Finite-sample inference in econometrics and statistics

Jean-Marie Dufour

First version: December 1998
This version: November 29, 2006, 4:04pm
1. **Hypothesis testing and nuisance parameters**

Testing an hypothesis $H_0$ usually involves finding a test statistic $T(H_0)$ with 2 characteristics:

1. the stochastic behavior (distribution) of $T(H_0)$ under $H_0$ must be known;

2. the way the distribution of $T(H_0)$ is affected under the alternative must also be known [e.g. $T(H_0)$ may tend to take large or small values with greater possibilities under the alternative].

→ Fundamental that the quantiles of the distribution function of $T(H_0)$ be either **uniquely defined** or (at least) **bounded**.

Otherwise, the behavior of $T(H_0)$ under $H_0$ is **not interpretable** and $T(H_0)$ **cannot** be the basis of a **valid** test of $H_0$.

Common difficulty: nuisance parameters

\[ \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \] \hspace{1cm} (1.1)

\[ H_0 : \theta_1 = \theta_1^0 \] \hspace{1cm} (1.2)

Test: $T(\theta_1^0) > c(\alpha)$ \hspace{1cm} (1.3)
2 basic cases:

1. the distribution of $T(\theta_1^0)$ does not depend on $\theta_2$: it is uniquely determined;

2. the distribution of $T(\theta_1^0)$ depends on $\theta_2$: it is not uniquely determined: $\theta_2$ is a nuisance parameter.

In many econometric and statistical problems, it is difficult to find the exact distribution of test statistics and confidence sets. Two basic reasons:

1. deriving the relevant distributions may require complex calculations (even when there is no nuisance parameter); and / or

2. distribution may involve nuisance parameters.
Most common approach to such distributional problems: use a **large-sample approximation**.

Important characteristic of such approximations in many situations, the asymptotic distribution does not involve nuisance parameters [e.g., N(0,1), chi-square]

→ great flexibility.

Main interest of asymptotic approximations: generate approximations useful in finite-samples

Shortcomings:

1. Finite-sample distribution may involve nuisance parameters

2. Accuracy of the approximation is typically unknown and may be **arbitrarily bad** especially with nuisance parameters (non-uniform convergence)

Approximation arbitrarily bad → Tests statistic not interpretable
2. **Distributions without nuisance parameters**

Deriving analytically the distribution of a test statistic typically involves complex calculations and remain feasible only in special cases. 

$t$ and $F$ distributions in the classical linear model.

Nowadays, it is often possible to simulate the relevant test statistic under the null hypothesis.

1. The distribution of the test statistic – and thus the relevant critical values – can be evaluated to any degree of precision (with a sufficiently large number of replications).

2. A Monte Carlo (MC) test can performed: the size of the test can be perfectly controlled, even with a small number replications [Dwass (1957), Barnard (1963)].
Work exploiting the technique of MC tests in econometrics:

- Dufour and Khalaf (2001, Baltagi, eds, Blackwell)
3. Basic techniques to deal with nuisance parameters

1. **Transforming:**
   Find a transformation that reduces the data for a statistic $T(\theta_0^0)$ whose distribution does not depend on $\theta_2$ [e.g. reduction to a maximal invariant statistic]
   - $t$ and $F$-statistics in classical linear regression
   - reduction of observations in cash or signs

2. **Conditioning**
   on a statistic $S$ such that the conditional distribution of $T(\theta_1^0)$ given $S$ does not depend on $\theta_2$:
   - conditioning on explanatory variables.
   - tests with Neyman structure;
   - permutation tests;

3. **Bounding:**
   find a bound on the distribution of $T(\theta_1^0)$ which is valid irrespective of the unknown value of $\theta_2$:
   \[
   \sup_{\theta_2} P_{(\theta_1^0, \theta_2)} \left[ T(\theta_1^0) > x \right] \leq B_{\theta_1^0}(x)
   \]
   \[
   \inf_{\theta_2} P_{(\theta_1^0, \theta_2)} \left[ T(\theta_1^0) > x \right] \geq C_{\theta_1^0}(x)
   \]
4. Approaches for building bounds procedures

Four approaches:

1. Bounding the statistics of interest by other random variables with more tractable distributions \( \rightarrow \) Bounds on distribution functions

2. Bounding directly the distribution function of interest (or its tail areas) by some function (not necessarily obtained as the distribution function of random variable)

3. Sequential confidence procedures

4. Projection techniques
4.1. **Bounding the statistic of interest by other statistics**

Given a statistic $T$ used in building a test on confidence set with a complicated distribution (possibly involving nuisance parameters), one tries to find other statistics $T_1$ and $T_2$ with more tractable distributions and such that

$$T_1 \leq T \leq T_2,$$

$$P[T_1 \geq x] \leq P[T \geq x] \leq P[T_2 \geq x].$$

Approach applied in:

- Dufour (1989, Econometrica)
- Dufour (1990, Econometrica)
4.2. Bounding tail areas by some function

\[ P_\theta [T \geq x] \leq G(x) \]

where \( G(x) \) is not necessarily obtained from the distribution of a random variables.

