On likelihood ratios for partially identified models

Boswijk, H.P.

Publication date
1995

Document Version
Author accepted manuscript

Published in
Proceedings of the Osaka Econometrics conference

On Likelihood Ratios for Partially Identified Models

H. Peter Boswijk*
Tinbergen Institute & University of Amsterdam†

May 1995; Revised June 1995.

Abstract

This paper studies asymptotic properties of likelihood-based estimators and test statistics for models which are partially identified. We concentrate on local or global identification problems which are linked to ("structural") reparameterisations, where the original ("reduced form") parameters are identified. The problem is approached via local asymptotic analysis of likelihood ratio functions. Applications include simultaneous equation models under limited information, testing for cointegration in VAR models, and testing structural hypotheses on cointegrating vectors.

1 Introduction

One of the central themes in econometric analysis is the identification problem. For example, the Cowles Commission's work focused on identification and estimation of simultaneous equations models; and more recently, identification of long-run relationships in cointegration models has attracted much attention. This emphasis may be explained by the fact that econometrics mostly pertains to non-experimental phenomena, so that the observed variation in the data is a reduced form outcome of the interaction of various underlying structural relationships. Thus, we are faced with the task of making inference on structural parameters, which only indirectly, and often non-linearly, determine the joint distribution of the observables.

If the relation between structural and reduced-form parameters is non-linear, then there will often be specific ranges of the parameter space where part of the structural parameter vector is not identified. If this range is essentially the entire parameter space, then this is a global lack of identification; on the other hand, if it is a subspace (of Lebesgue measure zero), then the identification problem is local. Although local identification problems have always been recognised

*I would like to thank, without implicating, Søren Johansen for helpful comments. Financial support from the Tinbergen Institute and the Royal Netherlands Academy of Arts and Sciences is gratefully acknowledged.

†Address for correspondence: Department of Actuarial Science and Econometrics, University of Amsterdam, Roetersstraat 11, NL-1018 WB Amsterdam, The Netherlands. E-mail: peterb@fee.uva.nl.

in econometric modelling, usually no specific attention is given to their effect on estimation and inference. Rather, the problem is avoided by assuming that the true value of the parameter is outside the problematic region of the parameter space. This common approach was questioned by Phillips (1989), who argued that by concentrating on well-behaved criterion functions, we exclude interesting and empirically relevant phenomena such as spurious regressions and simultaneous equations models under a (local) rank condition violation. More recently, Dufour (1994) showed that if parameters are locally non-identified, then confidence sets for these parameters cannot be bounded with probability one; he also proved that confidence sets based on Wald tests (as opposed to likelihood ratio tests) do not satisfy this property, and thus can yield very unreliable inferences.

These results suggest that analyzing the behaviour of maximum likelihood estimators and likelihood ratio test statistics under local unidentification is not only of theoretical interest, but also of empirical relevance. The present paper seeks to provide such an analysis for a rather general class of parametric models. Following Phillips (1989), we analyse the local asymptotic behaviour of criterion functions, from which the asymptotic distributions of the relevant estimators and test statistics are readily obtained; here we focus on the likelihood ratio as the criterion function. We extend Phillips' analysis to a more general class of models, where estimators need not have an explicit closed form expression. In that case the advantages of studying the likelihood ratio function are even more prominent: the conventional "delta method" to obtain the asymptotic distributions of statistics of interest will break down, since unidentification implies that (some of the) parameters are not estimated consistently, and the second derivative matrix of the criterion function is singular. Our analysis covers not only the classical case of root-n consistent and asymptotically Gaussian estimation, but also non-ergodic models with limiting mixed Gaussian distributions or Brownian motion functionals.

The plan of the paper is as follows. Section 2 reviews local asymptotic analysis of likelihood ratios (LR's). Following Jeganathan (1988), we define locally asymptotically quadratic (LAQ) LR's, as well as two special cases thereof, viz. locally asymptotically mixed normal (LAMN) and locally asymptotically normal (LAN) LR's. In Section 3, we analyse likelihood-based estimation and inference on structural parameters for two cases: the globally (partially) unidentified case, and the locally (partially) unidentified case. In Section 4 the results are applied to simultaneous equation models under limited information, testing for cointegration in VAR models, and testing structural hypotheses on cointegrating vectors. Section 5 contains some concluding remarks.

Throughout the paper, the Euclidean matrix norm $(\mathrm{tr}(A'A))^{1/2}$ is denoted by $\|A\|$. Positive definiteness of a matrix A is denoted by $A > 0$. For any $n \times m$ matrix A of full column rank, $A_\perp$ denotes an $n \times (n-m)$ matrix of full column rank such that $A_\perp' A = 0$. We use $\mathcal{L}(X)$ or $\mathcal{L}(X|P)$ to signify the distribution of X (under a probability measure P); convergence in distribution is denoted by $\xrightarrow{\mathcal{L}}$.

2 LAQ Likelihood Ratios

In this section we shall review the theory of locally asymptotically quadratic (LAQ) likelihood ratios (LR's). This theory was developed recently by Jeganathan (1988), building on earlier work of Le Cam (1960) on locally asymptotically normal (LAN) LR's and of Jeganathan (1980) on locally asymptotically mixed normal (LAMN) LR's. An overview of the material in this section is given in Le Cam and Yang (1990); see also Phillips (1988, 1989) for a lucid introduction.

Consider a p-vector stochastic process $\{X_t, t = 1, 2, \ldots\}$, defined on a family of probability spaces $\{(\Omega, \mathcal{F}, \mathsf{P}_\theta), \theta \in \Theta\}$, where θ is a k-dimensional parameter vector, with parameter space $\Theta \subseteq \mathbb{R}^k$. For fixed n, $X^n \equiv (X_1, \ldots, X_n)$ defines a Borel-measurable mapping from Ω into $\mathbb{R}^{kn}$, inducing a family of probability measures $\{P_{\theta,n}, \theta \in \Theta\}$ on $(\mathbb{R}^{kn}, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel σ-field. We shall assume that for all $n > 0$ and $\theta \in \Theta$, the measure $P_{\theta,n}$ is dominated by a σ-finite measure μ; the density of $P_{\theta,n}$ with respect to μ is defined as the Radon-Nykodym derivative $f_n(x^n; \theta) \equiv \mathrm{d}P_{\theta,n}/\mathrm{d}\mu$. It will be most convenient to think of X^n as a continuous random vector, so that we may take the Lebesgue measure for μ, in which case $f_n(x^n; \theta)$ is the probability density function of X^n. However, the following results may also be applied to distributions with discrete or mixed discrete/continuous supports. What is required is that the support of $P_{\theta,n}$ does not depend on θ, so that any two measures $P_{\theta_0,n}$ and $P_{\theta_1,n}$ are mutually absolutely continuous.

Define the likelihood function by $L_n(\theta) = f_n(X^n; \theta)$; note that $L_n(\theta)$ is viewed as a random function of θ. The log-likelihood ratio for two parameter points $\theta_0, \theta_1 \in \Theta$ is

$$\Lambda_n(\theta_1, \theta_0) \equiv \ln \frac{\mathrm{d}P_{\theta_1,n}}{\mathrm{d}P_{\theta_0,n}} = \ln \frac{f_n(X^n; \theta_1)}{f_n(X^n; \theta_0)} = \ln \frac{L_n(\theta_1)}{L_n(\theta_0)}. \tag{1}$$

Suppose that we consider $\Lambda_n(\theta, \theta_0)$ as a function of θ, with θ_0 fixed at a specific value. Then this function has the same maximand as the likelihood function $L_n(\theta)$, so that the maximum likelihood estimator (MLE) $\hat{\theta}_n$, if it exists, satisfies

$$\hat{\theta}_n \equiv \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\max_{\theta \in \Theta} \Lambda_n(\theta, \theta_0). \tag{2}$$

Thus, the distribution (under $P_{\theta_0,n}$) of the MLE (assuming measurability of $\hat{\theta}_n$) may be derived from the distribution of the log-likelihood ratio function $\Lambda_n(\theta, \theta_0)$.

