Identification, weak instruments and statistical inference in econometrics: problems and solution *

Jean-Marie Dufour †
Université de Montréal

May 2003

* 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

Contents

1. Introduction

Recent developments in econometrics over the last 25 years:

1. new fields of applications linked to the availability of new data, financial data, micro-data, panels, qualitative variables, etc.;
2. new models: multivariate time series models, GARCH, etc.;
3. greater ability to estimate nonlinear models which require an important computational capacity;
4. methods based on simulation: bootstrap, indirect inference, etc.;
5. methods based on weak distributional assumptions: nonparametric methods, asymptotic distributions based on "weak regularity conditions", etc.;
6. discovery of various nonregular problems which require nonstandard distributional theories: unit roots, unidentified (or poorly identified) models, etc.

Last year, James MacKinnon spoke about the use of simulation-based inference methods in econometrics, specifically **bootstrap** methods.
This year, I wish to focus on another issue which has attracted a lot of recent attention recently: **weak instruments**.

The problem of weak instruments is a problem associated with statistical inference in structural models: nonstandard asymptotic theory shows up.
By a structural model, I mean a model where identification problems may arise.
The problem has interest that goes far beyond simultaneous equations and IV regressions, because it underscores:

1. the pitfalls of using large-sample approximations;
2. the importance of going back to basic statistical theory when developing econometric methods.

One of the basic objective of work on bootstrapping and weak instruments is to develop **more reliable test and confidence set** procedures.

Most economists would agree (I hope) that **testability** or the formulation of **testable restrictions** should constitute a central feature of any empirical model.
What is **testability**?

Identification is closely related with testability:

- identification has to do with the possibility of distinguishing different values on the basis of the corresponding data distributions;
- testability deals with the possibility of designing procedures for distinguishing different subsets of parameter values.

Problems of non-identification and non-testability are associated with **bad parameterizations**, **bad choices of parameter representations**.

Two main topics will be discussed

1. testability and the limitations of asymptotic distributional theory;
2. inference in models where weak instruments may appear.

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 must have two basic properties:

1. to restrict the expected behavior of observations: to be **informative;**
   a non-restrictive hypothesis says nothing and, consequently, does not learn us anything: it is

   empirically empty ,
   void **of empirical content ;**

   the more restrictive a model is, the more informative it is, the more interesting it is;
2. to be **compatible with available data**;
   ideally, we would like it to be **true**.

These two criteria are in conflict:

1. **information criterion** $\longrightarrow$ **parsimony** $\longrightarrow$ parametric models, strong assumptions;
2. **compatibility with observed data** $\longrightarrow$ **vague models**, little restrictive $\longrightarrow$ nonparametric models, weak hypotheses.

There is a wide current of thought in philosophy of science that emphasizes

falsifiability as a criterion
for the scientific character of a theory (Popper).

Deterministic models (claim to make arbitrarily precise predictions):

- highly falsifiable;
- always inconsistent with observed data.

Probabilistic models

Most models used in econometrics are **probabilistic**, which has two consequences:

1. **unverifiable**:
   as for any theory that makes an indefinite number of predictions, we can never be sure that the model will not be jeopardized by new data;
2. **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.

3. Basic statistical notions

3.1. Hypotheses

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

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

where X_i takes real values, and let

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

its distribution, where $x = (x_1\,, \, \dots\,,\; x_n)$. We assume: $\bar{F}_n(\cdot) \in \mathcal{F}_n$.
For various reasons, we like to represent distributions in terms of **parameters**.
Two ways of introducing parameters in a model.

Function from a set of probability distributions to vector in some Euclidean space

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

Examples: moments (mean, variance, kurtosis, etc.), quantiles (median, quartiles, etc.) of a distribution.

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 function with a specific form [e.g., corresponding to a Gaussian law]: $\theta = (\mu,\, \sigma^2)$.

A **parametric model** is a model such the distribution of the data is specified up to finite number of (scalar) parameters. Otherwise, it is **nonparametric**.

An **hypothesis** H_0 on $X^{(n)}$ is an assertion which states that

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

where $\mathcal{H}_0$ is a subset of all possible distributions on F_n.
The set $\mathcal{H}_0$ may contain:

- a single distribution (**simple hypothesis**);
- several distributions (**composite hypothesis**).

