Asymptotic theory for linear regression and IV estimation *

Jean-Marie Dufour †
McGill University

First version: November 2011
Revised: December 2011
This version: December 2011
Compiled: December 13, 2011, 11:11

*This work was supported by the William Dow Chair in Political Economy (McGill University), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

† William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

Contents

1. Estimator consistency

Let $y_1, y_2, \ldots$ be a sequence of observations and

$$\hat{\theta}_T = \hat{\theta}_T(y_1, y_2, \ldots, y_T) \tag{1.1}$$

an estimator for a $k \times 1$ parameter vector θ. We say that $\hat{\theta}_T$ is *consistent* (or *weakly consistent*) for θ when

$$\hat{\theta}_T \xrightarrow[T\to\infty]{p} \theta \, . \tag{1.2}$$

This is also written:

$$\underset{T\to\infty}{\mathrm{plim}}\, \hat{\theta}_T = \theta \, . \tag{1.3}$$

This means that

$$\lim_{T\to\infty} \mathsf{P}\left[\left\|\hat{\theta}_T - \theta\right\| > \varepsilon\right] = 0 \, , \, \forall \varepsilon > 0 \tag{1.4}$$

where $\|\cdot\|$ represents the Euclidean distance.

We say that $\hat{\theta}_T$ is strongly *consistent consistent* (or *weakly consistent*) for θ when

$$\hat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta \, , \tag{1.5}$$

i.e., when

$$\mathsf{P}\left[\lim_{T\to\infty} \hat{\theta}_T = \theta\right] = 1 \, . \tag{1.6}$$

It is easy to see that strong consistency entails weak consistency.

We say that $\hat{\theta}_T$ is *asymptotically unbiased* for θ when

$$\lim_{T\to\infty} \mathsf{E}(\hat{\theta}_T) = \theta \, . \tag{1.7}$$

In general, a consistent estimator is not necessarily asymptotically unbiased, for example when the estimator does not have a finite mean. Similarly an asymptotically unbiased estimator may not be consistent, for example if it unbiased but not consistent. In the following proposition, we give a general condition under which asymptotic unbiasedness entails consistency.

1.1 Proposition *If the estimator $\hat{\theta}_T$ satisfies*

$$\lim_{T\to\infty} \mathsf{E}(\hat{\theta}_T) = \theta \tag{1.8}$$

and

$$\lim_{T\to\infty} \mathsf{V}(\hat{\theta}_T) = 0 \, , \tag{1.9}$$

then $\hat{\theta}_T \xrightarrow[T\to\infty]{p} \theta$.

2. Consistency of least squares in linear regression

Let us now consider a linear regression model of the form

$$y = X\beta + \varepsilon \tag{2.1}$$

where β is a fixed $k \times 1$ parameter vector, y and ε are $T \times 1$ vectors, X is a $T \times k$ matrix,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, \; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}, \tag{2.2}$$

$$X = [x_1, x_2, \ldots, x_k] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ x_{T1} & x_{T2} & \cdots & x_{Tk} \end{bmatrix}.$$

Instead of the finite-sample assumptions of the classical linear model, we make the following "asymptotic" assumptions:

$$X'X \text{ is nonsingular with probability one for all } T \geq k \tag{2.3}$$

$$\frac{1}{T}X'X \xrightarrow[T\to\infty]{p} \Sigma_X \text{ where } \det(\Sigma_X) \neq 0\,, \tag{2.4}$$

$$\frac{1}{T}X'\varepsilon \xrightarrow[T\to\infty]{p} 0\,, \tag{2.5}$$

$$\frac{1}{T}\varepsilon'\varepsilon \xrightarrow[T\to\infty]{p} \sigma^2 > 0\,. \tag{2.6}$$

Then, we have:

$$\hat{\beta}_T = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'\varepsilon \tag{2.7}$$

$$= \beta + \left(\frac{1}{T}X'X\right)^{-1}\frac{1}{T}X'\varepsilon \xrightarrow[T\to\infty]{p} \beta + \Sigma_X^{-1}0 = \beta \tag{2.8}$$

and the least squares estimator is (weakly) consistent.

