ is an interior point of $B(\Psi)$ in the relative topology of $\overline{B(\Psi)}$, and in some neighborhood of ψ_0 the function $B(\psi)$ is bijective and has a continuous inverse in some neighborhood of $B(\psi_0)$.

Of course, Assumption 3 is entirely similar to Assumption 2 so that the remarks made regarding the latter also apply here. Assumption 3 implies that our model allows for very general restrictions on the short-run dynamics. Such restrictions have received attention in the recent literature on cointegrated VAR models. For instance, they may result from weak or strong exogeneity assumptions, as discussed by Boswijk (1995a), Hsiao (1997), and Pesaran, Shin, and Smith (1997) (for a general formulation see also Johansen and Juselius, 1994). The so-called conditional model obtained in this setup is similar to a traditional simultaneous equations model so that, in the same way as in simultaneous equations models, it can be transformed to a reduced form whose parameters are nonlinear functions of underlying structural parameters. A very different motivation for restricting short-run parameters is given by Reinsel and Ahn (1992)

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

and Ahn (1997), who assume a nested reduced rank structure for the parameter matrices $B_1,\ldots,B_{p-1}$ and thereby restrict them in a nonlinear fashion. Yet another example, which can also involve restrictions on the cointegrating vectors, is obtained by combining individual models specified for different sectors of an economy. This possibility was recently discussed by Granger and Swanson (1997, Sect. IV), who also provide further references.

Later we use the obvious notation $\alpha(\psi)$ and $B_j(\psi)(\,j=1,\ldots,p-1)$. However, we continue to use the reduced form parameters $B,\alpha,B_j(\,j=1,\ldots,p-1)$ and also A and β when there is no need to be explicit about their dependence on the structural parameters ψ and ϕ.

2.2. Specification of the Deterministic Part of the Model

In the applications of the preceding model it has so far been customary to assume that the deterministic sequence d_t contains an intercept term and conventional dummy variables. Sometimes a linear time trend has also been included. In this paper we are mainly interested in nonlinear deterministic trends but only of a certain rather specific type. The main idea is to allow for the possibility that a standard model may fail because the mean of the cointegrating relations $\beta' y_t$ is not constant. Reasons for this may be structural changes caused by interventions or changes in institutions, policy, etc.

In this paper we shall assume that the deterministic trend term d_t can be absorbed into the cointegrating relations so that

$$d_t = -\alpha g_t(\mu). \tag{2.3}$$

Here $g_t(\mu)(r\times 1)$ is a generally nonlinear deterministic function of time and the parameter vector μ. Thus, inserting (2.2) and (2.3) into (2.1) shows that the observed series is generated by

$$\Delta y_t = \alpha(\beta' y_{t-1} - g_t(\mu)) + \sum_{j=1}^{p-1} B_j \Delta y_{t-j} + \varepsilon_t, \qquad t=1,\ldots,T. \tag{2.4}$$

A convenient way to specify the sequence $g_t(\mu)$ is to follow the approach used in nonparametric regression and assume that $g_t(\mu)=g(t/T;\mu)$ where $g(\cdot;\mu)$ is a suitable function defined on the interval $[0,1]$. Note that, for ease of notation, we usually suppress the dependence of the sequence $g_t(\mu)$ on the sample size T and use the same notational convention also for other related quantities. The preceding approach of specifying the sequence $g_t(\mu)$ is discussed in Ripatti and Saikkonen (1998), where previous references can also be found. Here we only mention that in the context of cointegrating regressions deterministic trends with the scaled time index t/T have previously been used by Phillips and Hansen (1990) and Park and Hahn (1999). However, unlike ourselves, the former authors only deal with linear models, whereas the latter use a nonpara-

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

metric approach to model deterministic structural changes in cointegrating vectors.

To give an idea of how the function $g(x;\mu)$ might be specified in practice, consider the functions

$$g(x;\mu) = \mu_1 + [1 + \exp\{-\mu_2(x - \mu_3)\}]^{-1}\mu_4 \tag{2.5}$$

and

$$g(x;\mu) = \mu_1 + \exp\{-\mu_2(x - \mu_3)^2\}\mu_4, \tag{2.6}$$

where $\mu = [\mu_1' \ \mu_2 \ \mu_3 \ \mu_4']'$ with μ_1 and μ_4 $r \times 1$ parameter vectors and μ_2 and μ_3 scalar parameters with $\mu_2 > 0$ and $0 < \mu_3 < 1$. If (2.5) is employed a smooth change of the logistic form in the mean of the cointegrating relations is assumed, whereas a specification based on (2.6) assumes that such a smooth change can essentially be modeled by the density function of the normal distribution. To see the idea of (2.5) more clearly, suppose that the components of μ_4 are positive. Then the parameter vector μ_1 describes the mean of the cointegrating relations at the beginning of the sample, and $\mu_1 + \mu_4$ is a new mean value that applies after a smooth increasing change. The parameter μ_3 determines the average location of the change, whereas μ_2 is a slope parameter that indicates how rapid the change is. The smaller the value of the parameter μ_2 is the longer it takes before the mean value reaches its new level. When the value of μ_2 is "large" we are close to the case of a single structural break. This case is actually obtained in the limit by letting $\mu_2 \to \infty$ because then (2.5) approaches $\mu_1 + I(x \geq \mu_3)\mu_4$ where $I(\cdot)$ is the indicator function. If the value of the parameter μ_3 is known this amounts to using a conventional step dummy. If (2.6) is employed the assumed change in the mean of the cointegrating relations is nonmonotonic and symmetric about μ_3. For instance, when the components of μ_4 are positive the mean is first assumed to increase smoothly from μ_1 to $\mu_1 + \mu_4$ and then decrease back to μ_1. Again, μ_2 is a slope parameter that determines how rapid these changes are. One can think of (2.6) as a smooth analog of the function $\mu_1 + I(\mu_2 \leq x < \mu_3)\mu_4$ $(0 < \mu_2 < \mu_3 < 1)$ that defines a conventional dummy variable when the values of the parameters μ_2 and μ_3 are known.

In our general setting we partition $\mu = [\nu' \ \gamma']'$ and, for theoretical reasons, make the additional structural assumption

$$g_t(\mu) = d(\nu)f_t(\gamma), \tag{2.7}$$

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 $(q \times 1)$. As already indicated, the sequence $f_t(\gamma)$is supposed to be of the form $f_t(\gamma) = f(t/T;\gamma)$ where $f(\cdot;\gamma)$ is a suitable function defined on the interval [0,1]. If the sequence $f_t(\gamma)$ is independent of the parameter vector γ and hence the sequence $g_t(\mu)$ is nonlinear only in parameters, the estimation theory simplifies. This is the case, for example, when the sequence

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

$f_t(\gamma)$ is defined by conventional dummy variables. However, in the general case where the sequence $f_t(\gamma)$ is unknown and nonlinear in the time variable t the situation becomes more complicated. Contrary to the other parameters of the model we have to assume that the parameter space of γ is compact.

Because our trend model is fairly general the number of required assumptions is somewhat large. Therefore, these assumptions are divided into several parts. First we give the following assumption about the matrix function $d(\nu)$. The parameter space of μ is denoted by $M = N \times \Gamma$ where $N \subset \mathbb{R}^n$ and $\Gamma \subset \mathbb{R}^q$ are the parameter spaces of ν and γ, respectively.

Assumption 4.

(a) $d(\nu)$ is a continuous function from a set $N \subset \mathbb{R}^n$ to $\mathbb{R}^{r \times r_1}$.
(b) $d(\nu_0)$ is an interior point of $d(N)$ in the relative topology of $\overline{d(N)}$, and in some neighborhood of ν_0 the function $d(\nu)$ is bijective and has a continuous inverse in some neighborhood of $d(\nu_0)$.

The two conditions in Assumption 4 are similar to the two conditions in Assumptions 2 and 3, and their motivation is also similar (see the discussion following Assumption 2).

Before presenting assumptions about the function $f(x;\gamma)$ used to define the sequence $f_t(\gamma)$ we shall give the following general condition, which can also be used for other similar functions, such as partial derivatives of $f(\cdot;\gamma)$. Recall that functions belonging to $D[0,1]$ have at most countably many discontinuities.

Condition 1. Let $G(x;\kappa)$ be a vector function from $[0,1] \times K$ to $\mathbb{R}^a$ where K is a compact subset of an Euclidean space. We say that the function $G(x;\kappa)$ satisfies Condition 1 if the following three requirements hold.

(a) The collection of functions $\{G(x;\cdot);\ x \in [0,1]\}$ is equicontinuous.
(b) $G(\cdot;\kappa) \in D[0,1]^a$ for all $\kappa \in K$, and for every $\varepsilon > 0$ and every $x_0 \in [0,1]$ at which the function $G(\cdot;\kappa)$ is continuous there exists a constant $\delta > 0$ such that $\sup_{\kappa \in K} \|G(x;\kappa) - G(x_0;\kappa)\| < \varepsilon$ as $\|x_0 - x\| < \delta$.
(c) $\sup_T \sup_{\kappa \in K} \sum_{t=2}^{T} \|G(t/T;\kappa) - G((t-1)/T;\kappa)\| < \infty$.

We shall discuss Condition 1 after the following assumption, which contains conditions required from the function $f(x;\gamma)$ used to define the sequence $f_t(\gamma)$. Note that, in the same way as with other parameters, we use the notation $d = d(\nu)$ and $d_0 = d(\nu_0)$.

Assumption 5. Let $f(x;\gamma)$ be a function from $[0,1] \times \Gamma$ to $\mathbb{R}^{r_1}$. Assume that the sequence $f_t(\gamma)$ is of the form $f_t(\gamma) = f(t/T;\gamma)$, $t = 1,\ldots,T$ and that the following conditions hold.

(a) Γ is a compact subset of $\mathbb{R}^q$.
(b) $f(x;\gamma)$ satisfies Condition 1 with $K = \Gamma$.
(c) $\int_0^1 f(x;\gamma) f(x;\gamma)'\, dx > 0$ for all $\gamma \in \Gamma$.
(d) There exists an $x \in [0,1]$ such that $df(x;\gamma) \neq d_0 f(x;\gamma_0)$ for $(d,\gamma) \neq (d_0,\gamma_0)$.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

Because the function $f(x;\gamma)$ may be nonlinear in both of its arguments the compactness of the parameter space Γ is a standard assumption, and it seems to be more difficult to avoid it than in the case of the parameter spaces Φ, Ψ, and N where the associated nonlinearity only appears in parameters.

Assumption 5(b) says that the function $f(x;\gamma)$ satisfies the three requirements of Condition 1 in the whole parameter space Γ. The assumed compactness of the parameter space Γ implies that the equicontinuity required in the first condition can actually be strengthened to uniform equicontinuity (see Dudley, 1989, Theorem 2.4.5). This means that for every $\varepsilon > 0$ there is $\delta > 0$ such that $\|\gamma_1 - \gamma_2\| < \delta$ implies $\|f(x;\gamma_1) - f(x;\gamma_2)\| < \varepsilon$ for all $\gamma_1, \gamma_2 \in \Gamma$ and all $x \in [0,1]$. (Here δ is independent of both x and γ_1, γ_2.) Continuity is not required in the former argument of $f(x;\gamma)$ so that, for example, conventional step dummies and trends with multiple breaks at known points of time are included. Trends with breaks at unknown points of time are excluded, however.

The second requirement of Condition 1 implies that at the continuity points of the function $f(\cdot;\gamma)$ continuity obtains uniformly over γ. Of course, the first two conditions hold if the function $f(x;\gamma)$ is continuous on $[0,1] \times \Gamma$. Notice that from these two requirements and the compactness of Γ one can readily conclude that the function $f(x;\gamma)$ is bounded on $[0,1] \times \Gamma$. This implies that the sequence $f_t(\gamma)$, $t = 1,\ldots,T$ is bounded uniformly in both $\gamma \in \Gamma$ and T. Thus, in some sense we assume that the considered trends are "small" (cf. Andrews and McDermott, 1995, p. 347).

According to the third requirement of Condition 1 differences of the trend variable $d_t = \alpha' f_t(\gamma)$ are asymptotically negligible. This assumption essentially means that the functions $f(\cdot;\gamma)$, $\gamma \in \Gamma$ are of bounded variation and their total variations have an upper bound independent of γ. A sufficient condition for this is that the function $f(x;\gamma)$ is continuously differentiable with respect to x and that $\partial f(x;\gamma)/\partial x$ is bounded on $[0,1] \times \Gamma$. The examples in (2.5) and (2.6) clearly satisfy this property.

Assumptions 5(c) and (d) are identification conditions for the parameters α and γ. Equation (2.7) implies that, from the viewpoint of estimating α, Assumption 5(c) is analogous to the usual full rank condition in least squares estimation. An analog of Assumption 5(d) appears in Assumption R of Andrews and McDermott (1995). Note that in the special case where $f_t(\gamma)$ does not depend on unknown parameters, so that $f_t(\gamma) = f_t(\gamma_0)$ can be assumed, Assumption 5(d) follows from Assumption 5(c). When Assumptions 5(c) and (d) are combined with Assumption 4(b), they yield an identification condition for the parameter vector μ.

