Identification, weak instruments and statistical inference in econometrics *

Jean-Marie Dufour †
Université de Montréal

First version: May 2003
Revised: August 2003
This version: August 17, 2003
Compiled: August 17, 2003, 5:13pm

* This paper is based on the author's Presidential Address to the *Canadian Economics Association* given on May 31, 2003, at Carleton University (Ottawa). The author thanks Bryan Campbell, Tarek Jouini, Lynda Khalaf, William McCausland, Nour Meddahi, Benoît Perron and Mohamed Taamouti for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 514 343 2400; FAX: 1 514 343 5831; e-mail: jean.marie.dufour@umontreal.ca. Web page: http://www.fas.umontreal.ca/SCECO/Dufour .

ABSTRACT

We discuss statistical inference problems associated with identification and testability in econometrics, and we emphasize the common nature of the two issues. After reviewing the relevant statistical notions, we consider in turn inference in nonparametric models and recent developments on weakly identified models (or weak instruments). We point out that many hypotheses, for which test procedures are commonly proposed, are not testable at all, while some frequently used econometric methods are fundamentally inappropriate for the models considered. Such situations lead to *ill-defined* statistical problems and are often associated with a misguided use of asymptotic distributional results. Concerning nonparametric hypotheses, we discuss three basic problems for which such difficulties occur: (1) testing a mean (or a moment) under (too) weak distributional assumptions; (2) inference under heteroskedasticity of unknown form; (3) inference in dynamic models with an unlimited number of parameters. Concerning weakly identified models, we stress that valid inference should be based on proper pivotal functions — a condition not satisfied by standard Wald-type methods based on standard errors — and we discuss recent developments in this field, mainly from the viewpoint of building valid tests and confidence sets. The techniques discussed include alternative proposed statistics, bounds, projection, split-sampling, conditioning, Monte Carlo tests. The possibility of deriving a finite-sample distributional theory, robustness to the presence of weak instruments, and robustness to the specification of a model for endogenous explanatory variables are stressed as important criteria assessing alternative procedures.

Key-words : hypothesis testing; confidence set; confidence interval; identification; testability; asymptotic theory; exact inference; pivotal function; nonparametric model; Bahadur-Savage; heteroskedasticity; serial dependence; unit root; simultaneous equations; structural model; instrumental variable; weak instrument; weak identification; simultaneous inference; projection; split-sample; conditional test; Monte Carlo test; bootstrap.

JEL classification numbers: C1, C12, C14, C15, C3, C5.

RÉSUMÉ

Nous analysons les problèmes d'inférence associés à l'identification et à la testabilité en économétrie, en soulignant la similarité entre les deux questions. Après une courte revue des notions statistiques requises, nous étudions tour à tour l'inférence dans les modèles non-paramétriques ainsi que les résultats récents sur les modèles structurels faiblement identifiés (ou les instruments faibles). Nous remarquons que beaucoup d'hypothèses, pour lesquelles des tests sont régulièrement proposés, ne sont pas en fait testables, tandis que plusieurs méthodes économétriques fréquemment utilisées sont fondamentalement inappropriées pour les modèles considérés. De telles situations conduisent à des problèmes statistiques *mal posés* et sont souvent associées à un emploi mal avisé de résultats distributionnels asymptotiques. Concernant les hypothèses non-paramétriques, nous analysons trois problèmes de base pour lesquels de telles difficultés apparaissent: (1) tester une hypothèse sur un moment avec des restrictions trop faibles sur la forme de la distribution; (2) l'inférence avec hétéroscédasticité de forme non spécifiée; (3) l'inférence dans les modèles dynamiques avec un nombre illimité de paramètres. Concernant les modèles faiblement identifiés, nous insistons sur l'importance d'utiliser des fonctions pivotales — une condition qui n'est pas satisfaite par les méthodes usuelles de type Wald basées sur l'emploi d'écart-types — et nous passons en revue les développements récents dans ce domaine, en mettant l'accent sur la construction de test et régions de confiance valides. Les techniques considérées comprennent les différentes statistiques proposées, l'emploi de bornes, la subdivision d'échantillon, les techniques de projection, le conditionnement et les tests de Monte Carlo. Parmi les critères utilisés pour évaluer les procédures, nous insistons sur la possibilité de fournir une théorie distributionnelle à distance finie, sur la robustesse par rapport à la présence d'instruments faibles ainsi que sur la robustesse par rapport la spécification d'un modèle pour les variables explicatives endogènes du modèle.

Mots clés : test d'hypothèse; région de confiance; intervalle de confiance; identification; testabilité; théorie asymptotique; inférence exacte; fonction pivotale; modèle non-paramétrique; Bahadur-Savage; hétéroscédasticité; dépendance sérielle; racine unitaire; équations simultanées; modèle structurel; variable instrumentale; instrument faible; inférence simultanée; projection; subdivision d'échantillon; test conditionnel; test de Monte Carlo; bootstrap.

Classification JEL: : C1, C12, C14, C15, C3, C5.

Contents

List of Propositions and Theorems **iii**

1. Introduction **1**

2. Models **4**

3. Statistical notions **5**
3.1. Hypotheses 5
3.2. Test level and size 7
3.3. Confidence sets and pivots 8
3.4. Testability and identification 9

4. Testability, nonparametric models and asymptotic methods **9**
4.1. Procedures robust to nonnormality 9
4.2. Procedures robust to heteroskedasticity of unknown form 12
4.3. Procedures robust to autocorrelation of arbitrary form 13

5. Structural models and weak instruments **14**
5.1. Standard simultaneous equations model 14
5.2. Statistical problems associated with weak instruments 16
5.3. Characterization of valid tests and confidence sets 17

6. Approaches to weak instrument problems **18**
6.1. Anderson-Rubin statistic 18
6.2. Projections and inference on parameter subsets 20
6.3. Alternatives to the AR procedure 21

7. Extensions **26**
7.1. Multivariate regression, simulation-based inference and nonnormal errors 26
7.2. Nonlinear models 27

8. Conclusion **28**

List of Propositions and Theorems

4.1 Theorem : Mean non-testability in nonparametric models 10
4.2 Theorem : Characterization of heteroskedasticity robust tests 12
4.4 Theorem : Unit root non-testability in nonparametric models 13
5.1 Theorem : Characterization of valid confidence set for a possibly unidentified parameter 17

1. Introduction

The main objective of econometrics is to supply methods for analyzing economic data, building models, and assessing alternative theories. Over the last 25 years, econometric research has led to important developments in many areas, such as: (1) new fields of applications linked to the availability of new data, financial data, micro-data, panels, qualitative variables; (2) new models: multivariate time series models, GARCH-types processes; (3) a greater ability to estimate nonlinear models which require an important computational capacity; (4) methods based on simulation: bootstrap, indirect inference, Markov chain Monte Carlo techniques; (5) methods based on weak distributional assumptions: nonparametric methods, asymptotic distributions based on "weak regularity conditions"; (6) discovery of various nonregular problems which require nonstandard distributional theories, such as unit roots and unidentified (or weakly identified) models.

An important component of this work is the development of procedures for *testing* hypotheses (or models). Indeed, a view widely held by both scientists and philosophers is that *testability* or the formulation of *testable hypotheses* constitutes a central feature of scientific activity — a view we share. With the exception of mathematics, it is not clear a discipline should be viewed as scientific if it does not lead to empirically testable hypotheses. But this requirement leaves open the question of formulating operational procedures for testing models and theories. To date, the only coherent — or, at least, the only well developed — set of methods are those supplied by statistical and econometric theory.

Last year, on the same occasion, MacKinnon (2002) discussed the use of simulation-based inference methods in econometrics, specifically *bootstrapping*, as a way of getting more reliable tests and confidence sets. In view of the importance of the issue, this paper also considers questions associated with the development of reliable inference procedures in econometrics. But our exposition will be, in a way, more specialized, and in another way, more general — and critical. Specifically, we shall focus on general statistical issues raised by *identification* in econometric models, and more specifically on *weak instruments* in the context of structural models [*e.g.*, simultaneous equations models (SEM)]. We will find it useful to bring together two separate streams of literature: namely, results (from mathematical statistics and econometrics) on *testability in nonparametric models*, and the recent econometric research on weak instruments.[1] In particular, we shall emphasize that identification problems arise in both literatures and have similar consequences for econometric methodology. Further, the literature on nonparametric testability sheds light on various econometric problems and their solutions.

Simultaneous equations models (SEM) are related in a natural way to the concept of equilibrium postulated by economic theory, both in microeconomics and macroeconomics. So it is not surprising that SEM were introduced and most often employed in the analysis of economic data. Methods for estimating and testing such models constitute a hallmark of econometric theory and represent one of its most remarkable achievements. The problems involved are difficult, raising among various issues the possibility of observational equivalence between alternative parameter values (*non-identification*) and the use of *instrumental variables* (IV). Further, the finite-sample

[1]By a *nonparametric* model (or hypothesis), we mean a set of possible data distributions such that a distribution [e.g., the "true" distribution] cannot be singled out by fixing a *finite* number of parameter values.

distributional theory of estimators and test statistics is very complex, so inference is typically based on large-sample approximations.[2]

IV methods have become a routine part of econometric analysis and, despite a lot of loose ends (often hidden by asymptotic distributional theory), the topic of SEM was dormant until a few years ago. Roughly speaking, an instrument should have two basic properties: first, it should be independent (or, at least, uncorrelated) with the disturbance term in the equation of interest (*exogeneity*); second, it should be correlated with the included endogenous explanatory variables for which it is supposed to serve as an instrument (*relevance*). The exogeneity requirement has been well known from the very beginning of IV methods. The second one was also known from the theory of identification, but its practical importance was not well appreciated and often hidden from attention by lists of instruments relegated to footnotes (if not simply absent) in research papers. It returned to center stage with the discovery of so-called weak instruments, which can be interpreted as instruments with little relevance (*i.e.*, weakly correlated with endogenous explanatory variables). Weak instruments lead to poor performance of standard econometric procedures and cases where they have pernicious effects may be difficult to detect.[3] Interest in the problem also goes far beyond IV regressions and SEM, because it underscores the pitfalls in using large-sample approximations, as well as the importance of going back to basic statistical theory when developing econometric methods.

A parameter (or a parameter vector) in a model is not identified when it is not possible to distinguish between alternative values of the parameter. In parametric models, this is typically interpreted by stating that the postulated distribution of the data — as a function of the parameter vector (the *likelihood function*) — can be the same for different values of the parameter vector.[4] An important consequence of this sort of situation is a statistical impossibility: we cannot design a data-based procedure for distinguishing between equivalent parameter values (unless additional information is introduced). In particular, no reasonable test can be produced.[5] In nonparametric models, identification is more difficult to characterize because a likelihood function (involving a finite number of parameters) is not available and parameters are often introduced through more abstract techniques (*e.g.*, functionals of distribution functions). But the central problem is the same: can we distinguish between alternative values of the parameter? So, quite generally, an *identification problem* can be viewed as a special form of *non-testability*. Specifically,

- identification involves the possibility of distinguishing different parameter values on the basis of the corresponding data distributions, while
- testability refers to the possibility of designing procedures that can discriminate between subsets of parameter values.

[2] For reviews, see Phillips (1983) and Taylor (1983).

[3] Early papers which called attention to the problem include: Nelson and Startz (1990a, 1990b), Buse (1992), Choi and Phillips (1992), Maddala and Jeong (1992), and Bound, Jaeger, and Baker (1993, 1995).

[4] For general expositions of the theory of identification in econometrics and statistics, the reader may consult Rothenberg (1971), Fisher (1976), Hsiao (1983), Prakasa Rao (1992), Bekker, Merckens, and Wansbeek (1994) and Manski (1995, 2003).

[5] By a reasonable test, we mean here a test that both satisfies a level constraint and may have power superior to the level when the tested hypothesis (the null hypothesis) does not hold. This will be discussed in greater detail below.

Alternatively, a problem of *non-testability* can be viewed as a form of *non-identification* (or *underidentification*). These problems are closely related. Furthermore, it is well known that one can create a non-identified model by introducing redundant parameters, and conversely identification problems can be eliminated by transforming the parameter space (*e.g.*, by reducing the number of parameters). Problems of non-identification are associated with *bad parameterizations*, *inappropriate choices of parameter representations*. We will see below that the same remark applies quite generally to non-testability problems, especially in nonparametric setups.

In this paper, we pursue two main objectives: *first*, we analyze the statistical problems associated with non-identification within the broader context of testability; *second*, we review the inferential issues linked to the possible presence of weak instruments in structural models. More precisely, regarding the issue of testability, the following points will be emphasized:

1. many models and hypotheses are formulated in ways that make them fundamentally non-testable; in particular, this tends to be the case in nonparametric setups;

2. such difficulties arise in basic apparently well-defined problems, such as: (a) testing an hypothesis about a mean when the observations are independent and identically distributed (i.i.d. ; (b) testing an hypothesis about a mean (or a median) with heteroskedasticity of unknown form; (c) testing the unit root hypothesis on an autoregressive model whose order can be arbitrarily large;

3. some parameters tend to be non-testable (badly identified) in nonparametric models while others are not; in particular, non-testability easily occurs for moments (*e.g.*, means, variances) while it does not for quantiles (*e.g.*, medians); from this viewpoint, moments are not a good way of representing the properties of distributions in nonparametric setups, while quantiles are;

4. these phenomena underscore *parametric nonseparability* problems: statements about a given parameter (often interpreted as the *parameter of interest*) are not empirically meaningful without supplying information about other parameters (often called *nuisance parameters*); but hypotheses that set both the parameter of interest and some nuisance parameters may well be testable in such circumstances, so that the development of appropriate inference procedures should start from a joint approach;

5. to the extent that asymptotic distributional theory is viewed as a way of producing statistical methods which are valid under "weak" distributional assumptions, it is fundamentally misleading because, under nonparametric assumptions, such approximations are arbitrarily bad in finite samples.