In particular, if

$$\theta = (\theta_1,\, \theta_2) \tag{3.6}$$

H_0 often takes the following form:

$$\mathcal{H}_0 = \mathcal{H}(F_0,\, \theta_1^0) \equiv \{F(\cdot) : F(x) = F_0(x \,|\, \theta_1, \theta_2) \text{ and } \theta_1 = \theta_1^0\} \tag{3.7}$$

Conventionally, we write in short form::

$$H_0 : \theta_1 = \theta_1^0 \;. \tag{3.8}$$

In such a case,

θ_1 is called the *parameter of interest*,

θ_2 is a *nuisance parameter.*

Interpretation of H_0 : there is at least one distribution in H_0 that can be viewed as a representation compatible the observed "behavior" of $X^{(n)}$.

$$H_0 \text{ is acceptable } \iff \Big((\exists F \in \mathcal{H}_0)\ F \text{ is acceptable}\Big) \tag{3.9}$$

$$H_0 \text{ is inacceptable } \iff \Big((\forall F \in \mathcal{H}_0) \quad F \text{ is inacceptable}\Big) \tag{3.10}$$

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).
Usually, it takes the form:

$$\begin{array}{ll} \text{reject } H_0 & \text{if } S_n(X_1, \ldots, X_n) > c\,, \\ \text{do not reject } H_0 & \text{if } S_n(X_1, \ldots, X_n) \leq c\,. \end{array} \tag{3.11}$$

The test has *level* α iff

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

or, equivalently,

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

The test has *size* α if

$$\sup_{F\in\mathcal{H}_0} \mathsf{P}_F[\text{Rejecting } H_0] = \alpha\,. \tag{3.14}$$

H_0 **is non-testable** testable if we can find a finite number c that satisfies the level restriction.
As the set $\mathcal{H}_0$ gets larger, the test procedure must satisfy more restrictions.
There may be a point where this is not anymore feasible, 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**.

3.3. Confidence sets

If we consider an hypothesis of the form

$$H_0(\theta_1^0) : \theta_1 = \theta_1^0 \, . \tag{3.15}$$

and if we can build a different test for each possible value of θ_1^0,

$$\begin{array}{ll} \text{reject } H_0 & \text{if } S_n(\theta_1^0; X_1, \ldots, X_n) > c(\theta_1^0) \\ \text{do not reject } H_0 & \text{if } S_n(\theta_1^0; X_1, \ldots, X_n) \leq c(\theta_1^0) \end{array} \tag{3.16}$$

we can determine the set of values which are can be viewed as compatible with the data according to the tests considered:

$$C = \left\{ \theta_1^0 : S_n(\theta_1^0; X_1, \ldots, X_n) \leq c(\theta_1^0) \right\} \, . \tag{3.17}$$

If

$$\mathsf{P}_F\left[\text{Rejecting } H_0(\theta_1^0)\right] \leq \alpha \quad \text{for all} \quad F \in \mathcal{H}(F_0, \theta_1^0) \tag{3.18}$$

we have

$$\inf_{\theta_1, \theta_2} \mathsf{P}[\theta_1 \in C] \geq 1 - \alpha \, . \tag{3.19}$$

C is a confidence region with level $1 - \alpha$ for θ_1 .

3.4. Pivots

In practice, *confidence regions* (or *confidence intervals*) were made possible by the discovery of **pivotal functions** (or **pivots**):

$S_n(\theta_1; X_1, \ldots, X_n) \sim$ Distribution without nuisance parameters (or boundable).

We can find a point c such that:

$$\mathsf{P}[S_n(\theta_1; X_1, \ldots, X_n) \geq c] \leq \alpha \, , \; \forall \theta_1 \, . \tag{3.20}$$

3.5. Identification

A parameter θ is **identifiable** iff

$$\theta(F_1) \neq \theta(F_2) \Longrightarrow F_1 \neq F_2 \tag{3.21}$$

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.
A parametric transformation $g(\theta)$ is identifiable iff

$$g[\theta(F_1)] \neq g[\theta(F_2)] \Longrightarrow F_1 \neq F_2 \tag{3.22}$$

3.6. Difficulties

Difficulties: there are problems where:

1. there is no valid test that satisfies reasonable properties [e.g., to depend upon the data]:

   non testable hypothesis ,
   empirically **empty hypothesis** ;

2. the proposed statistics cannot be pivotal.