As a whole, Assumption 5 appears reasonably general to cover a number of interesting cases, including (2.5) and (2.6) and also their extensions discussed in Ripatti and Saikkonen (2001).

The preceding assumptions are sufficient to develop an asymptotic theory of estimation and statistical inference when all the structural parameters of the model are identified. However, in Saikkonen (2001) we are also interested in asymptotic inference when some of the structural parameters, treated as nui-

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

sance parameters, are not identified. For that purpose the consistency results obtained with the foregoing assumptions are not sufficient because appropriate orders of consistency are not obtained. To be able to prove stronger consistency results additional assumptions are needed. Define the parameter vector $\underline{\mu} = [\text{vec}(d)' \ \gamma']'$ and let $\underline{M}$ denote the space of its permissible values (see Assumptions 4 and 5). Instead of the function $g(x;\mu)$ we now consider $g(x;\underline{\mu}) = (f(x;\gamma)' \otimes I_r)\text{vec}(d) = df(x,\gamma)$, that is, $g(x;\mu)$ treated as a function of the structural parameter γ and the reduced form parameter d.

Assumption 6.

(a) The true parameter value $\underline{\mu}_0$ is such that γ_0 is an interior point of Γ and in some neighborhood of γ_0 the function $f(x;\gamma)$ is continuously differentiable with respect to γ.
(b) The function formed from $\partial f(x;\gamma)/\partial\gamma_i$ $(i = 1,\ldots,q)$ satisfies Condition 1 with K some compact neighborhood of γ_0.
(c) The matrix $\int_0^1[\partial g(x;\underline{\mu}_0)'/\partial\mu][\partial g(x;\underline{\mu}_0)/\partial\mu']dx$ is of full rank.

As already indicated, Assumption 6 will be used to obtain consistency results needed to derive limiting distributions of estimators and test statistics. Therefore, it may be no wonder that Assumption 6 is similar to what one sees in treatments of asymptotic inference. In particular, in that context the differentiability assumption imposed in the first condition, is standard although we can do without assuming the existence of second partial derivatives. Because Assumption 6 will only be used when Assumption 5(b) is also assumed the second condition in Assumption 6 also holds for the function formed from $f(x;\gamma)$ and $\partial f(x;\gamma)/\partial\gamma_i$ $(i = 1,\ldots,q)$. This implies that sample moments that appear in the derivation of limiting distributions converge in an appropriate way. Because this assumption is only used when a consistent estimator of γ is available it suffices to assume that Condition 1 holds in some (small) neighborhood of the true parameter value. The third condition in Assumption 6 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 $\partial g/(x;\underline{\mu})'/\partial\underline{\mu} = (\partial g(x;\underline{\mu})/\partial\underline{\mu}')'$ (cf. Lütkepohl, 1996, p. 173).

2.3. A Representation Theorem

To prove the results of this paper and also those in Saikkonen (2001), we need a version of Granger's representation theorem or a solution of (2.4) for y_t. Because the deterministic sequence $g_t(\mu)$ and hence the series y_t are assumed to depend on the sample size T the previous representation theorem in Johansen (1995, p. 49) is not directly applicable. Fortunately, however, a similar result can readily be obtained and will be given in the following theorem. Here it is reasonable to include the sample size in the notation and write $g_{t,T}(\mu) = d(\nu)f_{t,T}(\mu)$ and $y_{t,T}(t = 1,\ldots,T)$.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

THEOREM 2.1. *Suppose that* $y_{1,T},\ldots,y_{T,T}$ *are generated by* (2.4) *and that Assumption* 1 *holds. Then there exists a choice of initial values* $y_{j,T} = y_j$, $j = -p+1,\ldots,0$, *whose distribution is independent of* T, *and such that, for* $t = 1,\ldots,T$,

$$y_{t,T} = \Lambda \sum_{j=1}^{t} \varepsilon_j + \sum_{j=0}^{\infty} K_j \varepsilon_{t-j} - \sum_{j=0}^{t} K_j \alpha g_{t-j,T}(\mu) + P_{\beta_\perp} y_0, \tag{2.8}$$

where $\Lambda = \beta_\perp[\alpha_\perp'(I_s - \sum_{j=1}^{p-1}\Gamma_j)\beta_\perp]^{-1}\alpha_\perp'$, $P_{\beta_\perp} = \beta_\perp(\beta_\perp'\beta_\perp)^{-1}\beta_\perp'$, *and the coefficient matrices* K_j *are defined by* $K(z) = \sum_{j=0}^{\infty} K_j z^j$, *a rational matrix function such that its power series representation converges absolutely in an open disk containing the unit disk.*

For later purposes we introduce the process

$$x_t = \Lambda \sum_{j=1}^{t} \varepsilon_j + K(L)\varepsilon_t. \tag{2.9}$$

As representation (2.8) shows, x_t is the stochastic part of $y_{t,T}$. It is independent of the sample size T and in fact identical to its counterpart in the corresponding representation in Johansen (1995, p. 49). Because the initial value y_0 is also independent of T the dependence of $y_{t,T}$ on the sample size is only due to the deterministic sequence $g_{t,T}(\mu)$ or $f_{t,T}(\gamma)$. This fact simplifies asymptotic derivations and makes them in many respects similar to those in previous standard cases. For this reason our usual convention to exclude the symbol T from the notation should not cause any confusion.

Because $g_{t,T}(\mu) = g(t/T;\mu)$ and the function $g(\cdot;\mu)$ is defined on the interval $[0,1]$ we have not defined the sequence $g_{t,T}(\mu)$ for $t < 0$. For this reason the third sum on the right hand side of (2.8) is truncated and therefore different from its counterpart in Johansen (1995, Theorems 2.2 and 4.2). However, from the proof of Theorem 2.1 it can be seen that, with suitable definitions of $g_{t,T}(\mu)$, $t < 0$, we could replace this sum by $\sum_{j=0}^{\infty} K_j \alpha g_{t-j,T}(\mu) = K(L)\alpha g_{t,T}(\mu)$. Assuming that the series is convergent is obviously required, and, if the distribution of the initial values is to be independent of T, it is also necessary that $g_{t,T}(\mu)$, $t < 0$ are independent of T. The present form of Theorem 2.1 can be formulated in this way by assuming $g_{t,T}(\mu) = 0$ for $t < 0$. If convenient this assumption will also be explicitly used.

Because we will assume that differences of the sequence $g_{t,T}(\mu)$ are asymptotically negligible in the sense implied by Condition 1(c) the series $\Delta y_{t,T}$ and $\beta' y_{t-1,T} - g_{t,T}(\mu)$ are nearly stationary or asymptotically stationary. For $\Delta y_{t,T}$ this follows directly from (2.8), which implies that $\Delta y_{t,T} = \Delta x_t - K(L)\alpha\Delta g_{t,T}(\mu) \approx \Delta x_t$. For $\beta' y_{t-1,T} - g_{t,T}(\mu)$ we can use the decomposition $K(L) = K(1) - K(L)^*\Delta$ and observe that $\beta' K(1)\alpha = -I_r$ (cf. Johansen, 1995, the proof of Theorem 4.2). From this and (2.8) one then obtains $\beta' y_{t-1,T} - g_{t,T}(\mu) = \beta' x_{t-1} - \Delta g_{t,T}(\mu) + \beta' K(L)^*\alpha\Delta g_{t-1}(\mu) \approx \beta' x_{t-1}$.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

3. MAIN RESULTS

3.1. Existence of ML Estimators

As already mentioned, some complications in our estimation theory arise because we do not assume compactness of the parameter space. This particularly holds for proving the existence of ML estimators, and for that purpose it appears convenient to write the model in (2.4) as

$$\Delta y_t = \alpha u_{t-1}(\gamma) + \sum_{j=1}^{p-1} B_j \Delta y_{t-j} + C y_{2,t-1} + D f_t(\gamma) + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where $C = C(\alpha, A) = -\alpha(A - A_0)$, $D = D(\alpha, d) = -\alpha(d - d_0)$, and $u_{t-1}(\gamma) = y_{1,t-1} - A_0 y_{2,t-1} - d_0 f_t(\gamma)$ with $y_t = [y_{1t}' \; y_{2t}']'$ partitioned in an obvious way. Here we have expressed the model in terms of the reduced form parameters discussed following Assumption 3. Because this form of the model will be employed later we write it more compactly as

$$\Delta y_t = B z_t(\gamma) + C y_{2,t-1} + D f_t(\gamma) + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{3.1}$$

where $z_t(\gamma) = [\Delta y_{t-1}' \ldots \Delta y_{t-p+1}' u_{t-1}(\gamma)']'$ and B is as in Assumption 3.

Our estimation theory for the structural parameters ϕ, ψ, and ν is based on a coordinate free approach in which existence and consistency of ML estimators of the reduced form parameters B, C, and D are first established. The rationale of this is that the likelihood function depends on these structural parameters only through the reduced form parameters B, C, and D. In cointegrated systems the same idea was recently used by Johansen (1997), and it has also been common in the estimation theory of stationary autoregressive moving average (ARMA) models where the transfer function of the infinite order moving average representation plays the role of reduced form parameters (for a brief discussion see Potscher, 1991, Appendix A, for a comprehensive account see Hannan and Deistler, 1988). Thus, we consider the ML estimation of the parameters B, C, D, and γ in (3.1) and, of course, also the estimation of the error covariance matrix Ω. At this stage the relation between this estimation problem and that concerning the structural parameters is that Assumptions 2–4 are used to define the parameter spaces of B, C, and D, which are $\mathcal{B} = B(\Psi)$, $\mathcal{C} = C(\alpha(\Psi), A(\Phi))$, and $\mathcal{D} = D(\alpha(\Psi), d(N))$, respectively.

Denote $\theta = (C, D, \gamma, B)$ and let $\Theta = \mathcal{C} \times \mathcal{D} \times \Gamma \times \mathcal{B}$ contain the permissible values of θ. Then, conditioning on the initial values $y_{-p+1}, \ldots, y_0$ we can write minus two times the logarithm of the likelihood function as

$$\ell_T(\theta, \Omega) = T \log \det(\Omega) + \operatorname{tr}\left\{\Omega^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta) \varepsilon_t(\theta)'\right\}, \tag{3.2}$$

where

$$\varepsilon_t(\theta) = \Delta y_t - B z_t(\gamma) - C y_{2,t-1} - D f_t(\gamma).$$

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

Note that even though model (3.1) is infeasible $\varepsilon_t(\theta)$ is observable for given values of A, B, d, and γ because its dependence on the unknown true parameter values A_0 and d_0 cancels. ML estimators of θ and Ω, if they exist, are obtained by minimizing the function $\ell_T(\theta,\Omega)$ over $\theta \in \Theta$ and $\Omega > 0$. The reason why we do not consider the likelihood function $-2\ell_T(\theta,\Omega)$ is that we make use of ideas similar to those in Saikkonen (1995, Sect. 4.3), where estimators are obtained as a solution of a minimization problem. Nevertheless, we shall refer to $\ell_T(\theta,\Omega)$ as the likelihood function.

We have not been able to establish the existence of a ML estimator of θ without slightly extending the parameter space Θ. Therefore we consider the likelihood function over $\theta \in \bar{\Theta}$. Of course because Γ is compact by assumption this only means that we have to extend the parameter spaces $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ by including their boundaries and, if these parameter spaces are closed, as assumed in Johansen (1997), no extension is needed. Note that $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ would be closed if the parameter spaces Θ, Ψ, and N were assumed compact. The following result shows that, when the sample size is large enough, ML estimators of θ and Ω exist over the extended parameter space $\{(\theta,\Omega):\theta \in \bar{\Theta}$ and $\Omega > 0\}$. These estimators are denoted by $\hat{\theta}_T$ and $\hat{\Omega}_T$, and they are of course related by

$$\hat{\Omega}_T = T^{-1}\sum_{t=1}^{T}\varepsilon_t(\hat{\theta}_T)\varepsilon_t(\hat{\theta}_T)'. \tag{3.3}$$

A similar notation will also be used for the components of $\hat{\theta}_T$ and also ML estimators of A, d and the structural parameters ϕ, ψ, and γ. Now we can prove the following proposition. Unless otherwise stated, all limits in this paper are taken as $T \to \infty$.

PROPOSITION 3.1. *Let* $y_1,\ldots,y_T$ *be generated by* (2.4), *where* ε_t *is as in* (2.1). *Suppose that Assumptions* 1, 2(a), 3(a), 4(a), *and* 5(a)–(c) *hold*. *Then*, ML *estimators* $\hat{\theta}_T$ *and* $\hat{\Omega}_T$ *exist with probability approaching one and*, *moreover*, $\hat{C}_T = O_p(T^{-1/2})$.

