Econometrica, Vol. 65, No. 6 (November, 1997), 1365–1387

SOME IMPOSSIBILITY THEOREMS IN ECONOMETRICS WITH APPLICATIONS TO STRUCTURAL AND DYNAMIC MODELS

BY JEAN-MARIE DUFOUR[1]

General characterizations of valid confidence sets and tests in problems which involve locally almost unidentified (LAU) parameters are provided and applied to several econometric models. Two types of inference problems are studied: (i) inference about parameters which are not identifiable on certain subsets of the parameter space, and (ii) inference about parameter transformations with discontinuities. When a LAU parameter or parametric function has an unbounded range, it is shown under general regularity conditions that any valid confidence set with level $1-\alpha$ for this parameter must be unbounded with probability close to $1-\alpha$ in the neighborhood of nonidentification subsets and will have a nonzero probability of being unbounded under any distribution compatible with the model: no valid confidence set which is almost surely bounded does exist. These properties hold even if "identifying restrictions" are imposed. Similar results also obtain for parameters with bounded ranges. Consequently, a confidence set which does not satisfy this characterization has zero coverage probability (level). This will be the case in particular for Wald-type confidence intervals based on asymptotic standard errors. Furthermore, Wald-type statistics for testing given values of a LAU parameter cannot be pivotal functions (i.e., they have distributions which depend on unknown nuisance parameters) and even cannot be usefully bounded over the space of the nuisance parameters. These results are applied to several econometric problems: inference in simultaneous equations (instrumental variables (IV) regressions), linear regressions with autoregressive errors, inference about long-run multipliers and cointegrating vectors. For example, it is shown that standard "asymptotically justified" confidence intervals based on IV estimators (such as two-stage least squares) and the associated "standard errors" have zero coverage probability, and the corresponding t statistics have distributions which cannot be bounded by any finite set of distribution functions, a result of interest for interpreting IV regressions with "weak instruments." Furthermore, expansion methods (e.g., Edgeworth expansions) and bootstrap techniques cannot solve these difficulties. Finally, in a number of cases where Wald-type methods are fundamentally flawed (e.g., IV regressions with poor instruments), it is observed that likelihood-based methods (e.g., likelihood-ratio tests and confidence sets) combined with projection techniques can easily yield valid tests and confidence sets.

KEYWORDS: Cointegration, confidence set, dynamic model, finite-sample theory, identification, structural model, testing, weak instruments.

[1] The author thanks Josh Angrist, Russell Davidson, Jean-Pierre Florens, John Galbraith, Ronald Gallant, Nick Kiefer, Thomas Lemieux, Helmut Lütkepohl, Doug McManus, Pierre Perron, Eric Renault, George Tauchen, a co-editor and three anonymous referees for several useful comments and discussions. Earlier versions of this paper were presented at the European Econometric Society Meetings in Maastricht (invited session, August 1994), the Humboldt University (Berlin), the Université de Lausanne, Cornell University, the Research Triangle Institute, and the Canadian Econometric Study Group. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Government of Québec (Fonds FCAR). Part of this research was also carried out at the Université Libre de Bruxelles (Institut de Statistique) and the Humboldt University Berlin (within Sonderforschungsbereich 373).

1. INTRODUCTION

A COMMON PROBLEM IN STATISTICS and econometrics consists in building confidence sets for the parameters of a statistical model. Since they report all parameter values acceptable at a given level (see Lehmann (1986)), confidence sets give considerably more information than significance tests for particular parameter values. For a scalar parameter θ, a confidence set often takes the form of an interval, such as $[\hat{\theta} - c_1 \hat{\sigma}_{\hat{\theta}},\ \hat{\theta} + c_2 \hat{\sigma}_{\hat{\theta}}]$ where $\hat{\theta}$ is an estimate of θ, $\hat{\sigma}_{\hat{\theta}}$ a "standard error," and c_1, c_2 are constants obtained from the distribution of $(\hat{\theta} - \theta)/\hat{\sigma}_{\hat{\theta}}$ to yield the desired level. This approach is justified when the latter distribution does not depend on unknown nuisance parameters or can be approximated by such a distribution (e.g., the $N(0,1)$). When θ is a vector, one would typically find a "covariance matrix" for $\hat{\theta}$ and build a confidence ellipsoid. Below we call such confidence sets *Wald-type* confidence sets. More generally, confidence set construction depends on the availability of *pivotal functions* (i.e., functions $\phi(Y, \theta)$ of both the data Y and the parameter vector θ whose distributions do not depend on unknown parameters), or at least of *boundedly pivotal functions* (i.e., functions $\phi(Y, \theta)$ whose distribution can be *bounded* over the parameter space by probabilities in the *open* interval (0, 1)). The notion of pivotal quantity was introduced by Fisher (1934) and lies at the heart of "classical" hypothesis testing and confidence set methods.

Many models in econometrics are not identified over the full parameter space, i.e., they contain subsets of observationally equivalent parameter values. Prominent examples include structural models, such as simultaneous equation models and errors-in-variables models, various nonlinear regression models, ARMA models (univariate or multivariate), and models of cointegrating relations. For general discussions of identification, see Rothenberg (1971), Bowden (1973), Fisher (1976), Deistler and Seifert (1978), Hsiao (1983), Breusch (1986), Heckman and Robb (1986), and Prakasa Rao (1992). Problems similar to nonidentification also occur when a discontinuous transformation of a parameter vector (e.g., a parameter ratio) is considered.

The typical approach to identification problems is to assume them away by imposing "identification restrictions" and then derive the asymptotic theory for the fully identified case. Although this leads to distributional simplifications, it also hides many important complications. "Identifiability restrictions" can be very real and rule out plausible data distributions: in no way can they be taken as granted (see Sims (1980)). Furthermore, both finite sample and asymptotic distributions for estimators and tests can be strongly affected if identifiability conditions are not satisfied (see Sargan (1983), Phillips (1984, 1985, 1989), Hillier (1990), Choi and Phillips (1992), Staiger and Stock (1997), McManus, Nankervis, and Savin (1994)), which suggests that asymptotic approximations can be very unreliable under conditions close to nonidentification. In particular, when appropriate identification conditions do not hold, certain parameters of interest (although not necessarily all of them) may not be "estimable" (see Bowden

(1973), Bunke and Bunke (1974), Deistler and Seifert (1978), Phillips (1989), and Hillier (1990)), and hypotheses about possibly nonidentifiable parameters (although not necessarily all hypotheses of interest) may not be "testable" in the sense that they are not "refutable" (see the discussion of Breusch (1986)). For equations estimated by instrumental variables (IV) methods, the distributional complications associated with (near) nonidentification are especially relevant because of the serious possibility of "weak instruments," a problem which has received renewed attention recently; see, e.g., Nelson and Startz (1990a, 1990b), Buse (1992), Maddala and Jeong (1992), Angrist and Krueger (1994), Staiger and Stock (1997), Bound, Jaeger, and Baker (1995), and Hall, Rudebusch, and Wilcox (1996).

Although the available analytical results indicate that distributions of IV-based estimators and test statistics can be strongly affected in nonidentified models, they do not throw much light on the properties of confidence procedures, in particular on whether we can bound the distributions of test statistics to obtain valid tests and confidence sets, *even if identifying restrictions are imposed*. The main purpose of this paper is to throw more light on these issues by extending *finite-sample* results and methods due to Gleser and Hwang (1987, henceforth GH) and Koschat (1987) in a number of special problems. For inference on errors-in-variables models, principal components and ratios of regression parameters, GH showed that no valid confidence interval for a parameter can have finite expected length if this parameter is not identifiable on a subset of the parameter space. Koschat (1987) independently gave a similar result for confidence intervals on the ratio of the means of two normal distributions (the Fieller (1954) problem).

Here we extend the results of GH, e.g., by allowing for less restricted models (including possibly discrete distributions, parameters in general metric spaces, and less restricted "troublesome" parameter subsets), and we apply them to some important econometric models. We consider first a general setup with a parameter vector θ and a parametric function of interest $\psi(\theta)$. The parameter space contains a subset Θ_0 near which the function $\psi(\theta)$ can take any value in a (typically large) set Ψ_0. This setup covers both cases where $\psi(\theta)$ has discontinuities at Θ_0 and where the points in Θ_0 correspond to the same data distribution (in which case Θ_0 is a *nonidentification subset*). When such conditions obtain, we say $\psi(\theta)$ is *locally almost unidentified* (LAU) near Θ_0. The main facts demonstrated here under general conditions include: (i) when $\psi(\theta)$ is LAU near Θ_0 and $\theta \in \Theta_0$, a level $1-\alpha$ confidence set $C_\psi(Y)$ for $\psi(\theta)$ must cover with probability $1-\alpha$ (at least) *any* value in the set Ψ_0 of all the values of $\psi(\theta)$ that can be met "near" Θ_0; (ii) $C_\psi(Y)$ must have a diameter as large as the diameter of Ψ_0 with probability $1-\alpha$ (or greater); in particular, if Ψ_0 is unbounded, $C_\psi(Y)$ must be unbounded with probability $1-\alpha$ (or greater); (iii) by continuity, similar properties must also hold *outside* Θ_0, at least in the neighborhood of Θ_0; (iv) when the model has a density with the same support for all θ, $C_\psi(Y)$ must have diameter as large as the one of Ψ_0 with positive probability for all θ. If these properties do not hold for a proposed confidence set, its *true level* is *zero*: it is *impossible* to build a valid confidence set which is bounded with probability one. In particular, most Wald-type confidence sets in such models have zero confidence level, irrespective of their stated nominal levels, because they are almost surely bounded.

As a result, any approximation for the null distribution of a Wald-type statistic (e.g., an asymptotic approximation) for testing a hypothesis of the form $\psi(\theta) = \psi_0$ *must* be arbitrarily bad for some ψ_0 (unless it depends on θ). In other words, Wald statistics *do not constitute valid pivotal functions* in such models and it is even *impossible* to bound their distribution over the parameter space (except by the trivial bounds 0 and 1). Furthermore, there is no way of producing "corrected standard errors" that would avoid this problem. Expansion methods (e.g., Edgeworth expansions) and "bootstrap" techniques will also fail in such contexts, as long as they lead to almost surely bounded confidence sets. This of course supports earlier work on the unreliability of Wald tests because of noninvariance problems (see Breusch and Schmidt (1988), Dagenais and Dufour (1991), and Nelson and Savin (1990)).

These results are then applied to discuss inference in the context of more specific econometric models and problems, including: (i) ratios of regression coefficients; (ii) simultaneous equations models and IV regressions; (iii) linear regressions with autoregressive errors; (iv) inference about long-run multipliers; (v) cointegrating vectors. For example, in simultaneous equations and similar models, it is shown that usual "asymptotically justified" confidence intervals for structural coefficients based on IV estimators, such as two-stage least squares (2SLS), and their asymptotic standard errors have zero coverage probability, and the corresponding t statistics have distributions which cannot be bounded by a finite set of distributions. By contrast, for the same model, we show that LR statistics have null distributions which can be bounded by a nuisance-parameter-free distribution (derived from the Wilks Λ distribution), and so the inference methods based on such statistics do not have these problems. Furthermore we show that projection techniques can be used in such contexts to obtain valid tests for a large variety of hypotheses.

The basic notations, definitions, and assumptions used in the paper are presented in Section 2. The main results on confidence sets for LAU parameters are presented in Section 3. Section 4 discusses implications for testing and the validity of Wald-type confidence sets, while the applications to specific econometric models and problems are presented in Section 5. We conclude in Section 6.