Concerning weak instruments, we will review the associated problems and proposed solutions, with an emphasis on finite-sample properties and the development of tests and confidence sets which are robust to the presence of weak instruments. In particular, the following points will be stressed:

1. in accordance with basic statistical theory, one should always look for *pivots* as the fundamental ingredient for building tests and confidence sets; this principle appears to be especially important when identification problems are present;

2. parametric nonseparability arises in striking ways when some parameters may not be identified, so that proper pivots may easily involve many more parameters than the parameter of interest; this also indicates that the common distinction between parameters of interest and nuisance parameters can be quite arbitrary, if not misleading;

3. important additional criteria for evaluating procedures in such contexts include various forms of invariance (or robustness), such as: (a) robustness to weak instruments; (b) robustness to instrument exclusion; (c) robustness to the specification of the model for the endogenous explanatory variables in the equation(s) of interest;

4. weak instrument problems underscore in a striking way the limitations of large-sample arguments for deriving and evaluating inference procedures;

5. very few informative pivotal functions have been proposed in the context of simultaneous equations models;

6. the early statistic proposed by Anderson and Rubin (1949, AR) constitutes one of the (very rare) truly pivotal functions proposed for SEM; furthermore, it satisfies all the invariance properties listed above, so that it may reasonably be viewed as a fundamental building block for developing reliable inference procedures in the presence of weak instruments;

7. a fairly complete set of inference procedures that allow one to produce tests and confidence sets for all model parameters can be obtained through *projection* techniques;

8. various extensions and improvements over the AR method are possible, especially in improving power; however, it is important to note that these often come at the expense of using large-sample approximations or giving up robustness.

The literature on weak instruments is growing rapidly, and we cannot provide here a complete review. In particular, we will not discuss in any detail results on estimation, the detection of weak instruments, or asymptotic theory in this context. For that purpose, we refer the reader to the excellent survey recently published by Stock, Wright, and Yogo (2002).

The paper is organized as follows. In the next two sections, we review succinctly some basic notions concerning models (section 2) and statistical theory (section 3), which are important for our discussion. In section 4, we study testability problems in nonparametric models. In section 5, we review the statistical difficulties associated with weak instruments. In section 6, we examine a number of possible solutions in the context of linear SEM, while extensions to nonlinear or non-Gaussian models are considered in Section 7. We conclude in section 8.

2. Models

The purpose of econometric analysis is to develop mathematical representations of data, which we call *models* or *hypotheses* (models subject to restrictions). An hypothesis should have two basic features.

1. It must restrict the expected behavior of observations, be *informative*. A non-restrictive hypothesis says nothing and, consequently, does not teach us anything: it is *empirically empty*, void *of empirical content*. The more restrictive a model is, the more informative it is, and the more interesting it is.

2. It must be *compatible with available data*; ideally, we would like it to be *true*.

However, these two criteria are not always compatible:

1. the *information criterion* suggests the use of *parsimonious* models that usually take the form of parametric models based on *strong assumptions*; note the information criterion is emphasized by an influential view in philosophy of science which stresses **falsifiability** as a criterion for the **scientific character** of a theory [Popper (1968)];

2. in contrast, *compatibility with observed data* is most easily satisfied by *vague models* which impose few restrictions; vague models may take the form of parametric models with a large number of free parameters or *nonparametric* models which involve an infinite set of free parameters and thus allow for *weak assumptions*.

Models can be classified as being either *deterministic* or *stochastic*. *Deterministic models* , which claim to make arbitrarily precise predictions, are highly falsifiable but always inconsistent with observed data. Accordingly, most models used in econometrics are *stochastic*. Such models are *unverifiable*: as with any theory that makes an indefinite number of predictions, we can never be sure that the model will not be put in jeopardy by new data. Moreover, they are *logically unfalsifiable*: in contrast with deterministic models, a probabilistic model is usually logically compatible with all possible observation vectors.

Given these facts, it is clear any criterion for assessing whether an hypothesis is acceptable must involve a *conventional* aspect. The purpose of *hypothesis testing* theory is to supply a coherent framework for accepting or rejecting probabilistic hypotheses. It is a probabilistic adaptation of the falsification principle.[6]

3. Statistical notions

In this section, we review succinctly basic statistical notions which are essential for understanding the rest of our discussion. The general outlook follows modern statistical testing theory, derived from the Neyman-Pearson approach and described in standard textbooks, such as Lehmann (1986).

3.1. Hypotheses

Consider an observational experiment whose result can be represented by a vector of observations

$$\mathbf{X}^{(n)} = (X_1, \ldots, X_n)' \tag{3.1}$$

[6]For further discussion on the issues discussed in this section, the reader may consult Dufour (2000).

where X_i takes real values, and let

$$\bar{F}(x) = \bar{F}(x_1, \ldots, x_n) = \mathsf{P}[X_1 \le x_1, \ldots, X_n \le x_n] \tag{3.2}$$

be its distribution, where $x = (x_1, \ldots, x_n)$. We denote by $\mathcal{F}_n$ the set of possible distribution functions on $\mathbb{R}^n$ [$\bar{F} \in \mathcal{F}_n$].

For various reasons, we prefer to represent distributions in terms of *parameters*. There are *two ways of introducing parameters* in a model. The *first* is to define a function from a space of probability distributions to a vector in some Euclidean space:

$$\theta : \mathcal{F}_n \longrightarrow \mathbb{R}^p . \tag{3.3}$$

Examples of such parameters include: the moments of a distribution (mean, variance, kurtosis, etc.), its quantiles (median, quartiles, etc.). Such functions are also called *functionals*. The *second* approach is to define a family of distribution functions which are indexed by a parameter vector θ :

$$F(x) = F_0(x \mid \theta) \tag{3.4}$$

where F_0 is a distribution function with a specific form. For example, if $F_0(x \mid \theta)$ represents a Gaussian distribution with mean μ and variance σ^2 [*e.g.*, corresponding to a Gaussian law], we have $\theta = (\mu, \sigma^2)$.

A model is *parametric* if the distribution of the data is specified up to a finite number of (scalar) parameters. Otherwise, it is *nonparametric*. An hypothesis H_0 on $X^{(n)}$ is an assertion of the type

$$H_0 : \bar{F} \in \mathcal{H}_0 , \tag{3.5}$$

where $\mathcal{H}_0$ is a subset of $\mathcal{F}_n$, the set of all possible distributions $\mathcal{F}_n$. The set $\mathcal{H}_0$ may contain: a single distribution (*simple hypothesis*), or several distributions (*composite hypothesis*). In particular, if we can write $\theta = (\theta_1, \theta_2)$, H_0 often takes the following form:

$$\mathcal{H}_0 \equiv \{F(\cdot) : F(x) = F_0(x \mid \theta_1, \theta_2) \text{ and } \theta_1 = \theta_1^0\} . \tag{3.6}$$

We usually abbreviate this as:

$$H_0 : \theta_1 = \theta_1^0 . \tag{3.7}$$

In such a case, we call θ_1 the *parameter of interest*, and θ_2 a *nuisance parameter*: the parameter of interest is set by H_0 but the nuisance parameter remains unknown. H_0 may be interpreted as follows: there is at least one distribution in H_0 that can be viewed as a representation compatible with the observed "behavior" of $X^{(n)}$. Then we can say that:

$$H_0 \text{ is acceptable} \iff \Big((\exists F \in \mathcal{H}_0)\ F \text{ is acceptable}\Big) \tag{3.8}$$

or, equivalently,

$$H_0 \text{ is unacceptable} \iff \Big((\forall F \in \mathcal{H}_0)\ F \text{ is unacceptable}\Big) . \tag{3.9}$$

Showing that H_0 is unacceptable requires one to show that *all* distributions in $\mathcal{H}_0$ are incompatible with the observed data.

3.2. Test level and size

A test for H_0 is a rule by which one decides to reject or accept the hypothesis (or to view it as incompatible with the data). It usually takes the form:

$$\begin{array}{ll} \text{reject } H_0 & \text{if } S_n(X_1, \ldots, X_n) > c\,, \\ \text{do not reject } H_0 & \text{if } S_n(X_1, \ldots, X_n) \leq c\,. \end{array} \tag{3.10}$$

The test has *level* α iff

$$\mathsf{P}_F[\text{Rejecting } H_0] \leq \alpha \text{ for all } F \in \mathcal{H}_0 \tag{3.11}$$

or, equivalently,

$$\sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\text{Rejecting } H_0] \leq \alpha\,, \tag{3.12}$$

where $\mathsf{P}_F[\,\cdot\,]$ is the function (*probability measure*) giving the probability of an event when the data distribution function is F. The test has *size* α if

$$\sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\text{Rejecting } H_0] = \alpha\,. \tag{3.13}$$

H_0 is *testable* if we can find a finite number c that satisfies the level restriction. Probabilities of rejecting H_0 for distributions outside $\mathcal{H}_0$ (*i.e.*, for $F \notin \mathcal{H}_0$) define the *power* function of the test.[7] Power describes the ability of a test to detect a "false" hypothesis. Alternative tests are typically assessed by comparing their powers: between two tests with the same level, the one with the highest power against a given alternative distribution $F \notin \mathcal{H}_0$ is deemed preferable (at least, under this particular alternative). Among tests with the same level, we typically like to have a test with the highest possible power against "alternatives of interest".

As the set $\mathcal{H}_0$ gets larger, the test procedure must satisfy a bigger set of constraints: the larger is the set of distributions compatible with a null hypothesis, the stronger are the restrictions on the test procedure. In other words, the less restrictive an hypothesis is, the more restricted will be the corresponding test procedure. It is easy to understand that imposing a large set of restrictions on a test procedure may reduce its power against specific alternatives. There may be a point where the restrictions are no longer implementable, in the sense that no procedure which has some power can satisfy the level constraint: H_0 *is non-testable.* In such a case, we have an **ill-defined test problem**.

In a framework such as the one in (3.6), where we distinguish between a parameter of interest θ_1 and a nuisance parameter θ_2, this is typically due to heavy dependence of the distribution of S_n on the nuisance parameter θ_2. If the latter is specified, we may be able to find a (finite) critical value $c = c(\alpha,\, \theta_2)$ that satisfies the level constraint (3.11). But, in ill-defined problems, $c(\alpha,\, \theta_2)$

[7] More formally, the power function can be defined as the function: $P(F) = \mathsf{P}_F[\text{Rejecting } H_0]$ for $F \in \mathcal{H}_1 \setminus \mathcal{H}_0$, where $\mathcal{H}_1$ is an appropriate subset of the set of all possible distributions $\mathcal{F}_n$. Sometimes, it is also defined on the set $\mathcal{H}_1 \cup \mathcal{H}_0$, in which case it should satisfy the level constraint for $F \in \mathcal{H}_0$.

depends heavily on θ_2, so that it is not possible to find a useful (finite) critical value for testing H_0, *i.e.* $\sup_{\theta_2} c(\alpha, \theta_2) = \infty$. Besides, even if this is the case, this does not imply that an hypothesis that would fix *both* θ_1 and θ_2, is not testable, *i.e.* the hypothesis $H_0' : (\theta_1, \theta_2) = (\theta_1^0, \theta_2^0)$ may be perfectly testable. But only a complete specification of the vector (θ_1, θ_2) does allow one to interpret the values taken by the test statistic S_n (*nonseparability*).

3.3. Confidence sets and pivots

If we consider an hypothesis of the form

$$H_0(\theta_1^0) : \theta_1 = \theta_1^0 \tag{3.14}$$

and if we can build a different test $S_n(\theta_1^0; X_1, \ldots, X_n)$ for each possible value of θ_1^0, we can determine the set of values that can be viewed as compatible with the data according to the tests considered:

$$C = \left\{\theta_1^0 : S_n(\theta_1^0; X_1, \ldots, X_n) \le c(\theta_1^0)\right\} . \tag{3.15}$$

If

$$\mathsf{P}_F\left[\text{Rejecting } H_0(\theta_1^0)\right] \le \alpha \quad \text{for all} \quad F \in \mathcal{H}(F_0, \theta_1^0), \tag{3.16}$$

we have

$$\inf_{\theta_1, \theta_2} \mathsf{P}_F[\theta_1 \in C] \ge 1 - \alpha . \tag{3.17}$$

C is a *confidence set* with level $1-\alpha$ for θ_1. The set C covers the "true" parameter value θ_1 with probability at least $1-\alpha$. The minimal probability of covering the true value of θ_1, *i.e.* $\inf_{\theta_1, \theta_2} \mathsf{P}[\theta_1 \in C]$, is called the *size* of the confidence set.