1. Exponential inequalities;
2. Chebyshev inequalities (based on second and higher-order moments);

There are cases (e.g. in nonparametric statistics) where such bounds can be used and combined to get fairly tight bounds on tail areas.
Approach used in:
- Dufour and Mahseredjian (1993, Econometric Theory)
- Dufour and Hallin (1991, Econometric Theory)
- Dufour and Hallin (1992a, Econometric Theory)
- Dufour and Hallin (1992b, Journal of Statistical Planning and Inference)
- Dufour and Hallin (1993, JASA)
4.3. Projection techniques

Let
\[ \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}. \] (4.1)

Very often, it is easy to relatively easy to find a joint confidence set for \( \theta \):
\[ \Pr [\theta \in C_\theta(\alpha)] \geq 1 - \alpha, \] (4.2)
then, for any function \( g(\theta) \),
\[ \Pr [g(\theta) \in g[C_\theta(\alpha)]] \geq 1 - \alpha. \] (4.3)
\( g[C_\theta(\alpha)] \) is a confidence set with level \( 1 - \alpha \) for \( g(\theta) \).
For example, we can take
\[ g(\theta) = \theta_2. \] (4.4)
If \( \theta_2 \) is a scalar and \( C_\theta(\alpha) \) is a compact set, then \( g[C_\theta(\alpha)] \) must be an interval.
Approach applied in:

- Dufour (1990, Econometrica)
- Dufour (1997, Econometrica)
- Dufour and Khalaf (2002)
- Dufour and Taamouti (2005, Econometrica)
4.4. Sequential confidence procedure

Useful with nuisance parameters

\[ \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \]  \hspace{1cm} (4.5)

where

- \( \theta_1 \): vector of nuisance parameter
- \( \theta_2 \): vector of parameters of interest  \hspace{1cm} (4.6)

Problem: inference about \( \theta_2 \) (confidence set on test)

Suppose 2 conditions are satisfied:

1. it is possible to build exact confidence set \( C_1 \) for \( \theta_1 \)
   \[ P [ \theta_1 \in C_1] = 1 - \alpha_1 ; \]  \hspace{1cm} (4.7)
2. if \( \theta_1 \) is known, it is possible to build a confidence set \( C_2(\theta_1) \) for \( \theta_2 \) such that
   \[ P[\theta_2 \in C_2(\theta_1)] = 1 - \alpha_2 . \]  \hspace{1cm} (4.8)

Procedure:

1. Build on exact confidence set \( C_1 \) for \( \theta_1 \):
   \[ P [ \theta_1 \in C_1] \geq 1 - \alpha_1 \]
2. Build a simultaneous confidence set $C$ for $\theta_1$ and $\theta_2$:

$$C = \{(\theta_1, \theta_2) : \theta_1 \in C_1, \theta_2 \in C_2(\theta_1)\}$$

$$P[(\theta_1, \theta_2) \in C] \geq 1 - (\alpha_1 + \alpha_2)$$

3. Use a projection (or an intersection) method to deduce conservative (or a liberal) confidence set for $\theta_2$:

$$U = \{\theta_2 : (\theta_1, \theta_2) \in C \text{ for some } \theta_1 \in C_1\}$$

$$P[\theta_2 \in U] \geq 1 - (\alpha_1 + \alpha_2)$$

$$L = \{\theta_2 : (\theta_1, \theta_2) \in C \text{ for all } \theta_1 \in C_1\} \quad (4.9)$$

$$P[\theta_2 \in L] \leq (1 - \alpha_2) + \alpha_1 \quad (4.10)$$

4. Conservative and liberal critical regions can be deduced from their confidence sets:

$\theta_2^0 \notin U$ is a conservative critical region for $H_0: \theta_2 = \theta_2^0$ with level $\alpha \equiv \alpha_1 + \alpha_2$;

$\theta_2 \in L$ is a liberal continual region for $H_0: \theta_2 = \theta_2^0$ with level $\alpha = \alpha_1 - \alpha_2$.

5. By combining a conservative and a liberal confidence region with the same level one gets a gen-
eralized bounds tests

Approach applied to linear regression with AR(1) errors in:

- Dufour (1990, Econometrica)
5. Weak identification

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 not anymore identifiable:

Sargan (1983, Econometrica)
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)
Dufour and Jasiak (1993, CRDE)
Bound, Jaeger, and Baker (1995, Journal of the...
American Statistical Association
Dufour (1997, Econometrica)
Staiger and Stock (1997, Econometrica)
Perron (1999)
Stock and Wright (2000, Econometrica)
Dufour and Taamouti (2001)
Stock and Yogo (2002)
Dufour (2003, Canadian Journal of Economics)
Dufour and Taamouti (2005, Econometrica)
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 conditions do not hold, standard asymptotic theory 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).


   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.
Crucial to use finite-sample approaches to produce reliable inference.

Finite-sample approaches to inference on models involving weak identification

- Dufour (1997, Econometrica)
- Dufour and Taamouti (2005, Econometrica)
Applications

1. Tobin’s $q$

2. Students’ achievements and self-esteem

3. Education and earnings

4. Trade and growth

5. New Keynesian Phillips curves

6. Black’s CAPM
   [Beaulieu, Dufour, and Khalaf (2005)]
References


BEAULIEU, M.-C., J.-M. DUFOUR, AND L. KHALAF (2005): “Testing Black’s CAPM with Possibly Non-Gaussian Errors: An Exact Identification-Robust Simulation-Based Approach,” Discussion paper, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de
recherche en économie quantitative (CIREQ), Université de Montréal.


**Choi, I., and P. C. B. Phillips** (1992): “Asymp-


——— (2006b): “Structural Estimation and Evaluation of Calvo-Style Inflation Models,” Discussion paper, Centre de recherche et développement en économie (CRDE), Université de Montréal, and Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Montréal, Canada.


**Hall, A. R., G. D. Rudebusch, and D. W.**


MADDALA, G. S., AND J. JEONG (1992): “On the Ex-