2. FRAMEWORK

Consider a family of probability spaces $\{(\mathscr{Z}, \mathscr{A}_{\mathscr{Z}}, \bar{P}_\theta) : \theta \in \Omega\}$, where $\mathscr{Z}$ is a sample space, $\mathscr{A}_{\mathscr{Z}}$ is a σ-algebra of subsets of $\mathscr{Z}$, and $\bar{P}_\theta$ is a probability measure on the measurable space $(\mathscr{Z}, \mathscr{A}_{\mathscr{Z}})$ indexed by a *parameter* θ in Ω. The sets $\mathscr{Z}$, $\mathscr{A}_{\mathscr{Z}}$, and Ω are all nonempty. Further, we are interested by a transformation $\psi : \Omega_1 \to \Psi$, defined on a nonempty subset Ω_1 of Ω, on which we wish to test hypotheses and build confidence sets. We assume also the sets Ω and Ψ possess metric space structures. Inferences about θ will be based on an $\mathscr{A}_{\mathscr{Z}}$-measurable observation (vector) Y in a space $\mathscr{Y}$. For future reference, we summarize these assumptions as follows, where $\mathbb{R}_0^+$ refers to the set of the nonnegative real numbers.

(A) Basic Assumptions. (A.1): *$\{\bar{P}_\theta : \theta \in \Omega\}$ is a family of probability measures on a measurable space $(\mathscr{Z}, \mathscr{A}_{\mathscr{Z}})$, and $(\Omega, \bar{\rho})$ is a metric space with the metric $\bar{\rho} : \Omega \times \Omega \to \mathbb{R}_0^+$.* (A.2): *$\psi : \Omega_1 \to \Psi$ is a function on Ω such that (Ψ, ρ) is a metric space with the metric $\rho : \Psi \times \Psi \to \mathbb{R}_0^+$, where Ω_1 is a nonempty subset of Ω.* (A.3) *$Y : \mathscr{Z} \to \mathscr{Y}$ is an $\mathscr{A}_{\mathscr{Z}}$-measurable function. The complete measurable space $(\mathscr{Y}, \mathscr{A}_{\mathscr{Y}})$ induced by Y on $\mathscr{Y}$ is the same for all $\theta \in \Omega$, and the probability measure determined by $\bar{P}_\theta$ on $(\mathscr{Y}, \mathscr{A}_{\mathscr{Y}})$ is denoted by $P_\theta = P_\theta(y)$, for any $\theta \in \Omega$. Furthermore, there is a metric $\rho_y : \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}_0^+$ such that all the corresponding open sets of $(\mathscr{Y}, \rho_y)$ are $\mathscr{A}_{\mathscr{Z}}$-measurable.*

Let Γ_0 be a nonempty subset of Ψ, $\Omega_0 = \{\theta \in \Omega_1 : \psi(\theta) \in \Gamma_0\}$ and $0 \le \alpha \le 1$. Following the classical terminology of hypothesis testing (Lehmann (1986, Sections 3.1, 3.5)), we say that a subset R of $\mathscr{Y}$ is a critical region with *level* α for testing the hypothesis $H_0 : \theta \in \Omega_0$ if and only if $P_\theta[Y \in R] \le \alpha$, $\forall \theta \in \Omega_0$ (or equivalently, $\sup_{\theta \in \Omega_0} P_\theta[Y \in R] \le \alpha$); if $\sup_{\theta \in \Omega_0} P_\theta[Y \in R] = \alpha$, R has *size* α. Correspondingly, a random subset $C_\psi(Y)$ of Ψ is a confidence set with level $1 - \alpha$ for $\psi(\theta)$ if and only if $\inf_{\theta \in \Omega_1} P_\theta[\psi(\theta) \in C_\psi(Y)] \ge 1 - \alpha$; $C_\psi(Y)$ has *size* (or *coverage probability*) $1 - \alpha$ when $\inf_{\theta \in \Omega_1} P_\theta[\psi(\theta) \in C_\psi(Y)] = 1 - \alpha$. We study here situations where the following conditions hold. Below $\lim_{n \to \infty} \theta_n = \bar{\theta}$ means $\bar{\rho}(\theta_n, \bar{\theta}) \to_{n \to \infty} 0$.

(B) Indeterminacy of $\psi(\theta)$ in a Neighborhood. *For some nonempty subset Ψ_0 of Ψ, there is a subset Θ_0 of Ω such that, for each $\psi_0 \in \Psi_0$, we can find a sequence $(\theta_n)_{n=1}^\infty$ with the following properties*: (a) $\theta_n \in \Omega_1 \setminus \Theta_0$, $\forall n$; (b) $\psi(\theta_n) = \psi_0$, $\forall n$; (c) $\lim_{n \to \infty} \theta_n = \bar{\theta}$ *for some* $\bar{\theta} \in \Theta_0$. *The set of sequences which satisfy the conditions* (a), (b), *and* (c) *above will be denoted* $S(\Theta_0, \Omega_1)$.

(C) Observational Equivalence on Θ_0. *If the set Θ_0 contains more than one point, the measures P_θ are identical for all $\theta \in \Theta_0$. In this case, the set Θ_0 will be called an* ***observational equivalence*** (*or* ***nonidentification***) ***subset*** *of Ω.*

Assumption B states that any value $\psi_0 \in \Psi_0$ can be met near Θ_0. Ψ_0 will typically be a large set (e.g., $\Psi_0 = \Psi$ the set of all possible values). By Assumption C, if Θ_0 has more than one element, the parameter vectors in Θ_0 are observationally equivalent. When B and C hold with Ψ_0 containing more than one distinct value, we say the parametric function $\psi(\theta)$ is *locally almost unidentified (LAU) near Θ_0*. In addition (Assumption D below), we shall assume that the probability measures P_θ enjoy a continuity property, in the sense of weak convergence (see Billingsley (1968, Chapter 1)) with respect to the sequences in $S(\Theta_0, \Omega_1)$. This condition holds in particular if Y has a density which is continuous in θ (Assumption E).

(D) Weak Convergence with Respect to $S(\Theta_0, \Omega_1)$. *For any* $\psi_1 \in \Psi_0$, *there is a sequence* $(\theta_n)_{n=1}^{\infty}$ *in* $S(\Theta_0, \Omega_1)$ *such that* $\psi(\theta_n) = \psi_1$, *for all* n, *and* P_{θ_n} *converges weakly to* $P_{\bar{\theta}}$, *where* $\bar{\theta} = \lim_{n\to\infty} \theta_n$.

(E) Existence and Continuity of Densities for the Measures P_θ. (E.1): *The probability measures* $P_\theta(y)$, $\theta \in \Omega$, *are absolutely continuous with respect to a* σ*-finite measure* $d\mu(y)$ *on* $(\mathscr{Y}, \mathscr{A}_{\mathscr{Y}})$, *with densities* $f(y \mid \theta)$, $\theta \in \Omega$, *where* $y \in \mathscr{Y}$. (E.2): *For any* $\psi_1 \in \Psi_0$, *there is a sequence* $(\theta_n)_{n=1}^{\infty}$ *in* $S(\Theta_0, \Omega_1)$ *such that* $\lim_{n\to\infty} f(y \mid \theta_n) = f(y \mid \bar{\theta})$, *a.e.* μ, *where* $\bar{\theta} = \lim_{n\to\infty} \theta_n$.

3. CONFIDENCE SETS FOR ALMOST UNIDENTIFIED PARAMETERS

Consider a confidence set $C_\psi(Y)$ for $\psi(\theta)$ whose *level* (or *coverage probability*) is $1-\alpha$, *at least* on the set $\Omega_1 \setminus \Theta_0$, according to the following assumptions.

(F) Confidence Set with Level $1-\alpha$. (F.1): $C_\psi(Y)$ *is a confidence set for* $\psi(\theta)$ *such that the event* $\psi_1 \in C_\psi(Y)$ *is* $\mathscr{A}_{\mathscr{Z}}$*-measurable*, $\forall \psi_1 \in \Psi$.

(F.2) $P_\theta[\partial(A(\psi_1))] = 0$, $\forall \psi_1 \in \Psi_0$, $\forall \theta \in \Theta_0$, *where* $A(\psi_1) = \{y \in \mathscr{Y} : y \in Y(\mathscr{Z})$ *and* $\psi_1 \in C_\psi(y)\}$ *and* $\partial(A(\psi_1))$ *is the boundary of the set* $A(\psi_1)$ *in* $\mathscr{Y}$.

(F.3) $P_\theta[\psi(\theta) \in C_\psi(Y)] \geq 1-\alpha$, $\forall \theta \in \Psi_1 \setminus \Theta_0$, *where* $0 \leq \alpha \leq 1$.

Assumption F.2 means there is no probability mass on the boundary of $A(\psi_1)$, where $A(\psi_1)$ is the acceptance region for the hypothesis $\psi(\theta) = \psi_1$ (in the definition of $A(\psi_1)$, $y \in Y(\mathscr{Z})$ simply means y belongs to the set $Y(\mathscr{Z})$ containing all possible values of the observable random vector $Y(Z)$, while $\psi_1 \in C_\psi(y)$ means ψ_1 is deemed to be acceptable by the confidence set $C_\psi(y)$). In the sequel, the symbol $\partial(\cdot)$ will refer to the boundary of a set. F.2 will typically be met when the distributions P_θ are absolutely continuous (Assumption E). Then, we can show the following proposition, where $N(\bar{\theta}) = \{\theta \in \Omega : \bar{\rho}(\theta, \bar{\theta}) < \delta\}$, with $\delta > 0$, refers to an open neighborhood of $\bar{\theta} \in \Omega$.

Proposition 3.1: *Let the Assumptions A, B, C, D, and F hold. Then, for every* $\psi_1 \in \Psi_0$ *and every sequence* $(\theta_n)_{n=1}^{\infty}$ *in* $S(\Theta_0, \Omega_1)$, *we have*

$$P_\theta\left[\psi_1 \in C_\psi(Y)\right] = \lim_{n\to\infty} P_{\theta_n}\left[\psi_1 \in C_\psi(Y)\right] \geq 1-\alpha, \qquad \forall \theta \in \Theta_0, \tag{3.1}$$

and $\sup_{\theta \in \tilde{N}(\bar{\theta})} P_\theta[\psi_1 \in C_\psi(Y)] \geq 1-\alpha$, *for every neighborhood* $N(\bar{\theta})$ *of* $\bar{\theta} = \lim_{n\to\infty} \theta_n$, *where* $\tilde{N}(\bar{\theta}) = N(\bar{\theta}) \cap (\Omega_1 \setminus \Theta_0)$. *Furthermore, the above conclusions hold a fortiori if Assumption D is replaced by the stronger Assumption E.*

The proofs of the propositions and theorems are given in the Appendix. When Y follows the distribution associated with Θ_0 (which is unique by definition), the latter proposition entails that any point $\psi_1 \in \Psi_0$ must be covered by $C_\psi(Y)$ with probability at least $1-\alpha$. Furthermore, the probability of the event $\psi_1 \in C_\psi(Y)$ must get arbitrarily close to $1-\alpha$ (or larger) at points in $N(\bar{\theta}) \cap (\Omega_1 \setminus \Theta_0)$, even if P_θ is identified everywhere in the domain $\Omega_1 \setminus \Theta_0$.

