

The basic intuition of econometrics:
**Ordinary Least Squares and Instrumental
Variables Estimation**

Théophile T. Azomahou
UNU-MERIT, Maastricht University

DEIP, Mar 30 - Apr 3, 2009
Montevideo, Uruguay

Definition and aim

Ragnar Frisch, 'Editorial,' *Econometrica*, 1:1, January 1933, p.2

'Econometrics is by no means the same as economic statistics. Nor is it identical with what we call general economic theory, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous with the application of mathematics to economics. Experience has shown that each of these three view-points, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient condition for a real understanding of the **quantitative relations** in modern economic life. It is the *unification* of all three that is powerful. And it is this unification that constitutes econometrics.'

- Aim: understand some **quantitative relations** in economics
- Tools: statistics, probability, mathematics
- Methods: estimation, inference, prediction
- Objects: DATA (observation of economic behaviors)

References

- Basic

- ① Berndt, E.R. 1996. The Practice of Econometrics: Classic and Contemporary. Addison-Wesley.
- ② Damodar N.G. 2002. Basic Econometrics. McGraw-Hill (Chapters: 1-7,9,15,16)
- ③ Greene, W.H. 2003. Econometric Analysis. Prentice Hall. (Chapters: 2,3,13,17,21)

- Advanced

- ① Arellano, M. 2003. Panel Data Econometrics. Oxford University Press.
- ② Davidson, R., J.G. MacKinnon 2004. Econometric Theory and Methods. Oxford University Press.
- ③ Wooldridge, J.M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge MA, MIT Press (Chapters: 4, 5,7[section 7.8],10,13,15)

Outline of the Lecture

- 1 Set-up
- 2 The linear regression: model and assumptions
- 3 The Ordinary Least Squares Estimator (OLS)
- 4 The Instrumental Variable estimation (IV)

Set-up: some data structures in economics

- **Cross section data:** observations on individuals, households, firms, cities, countries within a period of time:

$$X_i \text{ for } i = 1, \dots, N(1, \dots, 1000)$$

- **Time series data:** observation of ONE household, or ONE firm, or ONE country, etc., over time:

$$X_t \text{ for } t = 1, \dots, T(1960, \dots, 2008)$$

- **Panel (or longitudinal) data:** repeated observations on N cross section (individuals, households, firms, cities, countries) over T periods:

$$X_{it} \text{ for } i = 1, \dots, N; \text{ and } t = 1, \dots, T$$

- **Qualitative or discrete data:** binary response (1 or 0). For example survey data (can be cross section, time series, or panel)

Set-up: examples of quantitative relations

- **with cross section data**

$$\begin{aligned}\ln(\text{output}_i) &= \alpha + \beta_1 \ln(\text{labor}_i) + \beta_2 \ln(\text{capital}_i) + \beta_3(\text{spillover}_i) \\ &+ \beta_4(\text{R\&D}_i) + \text{quality} + u_i, \quad i = 1, \dots, N\end{aligned}$$

- spillover: measure of foreign firms concentration in the region of obs.
- R&D: measurement of innovation input
- quality: contains unobserved factors (managerial or worker quality)
- u : error (unobserved shocks)

- **with panel data**

$$\begin{aligned}\ln(\text{output}_{it}) &= \alpha + \beta_1 \ln(\text{labor}_{it}) + \beta_2 \ln(\text{capital}_{it}) + \beta_3(\text{spillover}_{it}) \\ &+ \beta_4(\text{R\&D}_{it}) + \text{quality}_i + u_{it}, \quad i = 1, \dots, N; t = 1, \dots, T\end{aligned}$$

- quality: a firm specific term that is constant over time

- **Goal:** obtain approximations (estimations) of parameters (α , β) using real data. – **A theory is needed for approximation (estimation)** –

Set-up: theory of estimation, basic intuitions (a)

- **What is estimation?**

Consider the problem of innovation:

$$y_i = \begin{cases} 1 & \text{if firm } i \text{ innovates} \\ 0 & \text{elsewhere} \end{cases}$$

- Statistical model: $(\mathcal{Y} = \{0, 1\}^n, \mathcal{P} = \{B(1, p)^{\otimes n}, p \in [0, 1]\})$, where $B(1, p)^{\otimes n}$ denotes the **binomial distribution**.

- Natural approximation (estimation) of p is the *proportion of firms* that innovate (the sample mean):

$$T(y_1, \dots, y_n) = \hat{p}(y_1, \dots, y_n) = \frac{y_1 + \dots + y_n}{n} = \frac{1}{n} \sum y_i = \bar{y}_n.$$

If $n = 1000$ and we observe 700 innovative firms, then a natural approximation of \hat{p} is: $\hat{p}(y) = 0.7$

Set-up: theory of estimation, basic intuitions (b)

- **What estimation means**

Consider the normal distribution of a random variable X such that, $X \sim N(\mu, \sigma^2)$. The probability density is:

$$f(x | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}$$

Natural approximations of $\mu = \mathbb{E}(X)$ (expectation) and $\sigma^2 = \mathbb{V}(X)$ (variance) are the sample analogue:

- ▶ Empirical mean: $\frac{1}{n} \sum_{i=1}^n X_i = \bar{X} = \hat{\mu}$
- ▶ Empirical variance: $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = S^2 = \hat{\sigma}^2$

Set-up: theory of estimation, basic intuitions (c)

• What estimation means

- ▶ **Definition:** An estimator is a **decision rule** (mapping) from the set of observations (say, \mathcal{Y}) to the set of decision (\mathcal{D}).