In practice, *confidence regions* (or *confidence intervals*) were made possible by the discovery of *pivotal functions* (or *pivots*): a pivot for θ_1 is a function $S_n(\theta_1; X_1, \ldots, X_n)$ whose distribution does not depend on unknown parameters (nuisance parameters); in particular, the distribution does not depend on θ_2. More generally, the function $S_n(\theta_1; X_1, \ldots, X_n)$ is *boundedly pivotal* if its distribution function may depend on θ but is bounded over the parameter space [see Dufour (1997)]. When we have a pivotal function (or a boundedly pivotal function), we can find a point c such that:

$$\mathsf{P}_F[S_n(\theta_1; X_1, \ldots, X_n) \ge c] \le \alpha , \ \forall \theta_1 . \tag{3.18}$$

For example, if $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N[\mu, \sigma^2]$, the t-statistic

$$t_n(\mu) = \sqrt{n}(\bar{X}_n - \mu)/s_X \tag{3.19}$$

where $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ and $s_X = \sum_{i=1}^{n} (X_i - \bar{X}_n)/(n-1)$, follows a Student $t(n-1)$ distribution which does not depend on the unknown values of μ and σ; hence, it is a pivot. By contrast, $\sqrt{n}(\bar{X}_n - \mu)$ is not a pivot because its distribution depends on σ. More generally, in the classical linear model with several regressors, the t statistics for individual coefficients [say, $t(\beta_i) = \sqrt{n}(\hat{\beta}_i - \beta_i)/\hat{\sigma}_{\hat{\beta}_i}$]

constitute pivots because their distributions do not depend on unknown nuisance parameters; in particular, the values of the other regression coefficients disappear from the distribution.

3.4. Testability and identification

When formulating and trying to solve test problems, two types of basic difficulties can arise. First, there is no valid test that satisfies reasonable properties [such as depending upon the data]: in such a case, we have a *non-testable hypothesis*, an empirically *empty hypothesis*. Second, the proposed statistic cannot be pivotal for the model considered: its distribution varies too much under the null hypothesis to determine a *finite* critical point satisfying the level restriction (3.18).

If an hypothesis is non-testable, we are not able to design a reasonable procedure for deciding whether it holds (without the introduction of additional data or information). This difficulty is closely related to the concept of *identification* in econometrics. A parameter θ is *identifiable* iff

$$\theta(F_1) \neq \theta(F_2) \Longrightarrow F_1 \neq F_2\,. \tag{3.20}$$

For $\theta_1 \neq \theta_2$, we can, in principle, design a procedure for deciding whether $\theta = \theta_1$ or $\theta = \theta_2$. The values of θ are *testable*. More generally, a parametric transformation $g(\theta)$ is identifiable iff

$$g[\theta(F_1)] \neq g[\theta(F_2)] \Longrightarrow F_1 \neq F_2\,. \tag{3.21}$$

Intuitively, these definitions mean that different values of the parameter imply different distributions of the data, so that we may expect to be able to "tell" the difference by looking at the data. This is certainly the case when a unique distribution is associated with each parameter value [for example, we may use the Neyman-Pearson likelihood ratio test to make the decision], but this may not be the case when a parameter covers *several* distributions. In the next section, we examine several cases where this happens.

4. Testability, nonparametric models and asymptotic methods

We will now discuss three examples of test problems that look perfectly well defined and sensible at first sight, but turn out to be ill-defined when we look at them more carefully. These include: (1) testing an hypothesis about a mean when the observations are independent and identically distributed (i.i.d.); (2) testing an hypothesis about a mean (or a median) with heteroskedasticity of unknown form; (3) testing the unit root hypothesis on an autoregressive model whose order can be arbitrarily large.[8]

4.1. Procedures robust to nonnormality

One of the most basic problems in econometrics and statistics consists in testing an hypothesis about a mean, for example, its equality to zero. Tests on regression coefficients in linear regressions

[8] Further discussion on the issues discussed in this section is available in Dufour (2001). For rrelated discussions, see also Horowitz (2001), Maasoumi (1992) and Pötscher (2002).

or, more generally, on parameters of models which are estimated by the generalized method of moments (GMM) can be viewed as extensions of this fundamental problem. If the simplest versions of the problem have no reasonable solution, the situation will not improve when we consider more complex versions (as done routinely in econometrics).

The problem of testing an hypothesis about a mean has a very well known and neat solution when the observations are independent and identically (i.i.d.) distributed according to a normal distribution: we can use a t test. The normality assumption, however, is often considered to be too "strong". So it is tempting to consider a weaker (less restrictive) version of this null hypothesis, such as

$$H_0(\mu_0) : X_1 \,, \, \dots \,, \, X_n \text{ are i.i.d. observations with } E(X_1) = \mu_0 \,. \tag{4.1}$$

In other words, we would like to test the hypothesis that the observations have mean μ_0, under the general assumption that $X_1, \dots, X_n$ are i.i.d. Here $H_0(\mu_0)$ is a nonparametric hypothesis because the distribution of the data cannot be completely specified by fixing a *finite* number of parameters. The set of possible data distributions (or data generating processes) compatible with this hypothesis, *i.e.*,

$$\mathcal{H}(\mu_0) = \{\text{Distribution functions} \quad F_n \in \mathcal{F}_n \text{ such that } H_0(\mu_0) \text{ is satisfied}\} \,, \tag{4.2}$$

is much larger here than in the Gaussian case and imposes very strong restrictions on the test. Indeed, the set $\mathcal{H}(\mu_0)$ is so large that the following property must hold.

Theorem 4.1 MEAN NON-TESTABILITY IN NONPARAMETRIC MODELS. *If a test has level α for* $H_0(\mu_0)$, i.e.

$$\mathsf{P}_{F_n}[\textit{Rejecting } H_0(\mu_0)] \le \alpha \textit{ for all } F_n \in \mathcal{H}(\mu_0) \,, \tag{4.3}$$

then, for any $\mu_1 \neq \mu_0$,

$$\mathsf{P}_{F_n}[\textit{Rejecting } H_0(\mu_0)] \le \alpha \textit{ for all } F_n \in \mathcal{H}(\mu_1) \,. \tag{4.4}$$

Further, if there is at least one value $\mu_1 \neq \mu_0$ such that

$$\mathsf{P}_{F_n}\big[\textit{Rejecting } H_0(\mu_0)\big] \ge \alpha \textit{ for at least one } F_n \in \mathcal{H}(\mu_1) \,, \tag{4.5}$$

then, for all $\mu_1 \neq \mu_0$,

$$\mathsf{P}_{F_n}\big[\textit{Rejecting } H_0(\mu_0)\big] = \alpha \textit{ for all } F_n \in \mathcal{H}(\mu) \,. \tag{4.6}$$

PROOF. See Bahadur and Savage (1956). ∎

In other words [by (4.4)], if a test has level α for testing $H_0(\mu_0)$, the probability of rejecting $H_0(\mu_0)$ should not exceed the level irrespective how far the "true" mean is from μ_0. Further [by (4.6)], if "by luck" the power of the test gets as high as the level, then the probability of rejecting should be uniformly equal to the level α. Here, the restrictions imposed by the level constraint are so strong that the test cannot have power exceeding its level: it should be insensitive to cases where the null hypothesis does not hold! An *optimal test* (say, at level .05) in such a problem can be run as follows: (1) ignore the data; (2) using a random number generator, produce a realization of a variable U according to a uniform distribution on the interval $(0,\ 1)$, *i.e.*, $U \sim U(0,\ 1)$; (3) reject H_0 if $U \leq .05$. Clearly, this is not an interesting procedure. It is also easy to see that a similar result will hold if we add various nonparametric restrictions on the distribution, such as a *finite variance* assumption.

The above theorem also implies that tests based on the "asymptotic distribution" of the usual t statistic for $\mu = \mu_0$ [$t_n(\mu_0)$ defined in (3.19)] has size one under $H_0(\mu_0)$:

$$\sup_{F_n \in \mathcal{H}(\mu_0)} \mathsf{P}_{F_n}\big[|t_n(\mu_0)| > c\big] = 1 \tag{4.7}$$

for any finite critical value c. In other words, procedures based on the asymptotic distribution of a test statistic have size that deviate arbitrarily from their nominal size.

A way to interpret what happens here is through the distinction between *pointwise convergence* and *uniform convergence*. Suppose, to simplify, that the probability of rejecting $H_0(\mu_0)$ when it is true depends on a single nuisance parameter γ in the following way:

$$P_n(\gamma) \equiv \mathsf{P}_\gamma\big[|t_n(\mu_0)| > c\big] = 0.05 + (0.95)e^{-|\gamma|n} \tag{4.8}$$

where $\gamma \neq 0$. Then, for each value of γ, the test has level 0.05 asymptotically, *i.e.*

$$\lim_{n\to\infty} P_n(\gamma) = 0.05\,, \tag{4.9}$$

but the *size* of the test is one for all sample sizes:

$$\sup_{\gamma>0} P_n(\gamma) = 1\,,\ \text{for all } n\,. \tag{4.10}$$

$P_n(\gamma)$ converges to a level of 0.05 pointwise (for each γ), but the convergence is not uniform, so that the probability of rejection is arbitrarily close to one for γ sufficiently close to zero (for all sample sizes n).

Many other hypotheses lead to similar difficulties. Examples include:

1. hypotheses about various moments of X_t :

$$\begin{aligned} H_0(\sigma^2) &\ :\ X_1, \ldots, X_n \text{ are i.i.d. observations such that } \mathsf{Var}(X_t) = \sigma^2\,, \\ H_0(\mu_p) &\ :\ X_1, \ldots, X_n \text{ are i.i.d. observations such that } E(X_t^p) = \mu_p\,; \end{aligned}$$

2. most hypotheses on the coefficients of a regression (linear or nonlinear), a structural equation (as in SEM), or a more general estimating function [Godambe (1960)]:

$$H_0(\theta_0) : g_t(X_t\,,\,\theta_0) = u_t\ ,\ t = 1\,,\ \ldots\ ,\ T\,,\ \text{where } u_1\,,\ \ldots\ ,\ u_T \text{ are i.i.d.}$$

In econometrics, models of the form $H_0(\theta_0)$ are typically estimated and tested through a variant of the generalized method of moments (GMM), usually with weaker assumptions on the distribution of $u_1 , \ldots , u_T$; see Hansen (1982), Newey and West (1987a), Newey and McFadden (1994) and Hall (1999). To the extent that GMM methods are viewed as a way to allow for "weak assumptions", it follows from the above discussion that they constitute pseudo-solutions of ill-defined problems.

It is important to observe that the above discussion does not imply that all nonparametric hypotheses are non testable. In the present case, the problem of non-testability could be eliminated by choosing another measure of central tendency, such as a *median*:

$$\begin{aligned} H_0^{0.5}(m_0) : \quad & X_1, \ldots , X_n \text{ are i.i.d. continuous r.v.'s such that} \\ & \mathrm{Med}(X_t) = m_0 \ , \ t = 1 , \ldots , T \ . \end{aligned}$$

$H_0^{0.5}(m_0)$ can be easily tested with a sign test [see Pratt and Gibbons (1981, Chapter 2)]. More generally, hypotheses on the quantiles of the distribution of observations in random sample remain testable nonparametrically:

$$\begin{aligned} H_0^{p}(Q_{p0}) : \quad & X_1, \ldots , X_n \text{ are i.i.d. observations such that} \\ & \mathsf{P}\big[X_t \le Q_{p0}\big] = p \ , \ t = 1 , \ldots , T \, . \end{aligned}$$

Moments are not empirically meaningful functionals in nonparametric models (unless strong distributional assumptions are added), though quantiles are.

4.2. Procedures robust to heteroskedasticity of unknown form

Another common problem in econometrics consists in developing methods which remain valid in making inference on regression coefficients when the variances of the observations are not identical (heteroskedasticity). In particular, this may go as far as looking for tests which are "robust to heteroskedasticity of unknown form". But it is not widely appreciated that this involves very strong restrictions on the procedures that can satisfy this requirement. To see this, consider the problem which consists in testing whether n observations are independent with common zero median, namely:

$$\begin{aligned} H_0 : \quad & X_1, \ldots , X_n \text{ are independent random variables} \\ & \text{each with a distribution symmetric about zero.} \end{aligned} \tag{4.11}$$

Equivalently, H_0 states that the joint distribution F_n of the observations belongs to the (huge) set $\mathcal{H}_0 = \{F_n \in \mathcal{F}_n : F_n \text{ satisfies } H_0\}$: H_0 allows heteroskedasticity of unknown form. In such a case, we have the following theorem.

Theorem 4.2 CHARACTERIZATION OF HETEROSKEDASTICITY ROBUST TESTS. *If a test has level α for H_0, where $0 < \alpha < 1$, then it must satisfy the condition*

$$\mathsf{P}[\,\textit{Rejecting } H_0 \mid |X_1|, \ldots , |X_n|\,] \le \alpha \textit{ under } H_0 \ . \tag{4.12}$$

PROOF. See Pratt and Gibbons (1981, Section 5.10) and Lehmann and Stein (1949). ∎

In other words, a valid test with level α must be a *sign test* — or, more precisely, its level must be equal to α conditional on the absolute values of the observations (which amounts to considering a test based on the signs of the observations). From this, the following remarkable property follows.