Now, for any $\psi_1 \in \Psi$ and any subset $A \subseteq \Psi$, define

$$\rho_U[A, \psi_1] = \sup\{\rho(\psi_1, \psi_2) : \psi_2 \in A\}, \tag{3.2}$$
$$D[A] = \sup\{\rho(\psi_1, \psi_2) : \psi_1, \psi_2 \in A\}.$$

$\rho_U[A, \psi_1]$ is the maximal "distance" between any point of A and ψ_1, while $D[A]$ is the "diameter" of A (the maximal distance between two points of A). After making appropriate measurability assumptions, we will now establish some general properties of the variables $\rho_U[C_\psi(Y), \psi_1]$ and $D[C_\psi(Y)]$. Note $\rho_U[\cdot]$ and $D[\cdot]$ take their values in $\mathbb{R}_0^+ \cup \{+\infty\}$, so $D[C_\psi(Y)] = \infty$ is a well-defined event.

(G.1) ρ_U Measurability: *The event $\rho_U[C_\psi(Y), \psi_1] \geq x$ is $\mathscr{A}_{\mathscr{Z}}$-measurable, for any $x \in [0, \infty]$ and $\psi_1 \in \Psi$.*

(G.2) Diameter Measurability: *The event $D[C_\psi(Y)] \geq x$ is $\mathscr{A}_{\mathscr{Z}}$-measurable, for any $x \in [0, \infty]$.*

We first show that the distance $\rho_U[C_\psi(Y), \psi_1]$ will not be inferior to $\rho_U[\Psi_0, \psi_1]$ with probability at least $1-\alpha$ when θ is close to Θ_0, for any $\psi_1 \in \Psi$.

PROPOSITION 3.2: *Under the assumptions of Proposition* 3.1 (A, B, C, D, F) *and G.*1 *let* $(\theta_n)_{n=1}^\infty$ *be any sequence in* $S(\Theta_0, \Omega_1)$, $\psi_1 \in \Psi$, *and* $R_0 \equiv \rho_U[\Psi_0, \psi_1]$. *Then,* $\forall \epsilon \in (0, \infty)$, $\forall \Delta \in (0, \infty)$,

$$P_\theta\Big[\rho_U\big[C_\psi(Y), \psi_1\big] \geq R_0\Big] \geq 1-\alpha, \qquad \forall \theta \in \Theta_0, \tag{3.3}$$

$$\begin{aligned} &\liminf_{n\to\infty} P_{\theta_n}\Big[\rho_U\big[C_\psi(Y), \psi_1\big] \geq R_0 - \epsilon\Big] \geq 1-\alpha, \qquad \text{if } R_0 < \infty, \\ &\liminf_{n\to\infty} P_{\theta_n}\Big[\rho_U\big[C_\psi(Y), \psi_1\big] \geq \Delta\Big] \geq 1-\alpha, \qquad \text{if } R_0 = \infty. \end{aligned} \tag{3.4}$$

If furthermore $P_\theta[\partial(\rho_U[C_\psi(Y), \psi_1] \geq R_0)] = 0$, $\forall \theta \in \Theta_0$, *then*

$$\lim_{n\to\infty} P_{\theta_n}\Big[\rho_U\big[C_\psi(Y), \psi_1\big] \geq R_0\Big] = P_\theta\Big[\rho_U\big[C_\psi(Y), \psi_1\big] \geq R_0\Big] \geq 1-\alpha, \quad \forall \theta \in \Theta_0. \tag{3.5}$$

We can now study how the diameter of $C_\psi(Y)$ behaves under similar conditions.

THEOREM 3.3: *Under the assumptions of Proposition* 3.1 *and G.*2, *let* $(\theta_n)_{n=1}^\infty$ *be any sequence in* $S(\Theta_0, \Omega_1)$. *Then* $\forall \epsilon \in (0, \infty)$, $\forall \Delta \in (0, \infty)$,

$$\begin{aligned} &P_\theta\Big[D\big[C_\psi(Y)\big] \geq D[\Psi_0]\Big] \geq 1-2\alpha, \qquad \forall \theta \in \Theta_0, \quad \text{if } D[\Psi_0] < \infty, \\ &P_\theta\Big[D\big[C_\psi(Y)\big] = \infty\Big] \geq 1-\alpha, \qquad \forall \theta \in \Theta_0, \quad \text{if } D[\Psi_0] = \infty, \end{aligned} \tag{3.6}$$

$$\begin{aligned} &\liminf_{n\to\infty} P_{\theta_n}\Big[D\big[C_\psi(Y)\big] \geq D[\Psi_0] - \epsilon\Big] \geq 1 - 2\alpha, &&\text{if } D[\Psi_0] < \infty,\\ &\liminf_{n\to\infty} P_{\theta_n}\Big[D\big[C_\psi(Y)\big] \geq \Delta\Big] \geq 1 - 2\alpha, &&\text{if } D[\Psi_0] = \infty. \end{aligned} \tag{3.7}$$

If furthermore $P_\theta[\partial(D[C_\psi(Y)] \geq D[\Psi_0])] = 0$, $\forall\theta \in \Theta_0$, *then*

$$\lim_{n\to\infty} P_{\theta_n}\Big[D\big[C_\psi(Y)\big] \geq D[\Psi_0]\Big] = P_\theta\Big[D\big[C_\psi(Y)\big] \geq D[\Psi_0]\Big], \quad \forall\theta \in \Theta_0. \tag{3.8}$$

The latter theorem shows the confidence set $C_\psi(Y)$ must be as "large" as the entire domain Ψ_0 with probability near (or greater than) $1 - 2\alpha$, at least when the distribution of Y is close to Θ_0. Further, on combining (3.6) and (3.8), if Ψ_0 is unbounded and the "continuity" condition $P_\theta[\partial(D[C_\psi(Y)] \geq D[\Psi_0])] = 0$ holds for $\theta \in \Theta_0$, $C_\psi(Y)$ must also be unbounded with probability near $1 - \alpha$ in the neighborhood of Θ_0. Contrariwise, if this property does not hold, we can conclude (keeping the other assumptions of Theorem 3.3) that the confidence set $C_\psi(Y)$ cannot have level $1 - \alpha$ (Corollary 3.4). In particular, if $D[\Psi_0] = \infty$ and $C_\psi(Y)$ is almost surely bounded, $C_\psi(Y)$ has *zero* coverage probability.

COROLLARY 3.4: *Under the assumptions of Theorem* 3.3 *with the exception of F.*3 *(i.e., A, B, C, D, F.*1*, F.*2*, and G.*2*), suppose* $P_\theta[D[C_\psi(Y)] \geq D[\Psi_0]] < 1 - 2\alpha$, $\forall\theta \in \Theta_0$, *for some* $\alpha \in [0, 0.5)$. *Then*, $\inf_{\theta \in \Omega_1 \setminus \Theta_0} P_\theta[\psi(\theta) \in C_\psi(Y)] < 1 - \alpha$. *Furthermore, when* $D[\Psi_0] = \infty$, *the property*

$$P_\theta\Big[D\big[C_\psi(Y)\big] = \infty\Big] = 0, \quad \textit{for} \quad \theta \in \Theta_0, \tag{3.9}$$

entails $\inf_{\theta \in \Omega_1 \setminus \Theta_0} P_\theta[\psi(\theta) \in C_\psi(Y)] = 0$.

We will now show that $C_\psi(Y)$ must be unbounded with nonzero probability *everywhere* (i.e., under all the distributions P_θ, $\theta \in \Omega$), provided the support of $P_\theta(y)$ for $\theta \in \Theta_0$ is included in the support of $P_\theta(y)$ for all $\theta \in \Omega$ (e.g., when all the distributions $P_\theta(y)$, $\theta \in \Omega$, have common support). This result is obtained by using the following lemma (implicit in GH).

LEMMA 3.5: *Under the Assumptions A and E.*1, *suppose that the probability measures* $P_\theta(y)$, $\theta \in \Omega$, *have densities* $f(y \mid \theta)$ *with support* $\mathscr{S}(\theta)$, *which may depend on* θ, *and let* $\theta_0 \in \Omega$. *If* $\mathscr{S}(\theta_0) \subseteq \mathscr{S}(\theta)$ *for all* $\theta \in \Omega$, *then, for any event* A *in* $\mathscr{A}_{\mathscr{Y}}$, $P_{\theta_0}(A) > 0 \Rightarrow P_\theta(A) > 0$, $\forall\theta \in \Omega$.

This suggests considering the following assumption.

(H) MINIMAL SUPPORT ON Θ_0: *For any* $\theta_0 \in \Theta_0$, *we have* $\mathscr{S}(\theta_0) \subseteq \mathscr{S}(\theta)$, *for all* $\theta \in \Omega \setminus \Theta_0$, *where* $\mathscr{S}(\theta)$ *is the support of the density* $f(y \mid \theta)$ *defined in E.*1.

This condition obviously holds when the densities $f(y \mid \theta)$ have common support. It is then straightforward to see that the following extensions of Theorem 3.3 and Corollary 3.4 must hold for models with density functions.

THEOREM 3.6: *Under the assumptions of Theorem* 3.3, *suppose that E and H also hold. Then, provided* $0 \le \alpha < 0.5$,

(3.10) $$P_\theta\big[D\big[C_\psi(Y)\big] \ge D[\Psi_0]\big] > 0, \qquad \forall \theta \in \Omega.$$

If furthermore $D[\Psi_0] = \infty$ *and* $0 \le \alpha < 1$, *then* $P_\theta[D[C_\psi(Y)] = \infty] > 0$, $\forall \theta \in \Omega$.

COROLLARY 3.7: *Under the assumptions of Theorem* 3.3 *with the exception of F.*3 *(A, B, C, D, F.*1, *F.*2, *G.*2), *let E and H also hold,* $D[\Psi_0] = \infty$ *and suppose*

(3.11) $$P_\theta\big[D\big[C_\psi(Y)\big] = \infty\big] = 0, \qquad \textit{for some } \theta \in \Omega.$$

Then $\inf_{\theta \in \Omega_1 \setminus \Theta_0} P_\theta[\psi(\theta) \in C_\psi(Y)] = 0$.

Theorem 3.6 and Corollary 3.7 include as a special case the Theorem of Gleser and Hwang (1987). By Corollary 3.7, it is sufficient to show that (3.11) holds at a single point θ in Ω (possibly not in Θ_0) to conclude that the confidence set $C_\psi(Y)$ has zero coverage probability. This will be the case in particular when Ψ_0 is unbounded ($D[\Psi_0] = \infty$) but $C_\psi(Y)$ is almost surely bounded.

4. TESTING AND WALD CONFIDENCE SETS

The results of the previous section have important implications for the properties of tests associated with a given confidence procedure. Any confidence set for a parameter can be interpreted as the result of a collection of tests for each possible value of the parameter: the confidence set simply reports all the values of the parameter which cannot be rejected at a given level (see Lehmann (1986, Chapter 3)). In particular, the confidence set $C_\psi(Y)$ can be interpreted as resulting from tests for a null hypothesis of the form $H_0(\psi_0): \psi(\theta) = \psi_0$, where $\psi_0 \in \Psi$. The tests themselves can be defined as follows: $\varphi(Y; \psi_0) = 1$, if $\psi_0 \notin C_\psi(Y)$, and $\varphi(Y; \psi_0) = 0$, if $\psi_0 \in C_\psi(Y)$, where $\varphi(Y; \psi_0) = 1$ means $H_0(\psi_0)$ is "rejected" and $\varphi(Y; \psi_0) = 0$ means it is "accepted."