Since the distributions of estimators and test statistics are often quite complex (if not intractable) in finite samples, we are interested in their asymptotic properties, as $n \to \infty$. The approach taken here is to derive the asymptotic properties of these statistics via analyzing the limiting behaviour of $\Lambda_n(\theta, \theta_0)$. To guarantee that this function actually has a limit, it is necessary to study the behaviour of Λ_n in shrinking neighbourhoods of θ_0, i.e., for sequences $\theta_n = \theta_0 + D_n \tau$, where D_n is a sequence of $k \times k$ non-singular matrices such that $\|D_n\| \to 0$, and $\tau = D_n^{-1}(\theta_n - \theta_0) \in \mathcal{T}_n$ is a parameter vector measuring the "normed" deviation from θ_0, with parameter space

$$\mathcal{T}_n = \{\tau \in \mathbb{R}^k : \theta_0 + D_n \tau \in \Theta\}. \tag{3}$$

In the classical case of independent and identically distributed observations, we have $D_n = n^{1/2} I_k$. However, the present general setup allows for different rates of convergence, which may vary across different linear combinations of θ; note that D_n need not be diagonal. We shall assume that the dimension of Θ is k, and that θ_0 lies in the interior of Θ, so that $\lim_{n\to\infty} \mathcal{T}_n = \mathbb{R}^k$, i.e., the parameter θ is allowed to vary freely in a neigbourhood of θ_0.

Definition 1 *(Jeganathan, 1988) The sequence of families* $\{P_{\theta,n}; \theta \in \Theta; n = 1, 2, \ldots\}$ *is said to have locally asymptotically quadratic (LAQ) likelihood ratios at* $\theta_0 \in \Theta$ *if*

(i) there exist sequences of norming matrices D_n*, random* k*-vectors* S_n *and almost surely positive definite* $k \times k$ *random matrices* J_n *(all possibly depending on* θ_0*), such that under* $P_{\theta_0,n}$*,*

$$\Lambda_n(\theta_0 + D_n\tau_n, \theta_0) - \tau_n' S_n + \frac{1}{2}\tau_n' J_n \tau_n \ \xrightarrow{P}\ 0 \tag{4}$$

for every bounded sequence $\{\tau_n \in \mathcal{T}_n\}$*;*

(ii) under $P_{\theta_0,n}$*,*

$$(S_n, J_n) \ \xrightarrow{\mathcal{L}}\ (S, J), \tag{5}$$

where S *is a random vector and* J *is an almost surely positive definite random matrix. Moreover,* $\mathsf{E}[\exp \Lambda(\tau)] = 1$ *for all* $\tau \in \mathbb{R}^k$*, where*

$$\Lambda(\tau) \equiv \tau' S - \frac{1}{2}\tau' J \tau. \tag{6}$$

The last part of condition (*ii*) implies that the sequences $\{P_{\theta_0,n}\}$ and $\{P_{\theta_0+D_n\tau,n}\}$ are *contiguous*, see Jeganathan (1988), which may be loosely defined by the property that in the limit, the support of $P_{\theta_0+D_n\tau,n}$ does not depend on τ. Note that S is the first derivative of $\Lambda(\tau)$, evaluated in $\tau = 0$, and hence (since Λ_n is equal to the log-likelihood, up to an additive constant) the limit of the normed *score vector*, evaluated in θ_0; similarly, J equals minus the limiting Hessian matrix, or the limit of the normed *observed information matrix.*

Abbreviate $\Lambda_n(\theta_0 + D_n\tau, \theta_0)$ as $\Lambda_n(\tau)$. The conditions for LAQ imply that under $P_{\theta_0,n}$,

$$\Lambda_n(\tau) \ \xrightarrow{\mathcal{L}}\ \Lambda(\tau). \tag{7}$$

However, this convergence is pointwise, i.e., for fixed τ. For the purpose of the present paper this needs to be strengthened to weak convergence of the random *function* $\Lambda_n(.)$. Let $C(\mathbb{R}^k)$ denote the space of continuous functions on $\mathbb{R}^k$ under the uniform metric, and let $C(K)$ be defined analogously for any compact set $K \subset \mathbb{R}^k$. Observe that in $C(\mathbb{R}^k)$, Λ is not a continuous function of (S, J): however close two (non-identical) realisations of (S, J) are, the distance between the corresponding realisations of Λ will be infinite. Therefore, we can only hope to establish (7) uniformly on compact sets, i.e., in $C(K)$. Analogously to the analysis of Billingsley (1968) for

$C[0,1]$, this will require, in addition to pointwise convergence, some smoothness conditions on Λ_n, in particular *stochastic equicontinuity* (see, e.g., Newey, 1991, and Saikkonen, 1993):

Definition 2 *The sequence of functions Λ_n on $C(K)$ is said to be stochastically equicontinuous if for each $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$ and an n_0 such that*

$$P_{\theta_0,n}\left\{\sup_{t,\tau\in K,\|t-\tau\|<\delta}|\Lambda_n(t)-\Lambda_n(\tau)|>\varepsilon\right\}\le\eta,\quad n\ge n_0. \tag{8}$$

The sequence D_n, and hence $\mathcal{T}_n$, may always be chosen such that $K \subset \mathcal{T}_n$ for all n, so that Λ_n is defined on the whole of K. Let $\hat{\tau}_n = \arg\max_{\tau\in K}\Lambda_n(\tau)$, which can be seen as a truncated version of the sequence $D_n^{-1}(\hat{\theta}_n-\theta_0)$: if the latter lies outside K, then $\hat{\tau}_n$ will be on the boundary of K (denoted by ∂K). If $\Lambda_n \xrightarrow{\mathcal{L}} \Lambda$ in $C(K)$, then the continuous mapping theorem (see Billingsley, 1968) guarantees that

$$\hat{\tau}_n \ \xrightarrow{\mathcal{L}}\ \arg\max_{\tau\in K}\Lambda(\tau)=\mathbf{1}_K(J^{-1}S)\cdot J^{-1}S+\mathbf{1}_{K^c}(J^{-1}S)\cdot\arg\max_{\tau\in\partial K}\Lambda(\tau), \tag{9}$$

where $\mathbf{1}_A(.)$ is the indicator function of the set A. Note that the maximand is a continuous mapping of Λ almost surely, because Λ is a quadratic function with negative definite second derivative almost surely.