Corollary 4.3 *If, for all* $0 < \alpha < 1$, *the condition* (4.12) *is not satisfied, then the size of the test is equal to one,* i.e.

$$\sup_{F_n \in \mathcal{H}_0} \mathsf{P}_{F_n}[\textit{Rejecting } H_0] = 1 \; . \tag{4.13}$$

In other words, if a test procedure does not satisfy (4.12) for all levels $0 < \alpha < 1$, then its true size is *one* irrespective of its nominal size. Most so-called "heteroskedasticity robust procedures" based on "corrected" standard errors [see White (1980), Newey and West (1987b), Davidson and MacKinnon (1993, Chapter 16), Cushing and McGarvey (1999)] do not satisfy condition (4.12) and consequently have size one.[9]

4.3. Procedures robust to autocorrelation of arbitrary form

As a third illustration, let us now examine the problem of testing the unit root hypothesis in the context of an autoregressive model whose order is infinite or is not bounded by a prespecified maximal order:

$$X_t = \beta_0 + \sum_{k=1}^{p} \lambda_k X_{t-k} + u_t \; , \; u_t \overset{i.i.d.}{\sim} N[0\,,\,\sigma^2]\,,\; t = 1\,,\,\ldots\,,\,n\,, \tag{4.14}$$

where p is not bounded *a priori*. This type of problem has attracted a lot of attention in recent years.[10] We wish to test:

$$\tilde{H}_0 : \sum_{k=1}^{p} \lambda_k = 1 \tag{4.15}$$

or, more precisely,

$$\begin{aligned} \tilde{H}_0 : \quad & X_t = \beta_0 + \sum_{k=1}^{p} \lambda_k X_{t-k} + u_t \; , \; t = 1\,,\,\ldots\,,\,n\,, \text{ for some } p \geq 0\,, \\ & \sum_{k=1}^{p} \lambda_k = 1 \quad \text{and} \quad u_t \overset{i.i.d.}{\sim} N[0\,,\,\sigma^2]\,. \end{aligned} \tag{4.16}$$

About this problem, we can show the following theorem and corollary.

Theorem 4.4 UNIT ROOT NON-TESTABILITY IN NONPARAMETRIC MODELS. *If a test has level* α *for* $\tilde{H}_0$, i.e.

$$\mathsf{P}_{F_n}[\textit{Rejecting } \tilde{H}_0] \leq \alpha \quad \textit{for all } F_n \textit{ satisfying } \tilde{H}_0 \; , \tag{4.17}$$

then

$$\mathsf{P}_{F_n}[\textit{Rejecting } \tilde{H}_0] \leq \alpha \quad \textit{for all } F_n \; . \tag{4.18}$$

PROOF. See Cochrane (1991) and Blough (1992). ∎

[9] For examples of size distortions, see Dufour (1981) and Campbell and Dufour (1995, 1997).

[10] For reviews of this huge literature, the reader may consult: Banerjee, Dolado, Galbraith, and Hendry (1993), Stock (1994), Tanaka (1996), and Maddala and Kim (1998).

Corollary 4.5 *If, for all* $0 < \alpha < 1$, *the condition* (4.18) *is not satisfied, then the size of the test is equal to one,* i.e.

$$\sup_{F_n \in \mathcal{H}_0} \mathsf{P}_{F_n}[\textit{Rejecting } \tilde{H}_0] = 1$$

where $\mathcal{H}_0$ *is the set of all data distributions* F_n *that satisfy* $\tilde{H}_0$.

As in the mean problem, the null hypothesis is simply too "large" (unrestricted) to allow testing from a finite data set. Consequently, all procedures that claim to offer corrections for very general forms of serial dependence [*e.g.*, Phillips (1987), Phillips and Perron (1988)] are affected by these problems: irrespective of the nominal level of the test, the true size under the hypothesis $\tilde{H}_0$ is equal to one.

To get a testable hypothesis, it is essential to fix jointly the order of the AR process (*i.e.*, a numerical upper bound on the order) and the sum of the coefficients: for example, we could consider the following null hypothesis where the order of the autoregressive process is equal to 12:

$$\begin{aligned} H_0(12): \quad & X_t = \beta_0 + \sum_{k=1}^{12} \lambda_k X_{t-k} + u_t \,,\; t = 1\,,\; \ldots\;,\; n\,, \\ & \sum_{k=1}^{12} \lambda_k = 1 \quad \text{and} \quad u_t \overset{i.i.d.}{\sim} N[0\,,\, \sigma^2]\,. \end{aligned} \tag{4.19}$$

The order of the autoregressive process is an essential part of the hypothesis: it is not possible to separate inference on the unit root hypothesis from inference on the order of the process. Similar difficulties will also occur for most other hypotheses on the coefficients of (4.16). For further discussion of this topic, the reader may consult Sims (1971a, 1971b), Blough (1992), Faust (1996, 1999) and Pötscher (2002).

5. Structural models and weak instruments

Several authors in the past have noted that usual asymptotic approximations are not valid or lead to very inaccurate results when parameters of interest are close to regions where these parameters are no longer identifiable. The literature on this topic is now considerable.[11] In this section, we shall examine these issues in the context of SEM.

[11] See Sargan (1983), Phillips (1984, 1985, 1989), Gleser and Hwang (1987), Koschat (1987), Phillips (1989), Hillier (1990), Nelson and Startz (1990a, 1990b), Buse (1992), Choi and Phillips (1992), Maddala and Jeong (1992), Bound, Jaeger, and Baker (1993, 1995), Dufour and Jasiak (1993, 2001), McManus, Nankervis, and Savin (1994), Angrist and Krueger (1995), Hall, Rudebusch, and Wilcox (1996), Dufour (1997), Shea (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Hall and Peixe (2000), Stock and Wright (2000), Hahn and Hausman (2002a, 2002b, 2002c), Hahn, Hausman, and Kuersteiner (2001), Dufour and Taamouti (2000, 2001b, 2001a), Startz, Nelson, and Zivot (2001), Kleibergen (2001b, 2001a, 2002a, 2002b, 2003), Bekker (2002), Bekker and Kleibergen (2001), Chao and Swanson (2001, 2003), Moreira (2001, 2003a, 2003b), Moreira and Poi (2001), Stock and Yogo (2002, 2003), Stock, Wright, and Yogo (2002)], Wright (2003, 2002), Imbens and Manski (2003), Kleibergen and Zivot (2003), Perron (2003), and Zivot, Startz, and Nelson (2003).

5.1. Standard simultaneous equations model

Let us consider the standard simultaneous equations model:

$$y = Y\beta + X_1\gamma + u\,, \tag{5.1}$$
$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + V\,, \tag{5.2}$$

where y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables, X_1 and X_2 are $T \times k_1$ and $T \times k_2$ matrices of exogenous variables, β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown coefficients, $\mathit{\Pi}_1$ and $\mathit{\Pi}_2$ are $k_1 \times G$ and $k_2 \times G$ matrices of unknown coefficients, $u = (u_1, \ldots, u_T)'$ is a $T \times 1$ vector of structural disturbances, and $V = [V_1, \ldots, V_T]'$ is a $T \times G$ matrix of reduced-form disturbances. Further,

$$X = [X_1,\, X_2] \text{ is a full-column rank } T \times k \text{ matrix} \tag{5.3}$$

where $k = k_1 + k_2$. Finally, to get a finite-sample distributional theory for the test statistics, we shall use the following assumptions on the distribution of u :

$$u \text{ and } X \text{ are independent;} \tag{5.4}$$
$$u \sim N\big[0,\, \sigma_u^2\, I_T\big]\,. \tag{5.5}$$

(5.4) may be interpreted as the strict exogeneity of X with respect to u.

Note that the distribution of V is not otherwise restricted; in particular, the vectors $V_1, \ldots, V_T$ need not follow a Gaussian distribution and may be heteroskedastic. Below, we shall also consider the situation where the reduced-form equation for Y includes a third set of instruments X_3 which are not used in the estimation:

$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + X_3\mathit{\Pi}_3 + V \tag{5.6}$$

where X_3 is a $T \times k_3$ matrix of explanatory variables (not necessarily strictly exogenous); in particular, X_3 may be unobservable. We view this situation as important because, in practice, it is quite rare that one can consider all the relevant instruments that could be used. Even more generally, we could also assume that Y obeys a general nonlinear model of the form:

$$Y = g(X_1,\, X_2,\, X_3,\, V,\, \mathit{\Pi}) \tag{5.7}$$

where $g(\cdot)$ is a possibly unspecified nonlinear function and $\mathit{\Pi}$ is an unknown parameter matrix.

The model presented in (5.1) - (5.2) can be rewritten in reduced form as:

$$y = X_1\pi_1 + X_2\pi_2 + v\,, \tag{5.8}$$
$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + V\,, \tag{5.9}$$

where $\pi_1 = \Pi_1\beta + \gamma$, $v = u + V\beta$, and

$$\pi_2 = \Pi_2\beta \, . \tag{5.10}$$

Suppose now that we are interested in making inference about β.

(5.10) is the crucial equation governing identification in this system: we need to be able to recover β from the values of the regression coefficients π_2 and Π_2. The necessary and sufficient condition for identification is the well-known *rank condition* for the identification of β :

$$\beta \text{ is } \textit{identifiable} \quad \text{iff} \quad \operatorname{rank}(\Pi_2) = G \, . \tag{5.11}$$

We have a *weak instrument problem* when either $\operatorname{rank}(\Pi_2) < k_2$ (*non-identification*), or Π_2 is *close to having deficient rank* [*i.e.*, $\operatorname{rank}(\Pi_2) = k_2$ with strong linear dependence between the rows (or columns) of Π_2]. There is no compelling definition of the notion of *near-nonidentification*, but reasonable characterizations include the condition that $\det(\Pi_2'\Pi_2)$ is "close to zero", or that $\Pi_2'\Pi_2$ has one or several eigenvalues "close to zero".

Weak instruments are notorious for causing serious statistical difficulties on several fronts: (1) parameter estimation; (2) confidence interval construction; (3) hypothesis testing. We now consider these problems in greater detail.

5.2. Statistical problems associated with weak instruments

The problems associated with weak instruments were originally discovered through its consequences for estimation. Work in this area includes:

1. theoretical work on the *exact distribution* of two-stage least squares (2SLS) and other "consistent" structural estimators and test statistics [Phillips (1983), Phillips (1984), Rothenberg (1984), Phillips (1985), Phillips (1989), Hillier (1990), Nelson and Startz (1990a), Nelson and Startz (1990a), Buse (1992), Maddala and Jeong (1992), Choi and Phillips (1992), Dufour (1997)];

2. *weak-instrument* (local to non-identification) *asymptotics* [Staiger and Stock (1997), Wang and Zivot (1998), Stock and Wright (2000)];

3. empirical examples [Bound, Jaeger, and Baker (1995)].

The main conclusions of this research can be summarized as follows.

1. Theoretical results show that the distributions of various estimators depend in a complicated way on unknown nuisance parameters. Thus, they are difficult to interpret.

2. When identification conditions are not satisfied, standard asymptotic theory for estimators and test statistics typically collapses.

3. With weak instruments,

(a) the 2SLS estimator becomes heavily biased [in the same direction as ordinary least squares (OLS)];

(b) the distribution of the 2SLS estimator is quite far from the normal distribution (*e.g.*, bimodal).

4. A striking illustration of these problems appears in the reconsideration by Bound, Jaeger, and Baker (1995) of a study on returns to education by Angrist and Krueger (1991). Using 329000 observations, these authors found that replacing the instruments used by Angrist and Krueger (1991) with randomly generated (totally irrelevant) instruments produced very similar point estimates and standard errors. This result indicates that the original instruments were weak.

For a more complete discussion of estimation with weak instruments, the reader may consult Stock, Wright, and Yogo (2002).

5.3. Characterization of valid tests and confidence sets

Weak instruments also lead to very serious problems when one tries to perform tests or build confidence intervals on the parameters of the model. Consider the general situation where we have two parameters θ_1 and θ_2 [*i.e.*, $\theta = (\theta_1, \theta_2)$] such that θ_2 is no longer identified when θ_1 takes a certain value, say $\theta_1 = \theta_1^0$:

$$L(y \mid \theta_1, \theta_2) \equiv \overline{L}(y \mid \theta_1^0) \; . \tag{5.12}$$

Theorem 5.1 CHARACTERIZATION OF VALID CONFIDENCE SET FOR A POSSIBLY UNIDENTIFIED PARAMETER. *If θ_2 is a parameter whose value is not bounded, then the confidence region C with level $1-\alpha$ for θ_2 must have the following property:*

$$\mathsf{P}_\theta[C \text{ is unbounded}] > 0 \tag{5.13}$$

and, if $\theta_1 = \theta_1^0$,

$$\mathsf{P}_\theta[C \text{ is unbounded}] \geq 1-\alpha \, . \tag{5.14}$$

PROOF. See Dufour (1997). ∎

Corollary 5.2 *If C does not satisfy the property given in the previous theorem, its size must be zero.*

This will be the case, in particular, for any Wald-type confidence interval, obtained by assuming that

$$t_{\widehat{\theta}_2} = \frac{\widehat{\theta}_2 - \theta_2}{\widehat{\sigma}_{\theta_2}} \overset{approx}{\sim} N(0,1) \, , \tag{5.15}$$

which yields confidence intervals of the form $\widehat{\theta}_2 - c\widehat{\sigma}_{\theta_2} \leq \theta_2 \leq \widehat{\theta}_2 + c\widehat{\sigma}_{\theta_2}$, where $\mathsf{P}[|N(0,1)| > c] \leq \alpha$. By the above corollary, this type of interval has level zero, irrespective of the critical value c used:

$$\inf_\theta \mathsf{P}_\theta\left[\widehat{\theta}_2 - c\widehat{\sigma}_{\theta_2} \leq \theta_2 \leq \widehat{\theta}_2 + c\widehat{\sigma}_{\theta_2}\right] = 0 \; . \tag{5.16}$$

In such situations, the notion of standard error loses its usual meaning and does not constitute a valid basis for building confidence intervals. In SEM, for example, this applies to standard confidence intervals based on 2SLS estimators and their asymptotic "standard errors".