4. Inference on nonparametric models

4.1. Procedures robust to nonnormality

$$H_0(\mu) : \; X_1 , \dots , \; X_n \text{ are i.i.d. observations such that } E(X_1) = \mu \tag{4.1}$$

We wish to test the hypothesis that $X_1 , \dots , \; X_n$ have mean zero, under the general assumption that the observations $X_1 , \dots , \; X_n$ are i.i.d. Let

$$\mathcal{H}(\mu) = \{\text{Distribution functions} \quad F \in \mathcal{F}_n \text{ such that } H_0(\mu) \text{ is satisfied}\} \; . \tag{4.2}$$

Theorem 4.1 *If a test with level α for $H_0(\mu_0)$, i.e.*

$$\mathsf{P}_F[\textit{Rejecting } H_0(\mu_0)] \leq \alpha \textit{ for all } F \in \mathcal{H}(\mu_0) \; , \tag{4.3}$$

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

$$\mathsf{P}_F[\textit{Rejecting } H_0(\mu_0)] \leq \alpha \textit{ for all } F \in \mathcal{H}(\mu) \; . \tag{4.4}$$

PROOF. See Bahadur and Savage (1956). □

The test must behave in the following way (for a test of level $.05$):

1. we throw away all data to garbage;
2. using a random number generator, produce a realization of $U \sim U(0, \; 1)$;
3. reject H_0 if $U \leq .05$.

Procedures based on the asymptotic distribution of a test statistic have size that deviate arbitrarily from their nominal size.

This problem of non-testability can be viewed as a form of **non-identification** in a wide sense: unless relatively strong distributional assumptions are made, moments are not empirically meaningful.

In the present case, the problem of non-testability could be eliminated by choosing another measure of central tendency, such as a **median**.

$$H_0(m): \; X_1 , \ldots , X_n \text{ are i.i.d. continuous r.v.'s such that} \\ \mathrm{Med}(X_t) = m , \; t = 1 , \ldots , T .$$

$H_0(m)$ can be easily tested with a sign test [see Lehmann and D'Abrera (1975), Pratt and Gibbons (1981)].
Consequences for GMM modelling.

4.2. Procedures robust to autocorrelation of arbitrary form

Consider the problem of testing the unit root hypothesis in the context of an autoregressive model whose order is infinite or in not bounded by a prespecified maximal order:

$$X_t = \beta_0 + \sum_{k=1}^{p} \lambda_k X_{t-k} + u_t\ ,\ t = 1\,,\ \ldots\ ,\ T \tag{4.5}$$

$$u_t \overset{i.i.d.}{\sim} N[0\,,\ \sigma^2] \tag{4.6}$$

where p is not bounded *a priori*. We wish to test:

$$H_0 :\ \sum_{k=1}^{p} \lambda_k = 1\ . \tag{4.7}$$

$$\begin{gathered} H_0 :\ X_t = \beta_0 + \sum_{k=1}^{p} \lambda_k X_{t-k} + u_t\ ,\ .t = 1\,,\ \ldots\ ,\ T\,, \\ \sum_{k=1}^{p} \lambda_k = 1 \text{ and } u_t \overset{i.i.d.}{\sim} N[0\,,\ \sigma^2] \end{gathered} \tag{4.8}$$

Theorem 4.2 *If is a test with level* α *for* H_0, *i.e.*

$$\mathsf{P}_F[\textit{Rejecting } H_0] \le \alpha \quad \textit{for all } F \textit{ satisfying } H_0\ , \tag{4.9}$$

then,

$$\mathsf{P}_F[\textit{Rejecting } H_0] \le \alpha \textit{ for all } F \in \mathcal{H}_0\ . \tag{4.10}$$

PROOF. See Cochrane (1991, Journal of Economic Dynamics and Control) and Blough (1992, Journal of Applied Econometrics). □

Testable hypothesis:

$$
\begin{aligned}
H_0(6): \; X_t = \beta_0 + \sum_{k=1}^{6} \lambda_k X_{t-k} + u_t \,, \; .t = 1\,, \,\ldots\,, \,T\,, \\
\sum_{k=1}^{6} \lambda_k = 1 \text{ and } u_t \overset{i.i.d.}{\sim} N[0\,, \,\sigma^2]
\end{aligned}
\tag{4.11}
$$

Similar difficulties for most hypotheses on the coefficients of (4.8).

Other relevant references: Sims (1971a), Sims (1971b), Blough (1992), Faust (1996), Faust (1999).