Let Q denote the probability measure induced on $\mathbb{R}^k$ by $J^{-1}S$. Since $\mathbb{R}^k$ is separable and complete, Q is *tight*, see Billingsley (1968, Theorem 1.4), which entails that for any $\varepsilon > 0$ there exists a K such that $Q(K) > 1-\varepsilon$. This means that the distribution of the right-hand side of (9) is arbitrarily close to the distribution of $J^{-1}S$. Similar reasoning shows that the distribution, *for fixed* n, of $D_n^{-1}(\hat{\theta}_n-\theta_0)$ is arbitrarily close to that of $\hat{\tau}_n$. In order to make the same claim for the *sequence* $\{D_n^{-1}(\hat{\theta}_n-\theta_0)\}$ however, we need the sequence of probability measures Q_n of $D_n^{-1}(\hat{\theta}_n-\theta_0)$ to be *uniformly tight*, which entails that for any $\varepsilon > 0$ there exists a K such that $Q_n(K) > 1-\varepsilon$ for all n. Since the measures Q_n are defined on $\mathbb{R}^k$, this is nothing else than requiring that $D_n^{-1}(\hat{\theta}_n-\theta_0)$ is bounded in probability, or $\mathrm{O}_p(1)$. If this is satisfied, then we may conclude that

$$D_n^{-1}(\hat{\theta}_n-\theta_0)\ \xrightarrow{\mathcal{L}}\ \arg\max_{\tau\in\mathbb{R}^k}\Lambda(\tau)=J^{-1}S. \tag{10}$$

Conditions for this kind of uniform tightness of maximum likelihood estimators are given, e.g., in Basawa and Scott (1983) and Prakasa Rao (1987). Since $\|D_n\|\to 0$, they immediately imply that $\hat{\theta}_n$ is consistent. Its asymptotic distribution is determined by the joint distribution of S and J. Two special cases are considered in the following definition:

Definition 3 *(Jeganathan, 1980) Assume that the sequence of families $\{P_{\theta,n},\theta\in\Theta\}$ have LAQ likelihood ratios at θ_0; then the likelihood ratios are said to be locally asymptotically mixed normal (LAMN), if*

$$\mathcal{L}(S,J)=\mathcal{L}(J^{1/2}Z,J), \tag{11}$$

where Z is a standard normal random vector, independent of J; if in addition to (11), J is non-random, then the likelihood ratios are said to be locally asymptotically normal (LAN).

Under LAN, and if $D_n = n^{-1/2} I_k$, we have the classical result

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \stackrel{\mathcal{L}}{\to} J^{-1/2} Z \sim \mathsf{N}(0, J^{-1}), \tag{12}$$

where J^{-1} is the Cramér-Rao lower bound. Under LAMN, the limit distribution of $D_n^{-1}(\hat{\theta}_n - \theta_0)$ is $\mathsf{N}(0, J^{-1})$ conditionally on J, so that the unconditional distribution is a *mixture* of normals. For both LAN and LAMN, it can easily be shown that the likelihood ratio test statistic for the simple hypothesis $\theta = \theta_0$ converges in distribution to $Z'Z$, which has a $\chi^2(k)$ distribution.

Jeganathan (1980) has shown that a necessary and sufficient condition for LAMN is that the limiting information is an asymptotically ancillary statistic, in the sense that $\mathcal{L}(J | P_{\theta_0 + D_n \tau, n})$ does not depend on τ. Moreover, he has shown that if LAQ holds for all $\theta \in \Theta$, then LAMN (or possibly LAN) is satisfied for almost all $\theta \in \Theta$, i.e., outside a set of Lebesgue measure zero. For example, the Gaussian AR(1) model has LAN likelihood ratios if the autoregressive root is stable, and LAMN LR's if the root is explosive; in case of a unit root, the LR is LAQ but not LAMN or LAN; in that case

$$\mathcal{L}(S, J) = \mathcal{L}\left(\int_0^1 W(t) \mathrm{d}W(t), \int_0^1 W(t)^2 \mathrm{d}t \right), \tag{13}$$

where $W(t)$ is a standard Brownian motion process; hence, this is a special case of a so-called *locally asymptotically Brownian functional* (LABF) likelihood ratio (see Jeganathan, 1988; Phillips, 1989).

Because Definition 1 requires the limiting normed information J to be non-singular, θ will be fully identified. See Rothenberg (1971) for the connection between identification and non-singularity of the information matrix. In the next section, we look at structural reparameterisations of θ which may be globally or locally unidentified, and study the effect of this lack of identification on likelihood ratio test statistics and maximum likelihood estimators.

3 Partially Identified Models

Consider a statistical model $\{P_{\theta,n}, \theta \in \Theta, n = 1, 2, \ldots\}$ which satisfies the following assumptions:

Assumption 1 *Under $P_{\theta_0,n}$, the likelihood ratio sequence $\Lambda_n(\tau) = \ln \mathrm{d}P_{\theta_0 + D_n \tau, n} / \mathrm{d}P_{\theta_0, n}$ is locally asymptotically quadratic and stochastically equicontinuous.*

Assumption 2 *Under $P_{\theta_0,n}$, the sequence of normed and centered maximum likelihood estimators $D_n^{-1}(\hat{\theta}_n - \theta_0)$ exists, is Borel-measurable, and is uniformly tight.*

As discussed in the previous section, these assumptions allow the limiting distribution of $D_n^{-1}(\hat{\theta}_n - \theta_0)$ to be obtained from the limiting likelihood ratio $\Lambda(\tau)$ and the continuous mapping theorem. It will be useful to consider the *normed log-likelihood* (cf. Barndorff-Nielsen and Cox, 1994)

$$\Lambda_n(\theta, \hat{\theta}_n) = \Lambda_n(\theta, \theta_0) - \max_{\theta \in \Theta} \Lambda_n(\theta, \theta_0). \tag{14}$$

Letting $\bar{\Lambda}_n(\tau) \equiv \Lambda_n(\theta_0 + D_n\tau, \hat{\theta}_n)$, we have under Assumptions 1 and 2,

$$\begin{aligned} \bar{\Lambda}_n(\tau) &\xrightarrow{\mathcal{L}} \Lambda(\tau) - \max_{\tau \in \mathbb{R}^k} \Lambda(\tau) \\ &= -\frac{1}{2}(\tau - J^{-1}S)'J(\tau - J^{-1}S) \\ &\equiv \bar{\Lambda}(\tau), \end{aligned} \tag{15}$$

uniformly on compact sets.

Consider now the hypothesis that the parameter vector θ is related to another l-dimensional "structural" parameter vector $\phi \in \Phi \subseteq \mathbb{R}^l$:

$$\mathcal{H}_0 : \theta = h(\phi), \tag{16}$$

where $h : \mathbb{R}^l \mapsto \mathbb{R}^k$ is a continuously differentiable function with derivative matrix $H_\phi = \partial h(\phi)/\partial \phi'$. If, at a particular point $\phi_0 \in \Phi$, H_{ϕ_0} has full column rank, then $\phi_0 \in \Phi$ is fully identifiable, and $\mathcal{H}_0$ entails $k - l$ restrictions on θ. We denote the parameter space under the null hypothesis by

$$\Theta_0 = \{\theta \in \Theta : \theta = h(\phi), \phi \in \Phi\}. \tag{17}$$

The corresponding sequence of parameter spaces for τ is

$$\mathcal{T}_{0,n} \equiv \{\tau \in \mathbb{R}^k : \theta_0 + D_n\tau \in \Theta_0\}. \tag{18}$$

The likelihood ratio test statistic for $\mathcal{H}_0$ and the restricted MLE of θ (if it exists) are given by

$$LR_n \equiv -2 \max_{\theta \in \Theta_0} \Lambda_n(\theta, \hat{\theta}_n) = -2 \max_{\tau \in \mathcal{T}_{0,n}} \bar{\Lambda}_n(\tau), \tag{19}$$

$$\tilde{\theta}_n \equiv \arg\max_{\theta \in \Theta_0} \Lambda_n(\theta, \hat{\theta}_n) = \theta_0 + D_n \arg\max_{\tau \in \mathcal{T}_{0,n}} \bar{\Lambda}_n(\tau). \tag{20}$$

We are interested in the asymptotic behaviour of LR_n and $\tilde{\theta}_n$. Analogously to the previous section, the idea is to derive their properties from convergence of $\bar{\Lambda}_n$ and the continuous mapping theorem. This requires two further assumptions:

Assumption 3 *Under $P_{\theta_0,n}$, the sequence of normed and centered restricted maximum likelihood estimators $D_n^{-1}(\tilde{\theta}_n - \theta_0)$ exists, is Borel-measurable, and is uniformly tight.*