Correspondingly, if we wish to test an hypothesis of form $H_0 : \theta_2 = \theta_2^0$, the size of any test of the form

$$\left|t_{\widehat{\theta}_2}(\theta_2^0)\right| = \left|\frac{\widehat{\theta}_2 - \theta_2^0}{\widehat{\sigma}_{\theta_2}}\right| > c(\alpha) \tag{5.17}$$

will deviate arbitrarily from its nominal size. No unique large-sample distribution for $t_{\widehat{\theta}_2}$ can provide valid tests and confidence intervals based on the asymptotic distribution of $t_{\widehat{\theta}_2}$. From a statistical viewpoint, this means that $t_{\widehat{\theta}_2}$ is *not a pivotal function* for the model considered. More generally, this type of problem affects the validity of all *Wald-type* methods, which are based on comparing parameter estimates with their estimated covariance matrix.

By contrast, in models of the form (5.1) - (5.5), the distribution of the LR statistics for most hypotheses on model parameters can be bounded and cannot move arbitrarily: likelihood ratios are boundedly pivotal functions and provide a valid basis for testing and confidence set construction [see Dufour (1997)].

The central conclusion here is: *tests and confidence sets on the parameters of a structural model should be based on proper pivots.*

6. Approaches to weak instrument problems

What should be the features of a satisfactory solution to the problem of making inference in structural models? We shall emphasize here four properties: (1) the method should be based on proper pivotal functions (ideally, a finite-sample pivot); (2) robustness to the presence of weak instruments; (3) robustness to excluded instruments; (4) robustness to the formulation of the model for the explanatory endogenous variables Y (which is desirable in many practical situations).

In the light of these criteria, we shall *first* discuss the Anderson-Rubin procedure, which in our view is the reference method for dealing with weak instruments in the context of standard SEM, *second* the projection technique which provides a general way of making a wide spectrum of tests and confidence sets, and *third* several recent proposals aimed at improving and extending the AR procedure.

6.1. Anderson-Rubin statistic

A solution to testing in the presence of weak instruments has been available for more than 50 years [Anderson and Rubin (1949)] and is now center stage again [Dufour (1997), Staiger and Stock (1997)]. Interestingly, the AR method can be viewed as an alternative way of exploiting "instruments" for inference on a structural model, although it pre-dates the introduction of 2SLS methods in SEM [Theil (1953), Basmann (1957)], which later became the most widely used method for estimating linear structural equations models.[12] The basic problem considered consists in testing

[12]The basic ideas for using instrumental variables for inference on structural relationships appear to go back to Working (1927) and Wright (1928). For an interesting discussion of the origin of IV methods in econometrics, see Stock and Trebbi

the hypothesis

$$H_0(\beta_0) : \beta = \beta_0 \tag{6.1}$$

in model (5.1) - (5.4). In order to do that, we consider an *auxiliary regression* obtained by subtracting $Y\beta_0$ from both sides of (5.1) and expanding the right-hand side in terms of the instruments. This yields the regression

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{6.2}$$

where $\theta_1 = \gamma + \mathit{\Pi}_1(\beta - \beta_0)$, $\theta_2 = \mathit{\Pi}_2(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. Under the null hypothesis $H_0(\beta_0)$, this equation reduces to

$$y - Y\beta_0 = X_1\theta_1 + \varepsilon\,, \tag{6.3}$$

so we can test $H_0(\beta_0)$ by testing $H_0'(\beta_0) : \theta_2 = 0,$ in the auxiliary regression (6.2). This yields the following F-statistic — the Anderson-Rubin statistic — which follows a Fisher distribution under the null hypothesis:

$$AR(\beta_0) = \frac{[SS_0(\beta_0) - SS_1(\beta_0)]/k_2}{SS_1(\beta_0)/(T-k)} \sim F(k_2,\, T-k) \tag{6.4}$$

where $SS_0(\beta_0) = (y - Y\beta_0)'M(X_1)(y - Y\beta_0)$ and $SS_1(\beta_0) = (y - Y\beta_0)'M(X)(y - Y\beta_0)$; for any full-rank matrix A, we denote $P(A) = A(A'A)^{-1}A'$ and $M(A) = I - P(A)$. What plays the crucial role here is the fact that we have instruments (X_2) that can be related to Y but are excluded from the structural equation. To draw inference on the structural parameter β, we "hang" on the variables in X_2 : if we add X_2 to the constrained structural equation (6.3), its coefficient should be zero. For these reasons, we shall call the variables in X_2 *auxiliary instruments*.

Since the latter statistic is a proper pivot, it can be used to build confidence sets for β :

$$C_\beta(\alpha) = \{\beta_0 : AR(\beta_0) \le F_\alpha(k_2,\, T-k)\} \tag{6.5}$$

where $F_\alpha(k_2,\, T-k)$ is the critical value for a test with level α based on the $F(k_2,\, T-k)$ distribution. When there is only one endogenous explanatory variable $(G = 1)$, this set has an explicit solution involving a quadratic inequation, *i.e.*

$$C_\beta(\alpha) = \{\beta_0 : a\beta_0^2 + b\beta_0 + c \le 0\} \tag{6.6}$$

where $a = Y'HY$, $H \equiv M(X_1) - M(X)\,[1 + k_2F_\alpha(k_2,\, T-k)/(T-k)]\,$, $b = -2Y'Hy$, and $c = y'Hy$. The set $C_\beta(\alpha)$ may easily be determined by finding the roots of the quadratic polynomial in equation (6.6); see Dufour and Jasiak (2001) and Zivot, Startz, and Nelson (1998) for details.

When $G > 1$, the set $C_\beta(\alpha)$ is not in general an ellipsoid, but it remains fairly manageable by using the theory of *quadrics* [Dufour and Taamouti (2000)]. When the model is correct and its parameters are well identified by the instruments used, $C_\beta(\alpha)$ is a closed bounded set close to an ellipsoid. In other cases, it can be unbounded or empty. Unbounded sets are highly likely when the model is not identified, so they point to *lack of identification*. Empty confidence sets can occur (with a non-zero probability) when we have more instruments than parameters in the structural

(2003).

equation (5.1), *i.e.* the model is *overidentified*. An empty confidence set means that no value of the parameter vector β is judged to be compatible with the available data, which indicates that the model is misspecified. So the procedure provides as an interesting byproduct a *specification test*.[13]

It is also easy to see that the above procedure remains valid even if the extended reduced form (5.6) is the correct model for Y. In other words, we can leave out a subset of the instruments (X_3) and use only X_2 : the level of the procedure will not be affected. Indeed, this will also hold if Y is determined by the general — possibly nonlinear — model (5.7). The procedure is *robust to excluded instruments* as well as to the *specification of the model for* Y. The power of the test may be affected by the choice of X_2, but its level is not. Since it is quite rare an investigator can be sure relevant instruments have not been left out, this is an important practical consideration.

The AR procedure can be extended easily to deal with linear hypotheses which involve γ as well. For example, to test an hypothesis of the form

$$H_0(\beta_0,\, \gamma_0) : \beta = \beta_0 \text{ and } \gamma = \gamma_0\,, \tag{6.7}$$

we can consider the transformed model

$$y - Y\beta_0 - X_1\gamma_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon\,. \tag{6.8}$$

Since, under $H_0(\beta_0,\, \gamma_0)$,

$$y - Y\beta_0 - X_1\gamma_0 = \varepsilon\,, \tag{6.9}$$

we can test $H_0(\beta_0,\, \gamma_0)$ by testing $H_0'(\beta_0,\, \gamma_0) : \theta_1 = 0$ and $\theta_2 = 0$ in the auxiliary regression (6.8); see Maddala (1974). Tests for more general restrictions of the form

$$H_0(\beta_0,\, \nu_0) : \beta = \beta_0 \text{ and } R\gamma = \nu_0\,, \tag{6.10}$$

where R is a $r \times K$ fixed full-rank matrix, are discussed in Dufour and Jasiak (2001).

The AR procedure thus enjoys several remarkable features. Namely, it is: (1) pivotal in finite samples; (2) robust to weak instruments; (3) robust to excluded instruments; (4) robust to the specification of the model for Y (which can be nonlinear with an unknown form); further, (5) the AR method provides asymptotically "valid" tests and confidence sets under quite weak distributional assumptions (basically, the assumptions that cover the usual asymptotic properties of linear regression); and (6) it can be extended easily to test restrictions and build confidence sets which also involve the coefficients of the exogenous variables, such as $H_0(\beta_0,\, \nu_0)$ in (6.10).

But the method also has its drawbacks. The main ones are: (1) the tests and confidence sets obtained in this way apply only to the full vector β [or $(\beta',\, \gamma')'$]; what can we do, if β has more than one element? (2) power may be low if too many instruments are added (X_2 has too many variables) to perform the test, especially if the instruments are irrelevant; (3) the error normality assumption is restrictive and we may wish to consider other distributional assumptions; (4) the structural equations are assumed to be linear. We will now discuss a number of methods which have been proposed in order to circumvent these drawbacks.

[13] For further discussion of this point, see Kleibergen (2002b).

6.2. Projections and inference on parameter subsets

Suppose now that β [or $(\beta', \gamma')'$] has more than one component. The fact that a procedure with a finite-sample theory has been obtained for "joint hypotheses" of the form $H_0(\beta_0)$ [or $H_0(\beta_0, \gamma_0)$] is not due to chance: since the distribution of the data is determined by the full parameter vector, there is no reason in general why one should be able to decide on the value of a component of β independently of the others. Such a separation is feasible only in special situations, *e.g.* in the classical linear model (without exact multicollinearity). Lack of identification is precisely a situation where the value of a parameter may be determined only after various restrictions (*e.g.*, the values of other parameters) have been imposed. So parametric nonseparability arises here, and inference should start from a simultaneous approach. If the data generating process corresponds to a model where parameters are well identified, precise inferences on individual coefficients may be feasible. This raises the question how one can move from a joint inference on β to its components.

A general approach to this problem consists in using a projection technique. If

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1 - \alpha\,, \tag{6.11}$$

then, for any function $g(\beta)$,

$$\mathsf{P}\big[g(\beta) \in g\,[C_\beta(\alpha)]\,\big] \geq 1 - \alpha\,. \tag{6.12}$$

If $g(\beta)$ is a component of β or (more generally) a linear transformation $g(\beta) = w'\beta$, the confidence set for a linear combination of the parameters, say $w'\beta$ takes the usual form $[w'\tilde{\beta} - \hat{\sigma}z_\alpha,\ w'\tilde{\beta} + \hat{\sigma}z_\alpha]$ with $\tilde{\beta}$ a k-class type estimator of β; see Dufour and Taamouti (2000).[14]

Another interesting feature comes from the fact that the confidence sets obtained in this way are *simultaneous* in the sense of Scheffé. More precisely, if $\{g_a(\beta) : a \in A\}$ is a set of functions of β, then

$$\mathsf{P}\big[g_a(\beta) \in g\,[C_\beta(\alpha)]\ \text{ for all } a \in A\big] \geq 1 - \alpha\,. \tag{6.13}$$

If these confidence intervals are used to test different hypotheses, an unlimited number of hypotheses can be tested without losing control of the overall level.

6.3. Alternatives to the AR procedure

In view of improving the power of AR procedures, a number of alternative methods have been recently suggested. We will now discuss several of them.

a. Generalized auxiliary regression A general approach to the problem of testing $H_0(\beta_0)$ consists in replacing X_2 in the auxiliary regression

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{6.14}$$

[14] $g\,[C_\beta(\alpha)]$ takes the form of a bounded confidence interval as soon as the confidence set $g\,[C_\beta(\alpha)]$ is unbounded. For further discussion of projection methods, the reader may consult Dufour (1990, 1997), Campbell and Dufour (1997), Abdelkhalek and Dufour (1998), Dufour, Hallin, and Mizera (1998), Dufour and Kiviet (1998), and Dufour and Jasiak (2001).

by an alternative set of auxiliary instruments, say Z of dimension $T \times \bar{k}_2$. In other words, we consider the generalized auxiliary regression

$$y - Y\beta_0 = X_1\theta_1 + Z\bar{\theta}_2 + \varepsilon \tag{6.15}$$

where $\bar{\theta}_2 = 0$ under $H_0(\beta_0)$. So we can test $H_0(\beta_0)$ by testing $\bar{\theta}_2 = 0$ in (6.15). Then the problem consists in selecting Z so that the level can be controlled and power may be improved with respect to the AR auxiliary regression (6.14). For example, it is easy to see that the power of the AR test could become low if a large set of auxiliary instruments is used, especially if the latter are weak. So several alternative procedures can be generated by reducing the number of auxiliary instruments (the number of columns in Z).