Thus, in addition to the existence of the ML estimators $\hat{\theta}_T$ and $\hat{\Omega}_T$, Proposition 3.1 also shows the consistency and order of consistency of the long-run parameter estimator $\hat{C}_T$. This latter result is actually obtained directly when the existence of the estimators $\hat{\theta}_T$ and $\hat{\Omega}_T$ is proved (see the Appendix). It will be used later to prove the consistency of the other components of $\hat{\theta}_T$ and the consistency of $\hat{\Omega}_T$. It will also provide a first step for obtaining better orders of consistency, which will be needed to derive limiting distributions of structural parameter estimators (cf. the discussion in Saikkonen, 1995).

3.2. Consistency of ML Estimators

The prove the consistency results discussed previously, it seems convenient to follow Saikkonen (1995) and express the vector $\varepsilon_t(\theta)$ and thereby the likeli-

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

hood function as a sum of two components. Because the idea is basically that one of the components of $\varepsilon_t(\theta)$ should contain integrated variables the two components might be defined as $\Delta y_t - Bz_t(\gamma) - Df_t(\gamma)$ and $-Cy_{2,t-1}$. However, in addition to C, we have found it more convenient to also interpret D and γ as long-run parameters, and therefore we partition $\theta = (\theta_1, \theta_2)$ where $\theta_1 = (C, D, \gamma)$ and $\theta_2 = B$. The two components of $\varepsilon_t(\theta)$ are thus defined as

$$\varepsilon_{1t}(\theta) = -Cy_{2,t-1} + \alpha d_0[\, f_t(\gamma) - f_t(\gamma_0)] - Df_t(\gamma) \tag{3.4}$$

and

$$\varepsilon_{2t}(\theta_2) = \Delta y_t - Bz_t, \tag{3.5}$$

where $z_t = z_t(\gamma_0)$. Using the definitions one can check that $Bz_t = Bz_t(\gamma) + \alpha d_0[\, f_t(\gamma) - f_t(\gamma_0)]$ and furthermore that $\varepsilon_t(\theta) = \varepsilon_{1t}(\theta) + \varepsilon_{2t}(\theta_2)$, as desired. Because the processes $u_{t-1}(\gamma_0) = y_{1,t-1} - A_0 y_{2,t-1} - d_0 f_t(\gamma_0)$ and Δy_t are asymptotically stationary the same is true for z_t (see the discussion at the end of Section 2.3 and the definition of $z_t(\gamma)$ following (3.1)). Note also that $\varepsilon_{1t}(\theta) = -Cy_{2,t-1} + \alpha[df_t(\gamma) - d_0 f_t(\gamma_0)]$ and $df_t(\gamma) - d_0 f_t(\gamma_0) = g_t(\mu) - g_t(\mu_0)$ when the structural parameter μ is used (cf. (2.7)). However, at this stage we will not use these expressions because we have not established the existence of ML estimators of $d$and ν.

Using the identity $\varepsilon_t(\theta) = \varepsilon_{1t}(\theta) + \varepsilon_{2t}(\theta_2)$ we can write

$$\ell_T(\theta, \Omega) = \ell_{1T}(\theta, \Omega) + \ell_{2T}(\theta_2, \Omega), \tag{3.6}$$

where

$$\ell_{1T}(\theta, \Omega) = 2 \operatorname{tr}\left(\Omega^{-1} \sum_{t=1}^{T} \varepsilon_{1t}(\theta)\varepsilon_{2t}(\theta_2)'\right) + \operatorname{tr}\left(\Omega^{-1} \sum_{t=1}^{T} \varepsilon_{1t}(\theta)\varepsilon_{1t}(\theta)'\right)$$

and

$$\ell_{2T}(\theta_2, \Omega) = T \log\det(\Omega) + \operatorname{tr}\left(\Omega^{-1} \sum_{t=1}^{T} \varepsilon_{2t}(\theta_2)\varepsilon_{2t}(\theta_2)'\right).$$

Clearly, $\varepsilon_{1t}(\theta_{10}, \theta_2) = 0$, and consequently $\ell_{1T}(\theta_{10}, \theta_2, \Omega) = 0$. Because $\ell_{2T}(\theta_2, \Omega)$ is the likelihood function of a linear regression model with asymptotically stationary regressors and Gaussian white noise errors this component of the likelihood function $\ell_T(\theta)$ is easy to deal with because standard methods apply. This is a major reason why we have found the preceding decomposition of $\varepsilon_t(\theta)$ convenient. The other component of the likelihood function, namely, $\ell_{1T}(\theta, \Omega)$, contains the integrated processes of the system, which make the estimation problem nonstandard.

The only identification condition assumed in Proposition 3.1 was Assumption 5(c). It is of course clear that the structural parameters ϕ, ψ, ν, and γ cannot be consistently estimated without further identification conditions. How-

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

ever, some consistency results can still be obtained, as the following theorem shows.

THEOREM 3.1. *Suppose that the assumptions of Proposition* 3.1 *hold. Then, with probability approaching one,* ML *estimators* $\hat{A}_T$ *and* $\hat{d}_T$ *also exist and, furthermore,*

(a) $\hat{C}_T = o_p(T^{-1/2})$ *and* $\hat{A}_T = \hat{A}_0 + o_p(T^{-1/2})$,
(b) $\hat{B}_T \xrightarrow{p} B_0$ *and* $\hat{\Omega}_T \xrightarrow{p} \Omega_0$
(c) $T^{-1}\sum_{t=1}^{T} \|\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)\|^2 \xrightarrow{p} 0$.

Theorem 3.1 shows that when the assumptions of Proposition 3.1 are not supplemented by other identification conditions one can only consistently estimate the reduced form parameters A and B in addition to the error covariance matrix Ω. The order of consistency obtained for the estimator $\hat{C}_T$ in Proposition 3.1 can also be slightly improved, and the same order of consistency can be obtained for the ML estimator of A. No consistency result can be proved for the reduced form parameter estimator $\hat{D}_T$. The identification condition given in Assumption 5(c) is only sufficient to prove the existence of this estimator and also the existence of the estimator $\hat{d}_T$. This is not surprising because the parameter d appears in the model in the product form $df_t(\gamma)$ so that, without further assumptions, it is not possible to separate the two factors of the product. Therefore it is also expected that the parameter γ cannot be consistently estimated without further assumptions. However, the last result of Theorem 3.1 can be seen as a reduced form consistency result for the trend model or the sequence $df_t(\gamma) = g_t(\mu)$, $t = 1,\ldots,T$.

When appropriate identification conditions are imposed further consistency results can be readily deduced from Theorem 3.1. Specifically we have the following corollary.

COROLLARY 3.1. *When the assumptions of Theorem* 3.1 *hold the following additional assumptions imply the existence* (*with probability approaching one*) *and consistency of the indicated* ML *estimators*.

(a) *If Assumption* 2(b) *holds, then* $\hat{\phi}_T \xrightarrow{p} \phi_0$.
(b) *If Assumption* 3(b) *holds, then* $\hat{\psi}_T \xrightarrow{p} \psi_0$.
(c) *If Assumption* 5(d) *holds, then* $\hat{d}_T \xrightarrow{p} d_0$ *and* $\hat{\gamma}_T \xrightarrow{p} \gamma_0$, *and, if Assumption* 5(d) *is supplemented by Assumption* 4(b), *then* $\hat{\nu}_T \xrightarrow{p} \nu_0$ *also holds.*

Each of the given conditions implies the stated consistency result regardless of the validity of the other conditions. For instance, Assumptions 2(b), 3(b), and 5(d) in conjunction with the assumptions of Theorem 3.1 imply that the structural parameters can be consistently estimated except for ν. Here no order of consistency is obtained for the long-run parameter estimator $\hat{\phi}_T$. It is easy to see that $\hat{\phi}_T = \phi_0 + o_p(T^{-1/2})$ if ϕ_0 is an interior point of Φ 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. This is a standard assumption in asymptotic inference on ϕ. When it holds the previously mentioned result can be jus-

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

tified by using a mean value expansion to show that, for some $\varepsilon > 0$, $\|A(\hat{\phi}_T) - A(\phi_0)\|^2 \geq \varepsilon \|\hat{\phi}_T - \phi_0\|^2$ with probability approaching one.

We have already mentioned that Assumption 5(c) was used to guarantee the existence of the estimators $\hat{D}_T$ and $\hat{d}_T$. It can be seen from the Appendix that this assumption is not needed if one considers the special case where the function $g_t(\mu)$ reduces to $f_t(\gamma)$. Then one can assume that the value of d is a priori known so that $D = 0$. On the other hand, we noticed in Section 2 that in the special case where the function $f_t(\gamma)$ does not depend on unknown parameters Assumption 5(d) is redundant because it follows from Assumption 5(c).

The consistency results of Theorem 3.1 and Corollary 3.1 are sufficient to derive limiting distributions of estimators and test statistics when identifiability of all structural parameters is assumed. However, when this is not done the situation gets more complicated, and it seems to be difficult to do without further results about order of consistency. Therefore we prove the following proposition.

PROPOSITION 3.2. *Suppose that, in addition to the assumptions of Proposition* 3.1, *Assumptions* 5(d) *and* 6 *also hold. Then*,

(a) $\hat{A}_T = A_0 + O_p(T^{-1})$,
(b) $\hat{d}_T = d_0 + O_p(T^{-1/2})$,
(c) $\hat{\gamma}_T = \gamma_0 + O_p(T^{-1/2})$,
(d) $\hat{B}_T = B_0 + O_p(T^{-1/2})$.

Proposition 3.2 improves the order of consistency obtained for the reduced form parameter estimator $\hat{A}_T$ in Theorem 3.1 and also provides orders of consistency for the estimators $\hat{d}_T$, $\hat{\gamma}_T$, and $\hat{B}_T$. For these results to hold it is not necessary that the structural parameters ϕ, ν, and ψ be identified, but appropriate identification conditions are of course needed for the parameters d and γ. Assumption 5(d) was already used in Corollary 3.1 to prove the consistency of the estimators $\hat{d}_T$ and $\hat{\gamma}_T$, and the additional requirements are similar to those needed to guarantee that these estimators have nonsingular limiting distributions. The first three results of Proposition 3.2 are used in Saikkonen (2001) when ψ is the parameter of interest and $A = A(\phi)$, $d = d(\nu)$, and γ are nuisance parameters with ϕ and ν not necessarily identified. The last result is used when the roles of these parameters are reversed so that ϕ, ν, and γ are parameters of interest and $B = B(\psi)$ is a nuisance parameter with ψ not necessarily identified.

4. CONCLUDING REMARKS

This paper provides a fairly complete treatment of consistent estimation in cointegrated VAR models with nonlinear time trends in cointegrating relations. However, there are still several potential extensions that are not covered by our analysis. For instance, one might wish to relax the assumption that the time trends can be totally absorbed in the cointegrating relations. An example of this would be a model with an unrestricted intercept term and nonlinear time trends in cointegrating relations. In this model the integrated components of the series

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

would have linear time trends. Note that linear time trends may appear in the cointegrating relations of our model. Although the time index is assumed to be standardized according to t/T it is not difficult to modify our treatment and allow for a nonstandardized time index in linear trend terms or even in higher order polynomials of time. One might also wish to follow Park and Hahn (1999) and allow for deterministic structural changes in cointegrating vectors that in our framework would be modeled by a parametric function similar to that used for the trend term. Finally, it might be of interest to consider models containing nonlinear functions of stochastic variables, which in some special cases have been studied by Wooldridge (1994) and Park and Phillips (1998).

We have not tried to cover the preceding extensions or even some of them in the present work because the considered model is already rather complex. However, the methods used to prove the results of this paper and Saikkonen (2001) may also be helpful if these extensions are considered. In some cases this may only require straightforward modifications to our treatment. In others a more elaborate analysis or additional assumptions may also be needed.

REFERENCES

Ahn, S.K. (1997) Inference in vector autoregressive models with cointegration and scalar components. *Journal of the American Statistical Association* 92, 350–356.

Andrews, D.W.K. & C.J. McDermott (1995) Nonlinear econometric models with deterministically trending variables. *Review of Economic Studies* 62, 343–360.

Boswijk, H.P. (1995a) Efficient inference on cointegration parameters in structural error correction models. *Journal of Econometrics* 69, 133–158.

Boswijk, H.P. (1995b) Identifiability of Cointegrated Systems. Working paper, University of Amsterdam.

Boswijk, H.P. (1996) Testing identifiability of cointegrating vectors. *Journal of Business and Economic Statistics* 14, 153–160.

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.

Deistler, M. & B.M. Pötscher (1984) The behaviour of the likelihood function for ARMA models. *Advances of Applied Probability* 16, 843–865.

Dudley, R.M. (1989) *Real Analysis and Probability.* Pacific Grove, California: Wadsworth and Brooks/Cole.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.