Assumption 4 *There exists a limiting local parameter space $\mathcal{T}_0$, i.e.,*

$$\lim_{n\to\infty} \mathcal{T}_{0,n} = \mathcal{T}_0. \tag{21}$$

Assumption 3 allows us to use weak convergence of $\bar{\Lambda}_n$ on compact subsets, together with the continuous mapping theorem, for deriving the limit distribution of LR_n and $D_n^{-1}(\tilde{\theta}_n - \theta_0)$. The fourth assumption is only related to the sequence of parameter spaces, and thus can be checked without making any probability argument. Observe that no assumption has been made regarding the *shape* of $\mathcal{T}_0$. We now obtain the following result:

Theorem 1 *Let Assumptions 1-4 hold for some θ_0 in the interior of Θ_0. Then, as $n \to \infty$,*

$$LR_n \xrightarrow{\mathcal{L}} \min_{\tau\in\mathcal{T}_0} (\tau - J^{-1}S)'J(\tau - J^{-1}S), \tag{22}$$

$$D_n^{-1}(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \arg\min_{\tau\in\mathcal{T}_0} (\tau - J^{-1}S)'J(\tau - J^{-1}S) \tag{23}$$

Proofs are given in the Appendix. From this theorem, it should be evident that the random function $\bar{\Lambda}(\tau)$ fully determines the asymptotic properties of the statistics of interest. Thus, once one has established the uniform LAQ property of the unrestricted model (together with uniform tightness of the restricted and unrestricted MLE sequences), no new probability arguments have to be made. What remains to be analysed is the shape of the null space $\mathcal{T}_0$, which determines the way in which the asymptotic distributions of LR_n and $\tilde{\theta}_n$ are defined from $\bar{\Lambda}(\tau)$. We now consider two special cases, which differ with respect to the properties of $\mathcal{T}_0$.

3.1 Partially Globally Unidentified Models

Let $B_r(\phi_0)$ denote a ball around a particular point $\phi_0 \in \Phi$ with radius r, such that $B_r(\phi_0) \subset \Phi$ (hence ϕ_0 is an interior point of Φ). We shall call a parameter point ϕ_0 partially globally unidentified if there is some $r > 0$ such that

$$\forall \phi \in B_r(\phi_0) : \text{rank} H_\phi = l_1, \tag{24}$$

where H_ϕ is the $k \times l$ derivative matrix of $h(\phi)$, so that $l_1 \leq \min(k, l)$. Let A be a particular orthogonal $l \times l$ matrix, such that

$$H_{\phi_0} A = [H_1 : 0], \quad A'\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \tag{25}$$

where H_1 is a full column rank matrix of order $m \times l_1$ and ϕ_1 and ϕ_2 are vectors of dimension l_1 and $l_2 = l - l_1$, respectively.

A Taylor series expansion of $h(\phi)$ around ϕ_0 (with $\theta_0 = h(\phi_0)$) gives

$$\begin{aligned} \theta - \theta_0 &= H_{\phi_0}(\phi - \phi_0) + \mathrm{o}(\|\phi - \phi_0\|) \\ &= H_1(\phi_1 - \phi_{1,0}) + \mathrm{o}(\|\phi - \phi_0\|). \end{aligned} \tag{26}$$

We see that in a neigbourhood of ϕ_0, θ only varies with ϕ_1, so that ϕ_2 is not identified, i.e., $\phi_{2,0}$ is not unique. The rank condition (24) implies that Θ_0 has dimension l_1 (in a neigbourhood of θ_0), which in turn implies that H_1 will not depend on the choice of $\phi_{2,0}$ (among all observationally equivalent points). Thus ϕ_2 does not enter (26), so that the remainder term may be replaced by $\mathrm{o}(\|\phi_1 - \phi_{1,0}\|)$.

Let C_n be a sequence of $l_1 \times l_1$ matrices such that $\|C_n\| \to 0$, and chosen such that $H_1 C_n = D_n \bar{H}$ for some $k \times l_1$ matrix $\bar{H}$ of full column rank. Note that if $D_n = n^{-\alpha} I_k$, then $C_n = n^{-\alpha} I_{l_1}$ and $\bar{H} = H_1$. Consider a local sequence $\phi_{1,n} = \phi_{1,0} + C_n \psi$, where ψ has the same relation to ϕ as τ has to θ. We now find

$$\begin{aligned} \tau &= D_n^{-1}(\theta_n - \theta_0) \\ &= D_n^{-1} H_1 C_n C_n^{-1}(\phi_{1,n} - \phi_{1,0}) + D_n^{-1}\mathrm{o}(\|\phi_{1,n} - \phi_{1,0}\|) \\ &= \bar{H}\psi + \mathrm{o}(1), \end{aligned} \tag{27}$$

which means that in the limit, the restricted parameter space of τ is the linear subspace

$$\begin{aligned} \mathcal{T}_0 &= \{\tau \in \mathbb{R}^k : \tau = \bar{H}\psi, \psi \in \mathbb{R}^{l_1}\} \\ &= \{\tau \in \mathbb{R}^k : \bar{H}_\perp' \tau = 0\}. \end{aligned} \tag{28}$$

Combining this with Theorem 1, we find

Theorem 2 *Let the assumptions of Theorem 1 be satisfied, and let ϕ_0 satisfy condition (24). Then, as $n \to \infty$,*

$$\begin{aligned} LR_n &\xrightarrow{\mathcal{L}} S' J^{-1} S - S' \bar{H} (\bar{H}' J \bar{H})^{-1} \bar{H}' S \\ &= S' J^{-1} \bar{H}_\perp (\bar{H}_\perp' J^{-1} \bar{H}_\perp)^{-1} \bar{H}_\perp' J^{-1} S, \end{aligned} \tag{29}$$

$$D_n^{-1}(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \bar{H}(\bar{H}' J \bar{H})^{-1} \bar{H}' S. \tag{30}$$

Corollary 1 *Under the assumptions of Theorem 2 and LAMN,*

$$LR_n \xrightarrow{\mathcal{L}} \chi^2(k - l_1). \tag{31}$$

We see that in globally partially unidentified models, the LAMN condition allows us to use conventional tables for testing hypotheses; the only effect of the partial identification is that the appropriate degrees of freedom is $k - l_1$ rather than $k - l$. In practice, the value of l_1 can be computed either as the rank of the Jacobian matrix $H_\phi = \partial\theta/\partial\phi'$, or as the rank of the observed information on ϕ, which is $H_\phi' J_{\theta,n} H_\phi$, with $J_{\theta,n}$ the observed information on θ; see Boswijk (1995) and Doornik (1995).

3.2 Partially Locally Unidentified Models

A parameter point ϕ_0 is said to be partially locally unidentified, if instead of (24), we have

$$\text{rank}H_{\phi_0} = l_1 < \max_{\phi \in B_r(\phi_0)} \text{rank}H_\phi = l, \tag{32}$$

where the maximal rank[1] l holds almost everywhere in $B_r(\phi_0)$. Thus ϕ_0 is a *singular point* of H_ϕ, see Rothenberg (1971). This implies that the rank of H_ϕ does not equal the dimension of the parameter space Θ_0, which is l.

Let H_1 and ϕ_1 be as defined before. The Taylor series expansion (26) is still valid, but now H_1 does depend on the choice of true value of the locally unidentified parameter ϕ_2. Explicitly, we now have

$$\theta - \theta_0 = H_1(\phi_2)(\phi_1 - \phi_{1,0}) + \text{o}(\|\phi_1 - \phi_{1,0}\|). \tag{33}$$

Thus the dependence of θ on ϕ_2 is non-linear, even in the limit. The following example illustrates this point.