Let $\Theta = \Omega_1 \setminus \Theta_0$. From the identity

$$\inf_{\theta \in \Theta} P_\theta\big[\psi(\theta) \in C_\psi(Y)\big] = 1 - \sup_{\theta \in \Theta} P_\theta\big[\psi(\theta) \notin C_\psi(Y)\big],$$

we see that $\inf_{\theta \in \Theta} P_\theta[\psi(\theta) \in C_\psi(Y)] = 0 \Leftrightarrow \sup_{\theta \in \Theta} P_\theta[\psi(\theta) \notin C_\psi(Y)] = 1$, and for $0 \le \alpha < 1$, $\inf_{\theta \in \Theta} P_\theta[\psi(\theta) \in C_\psi(Y)] < 1 - \alpha \Leftrightarrow \sup_{\theta \in \Theta} P_\theta[\psi(\theta) \notin C_\psi(Y)] > \alpha$. Consequently, when (3.9) or (3.11) holds, we can infer that, for any $0 \le \alpha_0 < 1$, there exists a parameter vector $\theta_0 \in \Omega_1 \setminus \Theta_0$ and a hypothetical value $\psi_0 = \psi(\theta_0)$ such that $E_{\theta_0}[\varphi(Y; \psi_0)] = P_{\theta_0}[\psi_0 \notin C_\psi(Y)] > \alpha_0$. In other words, for the family of tests $\varphi(Y; \psi_0)$, $\psi_0 \in \psi(\Theta)$, we can always find a hypothesis $H(\psi_0)$ such that the level of the corresponding test will exceed *any* nominal level. As a result, the statistics $\varphi(Y; \psi(\theta))$ cannot be *pivotal functions* for the family of distributions $\{P_\theta : \theta \in \Omega\}$, i.e., the distribution of $\varphi(Y; \psi(\theta))$ depends on θ. More importantly, for any significance level $0 \leq \alpha < 1$, there is *no way of bounding* the probability of the event $\varphi(Y; \psi(\theta)) = 1$ uniformly over $\theta \in \Theta$ (except trivially, by 0 and 1). Note also Ω_1 may be a fairly restricted subset of Ω. Furthermore, from (3.9), the presence of such problems can be assessed by looking at the properties of $C_\psi(Y)$ when $\theta \in \Theta_0$.

It is straightforward to see that the above results apply quite generally to Wald-type confidence sets. For example, suppose $\psi(\theta)$ is a scalar function of θ such that Ψ_0 is unbounded, let $\hat{\theta}$ be an estimate of θ and $\hat{\sigma}_\psi$ an estimate of the "standard error" of $\psi(\hat{\theta})$ which is positive with probability one. Then any confidence interval of the form $[\psi(\hat{\theta}) - c_1 \hat{\sigma}_\psi, \psi(\hat{\theta}) + c_2 \hat{\sigma}_\psi]$, where c_1 and c_2 are constants which depend on the "nominal level" of the interval, has true level zero. Similarly, when $\psi(\theta)$ is a vector in $\mathbb{R}^k$, any confidence ellipsoid $[\psi(\hat{\theta}) - \psi]' \hat{\Sigma}_\psi^{-1} [\psi(\hat{\theta}) - \psi] \leq c$ where c is a finite constant, will have true level zero whenever $\hat{\Sigma}_\psi$ is almost surely nonsingular. Correspondingly, the Wald statistic $W(\psi(\theta)) = [\psi(\hat{\theta}) - \psi(\theta)]' \hat{\Sigma}_\psi^{-1} [\psi(\hat{\theta}) - \psi(\theta)]$ cannot be a pivotal function. No distribution independent of the unknown parameter vector θ, e.g. an asymptotic distribution, can provide tests whose true levels would not deviate arbitrarily from their nominal levels. Or equivalently, there is no way to find a finite critical value $c(\alpha)$ (e.g., one derived from an approximating distribution, like an asymptotic distribution) such that all the hypotheses $H(\psi_0)$, $\psi_0 \in \psi(\Theta)$, would be testable at level α using a critical region of the form $W(\psi_0) > c(\alpha)$. Furthermore, no useful maximum value over a set of possible (approximating) distributions can be found. For example, if the supports of the distributions of the $W(\psi_0)$ statistics are the positive real line, the only critical value that can ensure a valid test of level α for all $\psi_0 \in \psi(\Theta)$ is $c(\alpha) = \infty$. Approximations based on expansion methods, such as Edgeworth expansions where unknown parameters have been replaced by estimates, will also face a similar problem because they would lead to confidence sets that are almost surely bounded. For similar reasons, "bootstrapping" the distribution of $W(\psi_0)$ cannot solve the problem either.

5. ECONOMETRIC APPLICATIONS

In this section, we apply the above results to a number of problems and models relevant to econometric practice and discuss possible solutions, including inference about parameters of simultaneous equations and dynamic models. Before studying those, however, we shall look at the problem of building a confidence set for the ratio of two regression coefficients in a linear regression. Even though this problem has been studied by GH, it will be illuminating to see how the more general results of Sections 3 and 4 apply to this relatively simple problem.

5.1. *Ratios of Parameters in Linear Regressions*

Consider the linear regression

$$(5.1)\qquad y = X\beta + u, \qquad u \sim N\left[0, \sigma^2 I_T\right]$$

where X is a $T \times k$ full-column rank fixed matrix ($2 \le k < T$), σ and $\beta = (\beta_1, \ldots, \beta_k)'$ are unknown coefficients ($\sigma > 0$). We wish to build a confidence set for the ratio $\psi(\theta) = \beta_2/\beta_1$ where $\theta = (\beta_1, \ldots, \beta_k, \sigma)' \in \Omega = \mathbb{R}^k \times \mathbb{R}^+$. By definition, $\psi(\theta)$ is the solution of the equation $\beta_2 = \psi(\theta)\beta_1$, unique except when $\beta_1 = 0$.

Here $\psi(\theta)$ has a discontinuity at every point of the set $\{\theta \in \Omega : \beta_1 = 0\}$. Consider the following (restricted) domain for $\psi(\theta)$: $\Omega_1 = \{\theta \in \Omega : \beta_1 \ne 0, \sigma = \bar{\sigma}, \beta_j = \bar{\beta}_j, j = 3, \ldots, k\}$, where $\bar{\sigma}$, $\bar{\beta}_j$, $j = 3, \ldots, k$ are fixed constants. Let also $\Theta_0 = \{\theta \in \Omega : \beta_1 = \beta_2 = 0, \sigma = \bar{\sigma}, \beta_j = \bar{\beta}_j, j = 3, \ldots, k\}$. Since Θ_0 contains only one vector, condition C is trivially satisfied. For any $\psi_0 \in \mathbb{R}$, we can define $\theta_n = (\beta_1^{(n)}, \psi_0\, \beta_1^{(n)}, \bar{\beta}_3, \ldots, \bar{\beta}_k, \bar{\sigma})'$, $n = 1, 2, \ldots,$ where $\beta_1^{(n)}$ is chosen so that $\beta_1^{(n)} \underset{n\to\infty}{\to} 0$ and $\beta_1^{(n)} \ne 0$, for all n. We see immediately that:

(a) $\theta_n \in \Omega_1 \setminus \Theta_0$, $\forall n$;
(b) $\psi(\theta_n) = \psi_0$, $\forall n$;
(c) $\lim_{n\to\infty} \theta_n = \bar{\theta}$, where $\bar{\theta} = (0, 0, \bar{\beta}_3, \ldots, \bar{\beta}_k, \bar{\sigma})' \in \Theta_0$.

Conditions A, B, C, and E are clearly satisfied here with $\Psi_0 = \mathbb{R}$. For $0 < \alpha < 1$, Theorems 3.3 and 3.6 entail that any level $1-\alpha$ confidence set for β_2/β_1 must have nonzero probability of being unbounded irrespective of the true value of θ, a probability that must get as high as $1-\alpha$ when $\theta = \bar{\theta}$. By varying $(\bar{\beta}_3, \ldots, \bar{\beta}_k, \bar{\sigma})'$, we see also this property must hold whenever $\beta_1 = \beta_2 = 0$.

As a result any confidence interval of the form $[(\hat{\beta}_2/\hat{\beta}_1) \pm c(\alpha/2)\hat{\sigma}_\psi]$, where (say) $\hat{\sigma}_\psi = G(\hat{\beta})\hat{\Sigma}_{\hat{\beta}} G(\hat{\beta})'$, $G(\beta) = \partial\psi(\theta)/\partial\beta'$, $\hat{\Sigma}_{\hat{\beta}} = s^2(X'X)^{-1} = [\hat{\sigma}_{ij}]_{i,j=1,\ldots,k}$ and $s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(T-k)$, has zero coverage probability. Furthermore, as shown in Section 4, we can always find a value ψ_0 such that the distribution of the Wald statistic $W(\psi_0) = [\psi(\hat{\theta}) - \psi_0]^2/\hat{\sigma}_\psi^2$ deviates arbitrarily from any "approximating distribution" (such as the $\chi^2(1)$ distribution).

By contrast, a valid confidence set for β_2/β_1 follows on "inverting" LR tests for the hypothesis $H_0(\psi_0): \beta_2/\beta_1 = \psi_0$. Since $H_0(\psi_0)$ is equivalent to $H_0(\psi_0)': \beta_2 - \psi_0\,\beta_1 = 0$, the LR test of $H_0(\psi_0)$ is equivalent to the Fisher test of $H_0(\psi_0)'$ based on $F(\psi_0) = (\hat{\beta}_2 - \psi_0\,\hat{\beta}_1)^2/(\hat{\sigma}_{11}\psi_0^2 - 2\hat{\sigma}_{12}\psi_0 + \hat{\sigma}_{22})$. Since $F(\psi_0) \sim F(1, T-k)$ under $H_0(\psi_0)$, $C_\psi(\alpha; y) = \{\psi_0 : F(\psi_0) \le F_\alpha(1, T-k)\}$ is a level $1-\alpha$ confidence set for β_2/β_1. This set can be put in explicit form by solving the quadratic inequality

$$(5.2)\qquad A\psi_0^2 + B\psi_0 + C \le 0,$$

where $A = \hat{\beta}_1^2 - F_\alpha\hat{\sigma}_{11}$, $B = -2(\hat{\beta}_1\hat{\beta}_2 - F_\alpha\hat{\sigma}_{12})$, $C = \hat{\beta}_2^2 - F_\alpha\hat{\sigma}_{22}$, and $F_\alpha = F_\alpha(1, T-k)$. This confidence set is unbounded when $A < 0$, an event with probability $P_\theta[A < 0] = P_\theta[\hat{\beta}_1^2/\hat{\sigma}_{11} < F_\alpha] = 1-\alpha$ when $\beta_1 = 0$. $C_\psi(\alpha; y)$ is a generalization of the well-known Fieller's (1954) confidence set for the ratio of two means (see Rao (1973, Section 4b)). The basic reason for the "smooth" behavior of tests and confidence sets based on this method is that the statistic $F(\psi)$ is a proper pivotal function for this problem, in contrast with Wald-type statistics.