4.3. Procedures robust to heteroskedasticity of unknown form

$$\begin{array}{ll} H_0: & X_1, \ldots, X_n \text{ are independent observations} \\ & \text{each one with a distribution symmetric about zero.} \end{array} \tag{4.12}$$

H_0 allows arbitrary heteroskedasticity. Let

$$\mathcal{H}_0 = \{F \in \mathcal{F}_n : F \text{ satisfies } H_0\} \tag{4.13}$$

Theorem 4.3 *If a test has level α for H_0, where $0 \leq \alpha < 1$, then it must satisfy the condition*

$$\mathsf{P}[\textit{Rejecting } H_0 \mid |X_1|, \ldots, |X_n|] \leq \alpha \textit{ under } H_0 . \tag{4.14}$$

PROOF. See Pratt and Gibbons (1981, Concepts of Nonparametric Theory, Section 5.10) and Lehmann and Stein (1949, Annals of Mathematical Statistics). □

The test must be a **sign test** (or, more generally, a sign test conditional on the absolute values of the observations).

Corollary 4.4 *If, for all $0 \leq \alpha < 1$, the condition (4.14) is not satisfied, then the size of the test is equal to one, i.e.*

$$\sup_{F \in \mathcal{H}_0} \mathsf{P}_F[\textit{Rejecting } H_0] = 1 . \tag{4.15}$$

All test procedures typically designated as "robust to heteroskedasticity" (White-type) do not satisfy condition (4.14) and consequently have size one:
For examples of size distortions, see:

Dufour (1981, Journal of Time Series Analysis),

Campbell and Dufour (1995, Review of Economics and Statistics),

Campbell and Dufour (1997, International Economic Review).

5. Inference on structural models and weak instruments

Several authors in the pas 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 not anymore identifiable:

Sargan (1983, Econometrica)
Phillips (1984, International Economic Review)
Phillips (1985, International Economic Review)
Gleser and Hwang (1987, Annals of Statistics)
Koschat (1987, Annals of Statistics)
Phillips (1989, Econometric Theory)
Hillier (1990, Econometrica)
Nelson and Startz (1990a, Journal of Business)
Nelson and Startz (1990b, Econometrica)
Buse (1992, Econometrica)
Maddala and Jeong (1992, Econometrica)
Choi and Phillips (1992, Journal of Econometrics)
Bound, Jaeger, and Baker (1993, NBER Discussion Paper)
Dufour and Jasiak (1993, CRDE)
Bound, Jaeger, and Baker (1995, Journal of the American Statistical Association)
McManus, Nankervis, and Savin (1994, Journal of Econometrics)
Hall, Rudebusch, and Wilcox (1996, International Economic Review)
Dufour (1997, Econometrica)
Shea (1997, Review of Economics and Statistics)
Staiger and Stock (1997, Econometrica)
Wang and Zivot (1998, Econometrica)
Zivot, Startz, and Nelson (1998, International Economic Review)
Startz, Nelson, and Zivot (1999, International Economic Review)
Perron (1999)

Stock and Wright (2000, Econometrica)
Dufour and Jasiak (2001)
Dufour and Taamouti (2001b)
Kleibergen (2001, 2002)
Moreira (2001, 2002)
Stock and Yogo (2002)
Stock, Wright, and Yogo (2002, Journal of Business and Economic Statistics)

5.1. Standard simultaneous equations model

$$y = Y\beta + X_1\gamma + u \tag{5.1}$$

$$Y = X_1\Pi_1 + X_2\Pi_2 + V \tag{5.2}$$

where:

y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables,

X_i is a $T \times k_i$ matrix of exogenous variables (instruments), $i = 1, 2, 3$:

X_1 : exogenous variables included in the structural equation;

X_2 : exogenous variables excluded from the structural equation ;

β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown coefficients;

Π_1 and Π_2 are $k_1 \times G$ and $k_2 \times G$ matrices of unknown coefficients;

u is a vector of structural disturbances;

V is a $T \times G$ matrix of reduced-form disturbances;

$X = [X_1, X_2]$ is a full-column rank $T \times k$ matrix, where $k = k_1 + k_2$;

$$u \text{ and } X \text{ are independent;} \tag{5.3}$$

$$u \sim N\left[0, \sigma_u^2 I_T\right]. \tag{5.4}$$

Variants

$$Y = X_1\Pi_1 + X_2\Pi_2 + X_3\Pi_3 + V \tag{5.5}$$

where $X_3 : T \times k_3$ matrix of explanatory variables (not necessarily strictly exogenous).