Example 1 *Consider the AR(1) model with constant*

$$X_t = \rho X_{t-1} + \alpha + \varepsilon_t, \tag{34}$$

and define the reduced form parameter vector $\theta = (\rho, \alpha)'$. Consider the "structural" reparameterisation in terms of $\phi = (\rho, \mu)'$:

$$\theta = \begin{pmatrix} \rho \\ \alpha \end{pmatrix} = \begin{pmatrix} \rho \\ \mu(1-\rho) \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \phi_2(1-\phi_1) \end{pmatrix} = h(\phi), \tag{35}$$

which corresponds to the model

$$(X_t - \mu) = \rho(X_{t-1} - \mu) + \varepsilon_t. \tag{36}$$

The derivative of $\theta = h(\phi)$ is

$$H_\phi = \frac{\partial \theta}{\partial \phi'} = \begin{bmatrix} 1 & 0 \\ -\phi_2 & (1-\phi_1) \end{bmatrix}. \tag{37}$$

If $\rho_0 = 1$ (in which case X_t is a random walk), then H_{ϕ_0} has rank 1, and μ_0 is not identified. However, the first column of H_{ϕ_0} obviously depends on ϕ_2. We may write this as

$$\theta - \theta_0 = \begin{pmatrix} 1 \\ -\phi_2 \end{pmatrix} (\phi_1 - \phi_{1,0}) = H_1(\phi_2)(\phi_1 - \phi_{1,0}). \tag{38}$$

[1] The maximal rank could also be a number l^* between l_1 and l, which would combine local and global identification problems; for simplicity this is not considered explicitly here.

The fact that ϕ_2 is not identified implies that we cannot in general expect its estimator $\tilde{\phi}_{2,n}$ to converge in probability to a unique limit. In some cases it will actually diverge, in the sense that it has to be divided by a power of n in order to have a non-degenerate limiting distribution. In the above example, it is well known that $\tilde{\theta}_{1,n} - \theta_{1,0} = \mathrm{O}_p(n^{-1})$ and $\tilde{\theta}_{2,n} - \theta_{2,0} = \mathrm{O}_p(n^{-1/2})$; since $\tilde{\phi}_{1,n} - \phi_{1,0} = \tilde{\theta}_{1,n} - \theta_{1,0}$, it follows that $\tilde{\phi}_{2,n} = -(\tilde{\theta}_{2,n} - \theta_{2,0})/(\tilde{\theta}_{1,n} - \theta_{1,0}) = \mathrm{O}_p(n^{1/2})$.

Therefore, consider the sequences $\phi_{1,n} = \phi_{1,0} + C_n\psi$ and $\phi_{2,n} = B_n\gamma$, where $\|C_n\| \to 0$, and were C_n and B_n are chosen such that

$$\lim_{n\to\infty} D_n^{-1} H_1(B_n\gamma)C_n = \bar{H}(\gamma), \tag{39}$$

where $\bar{H}(\gamma)$ is a $k \times l_1$ matrix function of full column rank (for all γ). Via similar derivations as in (27), this leads to the following restricted parameter space for τ (with $l_1 + l_2 = l$):

$$\mathcal{T}_0 = \{\tau \in \mathbb{R}^k : \tau = \bar{H}(\gamma)\psi, \psi \in \mathbb{R}^{l_1}, \gamma \in \mathbb{R}^{l_2}\}. \tag{40}$$

In contrast with the previous subsection, this is not a linear subspace. However, it is linear if we first fix γ. Maximizing over ψ yields the same results as in Theorem 2, but with LR_n, $\tilde{\theta}_n$ and $\bar{H}$ all functions of γ. This implies

Theorem 3 *Let the assumptions of Theorem 1 be satisfied, and let ϕ_0 satisfy condition (32). Then, as $n \to \infty$,*

$$LR_n \xrightarrow{\mathcal{L}} \min_{\gamma\in\mathbb{R}^{l_2}} S'J^{-1}\bar{H}(\gamma)_\perp[\bar{H}(\gamma)'_\perp J^{-1}\bar{H}(\gamma)_\perp]^{-1}\bar{H}(\gamma)'_\perp J^{-1}S, \tag{41}$$

$$D_n^{-1}(\tilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} \bar{H}(\bar{\gamma})[\bar{H}(\bar{\gamma})'J\bar{H}(\bar{\gamma})]^{-1}\bar{H}(\bar{\gamma})'S, \tag{42}$$

where $\bar{\gamma}$ is the minimand of (41).

Corollary 2 *Under the assumptions of Theorem 3 and LAMN,*

$$LR_n \xrightarrow{\mathcal{L}} \min_{\gamma\in\mathbb{R}^{l_2}} Z'A(\gamma)A(\gamma)'Z, \tag{43}$$

where Z is a standard normal vector and $A(\gamma) = J^{-1/2}\bar{H}(\gamma)_\perp[\bar{H}(\gamma)'_\perp J^{-1}\bar{H}(\gamma)_\perp]^{-1/2}$.

For fixed γ, the distribution of $A(\gamma)'Z$ is $\mathsf{N}(0, I_{k-l_1})$, which shows that the limiting distribution in (43) is the minimum over $\chi^2(k-l_1)$ random variables, so that it is stochastically dominated by the $\chi^2(k-l_1)$ distribution. Thus, we can use critical values from this χ^2 distribution to construct a conservative test based on the LR statistics. In the next section we shall encounter a number of examples where the right-hand side of (43) can in fact be bounded by a $\chi^2(k-l)$ random variable, which allows as to use tighter bounds for the appropriate critical values. No such results can however be established outside the LAMN framework.

4 Applications

In this section we shall apply the analysis developed above to a number of econometric models which are partially unidentified. We shall focus on the behaviour of the likelihood ratio statistic, although the asymptotic distribution of the restricted MLE can usually be obtained as a by-product. In all cases, the unrestricted reduced form model will be a (multivariate) Gaussian linear regression (possibly with lagged dependent regressors). For simplicity, we shall always assume that the error variance (matrix) is known, which implies that the likelihood ratio is exactly quadratic (the results can, however, be extended to the case of unknown variance parameters). This reduces proving LAQ to proving convergence in distribution of the normed score vector and observed information matrix. Also, stochastic equicontinuity will be trivially satisfied, as well as tightness of the unrestricted MLE. Assumption 3 (tightness of the restricted MLE) will be assumed rather than proved explicitly.

4.1 Limited Information Simultaneous Equations

Consider the multivariate linear regression model

$$Y_t = \Pi' x_t + V_t, \quad t = 1, \ldots, n, \tag{44}$$

where Y_t is a g-vector of dependent random variables, x_t is an m-vector of non-random explanatory variables, and V_t is a g-vector of i.i.d. $\mathsf{N}(0, \Omega)$ disturbances, with $\Omega > 0$ known. Let $\theta = \mathrm{vec}\Pi$ (of order $k = gm$); the unrestricted parameter space is $\Theta = \mathbb{R}^k$. The log-likelihood ratio is given by

$$\Lambda_n(\theta_0 + \Delta_n \tau, \theta_0) = \tau' D_n' \mathrm{vec}\left(\sum_{t=1}^n x_t V_t' \Omega^{-1}\right) - \frac{1}{2}\tau' D_n' \left(\Omega^{-1} \otimes \sum_{t=1}^n x_t x_t'\right) D_n \tau. \tag{45}$$

Assume that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^n x_t x_t' = M > 0, \tag{46}$$

which implies

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n x_t V_t' \Omega^{-1} \xrightarrow{\mathcal{L}} (\Omega^{-1} \otimes M)^{1/2} Z \sim \mathsf{N}(0, \Omega^{-1} \otimes M), \tag{47}$$

where Z is a standard normal vector. Thus the likelihood ratio is LAN, with $D_n = n^{-1/2} I_k$ and $J = (\Omega^{-1} \otimes M)$:

$$\Lambda_n(\theta_0 + D_n \tau, \theta_0) \xrightarrow{\mathcal{L}} \tau' J^{1/2} Z - \frac{1}{2}\tau' J \tau = \Lambda(\tau). \tag{48}$$