5.2. *Simultaneous Equations and Instrumental Variables Regressions*

Let us now consider the following structural model:

$$(5.3a)\qquad y = Y\beta + X_1\gamma_1 + u,$$

$$(5.3b)\qquad Y = X\Pi + V = X_1\Pi_1 + X_2\Pi_2 + V,$$

where y is a $T\times 1$ random vector, Y a $T\times G$ matrix of endogenous variables, X_1 and X_2 are $T\times k_1$ and $T\times k_2$ matrices of fixed (or strictly exogenous) variables, $X=[X_1, X_2]$ with $\operatorname{rank}(X)=k_1+k_2=k$, β and γ_1 are $G\times 1$ and $k_1\times 1$ vectors of unknown coefficients, Π_1 and Π_2 are $k_1\times G$ and $k_2\times G$ matrices of unknown coefficients, u and V are $T\times 1$ and $T\times G$ matrices of random disturbances; furthermore, we assume that the rows of the matrix $[u, V]$ are i.i.d. $N_{G+1}[0, \Sigma]$ where $\det(\Sigma)\neq 0$ and Σ does not depend on β, γ_1, and Π. Equation (5.3a) can be viewed as a typical relationship that would be estimated by IV methods.

Substituting (5.3b) into (5.3a), we obtain the reduced form equation for y:

$$(5.4)\qquad y = X_1\pi_1 + X_2\pi_2 + v,$$

where $v = u + V\beta$ and $\pi_1 = \Pi_1\beta + \gamma_1$, $\pi_2 = \Pi_2\beta$. If no restriction is imposed on γ_1 (which is typically the case), β is identifiable if and only if $\operatorname{rank}(\Pi_2)=G$. In other words, if the equation $\pi_2=\Pi_2\beta$ has a solution β_*, it is unique if and only if $\operatorname{rank}(\Pi_2)=G$ holds. The set of all possible solutions of $\pi_2=\Pi_2\beta$ is $\mathscr{B}(\pi_2, \Pi_2)=\{\beta\in\mathbb{R}^G : \beta=\beta_* + \delta,\ \delta\in\ker(\Pi_2)\}$, where $\ker(\Pi_2)$ is the set of all vectors $\delta\in\mathbb{R}^G$ such that $\Pi_2\delta=0$ ($\operatorname{rank}(\Pi_2)=G$ if and only if $\ker(\Pi_2)=\{0\}$). Instruments may be described as "weak" when the rank condition $\operatorname{rank}(\Pi_2)=G$ fails to hold or almost does not hold, a problem recently emphasized by several authors (e.g., Angrist and Krueger (1994), Bound, Jaeger, and Baker (1995), Hall, Rudebusch, and Wilcox (1996), Maddala and Jeong (1992), Nelson and Startz (1990a, 1990b), Staiger and Stock (1997)).

For any $k_2\times G$ matrix Π_2 whose rank is less than G, we can find a sequence $(\Pi_2^{(n)})_{n=1}^{\infty}$ of $k_2\times G$ matrices such that $\operatorname{rank}(\Pi_2^{(n)})=G$, $\forall n$, and $\Pi_2^{(n)} \underset{n\to\infty}{\to} \Pi_2$. If $\ker(\Pi_2)$ contains a vector whose jth component is unbounded ($1\leq j\leq G$), this jth component is also unbounded in $\mathscr{B}(\pi_2, \Pi_2)$. In particular, if $\Pi_2=0$, we have $\mathscr{B}(\pi_2, \Pi_2)=\mathbb{R}^G$, i.e., β is completely unrestricted. Further, in such a case, the same will hold for γ_1 provided the corresponding row of Π_1 is nonzero (for $\pi_1=\Pi_1\beta+\gamma_1$).

Here the complete parameter vector is $\theta=\operatorname{vec}(\beta, \gamma_1, \Pi_1, \Pi_2, \Sigma)$. We denote by p the dimension of θ and by Ω the subset of $\mathbb{R}^p$ whose elements θ satisfy the restrictions entailed by model (5.3). Taking $\beta=\psi(\theta)$ as the parametric function of interest, we have a case where $\psi(\theta)$ is a continuous function of θ but the parameter space contains subsets inside which the different parameters are observationally equivalent. Many different such subsets do exist. Define the $(G+1)\times(G+1)$ matrix

$$A(\beta)=\begin{bmatrix} 1 & \beta' \\ 0 & I_G \end{bmatrix},$$

which is easily seen to be nonsingular for all values of β. Then, for any given vector $\bar{\theta}=\text{vec}(\bar{\beta},\bar{\gamma}_1,\bar{\Pi}_1,\bar{\Pi}_2,\bar{\Sigma})$ such that $\text{rank}(\bar{\Pi}_2)<G$, we have the following observational equivalence subset: $\Theta_0=\{\theta\in\mathbb{R}^p : \Pi_1=\bar{\Pi}_1, \Pi_2=\bar{\Pi}_2, \bar{\Pi}_2\beta=\bar{\pi}_2, \bar{\Pi}_1\beta+\gamma_1=\bar{\pi}_1, A(\beta)\Sigma A(\beta)'=A(\bar{\beta})\bar{\Sigma}A(\bar{\beta})'\}$, where $\bar{\pi}_2=\bar{\Pi}_2\bar{\beta}$ and $\bar{\pi}_1=\bar{\Pi}_1\bar{\beta}+\bar{\gamma}_1$. The condition $A(\beta)\Sigma A(\beta)'=A(\bar{\beta})\bar{\Sigma}A(\bar{\beta})'$ ensures that the disturbances of the reduced form model associated with (5.3) have identical covariance matrices: β, γ_1, and Σ move together (in Θ_0) to ensure that the conditional distribution of $[y,Y]$ given X remains the same. Thus the set Θ_0 is a subset of $\mathbb{R}^p$ defined by imposing nonlinear constraints on θ, a case clearly not covered by the results of GH.

From Theorem 3.3, any confidence set for the vector β must be unbounded with probability $1-\alpha$ (at least) when $\text{rank}(\Pi_2)<G$. For components of β, the same will hold when θ belongs to a subset Θ_0 over which this component is unbounded. Again unbounded confidence sets must occur with probabilities close to $1-\alpha$ (or greater) in the neighborhood of these sets, and since the model has a density function, the probability of getting an unbounded confidence set is different from zero for any θ. Consequently, confidence sets which are bounded with probability one have zero coverage probability. In particular, this will be the case for any Wald-type confidence interval based on the 2SLS estimator of β, the usual 2SLS standard errors and a normal asymptotic distribution. Despite considerable theoretical work on the finite sample properties of 2SLS and other simultaneous equations estimators, as well as the associated inference procedures, this important property has not apparently been pointed out before (e.g., see the survey of Phillips (1983)).

It is of interest to note here that a valid confidence set for β in model (5.3) can be obtained by a method suggested long ago by Anderson and Rubin (1949, henceforth AR). Consider first the problem of testing $H_0(\beta_0):\beta=\beta_0$. On observing

$$y-Y\beta_0=X_1\pi_1^*+X_2\pi_2^*+u_* \tag{5.5}$$

where $\pi_1^*=\gamma_1+\Pi_1(\beta-\beta_0)$, $\pi_2^*=\Pi_2(\beta-\beta_0)$, and $u_*=u+V(\beta-\beta_0)$, we see that $H_0(\beta_0)$ can be tested by testing $\pi_2^*=0$ in the linear regression (5.5). This test can be interpreted as the *LR test of* $\pi_2^*=0$ in the regression (5.5) against the same regression with π_1^* and π_2^* unrestricted. An exact confidence set of level $1-\alpha$ for β is then provided by $C_\beta(\alpha;y,Y)=\{\beta_0 : F(\beta_0)\le F_\alpha(k_2,$

$T - k_1 - k_2)\}$ where

$$(5.6) \qquad F(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k_1 - k_2)},$$

where, for any nonsingular matrix A, we define $M(A) = I_T - A(A'A)^{-1}A'$. The confidence set $C_\beta(\alpha; y, Y)$ is similar at level $1 - \alpha$ *irrespective of the true value of* Π (it is *not conservative*) and so does not require an identifiability assumption, a remarkable feature. $C_\beta(\alpha; y, Y)$ is not generally an ellipsoid and, by the results of Section 3, we can conclude it is unbounded with positive probability (a property not apparently pointed out before). In particular, an unbounded confidence set will occur with probability at least $1 - \alpha$ when the rank of Π is deficient (so β is not identifiable), a natural outcome in this case. Note also $C_\beta(\alpha; y, Y)$ could be empty: specifically, this will occur when the smallest root of the usual LIML determinantal equation exceeds some constant. Since the probability that $\beta \in C_\beta(\alpha; y, Y)$ is $1 - \alpha$, the probability this occurs *cannot* be greater than α *under the model*. Thus the occurrence of an empty confidence set can be interpreted as a rejection of the model itself, e.g., because of *overidentifying restrictions* (i.e., a test of the restriction $\operatorname{rank}([\pi_2, \Pi_2]) = \operatorname{rank}(\Pi_2)$, or equivalently a test of the fact that $\pi_2 = \Pi_2 \beta$ for some vector β). We thus have a specification test.

$C_\beta(\alpha; y, Y)$ is a valid confidence set for β because $F(\beta)$ is a proper pivotal function for the model considered. More generally, any LR-type statistic for testing a hypothesis about some transformation $\delta = g(\beta, \gamma_1, \Pi_1, \Pi_2) \in \mathbb{R}^v$ of β, γ_1, and Π is boundedly pivotal. This can be shown by using an argument similar to the one in Dufour (1989) for bounding the distributions of LR statistics for nonlinear hypotheses in linear regressions. More precisely, consider the hypothesis

$$(5.7) \qquad H_0 : g(\beta, \gamma_1, \Pi_1, \Pi_2) \in \Delta_0,$$

where Δ_0 is a nonempty subset of $\mathbb{R}^v$, let $LR(H_0)$ be the LR statistic for testing H_0 against (5.3), and consider the multivariate linear regression model:

$$(5.8) \qquad Z = XB + W$$

where $Z = [y, Y]$, $W = [u, V]$, and the rows of W are i.i.d. $N[0, \Sigma_*]$ with $\det(\Sigma_*) \neq 0$. Model (5.3) is equivalent to a restricted version of (5.8) where B belongs to the set $\Gamma_1 = \{B \in M(k, G+1) : B = B(\beta, \gamma_1, \Pi_1, \Pi_2),\ \beta \in \mathbb{R}^G,\ \gamma_1 \in \mathbb{R}^{k_1},\ \Pi_1 \in M(k_1, G),\ \Pi_2 \in M(k_2, G)\}$,

$$(5.9) \qquad B(\beta, \gamma_1, \Pi_1, \Pi_2) = \begin{bmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{bmatrix}, \qquad \pi_1 = \Pi_1 \beta + \gamma_1, \qquad \pi_2 = \Pi_2 \beta,$$

and the symbol $M(m, n)$ denotes the set of the $m \times n$ real matrices. Γ_1 represents the restrictions imposed by the structural model (5.3) on the corresponding reduced form (5.8). Then the problem of testing H_0 against (5.3) is equivalent to testing

$$(5.10) \qquad H_0' : B \in \Gamma_0 \quad \text{against} \quad H_1' : B \in \Gamma_1$$

where $\Gamma_0 = \{B \in M(k, G+1): B = B(\beta, \gamma_1, \Pi_1, \Pi_2),\ g(\beta, \gamma_1, \Pi_1, \Pi_2) \in \Delta_0\}$. If we denote by $L(Z \mid B, \Sigma)$ the likelihood function of the regression (5.8) and

$$(5.11)\qquad L(\Gamma) = \sup\{L(Z \mid B, \Sigma_*): B \in \Gamma,\ \Sigma_* \in S_{G+1}\},$$
$$\varnothing \neq \Gamma \subseteq M(k, G+1),$$

where S_{G+1} is the set of the $(G+1) \times (G+1)$ positive definite matrices, we can establish the following theorem on the distribution of $LR(H_0)$.