Similarly, the "unbiased" least squares estimator of σ^2,

$$s_T^2 = \frac{1}{T-k}\hat{\varepsilon}'\hat{\varepsilon} \tag{2.9}$$

where $\hat{\varepsilon} = M(X)\varepsilon = [I_T - X\,(X'X)^{-1}\,X']\varepsilon$, satisfies

$$\begin{aligned}
s_T^2 &= \frac{1}{T-k}\varepsilon' M(X)\varepsilon \\
&= \frac{1}{T-k}\varepsilon'\left[I_T - X\left(X'X\right)^{-1}X'\right]\varepsilon \\
&= \frac{1}{T-k}\left[\varepsilon'\varepsilon - \varepsilon' X\left(X'X\right)^{-1}X'\varepsilon\right] \\
&= \frac{T}{T-k}\left[\frac{1}{T}\varepsilon'\varepsilon - \left(\frac{1}{T}X'\varepsilon\right)'\left(\frac{1}{T}X'X\right)^{-1}\frac{1}{T}X'\varepsilon\right] \qquad (2.10)
\end{aligned}$$

where

$$\frac{1}{T}\varepsilon'\varepsilon \xrightarrow[T\to\infty]{p} \sigma^2\,, \qquad (2.11)$$

$$\left(\frac{1}{T}X'\varepsilon\right)'\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'\varepsilon\right) \xrightarrow[T\to\infty]{p} 0'\Sigma_X^{-1}0 = 0\,, \qquad (2.12)$$

so that

$$s_T^2 \xrightarrow[T\to\infty]{p} \sigma^2\,. \qquad (2.13)$$

In other words, s_T^2 is a consistent estimator of σ^2.

If furthermore, $\frac{1}{\sqrt{T}}X'\varepsilon$ satisfies a central limit theorem, namely

$$\frac{1}{\sqrt{T}}X'\varepsilon \xrightarrow[T\to\infty]{L} N\left[0,\, \sigma^2\Sigma_X\right]\,, \qquad (2.14)$$

we have, using (2.7),

$$\begin{aligned}
\sqrt{T}[\hat{\beta}_T - \beta] &= \sqrt{T}\left(\frac{1}{T}X'X\right)^{-1}\frac{1}{T}X'\varepsilon \\
&= \left(\frac{1}{T}X'X\right)^{-1}\frac{1}{\sqrt{T}}X'\varepsilon \\
&\xrightarrow[T\to\infty]{L} N\left[0,\, \sigma^2\Sigma_X^{-1}\right]\,. \qquad (2.15)
\end{aligned}$$

In other words, the distribution of $\sqrt{T}[\hat{\beta}_T - \beta]$ is approximately normal for T large enough. This entails that the distributions of the t and F statistics can be approximated by the distributions obtained under the assumptions of the Gaussian classical linear model. [The details of the arguments to establish asymptotic distributions are not presented in this course.]

3. Instrumental variables

If X and ε are asymptotically correlated, *i.e.*

$$\frac{1}{T}X'\varepsilon \xrightarrow[T\to\infty]{p} \sigma_{X\varepsilon} \neq 0\,, \tag{3.1}$$

we have

$$\hat{\beta}_T = \beta + \left(\frac{1}{T}X'X\right)^{-1} \frac{1}{T}X'\varepsilon \xrightarrow[T\to\infty]{p} \beta + \Sigma_X^{-1}\sigma_{X\varepsilon} \neq \beta \tag{3.2}$$

and the least squares estimator is not consistent for β. Alternative estimation methods are typically required to deal with this problem.

The instrumental variables (IV) method is the simplest alternative to least squares when explanatory variables and disturbances are asymptotically correlated. Instrumental variables can be defined as variables which are (asymptotically) uncorrelated with the disturbance term but still correlated with the variables in X. More precisely, suppose with a $T \times l$ matrix Z of variables with the following properties:

$$\frac{1}{T}Z'\varepsilon \xrightarrow[T\to\infty]{p} 0\,, \tag{3.3}$$

$$Z'Z \text{ and } X'Z \text{ are full rank matrices with probability one for all } T, \tag{3.4}$$

$$\frac{1}{T}Z'Z \xrightarrow[T\to\infty]{p} \Sigma_Z \text{ where } \det(\Sigma_Z) \neq 0\,, \tag{3.5}$$

$$\frac{1}{T}Z'X \xrightarrow[T\to\infty]{p} \Sigma_{ZX} \text{ where } \mathrm{rank}(\Sigma_{ZX}) = k\,. \tag{3.6}$$

Assumption (3.3) means that Z and ε are (asymptotically) uncorrelated (*instrument validity*). Assumption (3.4) means that $Z'Z$ and $X'Z$ are full rank matrices, Assumption (3.5) means they are not (asymptotically) collinear, while Assumption (2.4) means the variables in Z contain information about all the variables in X (asymptotically)