Consider the following simultaneous model in structural form:

$$Y_{1t} = \beta' Y_{2t} + \gamma' x_{1t} + U_{1t}, \tag{49.a}$$

$$Y_{2t} = \Pi_{12}' x_{1t} + \Pi_{22}' x_{2t} + V_{2t}, \tag{49.b}$$

where Y_t and x_t are partitioned such that Y_{1t} is scalar, and x_{it} is of order $m_i, i = 1, 2$; Π is partitioned conformably. The reduced form of this model implies:

$$\Pi = \begin{bmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22} \end{bmatrix} = \begin{bmatrix} \Pi_{12}\beta + \gamma & \Pi_{12} \\ \Pi_{22}\beta & \Pi_{22} \end{bmatrix}. \tag{50}$$

Define $\theta = (\pi'_{11}, \pi'_{21}, (\text{vec}\Pi_{12})', (\text{vec}\Pi_{22})')'$ and $\phi = (\beta', \gamma', (\text{vec}\Pi_{12})', (\text{vec}\Pi_{22})')'$. The derivative matrix of $\theta = h(\phi)$ is

$$H_\phi = \begin{bmatrix} \Pi_{12} & I_{m_1} & (\beta' \otimes I_{m_1}) & 0 \\ \Pi_{22} & 0 & 0 & (\beta' \otimes I_{m_2}) \\ 0 & 0 & (I_{g-1} \otimes I_{m_1}) & 0 \\ 0 & 0 & 0 & (I_{g-1} \otimes I_{m_2}) \end{bmatrix}. \tag{51}$$

It is easily checked that H_{ϕ_0} is of full column rank if and only if Π_{22} is, which is the celebrated *rank condition.* This can only be satisfied if $m_2 \geq (g-1)$ (the *order condition*) which we assume to hold in the sequel.

As a polar case, suppose that $\Pi_{12,0} = 0$ and $\Pi_{22,0} = 0$. Then the first $g-1$ columns of H_{ϕ_0} are zero, whereas the remaining columns are linearly independent, so that β is completely (locally) unidentified, whereas γ, Π_{12} and Π_{22} are fully identified. The analysis of the previous section can now be directly applied with $C_n = n^{-1/2} I_{m_1+(g-1)m}$, $B_n = I_{g-1}$ and

$$\bar{H}(\beta)_\perp = \begin{pmatrix} 1 \\ -\beta \end{pmatrix} \otimes E, \tag{52}$$

where $E = [0 : I_{m_2}]'$, of order $m \times m_2$. Let $\delta = (1, -\beta')'$ and $a(\beta) = (\delta'\Omega\delta)^{-1/2}\delta'\Omega^{1/2}$, a unit length vector, and $B = M^{-1/2}E(E'M^{-1}E)^{-1/2}$, an $m \times m_2$ semi-orthogonal matrix. Then we have (with $A(\beta)$ as in Corollary 2)

$$A(\beta)'Z = (a(\beta)' \otimes B')Z = \text{vec}B'Xa(\beta), \tag{53}$$

where X is an $m \times g$ standard normal matrix such that $Z = \text{vec}X$. Letting $X_1 = B'X$, an $m_2 \times g$ standard normal matrix, we thus find that (43) reduces to

$$LR_n \xrightarrow{\mathcal{L}} \min_{a'a=1} a'X_1'X_1a = \lambda_{\min}(X_1'X_1), \tag{54}$$

where $\lambda_{\min}(.)$ denotes the minimal eigenvalue of the argument.

Using the same method, it can be shown that (54) also holds if $\text{rank}\Pi_{22} = r$, but with X_1 of order $(m_2 - r) \times (g - r)$. Thus, if the rank condition is satisfied, then LR_n has an asymptotic $\chi^2(m_2 - g - 1)$. Letting $\chi^2_\alpha(m_2 - g - 1)$ denote the $100(1-\alpha)$th percentile of this distribution, it can be shown (see Boswijk, 1994) that for all $r \leq g - 1$,

$$\mathsf{P}\{\lambda_{\min}(X_1'X_1) > \chi^2_\alpha(m_2 - g - 1)\} \leq \alpha. \tag{55}$$

This implies that the *size* (i.e., the maximum type I error probability) of the LR test converges to α if χ^2 critical values are used.

4.2 Testing for Cointegration in Vector Autoregressions

Consider the first-order vector autoregression in p dimensions, written in error correction form:

$$\Delta X_t = \Pi X_{t-1} + \varepsilon_t, \quad t = 1, \dots, n, \tag{56}$$

where $\{\varepsilon_t\}$ is an i.i.d. $\mathsf{N}(0, \Omega)$ sequence and X_0 is fixed. Assume that rank$\Pi = r$, $0 \le r \le p$, and that the characteristic equation $|(1-z)I_p - \Pi z| = 0$ has exactly $p - r$ roots equal to one and all other roots outside the unit circle. As shown by Johansen (1991), this implies that X_t is cointegrated of order (1,1): letting $\Pi = \alpha\beta'$, where α and β are $p \times r$ matrices of full column rank, $\beta' X_t$ is stationary, whereas $\beta'_\perp X_t$ is integrated of order 1.

Since Ω is considered as fixed, the log-likelihood ratio is given, for $\theta = \text{vec}(\Pi')$, by

$$\Lambda_n(\theta_0 + D_n\tau, \theta_0) = \tau' S_n - \frac{1}{2}\tau' J_n \tau, \tag{57}$$

where

$$S_n = D_n' \text{vec}\left(\sum_{t=1}^{n} X_{t-1}\varepsilon_t' \Omega^{-1}\right), \tag{58}$$

$$J_n = D_n' \left(\Omega^{-1} \otimes \sum_{t=1}^{n} X_{t-1}X_{t-1}'\right) D_n. \tag{59}$$

Defining

$$D_n' = \begin{bmatrix} n^{-1/2}(I_p \otimes \beta') \\ n^{-1}(\alpha' \otimes \beta_\perp') \\ n^{-1}(\alpha_\perp' \Omega \otimes \beta_\perp') \end{bmatrix}, \tag{60}$$

it can be shown (see Johansen, 1991) that

$$\begin{aligned} J_n &\xrightarrow{\mathcal{L}} \begin{bmatrix} J_{11} & 0 & 0 \\ 0 & J_{22} & 0 \\ 0 & 0 & J_{33} \end{bmatrix} \\ &= \begin{bmatrix} \Omega^{-1} \otimes \Sigma_{\beta\beta} & 0 & 0 \\ 0 & (\alpha'\Omega^{-1}\alpha) \otimes \int_0^1 W(t)W(t)'\text{d}t & 0 \\ 0 & 0 & (\alpha_\perp'\Omega\alpha_\perp) \otimes \int_0^1 W(t)W(t)'\text{d}t \end{bmatrix}, \end{aligned} \tag{61}$$

where $\Sigma_{\beta\beta} = \text{var}[\beta' X_t]$, and $W(t)$ is a $(p-r)$-dimensional vector Brownian motion process with variance matrix $\alpha_\perp' \Omega \alpha_\perp$. Furthermore,

$$S_n \xrightarrow{\mathcal{L}} \begin{pmatrix} J_{11}^{1/2} Z_1 \\ J_{22}^{1/2} Z_2 \\ \int_0^1 W(t)\text{d}W(t)' \end{pmatrix}, \tag{62}$$

where $Z = (Z_1', Z_2')'$ is a standard normal vector of dimension $pr + r(p-r)$, independent of $W(t)$. It follows that $\Lambda(\tau) = \sum_{i=1}^{3} \Lambda_i(\tau_i)$, where Λ_1 is LAN, Λ_2 is LAMN, and Λ_3 is LABF.