THEOREM 5.1: *Under the assumptions and notations* (5.3) *and* (5.7)–(5.11), *suppose* $B(\beta, \gamma_1, \Pi_1, \Pi_2) = \bar{B}$. *Then the likelihood ratio statistic* $LR(H_0)$ *for testing* H_0 *against* (5.3) *satisfies the inequality*

$$(5.12)\qquad LR(H_0) = L(\Gamma_1)/L(\Gamma_0) \leq L(M(k, G+1))/L(\{\bar{B}\}) \equiv \overline{LR}$$

where $\overline{LR} \sim [V_1 V_2 \cdots V_{G+1}]^{-T/2}$ *and* $V_1, V_2, \ldots, V_{G+1}$ *are independent random variables such that* V_i *follows a beta distribution with parameters*

$$((T-k-G-1+i)/2, (k/2)), \quad 1 \leq i \leq G+1.$$

The bound $\overline{LR}$ has a distribution which does not depend on $\bar{B}$ nor any nuisance parameter, and no identification condition is required. $\overline{LR}$ is a monotonic transformation of a Wilks Λ statistic, whose distribution has been extensively studied (see Anderson (1984, Chapter 8)). It can also be determined easily by simulation (see Dufour and Khalaf (1996a,b)). For any hypothesis H_0, e.g., a hypothesis of the form $H_0(\delta_0): g(\beta, \gamma_1, \Pi_1, \Pi_2) = \delta_0$, we have $P[LR(H_0) \geq x] \leq P[\overline{LR} \geq x]$, $\forall x$, so that the critical values of $LR(H_0)$ can be bounded from above by the quantiles of $\overline{LR}$. We do not claim this bound is very tight, but it shows clearly that LR-type statistics in simultaneous equations are boundedly pivotal, a property not shared by Wald-type statistics.

The AR procedure may be interpreted as an IV method in which the exogenous variables excluded from a structural equation of interest are added directly to the equation instead of being used to replace the endogenous explanatory variables by fitted values. The above discussion suggests this is a much sounder way of making inferences on structural coefficients than the more usual methods based on IV estimators and standard errors. Note also the AR statistic can yield "asymptotically valid" tests and confidence sets under much weaker assumptions on X and $[u, V]$; see, for example, Staiger and Stock (1997) and Dufour and Jasiak (1994). An important feature here is that finite and large sample validity results for the AR procedure are unaffected by the presence of identification problems. Further, the evidence available on the power of AR tests indicates their performance is excellent; see Maddala (1974) and Dufour and Jasiak (1994).

An apparent shortcoming of AR tests and confidence sets comes from the fact that they are designed to consider the complete vector β. When $G \geq 2$ (non-scalar β), we may still wish to build confidence sets for individual components of β or for some transformation $g(\beta) \in \mathbb{R}^m$. This can be done by using a projection approach similar to the one used in Dufour (1990) for a different problem. For any confidence set $C_\beta(\alpha)$ such that $P_\theta[\beta \in C_\beta(\alpha)] \geq 1-\alpha$, $\forall \theta \in \Omega$, e.g. $C_\beta(\alpha) = C_\beta(\alpha; y, Y)$, take the image set $g[C_\beta(\alpha)] = \{g(\beta) \in \mathbb{R}^m : \beta \in C_\beta(\alpha)\}$. Since $\beta \in C_\beta(\alpha)$ entails $g(\beta) \in g[C_\beta(\alpha)]$, we have

$$(5.13) \qquad P_\theta\left[g(\beta) \in g\left[C_\beta(\alpha)\right]\right] \geq P_\theta\left[\beta \in C_\beta(\alpha)\right] \geq 1-\alpha, \quad \forall \theta \in \Omega,$$

so that the confidence set $g[C_\beta(\alpha)]$ has level $1-\alpha$. Note there is no restriction on the dimension of $g(\beta)$. When $g(\beta) = \beta_i$, an individual element of β, $g[C_\beta(\alpha)]$ can be interpreted as the projection of $C_\beta(\alpha)$ on the β_i-axis. For a scalar function $g(\beta)$, this confidence set does not necessarily take the form of an interval, although this could easily be the case (e.g., if $g(\cdot)$ is continuous and the set $C_\beta(\alpha)$ is bounded and connected). If one wishes to have a confidence interval for any scalar function $g(\beta)$, this can be done by considering the variables $g_L(\alpha) = \inf\{g(\beta_0) : \beta_0 \in C_\beta(\alpha)\}$ and $g_U(\alpha) = \sup\{g(\beta_0) : \beta_0 \in C_\beta(\alpha)\}$ which are obtained by minimizing $g(\beta_0)$ subject to the restriction $F(\beta_0) \leq F_\alpha(k_2, T-k_1-k_2)$. Since $\beta \in C_\beta(\alpha) \Rightarrow g_L(\alpha) \leq g(\beta) \leq g_U(\alpha)$, we see again that

$$(5.14) \qquad P_\theta[g_L(\alpha) \leq g(\beta) \leq g_U(\alpha)] \geq 1-\alpha, \qquad \forall \theta \in \Omega.$$

The confidence sets $g[C_\beta(\alpha)]$ and $[g_L(\alpha), g_U(\alpha)]$ are obtained by first finding a joint confidence set for β and then deducing the corresponding set of $g(\beta)$ values. We call such sets *projection-based confidence sets*. These will typically be nonsimilar and conservative (at least at certain points of the parameter space), but no other valid procedure appears to exist in finite samples.

5.3. *Dynamic Models*

We will now examine a few dynamic models. As a first example, consider a linear regression with AR(1) disturbances:

$$(5.15) \qquad y_t = \beta_0 + x_t'\beta + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \qquad |\rho| \leq 1 \qquad (t = 1, \ldots, T),$$

where $x_1, \ldots, x_T$ are fixed $k \times 1$ vectors of explanatory variables, β_0, β, ρ, and σ are unknown coefficients, $\varepsilon_1, \ldots, \varepsilon_T$ are i.i.d. with a continuous distribution (say) which does not depend on the regression coefficients, and y_0 is taken as fixed (i.e., we consider the conditional distribution of $y_1, \ldots, y_T$ given y_0). For such a model, there is ample Monte Carlo evidence showing that usual asymptotic t and F tests based on generalized least squares (to correct serial correlation) can be quite unreliable in finite samples, especially when ρ is close to one and for inference about the intercept coefficient; see, for example, Park and Mitchell (1980) and Miyazaki and Griffiths (1984). Indeed, if we rewrite the model in the form

$$(5.16) \qquad y_t = \beta_0(1-\rho) + (x_t - \rho x_{t-1})'\beta + \rho y_{t-1} + \varepsilon_t \qquad (t = 1, \ldots, T),$$

we see that β_0 is not identified when $\rho = 1$. So by the results of Sections 3 and 4, any valid confidence set for β_0 must be unbounded with positive probability, and Wald-type tests for hypotheses on β_0 have distributions that will deviate arbitrarily from any uniform approximation. The same problems will occur even if we impose the restriction $|\rho| < 1$. For an example of a valid confidence set for the regression coefficients of the model just discussed (with normal disturbances), see Dufour (1990).

As a second example, consider inference on a long-run multiplier, which measures the long-run effect of a permanent unit change of an exogenous variable on some dependent variable. Take the simple first-order dynamic model:

$$(5.17) \qquad y_t = \lambda y_{t-1} + x_t'\beta + \varepsilon_t, \qquad |\lambda| \leq 1 \qquad (t = 1, \ldots, T),$$

where $x_1, \ldots, x_T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, and y_0 are defined as in (5.15). Then the long-run multiplier for the jth component of $x_t = (x_{1t}, \ldots, x_{kt})'$ is $\beta_{Lj} = \beta_j/(1-\lambda)$. Since β_{Lj} is a parameter ratio that becomes undefined (nonidentified) when $\lambda = 1$, the problem is similar to the one studied in Section 5.1 and the results of Sections 3 and 4 apply again. For an example of a valid confidence procedure for β_{Lk}, see Dufour and Kiviet (1994).

Thirdly, consider inference on the coefficients of a cointegrating relationship (for reviews, see Engle and Granger (1991) and Banerjee, Dolado, Galbraith, and Hendry (1993)). It is well known that such relationships can be uniquely determined only through identification restrictions (see Johansen and Juselius (1994)). Difficulties here are quite similar to those met in static simultaneous equations, but it will be useful to spell them out for a special case.

Take a bivariate time series $X_t = (X_{1t}, X_{2t})'$ which follows an autoregressive model of order p ($p \geq 1$), written in error-correction form:

$$(5.18) \qquad \Delta X_t = \mu + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + \Pi X_{t-p} + u_t \qquad (t = 1, \ldots, T),$$

where μ is a constant vector, $u_1, \ldots, u_T$ are i.i.d. $N[0, \Sigma]$ with $\det(\Sigma) \neq 0$, and the initial values $X_0, \ldots, X_{-p+1}$ are fixed. By the Engle-Granger representation theorem, X_{1t} and X_{2t} are cointegrated if and only if Π can be written $\Pi = \delta\beta'$, where δ and β are nonzero vectors of dimension two. The first step for identifying β is to impose a normalization constraint on β, e.g., by setting its first component equal to $1: \beta = (1, \beta_1)$. Then model (5.18) can be rewritten as

$$(5.19) \qquad \Delta X_t = \mu + \sum_{j=1}^{p-1} \Gamma_j \Delta X_{t-j} + \delta(X_{1,t-p} + \beta_1 X_{2,t-p}) + u_t \qquad (t = 1, \ldots, T),$$

and we see that β_1 cannot be identified when $\delta = 0$. The results of Sections 3 and 4 thus apply to inference about β_1. Note $\delta = 0$ corresponds to the usually quite plausible case where X_{1t} and X_{2t} are not cointegrated and a regression of X_{1t} on X_{2t} would be a "spurious regression."

Recent simulation experiments (Gonzalo (1994)) have shown that maximum likelihood in a fully specified error correction model (as suggested by Johansen (1988)) appears to be the best method for estimating cointegrating vectors.

Correspondingly, in the same context, our results suggest that more reliable tests and confidence sets for cointegrating vectors will be obtained by using LR-type tests and by building confidence sets through the inversion of such tests.

6. CONCLUSION

The results presented in this paper have important implications for econometric theory and practice. First, it is essential to remember that confidence sets should be based on proper *pivotal functions*, or at least on *boundedly pivotal functions*.

Second, the most commonly used method for building confidence sets, which is based on "inverting" Wald-type tests, does not rely on proper pivotal functions in situations involving LAU parameters: standard errors and covariance matrices largely lose their usual interpretation.

Thirdly, asymptotic arguments can be especially misleading in the models studied here. Even though a Wald-type statistic may be asymptotically pivotal at every point outside the nonidentification subset, convergence to the asymptotic distribution has to be arbitrarily slow at points outside the nonidentification subset (*nonuniform convergence*). Monte Carlo evidence strongly supporting this view is available in Dufour and Jasiak (1994), Hall, Rudebusch, and Wilcox (1996), and Nelson, Startz, and Zivot (1996).