Granger, C.W.J. & N. Swanson (1997) Future developments to the study of cointegrated variables. *Oxford Bulletin of Economics and Statistics* 59, 537–553.

Hannan, E.J. & M. Deistler (1988) *The Statistical Theory of Linear Systems.* New York: Wiley.

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

Hsiao, C. (1997) Cointegration and dynamic simultaneous equations models. *Econometrica* 65, 647–670.

Hu, S.T. (1964) *Elements of General Topology.* San Francisco: Holden Day.

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

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

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

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

Johansen, S. (1997) A likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics* 24, 433–462.

Johansen, S. & K. Juselius (1994) Identification of the long-run and short-run structure: An application to the ISLM model. *Journal of Econometrics* 63, 7–36.

Leybourne, S., P. Newbold, & D. Vougas (1998) Unit roots and smooth transitions. *Journal of Time Series Analysis* 19, 83–97.

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

Luukkonen, R., A. Ripatti, & P. Saikkonen (1999) Testing for a valid normalization of cointegrating vectors in vector autoregressive processes. *Journal of Business and Economic Statistics* 17, 195–204.

Park, J. & S. Hahn (1999) Cointegrating regressions with time varying coefficients. *Econometric Theory* 15, 664–703.

Park, J. & P.C.B. Phillips (1998) Nonlinear Regressions with Integrated Time Series. Mimeo, Yale University.

Pesaran, M.H. & Y. Shin (1997) Long-run Structural Modelling. Working paper, University of Cambridge.

Pesaran, M.H., Y. Shin, & R. Smith (1997) Structural Analysis of Vector Error Correction Models with Exogenous $I(1)$ Variables. Working paper, University of Cambridge.

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.

Phillips, P.C.B. & B. Hansen (1990) Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.

Pollard, D. (1990) *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2. Hayward, California: Institute of Mathematical Statistics.

Pötscher, B.M. (1991) Noninvertibility and pseudo-maximum likelihood estimation of misspecified ARMA models. *Econometric Theory* 7, 435–449.

Quintos, C.E. (1997) Stability tests in error correction models. *Journal of Econometrics* 82, 289–315.

Reinsel, G.C. & S.K. Ahn (1992) Vector autoregressive models with unit roots and reduced rank structure: Estimation, likelihood ratio test, and forecasting. *Journal of Time Series Analysis* 13, 353–375.

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

Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica* 39, 577–591.

Saikkonen, P. (1992) Estimation and testing of cointegrated systems by an autoregressive approximation. *Econometric Theory* 8, 1–27.

Saikkonen, P. (1993a) 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. (1993b) Estimation of cointegration vectors with linear restrictions. *Econometric Theory* 9, 19–35.

Saikkonen, P. (1995) Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems. *Econometric Theory* 11, 888–911.

Saikkonen, P. (2001) Statistical inference in cointegrated vector autoregressive models with nonlinear time trends in cointegrating relations. *Econometric Theory* 17, 327–356.

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

Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.

Wooldridge, J.M. (1994) Estimation and inference for dependent processes. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2641–2738. Amsterdam: North Holland.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

APPENDIX

A.1. Proof of Theorem 2.1. Define $\tilde{y}_{t,T} = [(\beta' y_{t,T})' \, (\beta_\perp' \Delta y_{t,T})']'(t = 1,\ldots,T)$. Then, proceeding exactly in the same way as in the proof of Theorem 4.2 Johansen (1995), we find that

$$\tilde{A}(L)\tilde{y}_{t,T} = \begin{bmatrix} -g_{t,T}(\mu) \\ 0 \end{bmatrix} + \begin{bmatrix} (\alpha'\alpha)^{-1}\alpha' \\ (\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp' \end{bmatrix}\varepsilon_t, \qquad t = 1,\ldots,T,$$

where $\tilde{A}(L) = \sum_{j=0}^{p} \tilde{A}_j L^j$ has no unit roots and all other roots are outside the unit circle by Assumption 1. Note that $\tilde{A}(L)$ is precisely the same lag polynomial matrix as in Johansen (1995, p. 51). Using Theorem 2.1 in Johansen (1995) we can now solve $\tilde{y}_{t,T}$ and obtain

$$\tilde{y}_{t,T} = \sum_{j=1}^{p} \tilde{F}_{t-j}(\tilde{A}_j \tilde{y}_{0,T} + \cdots + \tilde{A}_{j+p-1}\tilde{y}_{1-p,T}) + \sum_{j=0}^{t-1} \tilde{F}_j \begin{bmatrix} (\alpha'\alpha)^{-1}\alpha' \\ (\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp' \end{bmatrix} \varepsilon_{t-j} - \sum_{j=0}^{t-1} \tilde{F}_j \begin{bmatrix} g_{t-j,T}(\mu) \\ 0 \end{bmatrix}, \qquad t = 1,\ldots,T,$$

where the matrices $\tilde{F}_j$ are defined by $\tilde{F}(L) = \sum_{j=0}^{\infty} \tilde{F}_j L^j = \tilde{A}(L)^{-1}$. Next note that it is possible to choose the initial values such that

$$\sum_{j=1}^{p} \tilde{F}_{t-j}(\tilde{A}_j \tilde{y}_{0,T} + \cdots + \tilde{A}_{j+p-1}\tilde{y}_{1-p,T}) = \sum_{j=t}^{\infty} \tilde{F}_j \begin{bmatrix} (\alpha'\alpha)^{-1}\alpha' \\ (\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp' \end{bmatrix} \varepsilon_{t-j} - \tilde{F}_t \begin{bmatrix} g_{0,T}(\mu) \\ 0 \end{bmatrix}$$

(cf. Johansen, 1995, the proof of Theorem 2.2). Because this choice depends on the sequence $g_{t,T}(\mu)$ only through $g_{0,T}(\mu) = g(0;\mu)$ it is independent of T. Thus, if we set $g_{t,T}(\mu) = 0$, $t < 0$, we have

$$\tilde{y}_{t,T} = \tilde{F}(L) \begin{bmatrix} (\alpha'\alpha)^{-1}\alpha' \\ (\alpha_\perp'\alpha_\perp)^{-1}\alpha_\perp' \end{bmatrix} (\varepsilon_t - \alpha g_{t,T}(\mu)), \qquad t = 1,\ldots,T.$$

The final step is to recover $y_{t,T}$ from $\tilde{y}_{t,T}$. This, however, can be done in the same way as in the proof of Theorem 4.2 of Johansen (1995), and the result is easily seen to be as claimed in the theorem. This completes the proof. ■

A.2. Auxiliary Results. This section contains some auxiliary results that will be used in subsequent proofs and also in Saikkonen (2001). Proofs of these results make use of the representation given in Theorem 2.1. However, unlike in (2.8) we will not use the subscript T in the notation. Because the series y_t depends on the sample size T only through the deterministic sequence $f_t(\gamma)$ it is not difficult to check that the dependence on T has no effect on the validity of the employed arguments.

We start by introducing some notation. Let K be a compact set as in Condition 1 and denote by $C(K)^{a\times b}$ the space of all continuous functions from K to $\mathbb{R}^{a\times b}$ where $C(K)^{a\times b}$ is endowed with the uniform metric and K with the Euclidean metric. Thus, the distance between $h_1(\kappa)$ and $h_2(\kappa)$ in $C(K)^{a\times b}$ is given by $\sup_{\kappa\in K}\|h_1(\kappa) - h_2(\kappa)\|$.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

The notation $C(K)^{a\times b} = C(K)$ is used if $a = b = 1$. Next recall from Section 3.1 the notation $z_t = [\Delta y'_{t-1} \ldots \Delta y'_{t-p+1} u_{t-1}(\gamma_0)']'$ where $u_{t-1}(\gamma_0) = \beta'_0 y_{t-1} - g_t(\mu_0)$ and $\beta'_0 = [I_r \ -A(\phi_0)]$. In the following lemma Assumption 2 is not needed so that in the definition of z_t the matrix β_0 need not be of the preceding particular form. For this lemma we also introduce the vectors $\underset{\sim}{z}_t = [\Delta y'_t \ z'_t]'$ and $\underset{\sim}{\zeta}_t = [\Delta x'_t \ldots \Delta x'_{t-p+1} \ x'_{t-1}\beta_0]'$ where x_t is defined in (2.9). Further, let $W_0(x)$ be a Brownian motion with covariance matrix Ω_0 and define $W_2(x) = e'_2 \Lambda_0 W_0(x)$ with $e'_2 = [0 \ I_{s-r}]$ and Λ_0 defined in (2.8). Finally, we use $c, c_1, c_2, \ldots$ to signify finite positive constants not always the same ones.

LEMMA A.1. *Let $y_1, \ldots, y_T$ be generated by* (2.4) *where ε_t is as in* (2.1) *and $g_t(\mu)$ as in* (2.7). *Suppose that Assumption* 1 *holds and that the sequence $f_t(\gamma)$ in* (2.7) *is of the form $f_t(\gamma) = f(t/T;\gamma)$ where the function $f(x;\gamma)$ satisfies Condition* 1(c) *when $\gamma = \gamma_0$. Let $G(x;\kappa)$ be a function that satisfies Condition* 1. *Then, setting $G_t(\kappa) = G(t/T;\kappa)$ and using the notation defined previously*,

(a) $T^{-1}\sum_{t=1}^{T} \underset{\sim}{z}_t \underset{\sim}{z}'_t \xrightarrow{p} E\underset{\sim}{\zeta}_t \underset{\sim}{\zeta}'_t$,
(b) $T^{-1}\sum_{t=1}^{T} y_{2,t-1}[\underset{\sim}{z}'_t \ \varepsilon'_t] = O_p(1)$,
(c) $\sup_{\kappa\in K}\| T^{-1/2}\sum_{t=1}^{T} G_t(\kappa)[\underset{\sim}{z}'_t \ \varepsilon'_t]\| = O_p(1)$,
(d) $T^{-1}\sum_{t=1}^{T} G_t(\kappa_1)G_t(\kappa_1)' \to \int_0^1 G(x;\kappa_1)G(x;\kappa_1)'\,dx$,
(e) $T^{-2}\sum_{t=1}^{T} y_{2,t-1}y'_{2,t-1} \Rightarrow \int_0^1 W_2(x)W_2(x)'\,dx$
(f) $T^{-3/2}\sum_{t=1}^{T} y_{2,t-1}G_t(\kappa)' \Rightarrow \int_0^1 W_2(x)G(x;\kappa)'\,dx$.

Here (d) *holds uniformly over $(\kappa_1, \kappa_1) \in K \times K$, and in* (f) *the left hand side also has a weak limit in the function space $C(K)^{(s-r)\times a}$. Finally, joint weak convergence of* (e) *and* (f) *applies.*

Proof. First recall from the end of Section 2.3 that $\Delta y_t = \Delta x_t - K_0(L)\alpha_0 d_0 \Delta f_t(\gamma_0)$ and $u_{t-1}(\gamma_0) = \beta'_0 y_{t-1} - d_0 f_t(\gamma_0) = \beta'_0 x_{t-1} - d_0 \Delta f_t(\gamma_0) + \beta'_0 K_0(L)^* \alpha_0 d_0 \Delta f_{t-1}(\gamma_0)$ where $\Delta f_t(\gamma_0)$ is asymptotically negligible in the sense of Condition 1(c). It is not difficult to see that filtered versions of $\Delta f_t(\gamma_0)$, like those in the preceding expressions of Δy_t and $\beta'_0 y_{t-1} - d_0 f_t(\gamma_0)$, are asymptotically negligible in the same sense. Using this fact and the stationarity of the processes Δx_t and $\beta'_0 x_{t-1}$ it is further straightforward to show that the first and second sample moments obtained from Δy_t and $\beta'_0 y_{t-1} - d_0 f_t(\gamma_0)$ are asymptotically equivalent to those obtained from Δx_t and $\beta'_0 x_{t-1}$. The first result follows from this and the law of large numbers.