Consider now the reparameterisation

$$\begin{aligned} \Delta X_t &= \alpha\beta' X_{t-1} + \varepsilon_t \\ &= \alpha[I_r : -B']X_{t-1} + \varepsilon_t, \end{aligned} \tag{63}$$

where α is a $p \times r$ matrix of error correction coefficients, and $\beta = [I_r : -B']'$ is a $p \times r$ matrix of cointegrating vectors, so that B is of order $(p-r) \times r$. Letting $\phi = (\text{vec}(\alpha')', \text{vec}(B)')'$, and recalling that $\theta = \text{vec}(\Pi') = h(\phi)$, we have

$$H_\phi = \left[I_p \otimes \begin{pmatrix} I_r \\ -B \end{pmatrix} : -\alpha \otimes \begin{pmatrix} 0 \\ I_{n-r} \end{pmatrix} \right] = \left[(I_p \otimes \beta) : -(\alpha \otimes E) \right], \tag{64}$$

where $E = [0 : I_{p-r}]'$, of order $p \times (p-r)$. Given the normalization of β, the first pr columns of H_{ϕ_0} always have full rank, so that α is always identified (since it equals the first r columns of Π). However, if α_0 has less than full column rank, then part of B will not be identified.

Consider again the polar case where $\alpha_0 = 0$, so that $\Pi_0 = 0$ and the true cointegrating rank is zero. In (60)-(62), this implies that $\alpha_\perp = \beta_\perp = I_p$, $D_n = n^{-1}\Omega \otimes I_p$, and

$$(S, J) = \left(\text{vec} \int_0^1 W(t)\text{d}W(t)\ ,\ \Omega \otimes \int_0^1 W(t)W(t)'\text{d}t \right), \tag{65}$$

where $W(t)$ is a p-vector Brownian motion process with variance matrix Ω. Furthermore, $\bar{H}(B)_\perp = I_p \otimes (B : I_{p-r})'$, which after some manipulations implies

$$LR_n \xrightarrow{\mathcal{L}} \min_{C'C = I_{p-r}} C' \left[\left(\int_0^1 B(t)\text{d}B(t)' \right)' \left(\int_0^1 B(t)B(t)'\text{d}t \right)^{-1} \left(\int_0^1 B(t)\text{d}B(t)' \right) \right] C, \tag{66}$$

where $B(t)$ is a *standard* p-vector Brownian motion process. Thus the limit distribution of LR_n is characterised by the sum of the $(p-r)$ smallest eigenvalues of the matrix in square brackets in (66).

It can again be shown that if $\text{rank}\alpha_0 = s$ with $0 \leq s \leq r$, (66) still holds with $B(t)$ of dimension $p - s$; if $s = r$ this yields the trace of the matrix in square brackets. Quantiles of the distribution of this trace (for various values of $(p-r)$) are tabulated in Johansen (1988) and are used as critical values for the LR test, since this distribution holds almost everywhere in the restricted parameter space. However, unlike the example in the previous section, the distributions for $s < r$ are *not* stochastically dominated by the distribution for $s = r$, so that the asymptotic size of the test is *not* equal to α if the critical values for $s = r$ are used. The crucial difference with the previous example is that the likelihood ratio here is not LAMN or LAN, so that not only the distribution of J and S, but also that of $J^{-1/2}S$ changes with different values in the parameter space. A sequential testing procedure to control the asymptotic size was proposed by Johansen (1992). In this procedure, the hypothesis $\mathcal{H}_r : \text{rank}(\Pi) \leq r$ is rejected only if all subhypotheses $\mathcal{H}_s, s \leq r$ are rejected at their respective critical values.

4.3 Testing Structural Hypotheses on Cointegrating Vectors

Consider the same model as in the previous subsection, but now with the $(p-r)$ unit root restrictions implied by (63) imposed. That is, the *unrestricted* model is now formulated in terms of α and B. We consider a true value α_0 with full column rank, so that the problems mentioned above do not arise. Furthermore, it will simplify the analysis if we consider $B_0 = 0$; this can always be accomplished by suitable rotation of X_t. Under these assumptions, it can be checked from (60) and (64) that

$$D_n \begin{pmatrix} I_{pr+r(p-r)} \\ 0 \end{pmatrix} = H_{\phi_0} \begin{bmatrix} n^{-1/2} I_{pr} & 0 \\ 0 & n^{-1} I_{r(p-r)} \end{bmatrix}. \tag{67}$$

This implies that the limiting likelihood ratio for the cointegration model is given by

$$\begin{aligned} \Lambda(\tau) &= \tau_1' J_{11}^{1/2} Z_1 - \frac{1}{2}\tau_1' J_{11}\tau_1 \\ &\quad + \tau_2' J_{22}^{1/2} Z_2 - \frac{1}{2}\tau_2' J_{22}\tau_2, \end{aligned} \tag{68}$$

where J_{ii}, $i = 1, 2$, are as given in (61). Thus, the consequence of restricting the model is that the third (LABF) part of the likelihood ratio vanishes, so that the resulting likelihood ratio is LAMN.

Consider the hypothesis

$$\mathcal{H}_0 : \text{rank}(R'\beta) \le r - 1, \tag{69}$$

where R is a known $p \times m$ matrix of full column rank, with $m > r - 1$ (since otherwise $\mathcal{H}_0$ is not testable). The hypothesis entails that there is at least one vector β_1 in the cointegrating space $\text{sp}(\beta)$ satisfying $R'\beta_1 = 0$; this can be one of the columns of $\beta = [I_r : -B']'$, but also a more general linear combination of the columns.

Letting $G = R_\perp$, we find that $\beta_1 = G\gamma$, with $\gamma \in \mathbb{R}^{p-m}$. Assume that the first r components of X_t are ordered in such a way that the first component of β_1 may be normalized to one, and that this corresponds to the normalization $\gamma' = (1, \delta')'$. Then we have

$$\beta_1 = \begin{pmatrix} 1 \\ G_1\delta \\ G_2\delta \end{pmatrix} \begin{matrix} 1 \\ r-1 \\ p-r \end{matrix}, \tag{70}$$

where $\delta \in \mathbb{R}^{p-m-1}$ and $G = \text{diag}(1, [G_1' : G_2'])'$. Let $B = [b_1 : B_2]$, with b_1 a $(p-r)$-vector. Since β_1 is a linear combination of $\beta = [I_r : -B']'$, it follows that (70) implies $b_1 + B_2 G_1 \delta = G_2 \delta$, or

$$B = [b_1 : B_2] = [(G_2 - B_2 G_1)\delta : B_2]. \tag{71}$$

Thus, with $\theta = (\text{vec}(\alpha')', b_1', \text{vec}(B_2)')'$ and $\phi = (\text{vec}(\alpha')', \delta', \text{vec}(B_2)')'$, we have

$$H_\phi = \frac{\partial\theta}{\partial\phi'} = \begin{bmatrix} I_{pr} & 0 & 0 \\ 0 & (G_2 - B_2 G_1) & -(\delta' G_1' \otimes I_{p-r}) \\ 0 & 0 & I_{(p-r)(r-1)} \end{bmatrix}. \tag{72}$$

Since $B_0 = 0$, the column rank of H_{ϕ_0} is determined by $s = \text{rank}G_2$: if $s = p - m - 1$, then ϕ_0 is fully identified, and the likelihood ratio test has a limiting $\chi^2(m - r + 1)$ distribution. At the other extreme, if $G_2 = 0$, then δ is completely unidentified, and we find

$$\bar{H}(\delta)_\perp = \left[\begin{array}{c} 0 \\ \left(\begin{array}{c} 1 \\ G_1\delta \end{array} \right) \otimes I_{p-r} \end{array} \right] = \left[\begin{array}{c} 0 \\ \bar{G}_1\gamma \otimes I_{p-r} \end{array} \right], \tag{73}$$

where $\gamma = (1, \delta')'$ and $\bar{G}_1 = \text{diag}(1, G_1)$. From this is can again be shown that, analogously to (54),

$$LR_n \xrightarrow{\mathcal{L}} \min_{a'a=1} a'X_1'X_1 a = \lambda_{\min}(X_1'X_1), \tag{74}$$

where X_1 is now a $(p-r) \times (p-m)$ standard normal matrix. For general $s = \text{rank}G_2$, the same holds with X_1 of dimension $(p-r-s) \times (p-m-s)$, so that (74) also includes the fully identified case where $s = p - m - 1$ and the limit distribution is $\chi^2(m - r + 1)$. A result analogous to (55) applies to the present case.