Fourth, it appears that LR statistics behave relatively smoothly in the presence of identification problems, so that they have better chances of being bounded pivotal (for other illustrations of this phenomenon, see Dufour (1989)). Indeed, this is not surprising in view of the fact that the likelihood function is flat on a nonidentification subset. In the context of a standard simultaneous equations model, we showed explicitly that LR statistics for testing hypotheses about structural coefficients are boundedly pivotal, while Wald-type statistics are not. For Monte Carlo evidence showing that LR-type tests are indeed more reliable in such contexts, see Dufour and Jasiak (1994) and Nelson, Startz, and Zivot (1996).

Fifth, given a valid confidence set for a parameter vector, it is always possible to derive valid confidence sets for individual elements of the vector, or for any function of this vector, by using projection methods.

The examples analyzed in Section 5 by no way constitute an exhaustive list of the cases to which our general results apply. Other cases include: various nonlinear regressions, ARMA models both univariate and multivariate (e.g., because of common factors problems), inference in "structural" models derived from dynamic optimization models which are often estimated by the generalized method of moments, inference about structural change break dates, etc. To keep our exposition within limits, we emphasized here parametric models, i.e., models for which a *finite-dimensional* vector θ completely determines the data generating process. The results of Sections 2–4 however are sufficiently general to cover nonparametric models. Such models raise even stronger indeterminacies and "impossibilities." For example, on testing unit root hypotheses in time series models which allow for general forms of serial dependence, Blough (1992) and Cochrane (1991) showed that *any test* with level α should have power that does not exceed its level against any stationary alternative. Our results thus strongly concur with theirs by stressing the importance of finite-sample considerations for model formulation and inference.

Finally, it is important to remember not all Wald-type tests are problematic: when a Wald-type statistic is a pivotal function, as occurs for example when testing linear restrictions on the coefficients of linear regressions, there is no difficulty. The problems discussed above appear in models which contain LAU parameters. A question of interest here is whether it is possible to "salvage" Wald-type tests and confidence sets in such cases. We saw above it is totally insufficient to exclude the regions of the parameter space where the coefficient vector θ or the transformation $\psi(\theta)$ is not identifiable. Whether there is then a practical way of modifying Wald-type procedures remains doubtful. For example, in models estimated by IV, one may try to find methods for selecting "good" instruments. However, as the simulation results of Hall, Rudebusch, and Wilcox (1996) show, such procedures do not appear to work and may even make matters worse from the point of view of test reliability. Further, when it is possible to find alternative procedures that behave "smoothly" in the presence of identification difficulties (like the Anderson-Rubin procedure in simultaneous equations), there appears to be little motivation for sticking with Wald-type methods. Accepting the possibility of an unbounded confidence set for a structural coefficient is simply a matter of logic and scientific rigor: the data may simply be uninformative about such coefficients. Note this does not at all mean that the practice of building confidence sets should be abandoned for potentially unidentified models. Unbounded confidence sets do not necessarily occur for particular data sets and may indeed be very unlikely: if the data generating process is "far" from those cases where the structural parameter vector is not identified, we can expect any reasonably powerful confidence set procedure will yield unbounded confidence sets only with low probability. But unbounded confidence sets must occur with high probability when the parameters considered are not identified or are close to being so: the occurrence of such a set may be interpreted as a symptom of the fact that the parameter cannot be precisely evaluated from the available data.

C.R.D.E. and Dépt. de Sciences Économiques, Université de Montréal, C.P. 6128 succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada

Manuscript received April, 1995; final revision received November, 1996.

APPENDIX

PROOF OF PROPOSITION 3.1: For any $\psi_1 \in \Psi_0$, we can find a sequence $(\theta_n)_{n=1}^{\infty}$ in $S(\Theta_0, \Omega_1)$ such that $\psi(\theta_n) = \psi_1$, $\forall n$, and P_{θ_n} converges weakly to P_θ. Since $C_\psi(Y)$ is a level $1-\alpha$ confidence set for $\theta \in \Omega_1 \setminus \Theta_0$ (Assumption F), we have $P_{\theta_n}[\psi_1 \in C_\psi(Y)] \geq 1-\alpha$, $\forall n$. From B, D, F, and the portmanteau theorem for weak convergence (see Billingsley (1968, Chapter 1, Theorem 2.1)), we also have: $P_{\theta_n}[\psi_1 \in C_\psi(Y)] \to_{n\to\infty} P_{\bar{\theta}}[\psi_1 \in C_\psi(Y)]$, where $\bar{\theta} = \lim_{n\to\infty}\theta_n \in \Theta_0$. Then, using the fact that P_θ is the same for all $\theta \in \Theta_0$ (Assumption C), we see $P_\theta[\psi_1 \in C_\psi(Y)] = P_{\bar{\theta}}[\psi_1 \in C_\psi(Y)]$, $\forall\theta \in \Theta_0$, and (3.1) follows.

The inequality $\sup_{\theta\in\tilde{N}(\bar{\theta})} P_\theta[\psi_1 \in C_\psi(Y)] \ge 1-\alpha$ is a direct consequence of (3.1) and the fact that $\tilde{N}(\bar{\theta})$ contains all the elements (except possibly a finite number) of any sequence in $S(\Theta_0, \Omega_1)$. To complete this proof, we note that E entails D by the portmanteau theorem for weak convergence jointly with Scheffé's theorem (see Billingsley (1968, Appendix II)). *Q.E.D.*

PROOF OF PROPOSITION 3.2: To prove (3.3)–(3.4), we shall consider in turn three cases: (a) $R_0 = 0$; (b) $0 < R_0 < \infty$; (c) $R_0 = \infty$.

(a) $R_0 = 0$: The result is obvious since $\rho_U[C_\psi(Y), \psi_1] \ge 0$.

(b) $0 < R_0 < \infty$: For any $\epsilon \in (0,\infty)$, we can find $\psi_2 \in \Psi_0$ such that $\rho(\psi_1, \psi_2) \ge R_0 - \epsilon$. Since $\psi_2 \in C_\psi(Y) \Rightarrow \rho_U[C_\psi(Y), \psi_1] \ge \rho(\psi_2, \psi_1) \ge R_0 - \epsilon$, where $\Rightarrow$ is the implication operator, we see that

$$P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0 - \epsilon\,] \ge P_\theta[\psi_2 \in C_\psi(Y)], \qquad \forall\theta.$$

Then, for any $\epsilon > 0$ and $\theta \in \Theta_0$, we see from Proposition 3.1 that

$$P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0 - \epsilon\,] \ge 1-\alpha,$$

which entails $P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0 - \epsilon_m] \ge 1-\alpha$, for any sequence $(\epsilon_m)_{m=1}^\infty$ such that $\epsilon_m > 0$, $\epsilon_{m+1} < \epsilon_m$, and $\lim_{m\to\infty}\epsilon_m = 0$; hence

$$P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0] \ge 1-\alpha, \qquad \forall\theta \in \Theta_0.$$

Similarly, for all n and $\epsilon > 0$, we also have $P_{\theta_n}[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0 - \epsilon] \ge P_{\theta_n}[\psi_2 \in C_\psi(Y)]$; hence using again Proposition 3.1,

$$\liminf_{n\to\infty} P_{\theta_n}[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0 - \epsilon\,] \ge \liminf_{n\to\infty} P_{\theta_n}[\psi_2 \in C_\psi(Y)] \ge 1-\alpha.$$

(c) $R_0 = \infty$: For any $\Delta \in (0,\infty)$, we can find $\psi_2 \in \Psi_0$ such that $\rho(\psi_1, \psi_2) \ge \Delta$; hence $P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge \Delta] \ge P_\theta[\psi_2 \in C_\psi(Y)]$, $\forall\theta$. Thus, for any $\Delta > 0$ and $\theta \in \Theta_0$,

$$P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge \Delta] \ge 1-\alpha;$$

hence

$$P_\theta[\,\rho_U[C_\psi(Y), \psi_1] = \infty] = P_\theta[\,\rho_U[C_\psi(Y), \psi_1] \ge R_0] \ge 1-\alpha, \qquad \forall\theta \in \Theta_0.$$

Similarly, $P_{\theta_n}[\,\rho_U[C_\psi(Y), \psi_1] \ge \Delta] \ge P_{\theta_n}[\psi_2 \in C_\psi(Y)]$, $\forall n$; hence

$$\liminf_{n\to\infty} P_{\theta_n}[\,\rho_U[C_\psi(Y), \psi_1] \ge \Delta] \ge \liminf_{n\to\infty} P_{\theta_n}[\psi_2 \in C_\psi(Y)] \ge 1-\alpha, \qquad \forall\Delta \in (0,\infty).$$

(3.3) and (3.4) are thus established. (3.5) follows on applying the portmanteau theorem for weak convergence. *Q.E.D.*

PROOF OF THEOREM 3.3: Using the Boole-Bonferroni inequality, we first note

$$P_\theta[\{\psi_1, \psi_2\} \subseteq C_\psi(Y)] \ge 1 - P_\theta[\psi_1 \notin C_\psi(Y)] - P_\theta[\psi_2 \notin C_\psi(Y)], \qquad \forall\theta, \tag{A.1}$$

for any ψ_1, $\psi_2 \in \Psi_0$. Further, from Proposition 3.1, $P_\theta[\psi_i \in C_\psi(Y)] \ge 1-\alpha$, $i = 1, 2$, when $\theta \in \Theta_0$, so that $P_\theta[\{\psi_1, \psi_2\} \subseteq C_\psi(Y)] \ge 1 - 2\alpha$, $\forall\theta \in \Theta_0$. Now let $D_1 = D[\Psi_0]$. To prove (3.6) and (3.7), we proceed as in the proof of Proposition 3.2 and distinguish again three cases: (a) $D_1 = 0$; (b) $0 < D_1 < \infty$; (c) $D_1 = \infty$.

(a) $D_1 = 0$: In this case, $P_\theta[D(Y) \ge D_1] = P_\theta[D(Y) \ge 0] \ge 1 - 2\alpha$.

(b) $0 < D_1 < \infty$: For any $\epsilon \in (0,\infty)$, we can find ψ_1, $\psi_2 \in \Psi_0$ such that $\rho(\psi_1, \psi_2) \ge D_1 - \epsilon$. Then ψ_1, $\psi_2 \in C_\psi(Y) \Rightarrow D[C_\psi(Y)] \ge \rho(\psi_1, \psi_2) \ge D_1 - \epsilon \Rightarrow -D[C_\psi(Y)] \ge -D_1 + \epsilon$, and using (A.1),

$$P_\theta[-D[C_\psi(Y)] \le -D_1 + \epsilon] \ge P_\theta[\{\psi_1, \psi_2\} \subseteq C_\psi(Y)] \ge 1 - 2\alpha, \qquad \forall\theta \in \Theta_0.$$

Then, by the right continuity of distribution functions,

$$P_\theta[D[C_\psi(Y)] \geq D_1] = \lim_{\epsilon \to 0^+} P_\theta[-D[C_\psi(Y)] \leq -D_1 + \epsilon] \geq 1 - 2\alpha, \qquad \forall \theta \in \Theta_0.$$

Similarly, for all n and $\epsilon \in (0,\infty)$, we also have $P_{\theta_n}[D[C_\psi(Y)] \geq D_1 - \epsilon] \geq P_{\theta_n}[\{\psi_1, \psi_2\} \subseteq C_\psi(Y)] \geq 1 - P_{\theta_n}[\psi_1 \notin C_\psi(Y)] - P_{\theta_n}[\psi_2 \notin C_\psi(Y)]$, $\forall \epsilon \in (0,\infty)$; hence using Proposition 3.1, $\liminf_{n\to\infty} P_{\theta_n}[D[C_\psi(Y)] \geq D_1 - \epsilon] \geq 1 - 2\alpha$.