$$Y = g(X_1, X_2, X_3, V, \Pi) \tag{5.6}$$

This model can be rewritten in reduced form as:

$$
\begin{aligned}
y &= X_1(\mathit{\Pi}_1\beta+\gamma)+X_2\mathit{\Pi}_2\beta+(u+V\beta) \\
&= X_1\pi_1+X_2\pi_2+v && (5.7)\\
Y &= X_1\mathit{\Pi}_1+X_2\mathit{\Pi}_2+V && (5.8)
\end{aligned}
$$

where $\pi_1 = \mathit{\Pi}_1\beta+\gamma$, $v = u+V\beta$, and

$$\pi_2 = \mathit{\Pi}_2\beta\,. \tag{5.9}$$

Suppose we are interested by making inference about β.
Generalization of an old problem studied by Fieller (1940, 1954): inference on the ratio of two parameters:

$$
\begin{aligned}
q &= \frac{\mu_2}{\mu_1} && (5.10)\\
\mu_2 &= q\mu_1 && (5.11)
\end{aligned}
$$

.
(5.9) is the crucial equation which controls identification in this system: we need to be able to recuperate β from the values of the regression coefficients π_2 and $\mathit{\Pi}_2$.

Rank condition for the identification of β

$$\beta \text{ is } \textbf{identifiable} \text{ iff } \operatorname{rank}(\mathit{\Pi}_2) = k_2\,. \tag{5.12}$$

Weak instrument problem when:

1. $\operatorname{rank}(\mathit{\Pi}_2) < k_2$ (**nonidentification**)
2. or $\mathit{\Pi}_2$ is **close to being nonidentifiable**:
   (a) $\det(\mathit{\Pi}_2'\mathit{\Pi}_2)$ is “close to zero”;
   (b) $\mathit{\Pi}_2'\mathit{\Pi}_2$ has one or several eigenvalues “close to zero”.

Weak instruments have been notorious to cause serious statistical difficulties, form the viewpoints of

1. estimation;
2. confidence interval construction;
3. testing.

5.2. Problems

The problems associated with weak instruments were originally discovered through its consequences on estimation.

1. Theoretical work on the **exact distribution** of 2SLS and other "consistent" structural estimators and test statistics:
   Phillips (1983, Handbook of Econometrics 1), Phillips (1984, International Economic review), Rothenberg (1984), Phillips (1985), Phillips (1989, Econometric Theory), Hillier (1990, Econometrica), Nelson and Startz (1990a, Journal of Business), Nelson and Startz (1990a, Journal of Business), Buse (1992, Econometrica), Maddala and Jeong (1992, Econometrica), Choi and Phillips (1992, Journal of Econometrics), Dufour (1997, Econometrica).
2. **Weak-instrument** (local to non-identification) **asymptotics**:
   Staiger and Stock (1997, Econometrica), Stock and Wright (2000).
3. Empirical example:
   Bound, Jaeger, and Baker (1995, Journal of the American Statistical Association).

Results

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

2. When identification condition do not hold, standard asymptotic for estimators and test statistics typically collapses.

3. With weak instruments,

   (a) 2SLS becomes heavily biased (in the same direction as OLS)

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

4. Problems were strikingly illustrated by the reconsideration by Bound, Jaeger, and Baker (1995, Journal of the American Statistical Association) of a study on returns to education by Angrist and Krueger (1991, QJE):

   329000 observations;

   replacing the instruments used by Angrist and Krueger (1991, QJE) with randomly generated instruments (totally irrelevant) produced very similar point estimates and standard errors;

   indicates that the instruments originally used were weak.