Consider now equation (2.1) and multiply both sides by Z' :

$$Z'y = Z'X\beta + Z'\varepsilon\,. \tag{3.7}$$

If we then multiply by $\frac{1}{T}$, we get:

$$\frac{1}{T}Z'y = \frac{1}{T}Z'X\beta + \frac{1}{T}Z'\varepsilon\,. \tag{3.8}$$

Consider first the case where the number of instruments is equal to the number of explanatory variables $(l = k)$, so that $Z'X$ is a square invertible matrix. In view of assumption (3.3), we expect $\frac{1}{T}Z'\varepsilon$ to be close to zero for T large enough. This suggests to estimate β by solving the equation

$$\frac{1}{T}Z'y = \frac{1}{T}Z'X\beta\,, \tag{3.9}$$

which leads to the estimator:

$$\tilde{\beta} = (Z'X)^{-1}Z'y\,. \tag{3.10}$$

This estimator is called the IV estimator of β based on the instrument Z (in the case where $l = k$). It is easy to see that $\tilde{\beta}$ is consistent for β :

$$\begin{aligned} \tilde{\beta} &= \beta + (Z'X)^{-1}Z'\varepsilon \\ &= \beta + \left(\frac{1}{T}Z'X\right)^{-1}\frac{1}{T}Z'\varepsilon \xrightarrow[T\to\infty]{p} \beta + \Sigma_{ZX}^{-1}0 = \beta \end{aligned} \tag{3.11}$$

It is interesting to note the least squares estimator $\hat{\beta}$ can be viewed as a special case of the IV estimator obtained by taking $Z = X$. Of course, $\hat{\beta}$ will be consistent only if the orthogonality condition (2.5) holds.

Similarly, if we allow the number of instruments to be larger than the number of explanatory variables $(l \geq k)$, suppose temporarily that Z is fixed. Then the covariance matrix of the error term $Z'\varepsilon$ in (3.7) is:

$$\begin{aligned} V\left(Z'\varepsilon\right) &= E\left[Z'\varepsilon\varepsilon'Z\right] \\ &= Z'E(\varepsilon\varepsilon')Z = \sigma^2 Z'Z\,. \end{aligned} \tag{3.12}$$

This suggests to consider the following “generalized least squares” estimator:

$$\tilde{\beta} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y\,. \tag{3.13}$$

If $l = k$, $Z'X$ is a square invertible matrix, so that

$$\begin{aligned} \tilde{\beta} &= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'y \\ &= (Z'X)^{-1}Z'y \end{aligned} \tag{3.14}$$

reduces to the estimator in (3.10). So $\tilde{\beta}$ is also called the IV estimator of β based on the instrument Z (in the general case where $l \geq k$). Again, it is easy to see that $\tilde{\beta}$ is consistent for β :

$$\begin{aligned} \tilde{\beta} &= [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y \\ &= \beta + [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'\varepsilon \\ &= \beta + \left[\left(\frac{1}{T}X'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{T}Z'X\right)\right]^{-1}\left(\frac{1}{T}X'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{T}Z'\varepsilon\right) \\ &\xrightarrow[T\to\infty]{p} \beta + [\Sigma_{ZX}'\Sigma_Z^{-1}\Sigma_{ZX}]^{-1}\Sigma_{ZX}'\Sigma_Z^{-1}0 = \beta\,. \end{aligned} \tag{3.15}$$

If furthermore, $\frac{1}{\sqrt{T}}Z'\varepsilon$ satisfies a central limit theorem, namely

$$\frac{1}{\sqrt{T}}Z'\varepsilon \xrightarrow[T\to\infty]{L} N\left[0, \sigma^2\Sigma_Z\right], \tag{3.16}$$

we find

$$
\begin{aligned}
\sqrt{T}[\tilde{\beta}_T - \beta] &= \sqrt{T}\left[\left(\frac{1}{T}X'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{T}Z'X\right)\right]^{-1}\left(\frac{1}{T}X'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{T}Z'\varepsilon\right) \\
&= \left[\left(\frac{1}{T}X'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{T}Z'X\right)\right]^{-1}\left(\frac{1}{T}X'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{\sqrt{T}}Z'\varepsilon\right) \\
&\xrightarrow[T\to\infty]{L} N\left[0,\, \sigma^2[\Sigma_{ZX}'\Sigma_Z^{-1}\Sigma_{ZX}]^{-1}\right]. \qquad (3.17)
\end{aligned}
$$

Tests based on this distribution can also be derived. [The details are not presented in this course.]