To prove the second result, partition $x_t = [x'_{1t} \ x'_{2t}]'$ conformably with the partition of y_t and, for $0 \le j \le p-1$, consider

$$
\begin{aligned}
T^{-1}&\sum_{t=1}^{T} y_{2,t-1}\Delta y'_{t-j} \\
&= T^{-1}\sum_{t=1}^{T} [x_{2,t-1} - e'_2 K_0(L)\alpha_0 d_0 f_t(\gamma_0)][\Delta x_{t-j} - K_0(L)\alpha_0 d_0 \Delta f_{t-j}(\gamma_0)]' \\
&\quad + o_p(1) \\
&= T^{-1}\sum_{t=1}^{T} x_{2,t-1}\Delta x'_{t-j} - T^{-1}\sum_{t=1}^{T} [e'_2 K_0(L)\alpha_0 d_0 f_t(\gamma_0)]\Delta x'_{t-j} \\
&\quad - T^{-1}\sum_{t=1}^{T} x_{2,t-1}[K_0(L)\alpha_0 d_0 \Delta f_{t-j}(\gamma_0)]' + o_p(1). \qquad \textbf{(A.1)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

Here the first equality results from ignoring the term $e_2' P_{\beta_\perp} y_0$ from the expression of $y_{2,t-1}$ (see (2.8)), whereas the second one follows from the boundedness of the sequence $f_t(\gamma_0)$ and the asymptotic negligibility of the differences $\Delta f_t(\gamma_0)$. The first term in the last expression of (A.1) is of order $O_p(1)$ by well-known properties of stationary and integrated processes. Because Δx_t is a stationary zero mean ARMA process and the sequence $f_t(\gamma_0)$ is bounded one readily finds that the second term in the last expression of (A.1) obeys a weak law of large numbers and is therefore of order $o_p(1)$. Using partial summation and the same argument it can also be seen that the third term in the last expression of (A.1) is of order $o_p(1)$. Finally, similar arguments with Δy_{t-j} replaced by $u_{t-1}(\gamma_0)$ and ε_t can be used to complete the proof of the second result.

To prove the third result, we proceed in the same way as previously and, for $0 \leq j \leq p-1$, first consider

$$T^{-1/2} \sum_{t=1}^{T} G_t(\kappa) \Delta y_{t-j}' = T^{-1/2} G_T(\kappa) y_{T-j}' - T^{-1/2} \sum_{t=1}^{T} \Delta G_t(\kappa) y_{t-j-1}', \tag{A.2}$$

where $G_0(\kappa) = 0$ and the equality follows from partial summation. Arguments similar to those already used show that the first term on the right hand side is of order $O_p(1)$ uniformly in κ, whereas the replacement of y_{t-j-1} by x_{t-j-1} in the second one causes an error of order $o_p(1)$, again uniformly in κ. Because it is well known that $T^{-1/2} x_{[Tu]} \Rightarrow \Lambda_0 W_0(u) (0 \leq u \leq 1)$ we have $\max_{1 \leq t \leq T} T^{-1/2} \| x_t \| = O_p(1)$ and, because $G_t(\kappa)$ satisfies Condition 1(c), it follows that the second term on the right hand side of (A.2) is of order $O_p(1)$ uniformly in κ. Thus, we have shown that the left hand side of (A.2) is of order $O_p(1)$ uniformly in κ, and, because the same arguments apply with Δy_{t-j} replaced by $u_{t-1}(\gamma_0)$ and ε_t, the stated result follows.

The proof of the fourth result is based on arguments similar to those used in the proof of Lemma 1 of Andrews and McDermott (1995). Define $G_{(T)}(x;\kappa) = G(([Tx]+1)/T;\kappa)$ for $x \in [0,1)$ and $G_{(T)}(1;\kappa) = G(1;\kappa)$. Then, $G_{(T)}(x;\kappa) = G(t/T;\kappa)$ for $(t-1)/T \leq x < t/T$ $(t = 1,\ldots,T-1)$ and $G_{(T)}(x;\kappa) = G(1;\kappa)$ for $(T-1)/T \leq x \leq 1$. With these definitions we have

$$\begin{aligned}
&\lim_T \sup_{(\kappa_1,\kappa_2) \in K \times K} \left\| \int_0^1 G_{(T)}(x;\kappa_1) G_{(T)}(x;\kappa_2)' \, dx - \int_0^1 G(x;\kappa_1) G(x;\kappa_2)' \, dx \right\| \\
&\quad \leq \lim_T \int_0^1 \sup_{(\kappa_1,\kappa_2) \in K \times K} \| G_{(T)}(x;\kappa_1) G_{(T)}(x;\kappa_2)' - G(x;\kappa_1) G(x;\kappa_2)' \| \, dx \\
&\quad = \int_0^1 \lim_T \sup_{(\kappa_1,\kappa_2) \in K \times K} \| G_{(T)}(x;\kappa_1) G_{(T)}(x;\kappa_2)' - G(x;\kappa_1) G(x;\kappa_2)' \| \, dx \\
&\quad = 0.
\end{aligned}$$

Here the first equality follows from the dominated convergence theorem with dominating function $2 \sup_{0 \leq x \leq 1} \sup_{\kappa \in K} \| G(x;\kappa) \|^2$. The second equality is based on Condition 1(b), which implies that $G_{(T)}(x;\kappa_1) G_{(T)}(x;\kappa_2)' \to G(x;\kappa_1) G(x;\kappa_2)'$ uniformly over (κ_1,κ_2) and for (Lebesgue) almost all x. Because the former integral in the first expression of the given definition equals the left hand side in the assertion the stated result follows.

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

The fifth result is an immediate consequence of the boundedness of the sequence $f_t(\gamma_0)$ and well-known properties of integrated processes that imply that on the left hand side $y_{2,t-1}$ can be replaced by $x_{2,t-1}$. To prove the last assertion, first note that (2.8) and arguments used in the previous proofs readily show that

$$T^{-3/2}\sum_{t=1}^{T} y_{2,t-1}G_t(\kappa)' = e_2'\Lambda_0 T^{-3/2}\sum_{t=1}^{T}\sum_{j=1}^{t}\varepsilon_j G_t(\kappa)' + o_p(1)$$

uniformly over κ. Thus, we need to prove the relevant weak convergence for the first term on the right hand side. Denote this term by $X_T(\kappa)$ and, analogously to $G_{(T)}(x;\kappa)$, define $S_{(T)}(x) = T^{-1/2}(\varepsilon_1 + \cdots + \varepsilon_{[Tx]+1})$ for $x \in [0,1)$ and $S_{(T)}(1) = T^{-1/2}(\varepsilon_1 + \cdots + \varepsilon_T)$. Then we can write

$$X_T(\kappa) = e_2'\Lambda_0\int_0^1 S_{(T)}(x)G_{(T)}(x;\kappa)'\,dx.$$

In the same way as in the proof of (d) we have $G_{(T)}(x;\kappa) \to G(x;\kappa)$ for every fixed $\kappa \in K$ and at every continuity point of $G(x;\kappa)$. It is not difficult to check that this convergence also holds in the Skorohod topology. Thus, because $S_{(T)}(x) \Rightarrow W_0(x)$ in the Skorohod topology it follows from the continuous mapping theorem that, for every fixed $\kappa \in K$, $X_T(\kappa)$ converges weakly to the quantity obtained from the right hand side of (f). The corresponding weak convergence of finite dimensional distributions, that is, the joint weak convergence of $X_T(\kappa_1),\ldots,X_T(\kappa_b)$, is obtained similarly. Thus, from Theorem 10.2 of Pollard (1990) it follows that we only need to verify the stochastic equicontinuity of $X_T(\kappa)$. In other words, we have to show that for each $\varepsilon > 0$ and $\eta > 0$ there is a $\delta > 0$ such that

$$\limsup_T P\left\{\sup_{\|\kappa_1-\kappa_2\|<\delta}\|X_T(\kappa_1) - X_T(\kappa_2)\| > \eta\right\} < \varepsilon.$$

To show this, observe that

$$\begin{aligned}
&E\sup_{\|\kappa_1-\kappa_2\|<\delta}\|X_T(\kappa_1) - X_T(\kappa_2)\| \\
&\quad\le c_1 T^{-3/2}\sum_{t=1}^{T} E\left\|\sum_{j=1}^{t}\varepsilon_j\right\|\sup_{\|\kappa_1-\kappa_2\|<\delta}\|G_t(\kappa_1) - G_t(\kappa_2)\| \\
&\quad\le c_2 T^{-1}\sum_{t=1}^{T}\sup_{\|\kappa_1-\kappa_2\|<\delta}\|G(t/T;\kappa_1) - G(t/T;\kappa_2)\| \\
&\quad\le c_2\sup_{0\le x\le 1}\sup_{\|\kappa_1-\kappa_2\|<\delta}\|G(x;\kappa_1) - G(x;\kappa_2)\|.
\end{aligned}$$

Because the equicontinuity requirement of Condition 1(a) can be strengthened to uniform equicontinuity (see the discussion following Assumption 5) the last quantity can be made as small as desired by choosing δ small enough. Thus, by Markov's inequality, the stochastic equicontinuity of $X_T(\kappa)$ holds, and the last assertion follows. Because the stated joint weak convergence is straightforward to establish the proof of the lemma is complete. ■

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

Note that the third result of Lemma A.1 immediately implies that $G_t(\kappa)\underline{z}_t'$ and $G_t(\kappa)\varepsilon_t$ obey uniform laws of large numbers. In the last result we have not characterized the stochastic properties of the weak limit as a random element in the function space $C(K)^{(s-r)\times a}$. This characterization would not be difficult to obtain, but, because it is not needed in this paper or in Saikkonen (2001), we skip it. The following general result may be of interest when weak convergence results in the function space $C(K)^{a\times b}$ are applied.

LEMMA A.2. *Let $(Z_T(\kappa),\bar{\kappa}_T)$ and $(Z(\kappa),\kappa_*)$ be random elements in $C(K)^{a\times b}\times K$ and assume that $(Z_T(\kappa),\bar{\kappa}_T)\Rightarrow(Z(\kappa),\kappa_*)$ holds. Then, $Z_T(\bar{\kappa}_T)\Rightarrow Z(\kappa_*)$.*

Proof. Consider the function $(Z,\kappa)\to Z(\kappa)$ from $C(K)^{a\times b}\times K$ to $\mathbb{R}^{a\times b}$. This function is known as an evaluation mapping, and, by the continuous mapping theorem, it suffices to establish its continuity. This, however, is well known (see Hu, 1964, Proposition 5.2, and note that the boundedness of the metric assumed in this proposition is not needed here because the set K is compact). ■

Because of the rather strong requirements of Condition 1, the following lemma is simple to prove and yet very powerful in subsequent derivations and also in Saikkonen (2001).

LEMMA A.3. *Suppose that the function $G(x;\kappa)$ satisfies Condition 1 in a compact neighborhood of the point κ_0. Set $G_t(\kappa)=G(t/T;\kappa)$ and let Z_{tT}, $\bar{\kappa}_{tT}$, and $\bar{\kappa}_T$ be random vectors such that $\max_{1\le t\le T}\|Z_{tT}\|=O_p(1)$, $\bar{\kappa}_T\xrightarrow{p}\kappa_0$, and $\|\bar{\kappa}_{tT}-\kappa_0\|\le\|\bar{\kappa}_T-\kappa_0\|$ for all $t=1,\ldots,T$. Then,*

(a) $T^{-1}\sum_{t=1}^{T}[G_t(\bar{\kappa}_{tT})-G_t(\kappa_0)]Z_{tT}'\xrightarrow{p}0$,
(b) $\sum_{t=2}^{T}[G_t(\bar{\kappa}_T)-G_t(\kappa)]\Delta Z_{tT}'\xrightarrow{p}0$.

Proof. In case (a) it suffices to show that $\max_{1\le t\le T}\|G_t(\bar{\kappa}_{tT})-G_t(\kappa_0)\|=o_p(1)$. However, this maximum is dominated by $\sup_{0\le x\le 1}\max_{1\le t\le T}\|G(x;\bar{\kappa}_{tT})-G(x;\kappa_0)\|$, which is of order $o_p(1)$ because the equicontinuity in Condition 1(a) holds uniformly and because $\max_{1\le t\le T}\|\bar{\kappa}_{tT}-\kappa_0\|\le\|\bar{\kappa}_T-\kappa_0\|=o_p(1)$ by assumption. To prove (b), use partial summation to write

$$\sum_{t=2}^{T}[G_t(\bar{\kappa}_T)-G_t(\kappa_0)]\Delta Z_{tT}'=[G_T(\bar{\kappa}_T)-G_T(\kappa_0)]Z_{TT}'-[G_1(\bar{\kappa}_T)-G_1(\kappa_0)]Z_{1T}'$$
$$-\sum_{t=2}^{T}[\Delta G_t(\bar{\kappa}_T)-\Delta G_t(\kappa_0)]Z_{t-1,T}'.$$

The arguments used in case (a) show that the first two terms on the right hand side are of order $o_p(1)$. The norm of the third one is dominated by

$$\max_{1\le t\le T}\|Z_{tT}\|\sum_{t=2}^{T}\|\Delta G_t(\bar{\kappa}_T)-\Delta G_t(\kappa_0)\|$$
$$=O_p(1)\sum_{t=2}^{M}\|\Delta G_t(\bar{\kappa}_T)-\Delta G_t(\kappa_0)\|+O_p(1)\sum_{t=M+1}^{T}\|\Delta G_t(\bar{\kappa}_T)-\Delta G_t(\kappa_0)\|.$$

Here the first term on the right hand side is of order $o_p(1)$ for every fixed M. To analyze the second one, let K_1 be a compact neighborhood of the point κ_0 in which Condition 1 holds and suppose that $\bar{\kappa}_T\in K_1$. Then,

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

$$\sum_{t=M+1}^{T} \|\Delta G_t(\bar{\kappa}_T) - \Delta G_t(\kappa_0)\| \leq 2 \sup_{\kappa \in K_1} \sum_{t=M+1}^{T} \|\Delta G_t(\kappa)\|.$$

Using Condition 1(c) it can be shown that the right hand side can be made arbitrarily small by choosing M large enough. Hence, we have established the stated result in the special case $\bar{\kappa}_T \in K_1$. The general case is obtained in a straightforward manner because $P\{\bar{\kappa}_T \in K_1\} \to 1$. ■

The first result of Lemma A.3 makes it possible to apply mean value expansions termwise in sums containing quantities such as $f_t(\hat{\gamma}_T) - f_t(\gamma_0)$. In this way we can avoid using second order Taylor expansions, which simplifies the derivations considerably. In these applications Z_{tT} is typically $T^{-1/2}y_{2,t-1}$ or $f_t(\hat{\gamma}_T)$. The second result can be used in usual mean value expansions with ΔZ_{tT} typically being $T^{-1/2}\varepsilon_t$ or $T^{-1/2}z_t$.