At the outset, we should note that, if (5.2) were the correct model and $\Pi = [\Pi_1, \Pi_2]$ were known, then an optimal choice from the viewpoint of power consists in choosing $Z = X_2\Pi_2$; see Dufour and Taamouti (2001b). The practical problem, of course, is that Π_2 is unknown. This suggests that we replace $X_2\Pi_2$ by an estimate, such as

$$Z = X_2\tilde{\Pi}_2 \tag{6.16}$$

where $\tilde{\Pi}_2$ is an estimate of the reduced-form coefficient Π_2 in (5.2). The problem then consists in choosing $\tilde{\Pi}$. For that purpose, it is tempting to use the least squares estimator $\hat{\Pi} = (X'X)^{-1}X'Y$. However, $\hat{\Pi}$ and ε are not independent and we continue to face a simultaneity problem with messy distributional consequences. Ideally, we would like to select an estimate $\tilde{\Pi}_2$ which is independent of ε.

b. Split-sample optimal auxiliary instruments If we can assume that the error vectors $(u_t, V_t')'$, $t = 1, \ldots, T$, are independent, this approach to estimating Π may be feasible by using a split-sample technique: a fraction of the sample is used to obtain $\tilde{\Pi}$ and the rest to estimate the auxiliary regression (6.15) with $Z = X_2\tilde{\Pi}_2$. Under such circumstances, by conditioning on $\tilde{\Pi}$, we can easily see that the standard F test for $\bar{\theta}_2 = 0$ is valid. Further, this procedure is robust to weak instruments, excluded instruments as well as the specification of the model for Y [*i.e.*, under the general assumptions (5.6) or (5.7)], as long as the independence between $\tilde{\Pi}$ and ε can be maintained. Of course, using a split-sample may involve a loss of the effective number of observations and there will be a trade-off between the efficiency gain from using a smaller number of auxiliary instruments and the observations that are "sacrificed" to get $\tilde{\Pi}$. Better results tend to be obtained by using a relatively small fraction of the sample to obtain $\tilde{\Pi}$ — 10% for example — and the rest for the main equation. For further details on this procedure, the reader may consult Dufour and Jasiak (2001) and Kleibergen (2002a).[15]

A number of alternative procedures can be cast in the framework of equation (6.15).

c. LM-type GMM-based statistic If we take $Z = Z_{WZ}$ with

$$Z_{WZ} = P[M(X_1)X_2]Y = P[M(X_1)X_2]M(X_1)Y = [M(X_1)X_2]\hat{\Pi}_2\,, \tag{6.17}$$

[15] Split-sample techniques often lead to important distributional simplifications; for further discussion of this type of method, see Angrist and Krueger (1995) and Dufour and Torrès (1998, 2000).

$$\hat{\Pi}_2 \;=\; [X_2'M(X_1)X_2]^{-1}X_2'M(X_1)Y\,, \tag{6.18}$$

the F-statistic [say, $F_{GMM}(\beta_0)$] for $\bar{\theta}_2 = 0$ is a monotonic transformation of the LM-type statistic $LM_{GMM}(\beta_0)$ proposed by Wang and Zivot (1998). Namely,

$$F_{GMM}(\beta_0) = \left(\frac{T-k_1-G}{GT}\right)\frac{LM_{GMM}(\beta_0)}{1-(1/T)LM_{GMM}(\beta_0)} \tag{6.19}$$

where

$$LM_{GMM}(\beta_0) = \frac{(y-Y\beta_0)'P[Z_{WZ}](y-Y\beta_0)}{(y-Y\beta_0)'M(X_1)(y-Y\beta_0)/T}\,. \tag{6.20}$$

Note that $\hat{\Pi}_2$ above is the ordinary least squares (OLS) estimator of Π_2 from the multivariate regression (5.2), so that $F_{GMM}(\beta_0)$ can be obtained by computing the F-statistic for $\theta_2^* = 0$ in the regression

$$y - Y\beta_0 = X_1\theta_1^* + (X_2\hat{\Pi}_2)\theta_2^* + u\,. \tag{6.21}$$

When $k_2 \geq G$, the statistic $F_{GMM}(\beta_0)$ can also be obtained by testing $\theta_2^{**} = 0$ in the auxiliary regression

$$y - Y\beta_0 = X_1\theta_1^{**} + \hat{Y}\theta_2^{**} + u \tag{6.22}$$

where $\hat{Y} = X\hat{\Pi}$. It is also interesting to note that the OLS estimates of $\theta_1^{**}$and θ_2^{**}, obtained by fitting the latter equation, are identical to the 2SLS estimates of θ_1^{**} and θ_2^{**} in the equation

$$y - Y\beta_0 = X_1\theta_1^{**} + Y\theta_2^{**} + u\,. \tag{6.23}$$

The LM_{GMM} test may thus be interpreted as an approximation to the optimal test based on replacing the optimal auxiliary instrument $X_2\Pi_2$ by $X_2\hat{\Pi}_2$. The statistic $LM_{GMM}(\beta_0)$ is also numerically identical to the corresponding LR-type and Wald-type tests, based on the same GMM estimator (in this case, the 2SLS estimator of β).

As mentioned above, the distribution of this statistic will be affected by the fact that $X_2\hat{\Pi}_2$ and u are not independent. In particular, it is influenced by the presence of weak instruments. But Wang and Zivot (1998) showed that the distribution of $LM_{GMM}(\beta_0)$ is bounded by the $\chi^2(k_2)$ asymptotically. When $k_2 = G$ (usually deemed the "just-identified" case, although the model may be under-identified in that case), we see easily [from (6.21)] that $F_{GMM}(\beta_0)$ is (almost surely) identical with the AR statistic, *i.e.*

$$F_{GMM}(\beta_0) = AR(\beta_0) \text{ if } k_2 = G\,, \tag{6.24}$$

so that $F_{GMM}(\beta_0)$ follows an exact $F(G,\,T-k)$ distribution, while for $k_2 > G$,

$$G\,F_{GMM}(\beta_0) \leq \left(\frac{T-k_1-G}{T-k_1-k_2}\right)k_2\,AR(\beta_0)\,, \tag{6.25}$$

so that the distribution of $LM_{GMM}(\beta_0)$ can be bounded in finite samples by the distribution of a monotonic transformation of a $F(k_2,\,T-k)$ variable [which, for T large, is very close to the $\chi^2(k_2)$

distribution]. But, for T reasonably large, $AR(\beta_0)$ will always reject when $F_{GMM}(\beta_0)$ rejects (at a given level), so the power of the AR test is uniformly superior to that of the LM_{GMM} bound test.[16]

d. Kleibergen's K test If we take $Z = Z_K$ with

$$
\begin{aligned}
Z_K &= P(X)\left[Y - (y - Y\beta_0)\frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)}\right] = X\tilde{\Pi}(\beta_0) \equiv \tilde{Y}(\beta_0)\,, && (6.26)\\
\tilde{\Pi}(\beta_0) &= \hat{\Pi} - \hat{\pi}(\beta_0)\frac{s_{\varepsilon V}(\beta_0)}{s_{\varepsilon\varepsilon}(\beta_0)}\,, \quad \hat{\Pi} = (X'X)^{-1}X'Y\,, && (6.27)\\
\hat{\pi}(\beta_0) &= (X'X)^{-1}X'(y - Y\beta_0)\,, \quad s_{\varepsilon V}(\beta_0) = \frac{1}{T-k}(y - Y\beta_0)'M(X)Y\,, && (6.28)\\
s_{\varepsilon\varepsilon}(\beta_0) &= \frac{(y - Y\beta_0)'M(X)(y - Y\beta_0)}{T-k}\,, && (6.29)
\end{aligned}
$$

we obtain a statistic, which reduces to the one proposed by Kleibergen (2002a) for $k_1 = 0$. More precisely, with $k_1 = 0$, the F-statistic $F_K(\beta_0)$ for $\bar{\theta}_2 = 0$ is equal to Kleibergen's statistic $K(\beta_0)$ divided by G :

$$F_K(\beta_0) = K(\beta_0)/G\,. \tag{6.30}$$

This procedure tries to correct the simultaneity problem associated with the use of $\hat{Y}$ in the LM_{GMM} statistic by "purging" it from its correlation with u [by subtracting the term $\hat{\pi}(\beta_0)s_{\varepsilon V}(\beta_0)/s_{\varepsilon\varepsilon}(\beta_0)$ in Z_K] . In other words, $F_K(\beta_0)$ and $K(\beta_0) \equiv G\,F_K(\beta_0)$ can be obtained by testing $\bar{\theta}_2 = 0$ in the regression

$$y - Y\beta_0 = X_1\theta_1 + \tilde{Y}(\beta_0)\bar{\theta}_2 + \varepsilon \tag{6.31}$$

where the fitted values $\hat{Y}$, which appear in the auxiliary regression (6.22) for the LM_{GMM} test, have been replaced by $\tilde{Y}(\beta_0) = \hat{Y} - X\hat{\pi}(\beta_0)s_{\varepsilon V}(\beta_0)/s_{\varepsilon\varepsilon}(\beta_0)$, which are closer to being orthogonal with u.

If $k_2 = G$, we have $F_K(\beta_0) = AR(\beta_0) \sim F(G,\, T-k)$, while in the other cases ($k_2 \geq G$), we can see easily that the bound for $F_{GMM}(\beta_0)$ in (6.25) also applies to $F_K(\beta_0)$:

$$G\,F_K(\beta_0) \leq \left(\frac{T - k_1 - G}{T - k_1 - k_2}\right) k_2\, AR(\beta_0)\,, \tag{6.32}$$

Kleibergen (2002a) did not supply a finite-sample distributional theory but showed (assuming $k_1 = 0$) that $K(\beta_0)$ follows a $\chi^2(G)$ distribution asymptotically under $H_0(\beta_0)$, irrespective of the presence of weak instruments. This entails that the $K(\beta_0)$ test will have power higher than the one of LM_{GMM} test [based on the $\chi^2(k_2)$ bound], at least in the neighborhood of the null hypothesis, although not necessarily far away from the null hypothesis.

It is also interesting to note that the inequality (6.32) indicates that the distribution of $K(\beta_0) \equiv G\,F_K(\beta_0)$ can be bounded in finite samples by a $[k_2(T-k_1-G)/(T-k)]F(k_2,\, T-k)$ distribution.

[16]The $\chi^2(k_2)$ bound also follows in a straightforward way from (6.25). Note that Wang and Zivot (1998) do not provide the auxiliary regression interpretation (6.21) - (6.22) of their statistics. For details, see Dufour and Taamouti (2001b).

However, because of the stochastic dominance of $AR(\beta_0)$, there would be no advantage in using the bound to get critical values for $K(\beta_0)$, for the AR test would then have better power.

In view of the fact that the above procedure is based on estimating the mean of $X\Pi$ (using $X\hat{\Pi}$) and the covariances between the errors in the reduced form for Y and u [using $s_{\varepsilon V}(\beta_0)$], it can become quite unreliable in the presence of excluded instruments.

e. Likelihood ratio test The likelihood ratio (LR) statistic for $\beta = \beta_0$ was also studied by Wang and Zivot (1998). The LR test statistic in this case takes the form:

$$LR_{LIML} = T\left[\ln\bigl(\kappa(\beta_0)\bigr) - \ln\bigl(\kappa(\hat{\beta}_{LIML})\bigr)\right] \tag{6.33}$$

where $\hat{\beta}_{LIML}$ is the limited information maximum likelihood estimator (LIML) of β and

$$\kappa(\beta) = \frac{(y - Y\beta)'M(X_1)(y - Y\beta)}{(y - Y\beta)'M(X)(y - Y\beta)}\,. \tag{6.34}$$

Like LM_{GMM}, the distribution of LR_{LIML} depends on unknown nuisance parameters under $H_0(\beta_0)$, but its asymptotic distribution is $\chi^2(k_2)$ when $k_2 = G$ and bounded by the $\chi^2(k_2)$ distribution in the other cases [a result in accordance with the general LR distributional bound given in Dufour (1997)]. This bound can also be easily derived from the following inequality:

$$LR_{LIML} \leq \left(\frac{T}{T-k}\right) k_2\, AR(\beta_0)\,, \tag{6.35}$$

so that the distribution of LR_{LIML} is bounded in finite samples by the distribution of a $[Tk_2/(T - k)]F(k_2,\, T - k)$ variable; for details, see Dufour and Khalaf (2000). For T reasonably large, this entails that the $AR(\beta_0)$ test will have power higher than the one of LR_{LIML} test [based on the $\chi^2(k_2)$ bound], at least in the neighborhood of the null hypothesis. So the power of the AR test is uniformly superior to the one of the LR_{LIML} bound test. Because the LR test depends heavily on the specification of the model for Y, it is not robust to excluded instruments.

f. Conditional tests A promising approach was recently proposed by Moreira (2003a). His suggestion consists in conditioning upon an appropriately selected portion of the sufficient statistics for a gaussian SEM. On assuming that the covariance matrix of the errors is known, the corresponding conditional distribution of various test statistics for $H_0(\beta_0)$ does not involve nuisance parameters. The conditional distribution is typically not standard but may be established by simulation. Such an approach may lead to power gains. On the other hand, the assumption that error covariances are known is rather implausible, and the extension of the method to the case where the error covariance matrix is unknown is obtained at the expense of using a large-sample approximation. Like Kleibergen's procedure, this method yields an asymptotically similar test. For further discussion, see Moreira and Poi (2001) and Moreira (2003b).

g. Instrument selection procedures Systematic search methods for identifying relevant instruments and excluding unimportant instruments have been discussed by several authors; see Hall, Rudebusch, and Wilcox (1996), Hall and Peixe (2000), Dufour and Taamouti (2001a), and Donald and Newey (2001). In this setup, the power of AR-type tests depends on a function of model parameters called the *concentration coefficient*. One way to approach instrument selection is to maximize the concentration coefficient towards maximizing test power. Robustness to instrument exclusion is very handy in this context. For further discussion, the reader may consult Dufour and Taamouti (2001a).