5.3. Characterization of valid tests and confidence sets

Identification or weak instrument problems also lead to very serious problems when we try to perform tests or build confidence intervals on the parameters of a structural model

Consider a situation where we have two parameters θ_1 and θ_2 such that θ_2 stops being identifiable when θ_1 takes a certain value, say $\theta_1 = \theta_1^0$:

$$L(y \mid \theta_1, \theta_2) = \overline{L}(y \mid \theta_1^0) \quad \text{does not depend on } \theta_2 \text{ when} \quad \theta_1 = \theta_1^0 . \tag{5.13}$$

Theorem 5.1 *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:*

$$\mathsf{P}[C \textit{ is unbounded}] > 0 \tag{5.14}$$

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

$$\mathsf{P}[C \textit{ is unbounded}] \geq 1-\alpha \,. \tag{5.15}$$

PROOF. See Dufour (1997, Econometrica). □

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

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

$$t_{\widehat{\theta}_2} = \frac{\widehat{\theta}_2 - \theta_2}{\widehat{\sigma}_{\theta_2}} \overset{approx}{\sim} N(0,1) \quad \text{[or another distribution]} \tag{5.16}$$

hence an interval of the form

$$\widehat{\theta}_2 - c\widehat{\sigma}_{\theta_2} \leq \theta_2 \leq \widehat{\theta}_2 + c\widehat{\sigma}_{\theta_2} \tag{5.17}$$

where $\mathsf{P}[|N(0,1)| > c] \leq \alpha$. This interval has level:

$$\inf_{\theta} \mathsf{P}\left[\widehat{\theta}_2 - c\widehat{\sigma}_{\theta_2} \leq \theta_2 \leq \widehat{\theta}_2 + c\widehat{\sigma}_{\theta_2}\right] = 0 \; . \tag{5.18}$$

In many models, the notion of standard error loses its usual meaning and does not constitute a valid basis for building confidence intervals.
No unique large sample distribution for $t_{\widehat{\theta}_2}$ can provide valid tests and confidence intervals based on

Correspondingly, if we wish to test an hypothesis of form

$$H_0 : \theta_2 = \theta_2^0 \tag{5.19}$$

the size of any test of the form

$$\left|t_{\widehat{\theta}_2}(\theta_2^0)\right| = \left|\frac{\widehat{\theta}_2 - \theta_2^0}{\widehat{\sigma}_{\theta_2}}\right| > c(\alpha) \tag{5.20}$$

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 $t_{\widehat{\theta}_2}$.
$t_{\widehat{\theta}_2}$is not a pivotal function for the model considered.

Central message _ Tests and confidence sets on the parameters of a structural model should be based on proper pivots.

5.4. Solutions

What should the features of a satisfactory solution to the problem of making inference in structural models?

1. 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 (desirable in many desirable situations).

5.4.1. Anderson-Rubin statistic

A solution to the problem of testing in the presence of weak instruments has been available for more that 50 years [Anderson and Rubin (1949, Annals of Mathematical Statistics)].

$$H_0(\beta_0) : \beta = \beta_0 \tag{5.21}$$

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{5.22}$$

where $\theta_1 = \gamma + \mathit{\Pi}_1(\beta - \beta_0)$, and

$$\theta_2 = \mathit{\Pi}_2(\beta - \beta_0)\,, \quad \varepsilon = u + V(\beta - \beta_0). \tag{5.23}$$

$H_0(\beta_0)$ can be tested by testing

$$H_0' : \theta_2 = 0\,. \tag{5.24}$$

F-statistic for H_0'

$$AR(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)} \sim F(k_2, T - k)\,. \tag{5.25}$$

Confidence set for β :

$$C_{\beta}(\alpha) = \{\beta_0 : AR(\beta_0) \leq F_{\alpha}(k_2, T - k)\} . \tag{5.26}$$

This set is not in general an ellipsoid, but it remains fairly manageable by using the theory of **quadrics** [Dufour and Taamouti (2000)].

This set can be:

1. unbounded (lack of identification problem);
2. empty (specification error problem).

Features _ $AR(\beta_0)$ 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 (possibly nonlinear);
5. asymptotically "valid" under much weaker distributional assumptions (linear regression);
6. procedure can be extended easily to test restrictions which also involve the coefficients of the exogenous variables:

$$H_0(\beta_0, \gamma_0) : \beta = \beta_0 \text{ and } \gamma = \gamma_0 . \tag{5.27}$$

Drawbacks:

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. error normality assumption is restrictive;

4. extensions to nonlinear structural equations would be desirable.

5.4.2. Inference of subsets of structural parameters

Inference on individual coefficients can be performed by using a projection approach. If

$$\mathsf{P}[\beta \in C_\beta(\alpha)] \geq 1 - \alpha \tag{5.28}$$

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

$$\mathsf{P}\big[g(\beta) \in g\left[C_\beta(\alpha)\right]\big] \geq 1 - \alpha\,. \tag{5.29}$$

If $g(\beta)$ is a component of β [or a linear transformation $g(\beta) = w'\beta$], the projection-based confidence set can be obtained very easily [Dufour and Taamouti (2000)].