A.3. Proofs for Section 3.1.

A.3.1. Proof of Proposition 3.1. First note that the likelihood function defined in (3.2) can be written as

$$\ell_T(\theta, \Omega) = T \log \det(\Omega) + \operatorname{tr}\left\{\Omega^{-1} R_T \sum_{t=1}^{T} q_{tT}(\gamma) q_{tT}(\gamma)' R_T'\right\}, \tag{A.3}$$

where $R_T = [I_s \ -B \ -T^{1/2}C \ -D]$ and $q_{tT}(\gamma) = [\Delta y_t' \ z_t(\gamma)' \ T^{-1/2}y_{2,t-1}' \ f_t(\gamma)']'$. Using Lemma A.1 we now prove the following intermediate result.

LEMMA A.4. *Suppose that the assumptions of Lemma* A.1 *hold with* $G(x;k)$ *and* K *replaced by* $f(x;\gamma)$ *and* Γ, *respectively. Suppose also that Assumption* 5(c) *holds. Then, there exists an* $\varepsilon > 0$ *such that, with probability approaching one,*

$$\inf_{\gamma \in \Gamma} \lambda_{\min}\left(T^{-1} \sum_{t=1}^{T} q_{tT}(\gamma) q_{tT}(\gamma)'\right) \geq \varepsilon.$$

Proof. Let Σ_{11} be the covariance matrix of $\underline{\zeta}_t = [\Delta x_t' \ldots \Delta x_{t-p+1}' x_{t-1}' \beta_0]'$ and

$$\Sigma_{22}(\gamma) = \begin{bmatrix} d_0 \int_0^1 \bar{f}(x;\gamma)\bar{f}(x;\gamma)'\,dx d_0' & -d_0 \int_0^1 \bar{f}(x;\gamma)V(\gamma;x)'\,dx \\ -\int_0^1 V(x;\gamma)\bar{f}(x;\gamma)'\,dx d_0' & \int_0^1 V(x;\gamma)V(x;\gamma)'\,dx \end{bmatrix},$$

where $V(x;\gamma)' = [W_2(x)' \ f(x;\gamma)']$ and $\bar{f}(x;\gamma) = f(x;\gamma) - f(x;\gamma_0)$. From the definition of $q_{tT}(\gamma)$ it follows that

$$q_{tT}(\gamma) = \begin{bmatrix} \underline{z}_t \\ \cdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ -d_0 \bar{f}_t(\gamma) \\ \cdots \\ T^{-1/2}y_{2,t-1} \\ f_t(\gamma) \end{bmatrix},$$

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

where $\underline{z}_t = [\Delta y_t' \; z_t']'$ as before and $\bar{f}_t(\gamma) = \bar{f}(t/T;\gamma)$. Thus, an application of Lemma A.1 yields

$$T^{-1}\sum_{t=1}^{T} q_{tT}(\gamma)q_{tT}(\gamma)' \Rightarrow \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{22}(\gamma) \end{bmatrix}$$

$$\stackrel{\text{def}}{=} \Sigma(\gamma), \tag{A.4}$$

where the partitions on the right hand side are *not* conformable but the $r \times r$ lower right hand corner of Σ_{11} and the $r \times r$ upper left hand corner of $\Sigma_{22}(\gamma)$ overlap. Moreover, the left hand side also has a weak limit in $C(\Gamma)^{a\times a}$, $a = s(p+1) + r_1$. It is easy to see that $\Sigma_{11} > 0$ and $\Sigma_{22}(\gamma) \geq 0$ (a.s.) for all $\gamma \in \Gamma$. From this it can further be shown that the sum on the right hand side of Expression (A.4) is positive definite (a.s.) for all $\gamma \in \Gamma$. To demonstrate this, consider the quadratic form $b'\Sigma(\gamma)b = b_1'\Sigma_{11}b_1 + b_2\Sigma_{22}(\gamma)b_2$ where b_1 and b_2 are suitable subvectors of b and the last r components of b_1 and the first r components of b_2 are the same. Because $\Sigma_{11} > 0$ and $\Sigma_{22}(\gamma) \geq 0$ (a.s.) the equality $b'\Sigma(\gamma)b = 0$ implies $b_1 = 0$ so that b_2 is of the form $b_2 = [0' \; b_{22}']'$ where b_{22} is $(s - r + r_1) \times 1$. Thus, it follows that $b_{22}'\int_0^1 V(x;\gamma)V(x;\gamma)'\,dx b_{22} = 0$ and, because $W_2(x)$ is a full rank Brownian motion and $\int_0^1 f(x;\gamma)f(x;\gamma)'\,dx > 0$ for all γ (Assumption 5(c)), we must have $b_{22} = 0$. Hence, we have shown that $\lambda_{\min}(\Sigma(\gamma)) > 0$ (a.s.) for all γ and, because $\Sigma(\gamma)$ is a continuous function of γ, it follows from the continuity of eigenvalues and compactness of Γ that $\inf_{\gamma\in\Gamma}\lambda_{\min}(\Sigma(\gamma)) > 0$ (a.s.). This, (A.4), and the continuous mapping theorem imply the result of the lemma provided the functions $\lambda_{\min}\colon C(\Gamma)^{a\times a} \to C(\Gamma)$, $a = s(p+1) + r_1$, and $\inf_{\gamma\in\Gamma}\colon C(\Gamma) \to \mathbb{R}$ are continuous. These continuity results are straightforward to verify and therefore are omitted. This completes the proof. ■

Using Lemma A.4 we can now establish the existence of the estimators $\hat{\theta}_T$ and $\hat{\Omega}_T$. The first step is to restrict the parameter space by using the inequality in Lemma A.4. Thus, suppose that this inequality holds and note that $\text{tr}(A_1A_2) \geq \lambda_{\max}(A_1)\lambda_{\min}(A_2)$ for any $s \times s$ matrices A_1 and A_2 such that $A_1 > 0$ and $A_2 > 0$ (see, e.g., Deistler and Pötscher, 1984, p. 848). Then, because the first s columns of R_T form an identity matrix and, consequently, $\lambda_{\min}(R_TR_T') \geq 1$, one obtains from (A.3)

$$\begin{aligned} T^{-1}\ell_T(\theta,\Omega) &\geq s\log\lambda_{\min}(\Omega) + \lambda_{\max}(\Omega^{-1})\lambda_{\min}\left(R_T T^{-1}\sum_{t=1}^{T} q_{tT}(\gamma)q_{tT}(\gamma)'R_T'\right) \\ &\geq s\log\lambda_{\min}(\Omega) + \lambda_{\min}^{-1}(\Omega)\lambda_{\min}\left(T^{-1}\sum_{t=1}^{T} q_{tT}(\gamma)q_{tT}(\gamma)'\right) \\ &\geq s\log\lambda_{\min}(\Omega) + \varepsilon/\lambda_{\min}(\Omega). \end{aligned}$$

This shows that $T^{-1}\ell_T(\theta,\Omega) \to \infty$ if $\lambda_{\min}(\Omega) \to 0$. Next note that

$$T^{-1}\ell_T(\theta,\Omega) \geq \log\det(\Omega) \geq \log\lambda_{\max}(\Omega) + (s-1)\log\lambda_{\min}(\Omega),$$

where the second inequality is based on the fact that $\det(\Omega)$ is the product of the eigenvalues of Ω. Hence, if $\lambda_{\min}(\Omega) \geq c_1$ we have $T^{-1}\ell_T(\theta,\Omega) \to \infty$ when $\lambda_{\max}(\Omega) \to \infty$. From (A.3) one also obtains

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

$$T^{-1}\ell_T(\theta,\Omega) \geq s\log\lambda_{\min}(\Omega) + \lambda_{\min}(\Omega^{-1})\mathrm{tr}\left(R_T T^{-1}\sum_{t=1}^{T} q_{tT}(\gamma)q_{tT}(\gamma)'R_T'\right)$$

$$\geq s\log\lambda_{\min}(\Omega) + \lambda_{\max}^{-1}(\Omega)\lambda_{\min}\left(T^{-1}\sum_{t=1}^{T} q_{tT}(\gamma)q_{tT}(\gamma)'\right)\mathrm{tr}(R_T R_T')$$

(cf. Deistler and Pötscher, 1984, p. 850). Thus, if $\lambda_{\min}(\Omega) \geq c_1$ and $\lambda_{\max}(\Omega) \leq c_2$ we have $T^{-1}\ell_T(\theta,\Omega) \to \infty$ when $\mathrm{tr}(R_T R_T') \to \infty$. Using the definition of R_T we can now conclude that, for every T for which the inequality of Lemma A.4 holds, $T^{-1}\ell_T(\theta,\Omega) \to \infty$ if any of the following conditions holds: (i) $\lambda_{\min}(\Omega) \to 0$, (ii) $\lambda_{\max}(\Omega) \to \infty$, (iii) $\|B\| \to \infty$, (iv) $\|T^{1/2}C\| \to \infty$, (v) $\|D\| \to \infty$. Thus, because it is readily checked that $T^{-1}\ell_T(\theta_0,\Omega_0) = O_p(1)$ it follows that ML estimators $\hat{\theta}_T$ and $\hat{\Omega}_T$ exist with probability approaching one if the function $T^{-1}\ell_T(\theta,\Omega)$ has a minimum in the part of the parameter space where $\lambda_{\min}(\Omega) \geq c_1$, $\lambda_{\max}(\Omega) \leq c_2$, $\|B\| \leq c_3$, $\|T^{1/2}C\| \leq c_4$, and $\|D\| \leq c_5$ for suitable $c_1,\ldots,c_5$. These restrictions in conjunction with $\gamma \in \Gamma$ define a compact set whose intersection with the extended parameter space $\{(\theta,\Omega):\theta \in \bar{\Theta}$ and $\Omega > 0\}$ is also compact. Because the function $T^{-1}\ell_T(\theta,\Omega)$ is continuous, it has a minimum in a compact set. Thus, we have proved the first assertion of Proposition 3.1. The proof of the second assertion is immediate because we saw earlier that, with probability approaching one, the estimator $\hat{C}_T$ exists and satisfies $\|T^{1/2}\hat{C}_T\| \leq c_4$ for a suitable c_4.

Before turning to the proof of Theorem 3.1 we note that in the special case where the sequence $g_t(\mu)$ reduces to $f_t(\gamma)$ we have $R_T = [I_s \ -B \ -T^{1/2}C]$ and $q_{tT}(\gamma) = [\Delta y_t' \ z_t(\gamma)' \ T^{-1/2}y_{2,t-1}']'$ on the right hand side of (A.3) and $V(x;\gamma) = W_2(x)$ in the proof of Lemma A.4. From the given proofs it can then be seen that Lemma A.4 and, consequently, Proposition 3.1 can be proved without Assumption 5(c).