To summarize, in special situations, alternatives to the AR procedure may allow some power gains with respect to the AR test with an unreduced set of instruments. They themselves may have some important drawbacks. In particular, (1) only an asymptotic distributional theory is supplied, (2) the statistics used are not pivotal in finite samples, although Kleibergen's and Moreira's statistics are asymptotically pivotal, (3) they are not robust to instrument exclusion or to the formulation of the model for the explanatory endogenous variables. It is also of interest to note that finite-sample versions of several of these asymptotic tests may be obtained by using split-sample methods.

All the problems and techniques discussed above relate to sampling-based statistical methods. SEM can also be analyzed through a Bayesian approach, which may alleviate the indeterminacies associated with identification via the introduction of a prior distribution on the parameter space. Bayesian inferences always depend on the choice of prior distribution (a property viewed as undesirable in the sampling approach), but this dependence becomes especially strong when identification problems are present [see Gleser and Hwang (1987)]. This paper only aims at discussing problems and solutions which arise within the sampling framework, and it is beyond its scope to debate the advantages and disadvantages of Bayesian methods under weak identification. For additional discussion on this issue, see Kleibergen and Zivot (2003) and Sims (2001).

7. Extensions

We will discuss succinctly some extensions of the above results to multivariate setups (where several structural equations may be involved), models with non-Gaussian errors, and nonlinear models.

7.1. Multivariate regression, simulation-based inference and nonnormal errors

Another approach to inference on a linear structural equation model is based on observing that the structural model (5.1) - (5.4) can be put in the form of a multivariate linear regression (MLR):

$$\bar{Y} = XB + U \tag{7.1}$$

where $\bar{Y} = [y, Y]$, $B = [\pi, \Pi]$, $U = [u, V] = [\tilde{U}_1, \dots, \tilde{U}_T]'$, $\pi = [\pi_1', \pi_2']'$, $\Pi = [\Pi_1', \Pi_2']'$, $\pi_1 = \Pi_1\beta + \gamma$ and $\pi_2 = \Pi_2\beta$.[17] This model is linear except for the nonlinear restriction $\pi_2 = \Pi_2\beta$. Let us now make the assumption that the errors in the different equations for each observation, $\tilde{U}_t$, satisfy the property:

$$\tilde{U}_t = JW_t\,, \; t = 1, \dots, T\,, \tag{7.2}$$

[17] Most of this section is based on Dufour and Khalaf (2000, 2001, 2002).

where the vector $w = vec(W_1, \ldots, W_n)$ has a known distribution and J is an unknown nonsingular matrix (which enters into the covariance matrix Σ of the error vectors $\tilde{U}_t$). This distributional assumption is, in a way, more restrictive than the one made in section 5.1 — because of the assumption on V — and in another way, less restrictive, because the distribution of u is not taken to be necessarily $N\left[0, \sigma_u^2 I_T\right]$.

Consider now an hypothesis of the form

$$H_0 : RBC = D \tag{7.3}$$

where R, C and D are fixed matrices. This is called a *uniform linear (UL) hypothesis*; for example, the hypothesis $\beta = \beta_0$ tested by the AR test can be written in this form [see Dufour and Khalaf (2000)]. The corresponding gaussian LR statistic is

$$LR(H_0) = T \ln(|\hat{\Sigma}_0|/|\hat{\Sigma}|) \tag{7.4}$$

where $\hat{\Sigma} = \hat{U}'\hat{U}/T$ and $\hat{\Sigma}_0 = \hat{U}_0'\hat{U}_0/T$ are respectively the unrestricted and restricted estimates of the error covariance matrix. The AR test can also be obtained as a monotonic transformation of a statistic of the form $LR(H_0)$. An important feature of $LR(H_0)$ in this case is that its distribution under H_0 does not involve nuisance parameters and may be easily simulated (it is a pivot); see Dufour and Khalaf (2002). In particular, its distribution is completely invariant to the unknown J matrix (or the error covariance matrix). In such a case, even though this distribution may be complicated, we can use *Monte Carlo test techniques* — a form of parametric bootstrap — to obtain exact test procedures.[18] Multivariate extensions of AR tests, which impose restrictions on several structural equations, can be obtained in this way. Further, this approach allows one to consider any (possibly non-gaussian) distribution on w.

More generally, it is of interest to note that the LR statistic for about any hypothesis on B can be bounded by a LR statistic for an appropriately selected UL hypothesis: setting $b = vec(B)$ and

$$H_0 : Rb \in \Delta_0 \tag{7.5}$$

where R an arbitrary $q \times k(G+1)$ matrix and Δ_0 is an arbitrary subset of $\mathbb{R}^q$, the distribution of the corresponding LR statistic can be bounded by the LR statistic for a UL hypothesis (which is pivotal). This covers as special cases all restrictions on the coefficients of SEM (as long as they are written in the MLR form). To avoid the use of such bounds (which may be quite conservative), it is also possible to use *maximized Monte Carlo* tests [Dufour (2002)].

All the above procedures are valid for parametric models that specify the error distribution up to an unknown linear transformation (the J matrix) which allows an unknown covariance matrix. It is easy to see that these (including the exact procedures discussed in section 6) yield "asymptotically valid" procedures under much weaker assumptions than those used to obtain finite-sample results. However, in view of the discussion in section 4, the pitfalls and limitations of such arguments should be remembered: there is no substitute for a provably exact procedure.

If we aim at getting tests and confidence sets for nonparametric versions of the SEM (where the

[18] For further discussion of Monte Carlo test methods, see Dufour and Khalaf (2001) and Dufour (2002).

error distribution involves an infinite set of nuisance parameters), this may be achievable by looking at distribution-free procedures based on permutations, ranks or signs. There is very little work on this topic in the SEM. For an interesting first look, however, the reader should look at an interesting recent paper by Bekker (2002).

7.2. Nonlinear models

It is relatively difficult to characterize identification and study its consequences in nonlinear structural models. But problems similar to those noted for linear SEM do arise. Nonlinear structural models typically take the form:

$$f_t(y_t,\, x_t,\, \theta) = u_t\,, \quad \mathsf{E}_\theta[u_t \,|\, Z_t] = 0\,, \quad t,\, \dots\,,\, T, \tag{7.6}$$

where $f_t(\cdot)$ is a vector of possibly nonlinear relationships, y_t is a vector endogenous variables, x_t is a vector of exogenous variables, θ is vector of unknown parameters, Z_t is a vector of conditioning variables (or instruments) — usually with a number of additional distributional assumptions — and $\mathsf{E}_\theta[\,\cdot\,]$ refers to the expected value under a distribution with parameter value θ. In such models, θ can be viewed as identifiable if there is only one value of θ [say, $\theta = \bar{\theta}$] that satisfies (7.6), and we have *weak identification* (or weak instruments) when the expected values $\mathsf{E}_{\bar{\theta}}[f_t(y_t,\, x_t,\, \theta) \,|\, Z_t] = 0\,,$ $t,\, \dots\,,\, T$, are "weakly sensitive" to the value of θ.

Research on weak identification in nonlinear models remains scarce. Nonlinearity makes it difficult to construct finite-sample procedures even in models where identification difficulties do not occur. So it is not surprising that results in this area has been mostly based on large-sample approximations. Stock and Wright (2000) studied the asymptotic distributions of GMM-based estimators and test statistics under conditions of weak identification (and weak "high level" asymptotic distributional assumptions). While GMM estimators of θ have nonstandard asymptotic distributions, the objective function minimized by the GMM procedure follows an asymptotic distribution which is unaffected by the presence of weak instruments: it is asymptotically pivotal. So tests and confidence sets based on the objective function can be asymptotically valid irrespective of the presence of weak instruments. These results are achieved for the full parameter vector θ, *i.e.* for hypotheses of the form $\theta = \theta_0$ and the corresponding joint confidence sets. This is not surprising: parametric nonseparability arises here for two reasons, model nonlinearity and the possibility of non-identification. Of course, once a joint confidence set for θ has been built, inference on individual parameters can be drawn via projection methods. Other contributions in this area include papers by Kleibergen (2001a, 2003), who proposed an extension of the $K(\beta_0)$ test, and Wright (2003, 2002) proposed tests of underidentification and identification.

In view the discussion in section 4, the fact that all these methods are based on large-sample approximations without a finite-sample theory remains a concern. However, a first attempt at deriving finite-sample procedures is available in Dufour and Taamouti (2001b). Under parametric assumptions on the errors, the hypothesis $\theta = \theta_0$ is tested by testing $\gamma = 0$ in an auxiliary regression of the form:

$$f_t(y_t, x_t, \theta_0) = z_t(\theta_0,\, \theta_1)\gamma + \varepsilon_t\,, \quad t,\, \dots\,,\, T, \tag{7.7}$$

where the $z_t(\theta_0, \theta_1)$ are instruments in a way that maximizes power against a reference alternative (*point-optimal instruments*). In this way, one gets point-optimal tests [see King (1988) and Dufour and King (1991)]. Inference on nonlinear regressions are also covered by this setup. As in the case of linear SEM, sample-split techniques may be exploited to approximate optimal instruments, and projection methods can be used to draw inference on subvectors of θ.

8. Conclusion

By way of conclusion, we will summarize the main points made in this paper.

1. There are basic pitfalls and limitations faced in developing inference procedures in econometrics. If we are not careful, we can easily be led into *ill-defined problems* and find ourselves:

    (a) trying to *test a non-testable hypothesis*, such as an hypothesis on a moment in the context of an insufficiently restrictive nonparametric model, or an hypothesis (*e.g.*, a unit root hypothesis) on a dynamic model while allowing a dynamic structure with an unlimited (not necessarily infinite) number of parameters;

    (b) trying to solve an inference problem using a technique that cannot deliver a solution *because of the very structure of the technique*, as in (i) testing an hypothesis on a mean (or median) under heteroskedasticity of unknown form, via standard least-square-based "heteroskedasticity-robust" standard errors, or (ii) building a confidence interval for a parameter which is not identifiable in a structural model, via the usual technique based on standard errors. In particular, this type of difficulty arises for Wald-type statistics in the presence of *weak instruments* (or *weakly identified* models)

2. In many econometric problems (such as, inference on structural models), several of the intuitions derived from the linear regression model and standard asymptotic theory can easily be misleading.

    (a) Standard errors do not constitute a valid way of assessing parameter uncertainty and building confidence intervals.

    (b) In many models, such as structural models where parameters may be underidentified, individual parameters in statistical models are not generally meaningful, but parameter vectors can be (at least in parametric models). We called this phenomenon *parametric nonseparability*. As a result, restrictions on individual coefficients may not be testable, while restrictions on the whole parameter vector are. This feature should play a central role in designing methods for dealing with weakly identified models.

3. The above difficulties underscore the pitfalls of large-sample approximations, which are typically based on pointwise (rather than uniform) convergence results and may be arbitrarily inaccurate in finite samples.

4. Concerning solutions to such problems, and more specifically in the context of weakly identified models, we have emphasized the following points.

   (a) In accordance with basic statistical theory, one should always look for *pivots* as the fundamental ingredient for building tests and confidence sets.
   (b) Pivots are not generally available for individual parameters, but they can be obtained in a much wider set of cases for appropriately selected vectors of parameters.
   (c) Given a pivot for a parameter vector, we can construct valid tests and confidence sets for the parameter vector.
   (d) Inference on individual coefficients may then be derived through projection methods.

5. In the specific example of SEM, the following general remarks are in our view important.

   (a) Besides being pivotal, the AR statistic enjoys several remarkable robustness properties, such as robustness to the presence of weak instruments, to excluded instruments or to the specification of a model for the endogenous explanatory variables.
   (b) It is possible to improve the power of AR-type procedures (especially by reducing the number of instruments), but power improvements may come at the expense of using a possibly unreliable large-sample approximation or losing robustness (such as robustness to excluded instruments). As usual, there is a trade-off between power (which is typically increased by considering more restrictive models) and robustness (which involves considering a wider hypothesis).
   (c) Trying to adapt and improve AR-type procedures (without ever forgetting basic statistical principles) constitutes the most promising avenue for dealing with weak instruments.

References

ABDELKHALEK, T., AND J.-M. DUFOUR (1998): "Statistical Inference for Computable General Equilibrium Models, with Application to a Model of the Moroccan Economy," *Review of Economics and Statistics*, LXXX, 520–534.