Additional interesting feature _ The confidence sets obtained in this way are *simultaneous* in the sense of Scheffé [see Miller (1981), Savin (1984) and Dufour (1989)]]. Namely, if $\{g_a(\beta) : a \in A\}$ is a set of functions of β, then

$$\mathsf{P}\big[g_a(\beta) \in g\left[C_\beta(\alpha)\right] \text{ for all } a \in A\big] \geq 1 - \alpha\,. \tag{5.30}$$

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. Further discussion of the projection approach is available in Dufour (1997), Abdelkhalek and Dufour (1998) and Dufour and Jasiak (2001).

5.4.3. Other approaches

As mentioned above, the power of the AR procedure may be adversely affected if the number of auxiliary instruments (i.e., the number of columns in X_2) is large. This effect is similar to what happens when too many variables are included in a regression. Even if all the regressors are relevant, the decrease in the number of degrees of freedom induces a loss in efficiency for both estimation and testing. Of course, this deterioration will be relatively important if many irrelevant regressors are included. In the present case, the regressors correspond to instruments: including too many of them (for a given number of observations) can lead to substantial power losses. So many modifications or alternatives to the AR procedure aim at improving its power. I will now discuss succinctly some of these developments.

5.4.3.1. Instrument set reduction.

A unifying view on most of the methods proposed Replace X_2 in

$$y - Y\beta_0 = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{5.31}$$

by an instrumental matrix Z that contains less variables (columns):

$$y - Y\beta_0 = X_1\theta_1 + Z\bar{\theta}_2 + \varepsilon \tag{5.32}$$

$H_0(\beta_0)$ is tested by testing $\bar{\theta}_2 = 0$ with an $F-$test.
Possible choices include:

1. $Z = X\tilde{\Pi}$ where $X = [X_1, X_2]$
   and $\tilde{\Pi}$ is chosen so that way that $\tilde{\Pi}$.
   Sample-splitting can be used for that purpose [Dufour and Jasiak (2001)]. Basic robustness properties of this procedure similar to AR.
2. Approximately optimal instruments Z
   [Dufour and Taamouti (2001b)].
   It is possible to characterize what is the form of the instruments that will maximize the power of the tests (locally or against specific alternatives).

The optimal instruments typically depend on unknown nuisance parameters but it is possible to approximate them.
Split-sample techniques can be used to preserve exactness.

3. Systematic search methods for identifying relevant instruments and excluding unimportant instruments
[Hall, Rudebusch, and Wilcox (1996), Hall and Peixe (2000), Dufour and Taamouti (2001a), Donald and Newey (2001)].
Maximization of concentration parameter.
Robustness to instrument exclusion is very handy in this context.

5.4.3.2. **Pseudo-pivotal statistics.** Other authors have tried to obtain methods with better power properties. For that purpose, various authors

1. Modifications of conventional GMM statistics: LM-type and LR-type statistics [Wang and Zivot (1998, Econometrica), Zivot, Startz, and Nelson (1998, International Economic Review)]
2. Kleibergen's K statistic [Kleibergen (2002, Econometrica)].
3. Moreira's conditional LR statistic [Moreira (2002, Econometrica, forthcoming)].

In special situations, these statistics may allow some power gains with respect to AR test with an unreduced set of instruments.

Drawbacks:

1. Only an asymptotic distributional theory is supplied.
2. Not pivotal in finite samples (although Kleibergen's and Moreira's statistics are asymptotically pivotal).
3. Not robust to instrument exclusion.

4. Not robust to the formulation of the model for the explanatory endogenous variables.

5.5. Extensions

5.5.1. Nonnormal errors

To deal with alternative error distributions, one can take advantage of the pivotal nature of the AR statistic to obtain AR-type finite-sample tests that allow for alternative error distributions.

This can be done by using **Monte Carlo test techniques** [Dufour and Khalaf (1998)].

Because the AR statistic is based on linear regression methods, it is also relatively easy to apply various bootstrap methods.

Bootstrapping may fail if the asymptotic distribution is not pivotal and involves discontinuities.

5.5.2. Nonlinear models

Even though it is more difficult to characterize identification in nonlinear structural models. But the same type of problems does exist.

Work on this topic based large-sample methods may be found in Stock and Wright (2000), Wright (2000).