A.3.2. Proofs of Theorem 3.1 and Corollary 3.1. On the basis of Proposition 3.1 we can assume that the estimators $\hat{\theta}_T$ and $\hat{\Omega}_T$ exist and that $\hat{C}_T = O_p(T^{-1/2})$. The arguments used in the proof of Proposition 3.1 also show that we can assume that $\hat{\theta}_T$, $\hat{\Omega}_T$, and $\hat{\Omega}_T^{-1}$ are of order $O_p(1)$. From the definitions and (3.6) it thus follows that

$$0 \geq \ell_T(\hat{\theta}_T,\hat{\Omega}_T) - \ell_T(\theta_0,\Omega_0)$$

$$= \ell_{1T}(\hat{\theta}_T,\hat{\Omega}_T) + \ell_{2T}(\hat{\theta}_{2T},\hat{\Omega}_T) - \ell_{2T}(\theta_{20},\Omega_0)$$

$$\geq \ell_{1T}(\hat{\theta}_T,\hat{\Omega}_T) + \inf_{\{\theta_2,\Omega\}} \ell_{2T}(\theta_2,\Omega) - \ell_{2T}(\theta_{20},\Omega_0),$$

where the infimum is over unrestricted values of θ_2 and $\Omega > 0$. Recall that $\theta_2 = B$ and that $\ell_{2T}(\theta_2,\Omega)$ is the likelihood function of the linear regression model $\Delta y_t = Bz_t + \varepsilon_t$ where the regressors are asymptotically stationary in the sense that the matrix of second sample moments converges in probability to a nonstochastic positive definite limit. Thus, using standard arguments it can be seen that the difference between the second and third terms in the last preceding expression is of order $O_p(1)$. In fact, this difference is the negative of the likelihood ratio test statistic for the null hypothesis $\theta_2 = \theta_{20}$ and $\Omega = \Omega_0$. Hence, we have

$$0 \geq \ell_{1T}(\hat{\theta}_T,\hat{\Omega}_T) + O_p(1). \qquad \textbf{(A.5)}$$

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

This result will be used to show that

$$T^{-1}\ell_{1T}(\hat{\theta}_T, \hat{\Omega}_T) = O_p(T^{-1/2}). \tag{A.6}$$

The definition of $\ell_{1T}(\theta_1, \Omega)$ shows that the left hand side of (A.6) consists of two additive components. For the first one we consider

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} \varepsilon_{1t}(\hat{\theta}_T)\varepsilon_{2t}(\hat{\theta}_{2T})' &= -\hat{C}_T T^{-1}\sum_{t=1}^{T} y_{2,t-1}[\Delta y_t - \hat{B}_T z_t]' \\ &\quad + \hat{\alpha}_T d_0 T^{-1}\sum_{t=1}^{T}[\, f_t(\hat{\gamma}_T) - f_t(\gamma_0)][\Delta y_t - \hat{B}_T z_t]' \\ &\quad - \hat{D}_T T^{-1}\sum_{t=1}^{T} f_t(\hat{\gamma}_T)[\Delta y_t - \hat{B}_T z_t]' \\ &= O_p(T^{-1/2}). \end{aligned} \tag{A.7}$$

Here the latter equality is based on Proposition 3.1, Lemma A.1(b),(c), and the fact that $\hat{\theta}_T = O_p(1)$. Thus, because $\hat{\Omega}_T^{-1} = O_p(1)$ we can conclude from (A.5) and the definition of $\ell_{1T}(\theta, \Omega)$ that

$$\begin{aligned} 0 &\geq T^{-1}\ell_{1T}(\hat{\theta}_T, \hat{\Omega}_T) + O_p(T^{-1}) \\ &= \mathrm{tr}\left(\hat{\Omega}_T^{-1} T^{-1}\sum_{t=1}^{T} \varepsilon_{1t}(\hat{\theta}_T)\varepsilon_{1t}(\hat{\theta}_T)'\right) + O_p(T^{-1/2}). \end{aligned} \tag{A.8}$$

Because the first term in the last expression is nonnegative it must be at most of order $O_p(T^{-1/2})$, and so (A.6) follows.

The results obtained earlier will first be used to prove the consistency of the estimators $\hat{B}_T$ and $\hat{\Omega}_T$. Define

$$\bar{\ell}_2(B, \Omega) = \log\det(\Omega) + \mathrm{tr}\{\Omega^{-1}[\Omega_0^{-1} + (B - B_0)(E\zeta_t\zeta_t')^{-1}(B - B_0)']\},$$

where $\zeta_t = [\Delta x_{t-1}' \ldots \Delta x_{t-p+1}' x_{t-1}' \beta_0]'$ as before. Using the definitions and Lemma A.1(a) it is not difficult to check that $T^{-1}\ell_{2T}(B, \Omega) \xrightarrow{p} \bar{\ell}_2(B, \Omega)$ where the convergence is uniform over compact subsets of the parameter spaces of B and Ω. It is also easy to check (and actually well known) that the function $\bar{\ell}_2(B, \Omega)$ has a unique minimum at $(B, \Omega) = (B_0, \Omega_0)$. Thus, we have

$$\begin{aligned} \bar{\ell}_2(\hat{B}_T, \hat{\Omega}_T) &\geq \bar{\ell}_2(B_0, \Omega_0) \\ &= T^{-1}\ell_T(\theta_0, \Omega_0) + [\bar{\ell}_2(B_0, \Omega_0) - T^{-1}\ell_{2T}(B_0, \Omega_0)] \\ &\geq T^{-1}\ell_{1T}(\hat{\theta}_T, \hat{\Omega}_T) + T^{-1}\ell_{2T}(\hat{B}_T, \hat{\Omega}_T) + [\bar{\ell}_2(B_0, \Omega_0) - T^{-1}\ell_{2T}(B_0, \Omega_0)] \\ &= \bar{\ell}_2(\hat{B}_T, \hat{\Omega}_T) + o_p(1). \end{aligned}$$

Here the last equality is based on (A.6) and the previously mentioned uniform convergence of $T^{-1}\ell_{2T}(B, \Omega)$. As a whole, the preceding inequalities show that $\bar{\ell}_2(\hat{B}_T, \hat{\Omega}_T) \xrightarrow{p} \bar{\ell}_2(B_0, \Omega_0)$ and, because $\bar{\ell}_2(B, \Omega)$ is uniquely minimized at $(B, \Omega) = (B_0, \Omega_0)$, we must have $(\hat{B}_T, \hat{\Omega}_T) \xrightarrow{p} (B_0, \Omega_0)$. Indeed, because $\hat{B}_T = O_p(1)$ and $\hat{\Omega}_T = O_p(1)$ every subsequence of $(\hat{B}_T, \hat{\Omega}_T)$ contains a weakly convergent subsequence with a

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

limit (B_*, Ω_*), say. Thus, if T_j signifies this latter subsequence one obtains from the continuity of the function $\bar{\ell}_2(B, \Omega)$ that $\bar{\ell}_2(\hat{B}_{T_j}, \hat{\Omega}_{T_j}) \Rightarrow \bar{\ell}_2(B_*, \Omega_*)$, and, because we must have $\bar{\ell}_2(B_0, \Omega_0) = \bar{\ell}_2(B_*, \Omega_*)$, we have a contradiction unless $(B_*, \Omega_*) = (B_0, \Omega_0)$ (a.s.). This implies that $(\hat{B}_T, \hat{\Omega}_T)$ converges weakly to (B_0, Ω_0) and hence also in probability. This proves part (b) of Theorem 3.1.

The next step is to establish the existence of the estimators $\hat{A}_T$ and $\hat{d}_T$. Recall that $C = -\alpha(A - A_0)$ and $D = -\alpha(d - d_0)$. Because we have already proved the consistency of the estimator $\hat{\alpha}_T$ we can here assume that the parameter space of α is $\alpha(\Psi)$ instead of its closure. Then we have $A = A_0 - (\alpha'\alpha)^{-1}\alpha' C$ and $d = d_0 - (\alpha'\alpha)^{-1}\alpha' D$ so that there is a one to one correspondence between the parameters (α, C, D) and (α, A, d). In fact, this correspondence is even a homeomorphism, that is, a continuous bijection with a continuous inverse. Hence, it follows from Proposition 3.1 that, with probability approaching one, ML estimators of A and d exist and are related to the ML estimators $\hat{\alpha}_T$, $\hat{C}_T$, and $\hat{D}_T$ by $\hat{A}_T = A_0 - (\hat{\alpha}_T'\hat{\alpha}_T)^{-1}\hat{\alpha}_T'\hat{C}_T$ and $\hat{d}_T = d_0 - (\hat{\alpha}_T'\hat{\alpha}_T)^{-1}\hat{\alpha}_T'\hat{D}_T$, respectively. For later purposes we also note that this means that subsequently we can write $\hat{D}_T = -\hat{\alpha}_T(\hat{d}_T - d_0)$ and, furthermore, $\varepsilon_{1t}(\hat{\theta}_T) = -\hat{C}_T y_{2,t-1} + \hat{\alpha}_T[\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)]$ (cf. the discussion following (3.5)).

To prove the remaining parts of Theorem 3.1, recall that the first term in the last expression of (A.8) is of order $O_p(T^{-1/2})$. Thus, using the previously mentioned expression of $\varepsilon_{1t}(\hat{\theta}_T)$ yields

$$
\begin{aligned}
O_p(T^{-1/2}) = \operatorname{tr}&\left(\hat{\Omega}_T^{-1}\hat{C}_T T^{-1}\sum_{t=1}^{T} y_{2,t-1}y_{2,t-1}'\hat{C}_T'\right)\\
&- 2\operatorname{tr}\left(\hat{\Omega}_T^{-1}\hat{C}_T T^{-1}\sum_{t=1}^{T} y_{2,t-1}[\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)]'\hat{\alpha}_T'\right)\\
&+ \operatorname{tr}\left(\hat{\Omega}_T^{-1}\hat{\alpha}_T T^{-1}\sum_{t=1}^{T}[\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)][\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)]'\hat{\alpha}_T'\right)\\
&\stackrel{\text{def}}{=} Y_T. \qquad \textbf{(A.9)}
\end{aligned}
$$

Because $T^{1/2}\hat{C}_T$, $\hat{d}_T$, and $\hat{\gamma}_T$ are of order $O_p(1)$ we can extract from an arbitrary subsequence of $(T^{1/2}\hat{C}_T, \hat{d}_T, \hat{\gamma}_T)$ a weakly convergent subsequence so that $(T_j^{1/2}\hat{C}_{T_j}, \hat{d}_{T_j}, \hat{\gamma}_{T_j}) \Rightarrow (C_*, d_*, \gamma_*)$, say. We wish to show that $C_* = 0$ (a.s.), which implies that $T^{1/2}\hat{C}_T = o_p(1)$. To this end, we use Lemmas A.1(d)–(f) and A.2 in conjunction with the consistency of the estimators $\hat{\alpha}_T$ and $\hat{\Omega}_T$ to derive an expression for the weak limit of Y_{T_j}. It is easy to check that, redefining the subsequence if necessary, we can assume subsequently that all the needed joint weak convergencies apply. Thus, we find that

$$
\begin{aligned}
Y_{T_j} \Rightarrow \operatorname{tr}&\left(\Omega_0^{-1}\int_0^1 \{C_* W_2(x) - \alpha_0[d_* f(x;\gamma_*) - d_0 f(x;\gamma_0)]\}\right.\\
&\left.\times\{C_* W_2(x) - \alpha_0[d_* f(x;\gamma_*) - d_0 f(x;\gamma_0)]\}'\,dx\right)\\
\geq\ &\lambda_{\min}(\Omega_0^{-1})\int_0^1 \|C_* W_2(x) - \alpha_0[d_* f(x;\gamma_*) - d_0 f(x;\gamma_0)]\|^2\,dx. \qquad \textbf{(A.10)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

However, from (A.9) we see that the last expression must be zero (a.s.) so that the integrand must be zero for (Lebesgue) almost all x. Because $W_2(x)$ is a full rank Brownian motion this implies $C_* = 0$ (a.s.), as desired. Thus, we have proved that $T^{1/2}\hat{C}_T = o_p(1)$. Because $\hat{A}_T = A_0 - (\hat{\alpha}_T'\hat{\alpha}_T)^{-1}\hat{\alpha}_T'\hat{C}_T$ this and the consistency of $\hat{\alpha}_T$ give $T^{1/2}(\hat{A}_T - A_0) = o_p(1)$. This proves part (a) of Theorem 3.1.