(c) $D_1 = \infty$: For any $\Delta > 0$, we can find $\psi_1, \psi_2 \in \Psi_0$ such that $\rho(\psi_1, \psi_2) \geq \Delta$. Thus,

$$P_\theta[D[C_\psi(Y)] \geq \Delta] \geq P_\theta[\{\psi_1, \psi_2\} \subseteq C_\psi(Y)] \geq 1 - 2\alpha, \quad \forall \theta \in \Theta_0.$$

Since the latter inequality holds for any $\Delta \in (0,\infty)$ however large, we must have:

$$P_\theta[D[C_\psi(Y)] \geq D_1] = P_\theta[D[C_\psi(Y)] = \infty] \geq 1 - 2\alpha, \qquad \forall \theta \in \Theta_0.$$

$\liminf_{n\to\infty} P_{\theta_n}[D[C_\psi(Y)] \geq \Delta] \geq 1 - 2\alpha$ follows from (A.1) and Proposition 3.1.

Consequently, (3.6)–(3.7) are established. (3.8) follows on applying Proposition 3.2 and noting $D[C_\psi(Y)] = \infty \Leftrightarrow \rho_U[C_\psi(Y), \psi_1] = \infty$, where $\psi_1 \in \Psi$. *Q.E.D.*

Proof of Lemma 3.5: If $P_{\theta_0}(A) > 0$, we have

$$P_\theta(A) = \int_A dP_\theta(y) = \int_A f(y \mid \theta)\, dy = \int_{A \cap \mathscr{S}(\theta)} f(y \mid \theta)\, dy$$

$$\geq \int_{A \cap \mathscr{S}(\theta_0)} f(y \mid \theta)\, dy = \int_{A \cap \mathscr{S}(\theta_0)} \frac{f(y \mid \theta)}{f(y \mid \theta_0)} f(y \mid \theta_0)\, dy > 0$$

where the last inequality follows on observing that $P_{\theta_0}(A) = \int_{A \cap \mathscr{S}(\theta_0)} f(y \mid \theta_0)\, dy > 0$ and $f(y \mid \theta)/f(y \mid \theta_0) > 0$ for $y \in A \cap \mathscr{S}(\theta_0)$. *Q.E.D.*

Proof of Theorem 5.1: The fact that $LR(H_0) = L(\Gamma_1)/L(\Gamma_0)$ follows from the invariance of LR test statistics to model reparameterizations (see, for example, Dagenais and Dufour (1991)) and the observation that model (5.3) is equivalent to H_1' (model (5.8) with $B \in \Gamma_1$) while the hypothesis H_0 is equivalent to H_0' (model (5.8) with $B \in \Gamma_0$). Consider now the hypothesis $H_{00}: B \in \{\bar{B}\}$ and $H_{11}: B \in M(k, G+1)$. H_{00} is the reduced form model (5.8) restricted to the single "true" value $B = \bar{B}$, while H_{11} corresponds to a completely unrestricted reduced form. Under H_0, we have $\{\bar{B}\} \subseteq \Gamma_0 \subseteq \Gamma_1 \subseteq M(k, G+1)$, so that $L(\{\bar{B}\}) \leq L(\Gamma_0) \leq L(\Gamma_1) \leq L(M(k, G+1))$ and

$$LR(H_0) = L(\Gamma_1)/L(\Gamma_0) \leq L(M(k, G+1))/L(\{\bar{B}\}) = LR(H_{00} \mid H_{11}),$$

where $LR(H_{00} \mid H_{11})$ is the LR statistic for testing H_{00} against H_{11}. The null distribution of $LR(H_{00} \mid H_{11})$ is well known from the literature on multivariate statistical analysis, since it is a monotonic transformation of Wilks Λ statistic with parameters $(G+1, T-k, k)$. Hence we have $LR(H_{00} \mid H_{11}) = \Lambda^{-T/2}$, where $\Lambda \sim V_1 V_2 \cdots V_{G+1}$ and the variables V_i, $i = 1, \ldots, G+1$, are independent with beta distributions: $V_i \sim \mathrm{Beta}((T-k-G-1+i)/2, (k/2))$, $i = 1, \ldots, G+1$; see Rao (1973, Ch. 8, pp. 540–541 and 551) or Anderson (1984, Ch. 8). *Q.E.D.*

REFERENCES

Anderson, T. W. (1984): *An Introduction to Multivariate Statistical Analysis*, Second Edition. New York: John Wiley & Sons.

Anderson, T. W., and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Angrist, J. D., and A. B. Krueger (1994): "Split Sample Instrumental Variables," Technical Working Paper 150, N.B.E.R., Cambridge, MA.

BANERJEE, A., J. DOLADO, J. W. GALBRAITH, AND D. F. HENDRY (1993): *Co-integration, Error Correction, and the Econometric Analysis of Non-stationary Data*. New York: Oxford University Press.

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

BLOUGH, S. R. (1992): "The Relationship between Power and Level for Generic Unit Root Tests in Finite Samples," *Journal of Applied Econometrics*, 7, 295–308.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

BOWDEN, R. (1973): "The Theory of Parametric Identification," *Econometrica*, 41, 1069–1074.

BREUSCH, T. S. (1986): "Hypothesis Testing in Unidentified Models," *Review of Economic Studies*, 53, 635–651.

BREUSCH, T. S., AND P. SCHMIDT (1988): "Alternative Forms of the Wald Test: How Long is a Piece of String?" *Communications in Statistics, Theory and Methods*, 17, 2789–2795.

BUNKE, H., AND O. BUNKE (1974): "Identifiability and Consistency," *Mathematische Operationsforschung und Statistik*, 5, 223–233.

BUSE, A. (1992): "The Bias of Instrumental Variables Estimators," *Econometrica*, 60, 173–180.

CHOI, I., AND P. C. B. PHILLIPS (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113–150.

COCHRANE, J. H. (1991): "A Critique of the Application of Unit Root Tests," *Journal of Economic Dynamics and Control*, 15, 275–284.

DAGENAIS, M. G., AND J.-M. DUFOUR (1991): "Invariance, Nonlinear Models and Asymptotic Tests," *Econometrica*, 59, 1601–1615.

DEISTLER, M., AND H.-G. SEIFERT (1978): "Identifiability and Consistent Estimability in Econometric Models," *Econometrica*, 46, 969–980.

DUFOUR, J.-M. (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica*, 57, 335–355.

——— (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475–494.

DUFOUR, J.-M., AND J. JASIAK (1994): "Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors," Technical Report, C.R.D.E., Université de Montréal.

DUFOUR, J.-M., AND L. KHALAF (1996a): "Simulation Based Finite and Large Sample Inference Methods in Seemingly Unrelated Regressions," Technical Report, C.R.D.E., Université de Montréal.

——— (1996b): "Simulation Based Finite and Large Sample Inference Methods in Simultaneous Equations," Technical Report, C.R.D.E., Université de Montréal.

DUFOUR, J.-M., AND J. F. KIVIET (1994): "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models," *Econometrica*, forthcoming.

ENGLE, R. F., AND C. W. J. GRANGER (1991): *Long-Run Economic Relationships*. New York: Oxford University Press.

FIELLER, E. C. (1954): "Some Problems in Interval Estimation," *Journal of the Royal Statistical Society, Series B*, 16, 175–185.

FISHER, F. M. (1976): *The Identification Problem in Econometrics*. Huntington, New York: Krieger Publishing Company.

FISHER, R. (1934): "Two New Properties of Mathematical Likelihood," *Proceedings of the Royal Society of London A*, 144, 285–307.

GLESER, L. J., AND J. T. HWANG (1987): "The Nonexistence of $100(1-\alpha)$ Confidence Sets of Finite Expected Diameter in Errors in Variables and Related Models," *The Annals of Statistics*, 15, 1351–1362.

GONZALO, J. (1994): "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships," *Journal of Econometrics*, 60, 203–233.

HALL, A. R., G. D. RUDEBUSCH, AND D. W. WILCOX (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–298.

HECKMAN, J. J., AND R. ROBB (1986): "Alternative Identifying Assumptions in Econometric Models of Selection Bias," in *Advances in Econometrics, Volume 5, Innovations in Quantitative Economics: Essays in Honor of Robert L. Basmann*, ed. by D. J. Slottje and G. F. Rhodes Jr. Greenwich, Conn.: JAI Press, pp. 243–287.

HILLIER, G. H. (1990): "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica*, 58, 1181–1194.

HSIAO, C. (1983): "Identification," in *Handbook of Econometrics*, Volume 1, ed. by Z. Griliches and M. D. Intrilligator. Amsterdam: North-Holland, Chapter 4, pp. 223–283.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

JOHANSEN, S., AND K. JUSELIUS (1994): "Identification of the Long-Run and the Short-Run Structure: An Application to the ISLM Model," *Journal of Econometrics*, 63, 7–36.

KOSCHAT, M. A. (1987): "A Characterization of the Fieller Solution," *The Annals of Statistics*, 15, 462–468.

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses*, 2nd Edition. New York: John Wiley & Sons.

MADDALA, G. S. (1974): "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," *Econometrica*, 42, 841–851.

MADDALA, G. S., AND J. JEONG (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181–183.

MCMANUS, D. A., J. C. NANKERVIS, AND N. E. SAVIN (1994): "Multiple Optima and Asymptotic Approximations in the Partial Adjustment Model," *Journal of Econometrics*, 62, 91–128.

MIYAZAKI, S., AND W. E. GRIFFITHS (1984): "The Properties of Some Covariance Matrix Estimators in Linear Models with Autocorrelated Errors," *Economics Letters*, 14, 351–356.

NELSON, C. R., AND R. STARTZ (1990a): "The Distribution of the Instrumental Variable Estimator and its t Ratio When the Instrument is a Poor One," *Journal of Business*, 63, 125–140.

——— (1990b): "Some Further Results on the Exact Small Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

NELSON, C. R., R. STARTZ, AND E. ZIVOT (1996): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," Technical Report, Department of Economics, University of Washington.

NELSON, F. D., AND N. E. SAVIN (1990): "The Danger of Extrapolating Asymptotic Local Power," *Econometrica*, 58, 977–981.

PARK, R. E., AND B. M. MITCHELL (1980): "Estimating the Autocorrelated Error Model with Trended Data," *Journal of Econometrics*, 13, 185–201.

PHILLIPS, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, Volume 1, ed. by Z. Griliches and M. D. Intrilligator. Amsterdam: North-Holland, Chapter 8, pp. 449–516.

——— (1984): "The Exact Distribution of LIML: I," *International Economic Review*, 25, 249–261.

——— (1985): "The Exact Distribution of LIML: II," *International Economic Review*, 26, 21–36.

——— (1989): "Partially Identified Econometric Models," *Econometric Theory*, 5, 181–240.

PRAKASA RAO, B. L. S. (1992): *Identifiability in Stochastic Models: Characterization of Probability Distributions*. New York: Academic Press.

RAO, C. R. (1973): *Linear Statistical Inference and its Applications*, Second Edition. New York: John Wiley & Sons.

ROTHENBERG, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39, 577–591.

SARGAN, J. D. (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–1633.

SIMS, C. (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

STAIGER, D., AND J. H. STOCK (1994): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 67, 557–586.