Finite-sample work is available in Dufour and Taamouti (2001b).

6. Conclusion

1. Basic pitfalls and limitations that one must face in developing inference procedures in econometrics, especially in nonparametric setups.
   Two types of ill-defined problems are frequently met in econometrics.
   (a) Trying to solve a statistical problem for which no reasonable no reasonable solution can possibly exist:
      i. testing an hypothesis on a dynamic model, allowing a dynamic structure (under the null hypothesis) which involves an unlimited (not necessarily infinite) number of parameters;
      ii. testing an hypothesis on a mean in the context of a nonparametric model, *e.g.* assuming that the observations are i.i.d. with a finite mean.
   (b) Trying to solve an inference problem using a technique that cannot deliver a solution **because of the very structure of the technique**:
      two examples:
      i. to test an hypothesis on a mean under an assumption heteroskedasticity of unknown form, using the usual techniques based on correcting least square standard errors ("heteroskedasticity-robust methods").
      ii. in the context of a structural model, to build a confidence interval for a parameter which is not identifiable, using the usual technique based on standard errors. This happens with **weak instrument** problems.
2. In many econometric problems (such as, inference on structural models), several of the intuitions which people get from studying the linear regression model and using 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) Individual parameters in statistical models are not generally meaningful, but parameter vectors are.

i. In nonparametric models, statements about moments of distributions are empirically empty.
ii. In many parametric models (such as structural models with identification difficulties), restrictions on the values of individual coefficients are empirically empty, while restrictions on the whole parameter vector typically are. (**Parametric nonseparability**)

3. Solutions

   (a) Always look for **pivots**.
   (b) Pivots are not generally available for individual parameters but typically are 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.

4. For the IV regressions, the AR statistic is one of the rare pivots available.

   (a) Besides being pivotal, the AR statistic enjoys several remarkable robustness properties.
   (b) It is possible to improve the power of AR-type procedures (especially by reducing the number of instruments), but many recent in that respect do not qualify as satisfactory.
   (c) Trying to adapt and improve AR-type procedures (without ever forgetting basic statistical principles) constitutes the most promising avenue for dealing with weal 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.

BAHADUR, R. R., AND L. J. SAVAGE (1956): "The Nonexistence of Certain Statistical Procedures in Nonparametric Problems," *Annals of Mathematical Statistics*, 27(4), 1115–1122.

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.

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.

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.

——— (1989): “Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions,” *Econometrica*, 57, 335–355.

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

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 (1998): “Simulation-Based Finite and Large Sample Inference Methods in Simultaneous Equations,” Discussion paper, C.R.D.E., Université de Montréal.

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.

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.

FIELLER, E. C. (1940): “The Biological Standardization of Insulin,” *Journal of the Royal Statistical Society (Supplement)*, 7, 1–64.

——— (1954): “Some Problems in Interval Estimation,” *Journal of the Royal Statistical Society, Series B*, 16, 175–185.

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.

GRILICHES, Z., AND M. D. INTRILLIGATOR (eds.) (1984): *Handbook of Econometrics, Volume 2*. North-Holland, Amsterdam.

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.

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

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

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

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

LEHMANN, E. L., AND H. J. M. D'ABRERA (1975): *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco.

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

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

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.

MILLER, JR., R. G. (1981): *Simultaneous Statistical Inference*. Springer-Verlag, New York, second edn.

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

——— (2002): "A Conditional Likelihood Ratio Test for Structural Models," Discussion paper, Department of Economics, Harvard University, Cambridge, Massachusetts.

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.

PERRON, B. (1999): "Semi-Parametric Weak Instrument Regressions with an Application to the Risk Return Trade-Off," Discussion Paper 0199, C.R.D.E., Université de Montréal.

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

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

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

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

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

ROTHENBERG, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in Griliches and Intrilligator (1984), chap. 15, pp. 881–935.

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

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in Griliches and Intrilligator (1984), chap. 14, pp. 827–879.

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.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STARTZ, R., C. R. NELSON, AND E. ZIVOT (1999): "Improved Inference for the Instrumental Variable Estimator," Discussion paper, Department of Economics, University of Washington.

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.

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66(6), 1389–1404.

WRIGHT, J. H. (2000): "Detecting Lack of Identification in GMM," Discussion paper, International Finance Discussion Papers, Board of Governors of the Federal Reserve System, Washington, D.C.

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.