To complete the proof, note that when $T^{1/2}\hat{C}_T = o_p(1)$ holds we find from (A.9), Lemma A.1(e) and (f), and arguments used to derive (A.10) that

$$o_p(1) = \operatorname{tr}\left(\hat{\Omega}_T^{-1}\hat{\alpha}_T T^{-1}\sum_{t=1}^{T}[\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)][\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)]'\hat{\alpha}_T'\right)$$

$$\geq \lambda_{\min}(\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T)T^{-1}\sum_{t=1}^{T}\|\hat{d}_T f_t(\hat{\gamma}_T) - d_0 f_t(\gamma_0)\|^2.$$

The last result of the theorem follows from this and the consistency of the estimators $\hat{\alpha}_T$ and $\hat{\Omega}_t$. Thus, the proof of Theorem 3.1 is complete. ■

To prove Corollary 3.1, observe first that when $A_0 = A(\phi_0)$ is an interior point of $A(\Phi)$ in the relative topology of $\overline{A(\Phi)}$ it follows from the consistency of $\hat{A}_T$ that, with probability approaching one, $\hat{A}_T$ belongs to a neighborhood of A_0 contained in $A(\Phi)$. By the latter condition of Assumption 2(b) we can choose this neighborhood $\mathcal{N}_0$, say, such that the continuous inverse function $A^{-1}(\cdot)$ is defined on it. In the set $A^{-1}(\mathcal{N}_0)$ there then exists $\hat{\phi}_T$ such that $A(\hat{\phi}_T) = \hat{A}_T$ and the consistency of $\hat{\phi}_T$ follows from the consistency of $\hat{A}_T$ and the continuity of the function $A^{-1}(\cdot)$. The existence and consistency of the estimator $\hat{\psi}_T$ follow similarly from Assumption 3(b) and the properties of the estimator $\hat{B}_T$. To prove the third result, recall the argument used in (A.10) where we concluded that the integrand must be zero for (Lebesgue) almost all x and that $C_* = 0$ (a.s.) also holds. Thus, we must have $d_* f(x;\gamma_*) = d_0 f(x;\gamma_0)$ (a.s.) so that the identification condition of Assumption 5(d) and the right continuity of the function $f(\cdot;\gamma)$ imply that $d_* = d_0$ and $\gamma_* = \gamma_0$ (a.s.). This gives the consistency of the estimators $\hat{d}_T$ and $\hat{\gamma}_T$. Finally, the last assertion follows from the consistency of $\hat{d}_T$ and Assumption 4(b) by arguments similar to those used in the case of the estimator $\hat{\phi}_T$ earlier. ■

A.3.3. Proof of Proposition 3.2. Denote $\underline{\vartheta}_1 = (A, d, \gamma)$ and note that under the stated conditions ML estimators of $\underline{\vartheta}_1$, B, and Ω exist with probability approaching one and are consistent. Write the vector $\varepsilon_t(\theta)$ and the likelihood function $\ell_T(\theta, \Omega)$ in terms of the parameters $\underline{\vartheta}_1$, B, and Ω and denote the resulting quantities by $\varepsilon_t(\underline{\vartheta}_1, B) = \varepsilon_{1t}(\underline{\vartheta}_1, B) + \varepsilon_{2t}(B)$ and $\ell_T(\underline{\vartheta}_1, B, \Omega) = \ell_{1T}(\underline{\vartheta}_1, B, \Omega) + \ell_{2T}(B, \Omega)$, respectively (cf. (3.4)–(3.6)). Again, we have $\varepsilon_{1t}(\underline{\vartheta}_{10}, B) = 0$ and, consequently, $\ell_{1T}(\underline{\vartheta}_{10}, B, \Omega) = 0$. Thus, it follows that

$$\begin{aligned}
0 &\geq \ell_T(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T, \hat{\Omega}_T) - \ell_T(\underline{\vartheta}_{10}, \hat{B}_T, \hat{\Omega}_T) \\
&= \ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T, \hat{\Omega}_T) \\
&= 2\operatorname{tr}\left(\hat{\Omega}_T^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T)\varepsilon_{2t}(\hat{B}_T)'\right) \\
&\quad + \operatorname{tr}\left(\hat{\Omega}_T^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T)\varepsilon_{1t}(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T)'\right). \qquad \textbf{(A.11)}
\end{aligned}$$

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

Here we have $\varepsilon_{2t}(\hat{B}_T) = \Delta y_t - \hat{B}_T z_t$, whereas $\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T)$ can be written as

$$\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T) = \hat{\alpha}_T(y'_{2,t-1} \otimes I_r)\mathrm{vec}(\hat{A}_T - A_0) + \hat{\alpha}_T(f_t(\hat{\gamma}_T)' \otimes I_r)\mathrm{vec}(\hat{d}_T - d_0) + \hat{\alpha}_T d_0(f_t(\hat{\gamma}_T) - f_t(\gamma_0))$$

(cf. (3.4)). Inserting these expressions of $\varepsilon_{2t}(\hat{B}_T)$ and $\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T)$ into the first term of the last sum in (A.11) and using a mean value expansion to linearize $f_t(\hat{\gamma}_T) - f_t(\gamma_0)$ gives

$$\mathrm{tr}\left(\hat{\Omega}_T^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T)\varepsilon_{2t}(\hat{B}_T)'\right) = T^{-1/2}\sum_{t=1}^{T}[\Delta y_t - \hat{B}_T z_t]'\hat{\Omega}_T^{-1}\hat{\alpha}_T Q_{tT}(\hat{\gamma}_T, \bar{\gamma}_T)'\hat{\xi}_T \stackrel{\mathrm{def}}{=} \hat{\Xi}_{1T}\hat{\xi}_T, \tag{A.12}$$

where

$$\hat{\xi}_T = \begin{bmatrix} T\,\mathrm{vec}(\hat{A}_T - A_0) \\ T^{1/2}\,\mathrm{vec}(\hat{d}_T - d_0) \\ T^{1/2}(\hat{\gamma}_T - \gamma_0) \end{bmatrix}$$

and

$$Q_{tT}(\hat{\gamma}_T, \bar{\gamma}_T)' = [(T^{-1/2}y'_{2,t-1} \otimes I_r)\,(f_t(\hat{\gamma}_T)' \otimes I_r)\,d_0(\partial f_t(\bar{\gamma}_T)/\partial\gamma')].$$

Here $\partial f_t(\bar{\gamma}_T)/\partial\gamma'$ is due to the mean value expansion and signifies a matrix whose ith row equals the ith row of the matrix $\partial f_t(\bar{\gamma}_T^{(i)})/\partial\gamma'$ with $\bar{\gamma}_T^{(i)} = a_i\gamma_0 + (1 - a_i)\hat{\gamma}_T$, $0 \le a_i \le 1$ $(i = 1, \ldots, r_1)$.

Now consider the latter term of the last sum in (A.11). Here we linearize $f_t(\hat{\gamma}_T) - f_t(\gamma_0)$ in $\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T)$ by a termwise mean value expansion. Using the preceding notation we can then write $\varepsilon_{1t}(\underline{\vartheta}_{10}, \hat{B}_T) = \hat{\alpha}_T Q_{tT}(\hat{\gamma}_T, \bar{\gamma}_{tT})'\hat{\xi}_T$ where $\bar{\gamma}_{tT}$ indicates intermediate points $\bar{\gamma}_{tT}^{(i)}$ such that $\|\bar{\gamma}_{tT}^{(i)} - \gamma_0\| \le \|\hat{\gamma}_T - \gamma_0\|$ $(i = 1, \ldots, r_1)$. Thus,

$$\begin{aligned}\mathrm{tr}&\left(\hat{\Omega}_T^{-1}\sum_{t=1}^{T}\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T)\varepsilon_{1t}(\underline{\hat{\vartheta}}_{1T}, \hat{B}_T)'\right) \\ &= \hat{\xi}_T' T^{-1}\sum_{t=1}^{T} Q_{tT}(\hat{\gamma}_T, \bar{\gamma}_{tT})\hat{\alpha}_T'\hat{\Omega}_T^{-1}\hat{\alpha}_T Q_{tT}(\hat{\gamma}_T, \bar{\gamma}_{tT})'\hat{\xi}_T. \\ &\stackrel{\mathrm{def}}{=} \hat{\xi}_T'\hat{\Xi}_{2T}\hat{\xi}_T.\end{aligned} \tag{A.13}$$

Because our assumptions imply that the function formed from $f(x;\gamma)$ and $\partial f(x;\gamma)/\partial\gamma_i$ $(i = 1, \ldots, q)$ satisfies Condition 1 in some compact neighborhood of γ we can now use the consistency of the involved estimators and Lemmas A.3(a) and A.1(d)–(f) to conclude that

$$\begin{aligned}\hat{\Xi}_{2T} &= T^{-1}\sum_{t=1}^{T} Q_{tT}(\gamma_0, \gamma_0)\alpha_0'\Omega_0^{-1}\alpha_0 Q_{tT}(\gamma_0, \gamma_0)' + o_p(1) \\ &\Rightarrow \int_0^1 r(x;\gamma_0)\alpha_0'\Omega_0^{-1}\alpha_0 r(x, \gamma_0)'\,dx \stackrel{\mathrm{def}}{=} \Xi_{20},\end{aligned}$$

where $r(x;\gamma_0) = [(W_2(x)' \otimes I_r)\,(f(x;\gamma_0)' \otimes I_r)\,d_0(\partial f(x;\gamma_0)/\partial\gamma')]'$. In the same way as in the proof of Lemma A.4 it can be seen that $\Xi_{20} > 0$ (a.s.) when Assumption 6(c)

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press

holds. Hence, for some $c > 0$, $\lambda_{\min}(\hat{\Xi}_{2T}) \geq c$ with probability approaching one. As for $\hat{\Xi}_{1T}$ in (A.12), note that the consistency of the involved estimators and Lemma A.1(b) and (c) imply that $\hat{\Xi}_{1T} = O_p(1)$. Thus, from (A.11)–(A.13) it follows that, with probability approaching one,

$$0 \geq \lambda_{\min}(\hat{\Xi}_{2T})\|\hat{\xi}_T\|^2(1 - \hat{c}_{1T}/\|\hat{\xi}_T\|), \tag{A.14}$$

where $\hat{c}_{1T} = 2\|\hat{\Xi}_{1T}\|/\lambda_{\min}(\hat{\Xi}_{2T}) = O_p(1)$. Because $\lambda_{\min}(\hat{\Xi}_{2T}) \geq c$ holds with probability approaching one it is straightforward to conclude from this that $\|\hat{\xi}_T\| = O_p(1)$. By the definition of $\hat{\xi}_T$, this proves the first three results of Proposition 3.2.

Now consider the proof of the last result of Proposition 3.2 and notice that $\hat{A}_T = A_0 + O_p(T^{-1})$, $\hat{d}_T = d_0 + O_p(T^{-1/2})$, and $\hat{\gamma}_T = \gamma_0 + O_p(T^{-1/2})$ by the first three results just proved. These facts and the consistency of $\hat{B}_T$ imply that we can improve (A.7) by replacing the term $O_p(T^{-1/2})$ by $O_p(T^{-1})$. As far as the first and third expressions on the right hand side of the first equality of (A.7) are concerned, this follows directly from Lemma A.1(b) and (c) and the previously mentioned consistency results. For the second expression the same conclusion is obtained similarly after linearizing $f_t(\hat{\gamma}_T) - f_t(\gamma_0)$ by a mean value expansion. Thus, we can conclude that the first term in the last expression of (A.8) is of order $O_p(T^{-1})$, and therefore we also have

$$\ell_{1T}(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T, \hat{\Omega}_T) = O_p(1). \tag{A.15}$$

Thus, from the definitions and (A.15) we find that

$$\ell_{2T}(B_0, \hat{\Omega}_T) = \ell_T(\underline{\vartheta}_{10}, B_0, \hat{\Omega}_T) \geq \ell_T(\hat{\underline{\vartheta}}_{1T}, \hat{B}_T, \hat{\Omega}_T) = \ell_{2T}(\hat{B}_T, \hat{\Omega}_T) + O_p(1).$$

Taking the difference between the extreme quantities yields

$$\begin{aligned}
0 &\geq \operatorname{tr}\left\{\hat{\Omega}_T^{-1}\left(\sum_{t=1}^{T}\varepsilon_{2t}(\hat{B}_T)\varepsilon_{2t}(\hat{B}_T)' - \sum_{t=1}^{T}\varepsilon_{2t}(B_0)\varepsilon_{2t}(B_0)'\right)\right\} + O_p(1) \\
&= \operatorname{tr}\left\{\hat{\Omega}_T^{-1}\left((\hat{B}_T - B_0)\sum_{t=1}^{T} z_t z_t'(\hat{B}_T - B_0)'\right)\right\} - 2\operatorname{tr}\left\{\hat{\Omega}_T^{-1}\left((\hat{B}_T - B_0)\sum_{t=1}^{T} z_t \varepsilon_t'\right)\right\} \\
&\quad + O_p(1) \\
&\geq \lambda_{\min}(\hat{\Omega}_T^{-1})\lambda_{\min}\left(T^{-1}\sum_{t=1}^{T} z_t z_t'\right) T\|\hat{B}_T - B_0\|^2 \\
&\quad - 2\|\hat{\Omega}_T^{-1}\|\left\|T^{-1/2}\sum_{t=1}^{T} z_t \varepsilon_t'\right\| T^{1/2}\|\hat{B}_T - B_0\| + O_p(1) \\
&\overset{\text{def}}{=} \hat{c}_{2T} T\|\hat{B}_T - B_0\|^2 - \hat{c}_{3T} T^{1/2}\|\hat{B}_T - B_0\| + O_p(1).
\end{aligned}$$

Hence,

$$0 \geq \hat{c}_{2T} T\|\hat{B}_T - B_0\|^2(1 - \hat{c}_{3T}/\hat{c}_{2T} T^{1/2}\|\hat{B}_T - B_0\|) + O_p(1).$$

The definitions show that $\hat{c}_{3T} = O_p(1)$ and that, for some $c_2 > 0$, $\hat{c}_{2T} \geq c_2$ with probability approaching one. Thus, in the same way as in the case of (A.14) we can conclude that $\hat{B}_T - B_0 = O_p(T^{-1/2})$. ■

https://doi.org/10.1017/S0266466601172026 Published online by Cambridge University Press