5 Concluding Remarks

In this paper we have analyzed a particular class of partially identified models, viz. models with fully identified reduced form parameters. As we have shown, within this class the behaviour of the restricted maximum likelihood estimate and likelihood ratio test statistic can be analyzed in two steps: first establish that the unrestricted model has locally asymptotically quadratic likelihood ratios, and then determine the shape of the limiting local parameter space. The examples have illustrated this approach, and have suggested that under LAN or LAMN, an asymptotic size α LR test is obtained if critical values for the fully identified case are used; outside the LAMN class such results cannot be established, so special care has to be taken of the unidentified subsets of the parameter space.

Although the chosen setup is quite general, it excludes models with no fully identified reduced form parameters. An example of such a model is an ARMA model with a common factor; in that case the AR and MA parameters are not identified, and cannot be related to a fully identified *finite dimensional* parameter vector (although the *infinite order* autoregressive representation is unique). Another class of models which have recently attracted much attention are time series models with a structural break in the parameters at an unknown break point, see e.g. Andrews (1993), and Andrews and Ploberger (1994). In such models, the break point parameter is globally unidentified under the null hypothesis, but the analysis of Section 3.1 cannot be applied, because at the true parameter value the unrestricted model is unidentified as well.

Appendix

Proof of Theorem 1: Let Q_n and $\tilde{Q}_n$ be the probability measures of $D_n^{-1}(\hat{\theta}_n - \theta_0)$ and $D_n^{-1}(\tilde{\theta}_n - \theta_0)$, respectively; similarly, let Q and $\tilde{Q}$ be the probability measures of $J^{-1}S$ and $\arg\min_{\tau\in\mathcal{T}_0} \bar{\Lambda}(\tau)$, respectively. For any $\varepsilon > 0$, fix K_ε such that each of these four measures (for all n) assign a probability greater than $1-\varepsilon$ to K_ε. Let I_ε be a binary random variable with $\mathsf{P}\{I_\varepsilon = 0\} = 1 - \mathsf{P}\{I_\varepsilon = 1\} = \varepsilon$, such that $I_\varepsilon = 0$ if any of the above (sequences of) random variables have a realisation outside K_ε. What needs to be proved first is that the sequence of parameter spaces $\mathcal{T}_{0,n}$ can essentially be replaced by its limit $\mathcal{T}_0$, i.e.,

$$\left|I_\varepsilon \min\nolimits_{\tau\in\mathcal{T}_{0,n}\cap K_\varepsilon} \bar{\Lambda}_n(\tau) - I_\varepsilon \min\nolimits_{\tau\in\mathcal{T}_0\cap K_\varepsilon} \bar{\Lambda}_n(\tau)\right| \xrightarrow{P} 0, \tag{A.1}$$

$$\left\|I_\varepsilon \arg\min\nolimits_{\tau\in\mathcal{T}_{0,n}\cap K_\varepsilon} \bar{\Lambda}_n(\tau) - I_\varepsilon \arg\min\nolimits_{\tau\in\mathcal{T}_0\cap K_\varepsilon} \bar{\Lambda}_n(\tau)\right\| \xrightarrow{P} 0. \tag{A.2}$$

This follows from Assumption 4, together with (4), which is implied by Assumption 1.

Next, uniform weak convergence of $\bar{\Lambda}_n(\tau)$ on K_ε implies, via the continuous mapping theorem,

$$I_\varepsilon \min_{\tau\in\mathcal{T}_{0,n}\cap K_\varepsilon} \bar{\Lambda}_n(\tau) = I_\varepsilon \min_{\tau\in\mathcal{T}_0\cap K_\varepsilon} \bar{\Lambda}_n(\tau) + \mathrm{o}_p(1) \xrightarrow{\mathcal{L}} I_\varepsilon \min_{\tau\in\mathcal{T}_0\cap K_\varepsilon} \bar{\Lambda}(\tau), \tag{A.3}$$

and the analogous result for the minimand. Since ε can be made arbitrarily small, it follows that the left-hand side of (A.3) has the same distribution as LR_n, and the right-hand side limit has the same distribution as $\min_{\tau\in\mathcal{T}_0} \bar{\Lambda}(\tau)$. This proves (22); (23) follows analogously.

Proof of Theorem 2: Follows from Theorem 1 and the linearity of the null space $\mathcal{T}_0$ in (28).

Proof of Theorem 3: Follows from Theorem 1 and the shape of the null space $\mathcal{T}_0$ in (40).

References

Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61, 821-856.

Andrews, D.W.K and W. Ploberger (1994), "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative", *Econometrica*, 62, 1383-1414.

Barndorff-Nielsen, O.E. and D.R. Cox (1994), *Inference and Asymptotics.* London: Chapman & Hall.

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

Billingsley, P. (1968), *Convergence of Probability Measures.* New York: John Wiley & Sons.

Boswijk, H.P. (1994), "Testing Identifiability of Cointegrating Vectors", Working paper, University of Amsterdam.

Boswijk, H.P. (1995), "Identifiability of Cointegrated Systems", Tinbergen Institute Discussion Paper TI 95-78.

Doornik, J.A., (1995), "Testing General Restrictions on the Cointegrating Space", Working paper, Nuffield College, University of Oxford.

Dufour, J.-M. (1994), "Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models and Cointegration", Invited paper presented at the Econometric Society European Meeting, Maastricht.

Jeganathan, P. (1980), "An Extension of a Results of L. Le Cam Concerning Asymptotic Normality", *Sankhya A*, 42, 146-160.

Jeganathan, P. (1988), "Some Aspects of Asymptotic Theory with Application to Time Series Models", Technical Report #166, University of Michigan, forthcoming in *Econometric Theory.*

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, 59, 1551-1580.

Johansen, S. (1992), "Determination of Cointegration Rank in the Presence of a Linear Trend", *Oxford Bulletin of Economics and Statistics*, 54, 383-397.

Johansen, S. and K. Juselius (1992), "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and UIP for UK", *Journal of Econometrics*, 53, 211-244.

Le Cam, L. (1960), "Locally Asymptotically Normal Families of Distributions", *University of California Publications in Statistics*, 3, 37-89.

Le Cam, L. and G.L. Yang (1990), *Asymptotics in Statistics. Some Basic Concepts.* New York: Springer-Verlag.

Newey, W.K. (1991), "Uniform Convergence in Probability and Stochastic Equicontinuity", *Econometrica*, 59, 1161-1167.

Phillips, P.C.B. (1988), "Multiple Regression with Integrated Time Series", *Contemporary Mathematics*, 80, 79-105.

Phillips, P.C.B. (1989), "Partially Identified Econometric Models", *Econometric Theory*, 5, 181-240.

Prakasa Rao, B.L.S. (1987), *Asymptotic Theory of Statistical Inference.* New York: John Wiley & Sons.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica*, 39, 577-591.

Saikkonen, P. (1993), "Continuous Weak Convergence and Stochastic Equicontinuity Results for Integrated Processes with an Application to the Estimation of a Regression Model", *Econometric Theory*, 9, 155-188.