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 (1991): "Does Compulsory School Attendance Affect Schooling and Earning?," *Quarterly Journal of Economics*, CVI, 979–1014.

ANGRIST, J. D., AND A. B. KRUEGER (1995): "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225–235.

BAHADUR, R. R., AND L. J. SAVAGE (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics*, 27, 1115–1122.

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

BASMANN, R. L. (1957): "A General Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, 25, 77–83.

BEKKER, P. A. (2002): "Symmetry-Based Inference in an Instrumental Variable Setting," Discussion paper, Department of Economics, University of Groningen, Groningen, The Netherlands.

BEKKER, P. A., AND F. KLEIBERGEN (2001): "Finite-Sample Instrumental Variables Inference Using an Asymptotically Pivotal Statistic," Discussion Paper TI 2001-055/4, Tinbergen Institute, University of Amsterdam, Amsterdam, The Netherlands.

BEKKER, P. A., A. MERCKENS, AND T. J. WANSBEEK (1994): *Identification, Equivalent Models, and Computer Algebra*. Academic Press, Boston.

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. BAKER (1993): "The Cure can be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables," Technical Working Paper 137, National Bureau of Economic Research, Cambridge, MA.

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.

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

CAMPBELL, B., AND J.-M. DUFOUR (1995): "Exact Nonparametric Orthogonality and Random Walk Tests," *Review of Economics and Statistics*, 77, 1–16.

——— (1997): "Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter," *International Economic Review*, 38, 151–173.

CHAO, J., AND N. R. SWANSON (2001): "Bias and MSE Analysis of the IV Estimator Under Weak Identification with Application to Bias Correction," Discussion paper, Department of Economics, Rutgers University, New Brunswick, New Jersey.

——— (2003): "Alternative Approximations of the Bias and MSE of the IV Estimator Under Weak Identification With an Application to Bias Correction," Discussion paper, Department of Economics, Rutgers University, New Brunswick, New Jersey.

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.

CUSHING, M. J., AND M. G. MCGARVEY (1999): "Covariance Matrix Estimation," in Mátyás (1999), chap. 3, pp. 63–95.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press, New York.

DONALD, S. G., AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161–1191.

DUFOUR, J.-M. (1981): "Rank Tests for Serial Dependence," *Journal of Time Series Analysis*, 2, 117–128.

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

——— (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

——— (2000): "Économétrie, théorie des tests et philosophie des sciences," in *Présentations de l'Académie des lettres et des sciences humaines*, vol. 53, pp. 166–182. Royal Society of Canada/Société royale du Canada, Ottawa.

——— (2001): "Logique et tests d'hypothèses: réflexions sur les problèmes mal posés en économétrie," *L'Actualité économique*, 77(2), 171–190.

——— (2002): “Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics,” *Journal of Econometrics*, forthcoming.

DUFOUR, J.-M., M. HALLIN, AND I. MIZERA (1998): “Generalized Runs Tests for Heteroskedastic Time Series,” *Journal of Nonparametric Statistics*, 9, 39–86.

DUFOUR, J.-M., AND J. JASIAK (1993): “Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors,” Discussion paper, C.R.D.E., Université de Montréal, 38 pages.

DUFOUR, J.-M., AND J. JASIAK (2001): “Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors,” *International Economic Review*, 42, 815–843.

DUFOUR, J.-M., AND L. KHALAF (2000): “Simulation-Based Finite-Sample Inference in Simultaneous Equations,” Discussion paper, C.R.D.E., Université de Montréal.

——— (2001): “Monte Carlo Test Methods in Econometrics,” in *Companion to Theoretical Econometrics*, ed. by B. Baltagi, Blackwell Companions to Contemporary Economics, chap. 23, pp. 494–519. Basil Blackwell, Oxford, U.K.

——— (2002): “Simulation Based Finite and Large Sample Tests in Multivariate Regressions,” *Journal of Econometrics*, 111(2), 303–322.

DUFOUR, J.-M., AND M. L. KING (1991): “Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors,” *Journal of Econometrics*, 47, 115–143.

DUFOUR, J.-M., AND J. F. KIVIET (1998): “Exact Inference Methods for First-Order Autoregressive Distributed Lag Models,” *Econometrica*, 66, 79–104.

DUFOUR, J.-M., AND M. TAAMOUTI (2000): “Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments,” Discussion paper, C.R.D.E., Université de Montréal.

——— (2001a): “On Methods for Selecting Instruments,” Discussion paper, C.R.D.E., Université de Montréal.

——— (2001b): “Point-Optimal Instruments and Generalized Anderson-Rubin Procedures for Nonlinear Models,” Discussion paper, C.R.D.E., Université de Montréal.

DUFOUR, J.-M., AND O. TORRÈS (1998): “Union-Intersection and Sample-Split Methods in Econometrics with Applications to SURE and MA Models,” in *Handbook of Applied Economic Statistics*, ed. by D. E. A. Giles, and A. Ullah, pp. 465–505. Marcel Dekker, New York.

——— (2000): “Markovian Processes, Two-Sided Autoregressions and Exact Inference for Stationary and Nonstationary Autoregressive Processes,” *Journal of Econometrics*, 99, 255–289.

ENGLE, R. F., AND D. L. MCFADDEN (eds.) (1994): *Handbook of Econometrics, Volume 4*. North-Holland, Amsterdam.

FAUST, J. (1996): "Near Observational Equivalence Problems and Theoretical Size Problems with Unit Root Tests," *Econometric Theory*, 12, 724–732.

FAUST, J. (1999): "Theoretical confidence level problems with confidence intervals for the spectrum of a time series," *Econometrica*, 67, 629–637.

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

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.

GODAMBE, V. P. (1960): "An Optimum Property of Regular Maximum Likelihood Estimation," *The Annals of Mathematical Statistics*, 31, 1208–1212, Ackowledgement 32 (1960), 1343.

GRILICHES, Z., AND M. D. INTRILLIGATOR (eds.) (1983): *Handbook of Econometrics, Volume 1*. North-Holland, Amsterdam.

HAHN, J., AND J. HAUSMAN (2002a): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.

——— (2002b): "Notes on Bias in Estimators for Simultaneous Equation Models," *Economics Letters*, 75, 237–241.

——— (2002c): "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," Discussion paper, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

HAHN, J., J. HAUSMAN, AND G. KUERSTEINER (2001): "Higher Order MSE of Jackknife 2SLS," Discussion paper, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

HALL, A. R. (1999): "Hypothesis Testing in Models Estimated by GMM," in Mátyás (1999), chap. 4, pp. 96–127.

HALL, A. R., AND F. P. M. PEIXE (2000): "A Consistent Method for the Selection of Relevant Instruments," Discussion paper, Department of Economics, North Carolina State University, Raleigh, North Carolina.

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

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

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

HOROWITZ, J. L. (2001): "The Bootstrap and Hypothesis Tests in Econometrics," *Journal of Econometrics*, 100(1), 37–40.

HSIAO, C. (1983): "Identification," in Griliches and Intrilligator (1983), chap. 4, pp. 223–283.

IMBENS, G. W., AND C. F. MANSKI (2003): "Confidence Intervals for Partially Identified Parameters," Discussion paper, Department of Economics, University of California at Berkeley, Berkeley, California.

KING, M. L. (1988): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218, Comments and reply, 219-255.

KLEIBERGEN, F. (2001a): "Testing Parameters in GMM Without Assuming That They are Identified," Discussion Paper TI 01-067/4, Tinbergen Institute, Amsterdam, The Netherlands.

——— (2001b): "Testing Subsets of Structural Coefficients in the IV Regression Model," Discussion paper, Department of Quantitative Economics, University of Amsterdam.

——— (2002a): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

——— (2002b): "Two Independent Statistics That Test Location and Misspecification and Add-Up to the Anderson-Rubin Statistic," Discussion paper, Department of Quantitative Economics, University of Amsterdam.

——— (2003): "Expansions of GMM Statistics That Indicate their Properties under Weak and/or Many Instruments," Discussion paper, Department of Quantitative Economics, University of Amsterdam.

KLEIBERGEN, F., AND E. ZIVOT (2003): "Bayesian and Classical Approaches to Instrumental Variable Regression," *Journal of Econometrics*, 114(1), 29–72.

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*. John Wiley & Sons, New York.

LEHMANN, E. L., AND C. STEIN (1949): "On the Theory of some Non-Parametric Hypotheses," *Annals of Mathematical Statistics*, 20, 28–45.

MAASOUMI, E. (1992): "Fellow's Opinion: Rules of Thumb and Pseudo-Science," *Journal of Econometrics*, 53, 1–4.

MACKINNON, J. G. (2002): "Bootstrap Inference in Econometrics," *Canadian Journal of Economics*, 35(4), 615–645.

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.

MADDALA, G. S., AND I.-M. KIM (1998): *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge, U.K.

MANSKI, C. (1995): *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge and London.

MANSKY, C. (2003): *Partial Identification of Probability Distributions*, Springer Series in Statistics. Springer-Verlag, New York.

MÁTYÁS, L. (ed.) (1999): *Generalized Method of Moments Estimation*. Cambridge University Press, Cambridge, U.K.

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.

MOREIRA, M. J. (2001): "Tests With Correct Size When Instruments Can Be Arbitrarily Weak," Discussion paper, Department of Economics, Harvard University, Cambridge, Massachusetts.

——— (2003a): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

——— (2003b): "A General Theory of Hypothesis Testing in the Simultaneous Equations Model," Discussion paper, Department of Economics, Harvard University, Cambridge, Massachusetts.

MOREIRA, M. J., AND B. P. POI (2001): "Implementing Tests with Correct Size in the Simultaneous Equations Model," *The Stata Journal*, 1(1), 1–15.

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.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in Engle and McFadden (1994), chap. 36, pp. 2111–2245.

NEWEY, W. K., AND K. D. WEST (1987a): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

——— (1987b): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

PERRON, B. (2003): "Semiparametric Weak Instrument Regressions with an Application to the Risk Return Tradeoff," *Review of Economics and Statistics*, 85(2), 424–443.

PHILLIPS, P. C. B. (1983): "Exact Small Sample theory in the Simultaneous Equations Model," in Griliches and Intrilligator (1983), chap. 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.

——— (1987): "Time Series Regression with Unit Root," *Econometrica*, 55(2), 277–301.

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

PHILLIPS, P. C. B., AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

POPPER, K. (1968): *The Logic of Scientific Discovery*. Harper Torchbooks, New York, revised edn.

PÖTSCHER, B. (2002): "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots and Estimation of Long Memory Parameters," *Econometrica*, 70(3), 1035–1065.

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

PRATT, J. W., AND J. D. GIBBONS (1981): *Concepts of Nonparametric Theory*. Springer-Verlag, New York.

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

——— (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. D. Intrilligator, chap. 15, pp. 881–935. North-Holland, Amsterdam.

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

SHEA, J. (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, LXXIX, 348–352.

SIMS, C. (1971a): "Distributed Lag Estimation When the Parameter Space is Explicitly Infinite-Dimensional," *Annals of Mathematical Statistics*, 42, 1622–1636.

SIMS, C. A. (1971b): "Discrete Approximations to Continuous Time Distributed Lags in Econometrics," *Econometrica*, 39, 545–563.

——— (2001): "Thinking About Instrumental Variables," Discussion paper, Department of Economics, Priceton University, Princeton, New Jersey.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STARTZ, R., C. R. NELSON, AND E. ZIVOT (2001): "Improved Inference for the Instrumental Variable Estimator," Discussion paper, Department of Economics, University of Washington.

STOCK, J. H. (1994): "Unit Root, Structural Breaks and Trends," in Engle and McFadden (1994), chap. 46, pp. 2740–2841.

STOCK, J. H., AND F. TREBBI (2003): "Who Invented IV Regression?," *The Journal of Economic Perspectives*, forthcoming.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1097–1126.

STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–529.

STOCK, J. H., AND M. YOGO (2002): "Testing for Weak Instruments in Linear IV Regression," Discussion Paper 284, N.B.E.R., Harvard University, Cambridge, Massachusetts.

——— (2003): "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," Discussion paper, Department of Economics, Harvard University, Cambridge, Massachusetts.

TANAKA, K. (1996): *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*. John Wiley & Sons, New York.

TAYLOR, W. E. (1983): "On the Relevance of Finite Sample Distribution Theory," *Econometric Reviews*, 2, 1–39.

THEIL, H. (1953): "Repeated Least-Squares Applied to Complete Equation Systems," Discussion paper, Central Planing Bureau, The Hague, The Netherlands.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

WHITE, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

WORKING, E. J. (1927): "What Do Statistical Demand Curves Show?," *The Quarterly Journal of Economics*, 41, 212–235.

WRIGHT, J. H. (2002): "Testing the Null of Identification in GMM," Discussion Paper 732, International Finance Discussion Papers, Board of Governors of the Federal Reserve System, Washington, D.C.

——— (2003): "Detecting Lack of Identification in GMM," *Econometric Theory*, 19(2), 322–330.

WRIGHT, P. G. (1928): *The Tariff on Animal and Vegetable Oils*. Macmillan, New York.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.

——— (2003): "Inference in Partially Identified Instrumental Variables Regression with Weak Instruments," Discussion paper, Department of Economics, University of Washington.