Back to example of innovation: the decision set is $\mathcal{D} = \{0, 1\}$

- ▶ **Why estimation is important:** we cannot observed the master population (time and costly), but we can get a sampling of it.
- ▶ **Estimation vs. Exact calculus**
 - ★ Estimation (uses the sample): 'with approximation error'
 - ★ Exact calculus (uses the master population): 'no approximation error'

Set-up: theory of estimation, basic intuitions (d)

- **Estimation is an approximation: how good (precise) is it?**

Properties of an estimator:

- ▶ **Unbias**: if the expectation of the estimator equals the parameter of interest

$$\mathbb{E}(\hat{\beta}) = \beta$$

- ▶ **Consistency (convergence)**: if unbiased and the variance goes to zero asymptotically

$$\mathbb{V}(\hat{\beta}_n) = 0 \text{ for } n \rightarrow \infty$$

- **Example**: back to the normal distribution

- ▶ \bar{X} is consistent. Indeed, $\mathbb{E}(\bar{X}) = \mu$, and $\mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}$
- ▶ S^2 is not consistent (it is biased):

$$\mathbb{E}(S^2) = \sigma^2 \left(\frac{n-1}{n} \right) \leq \sigma^2$$

- ▶ The corrected variance $S'^2 = \left(\frac{n}{n-1} \right) S^2$ is consistent as $\mathbb{E}(S'^2) = \sigma^2$

Set-up: role of conditional expectations (con'd)

- Plays a crucial role in econometrics even if not always explicitly stated.
- Applied econometric: estimate and test hypotheses about the expectation of a variable called the **explained variable** (or **dependent variable** or **regressand** or **response variable**) denoted y , conditional on a set of **explanatory variables** (or **independent variables** or **regressors** or **control variables** or **covariates**) usually denoted:

$$\mathbf{x} \equiv (x_1, x_2, \dots, x_K), \quad \text{a } 1 \times K \text{ vector}$$

- If $\mathbb{E}(|y|) < \infty$, then there exists a function, say $m: \mathbb{R}^K \rightarrow \mathbb{R}$ such that

$$\mathbb{E}(y \mid x_1, x_2, \dots, x_K) = \mathbb{E}(y \mid \mathbf{x}) = m(\mathbf{x})$$

- The function $m(\mathbf{x})$ determines how the *average* value of y changes as elements of \mathbf{x} change.
- Linear regression: estimate some parametric functionals of $m(\mathbf{x})$:

$$y = \mathbb{E}(y \mid \mathbf{x}) + u$$

Set-up: role of conditional expectations (con'd)

Exercise

Let (X, Y) be a couple of random variables, the joint distribution of which is given in the table.

X	Y	1	2	3	4
1		0,1	0,1	0,2	0,1
2		0	0,1	0,1	0
3		0,1	0	0	0,2

- 1 Compute the (marginal) distribution of X and Y and indicate if X and Y are independent or not. Compute the covariance between X and Y , $\text{Cov}(X, Y)$.
- 2 Determine the distribution (probability) of the random variable $\mathbb{E}(Y|X)$ and compute its expectation, $\mathbb{E}[\mathbb{E}(Y|X)]$. Compare the result with $\mathbb{E}(Y)$.

Set-up: role of conditional expectations (con'd)

1. **Marginal distributions:** denote $\mathbb{P}_{x.}$ and $\mathbb{P}_{.y}$ the (marginal) probabilities of X and Y resp. We have:

X	Y	1	2	3	4	$\mathbb{P}_{x.}$
1		0.1	0.1	0.2	0.1	0.5
2		0	0.1	0.1	0	0.2
3		0.1	0	0	0.2	0.3
	$\mathbb{P}_{.y}$	0.2	0.2	0.3	0.3	1

Definition of independence: product of marginal probabilities equals the joint. Here: $\mathbb{P}(X = 3)\mathbb{P}(Y = 2) = 0.3 * 0.2 \neq \mathbb{P}[(X = 3) \cap (Y = 2)] = 0$
As a result, X and Y are **not** independent.

Covariance (degree of dependence between two variables):

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \text{ and } \mathbb{E}(XY) = \sum_x \sum_y xy \cdot \mathbb{P}_{xy}$$

- ▶ $\mathbb{E}(X) = 0.5 + 0.4 + 0.9 = 1.8$
- ▶ $\mathbb{E}(Y) = 0.2 + 0.4 + 0.9 + 1.2 = 2.7$
- ▶ $\mathbb{E}(XY) = 0.1 + 0.2 + 0.6 + 0.4 + 0.4 + 0.6 + 0.3 + 2.4 = 5$
- ▶ $\text{Cov}(X, Y) = 5 - 1.8 * 2.7 = 0.14$

Set-up: role of conditional expectations (end)

2. **The conditional expectation** $\mathbb{E}(Y|X)$ is a random variable which takes values $\mathbb{E}(Y|X = x)$ with probabilities $\mathbb{P}_{x.} = \mathbb{P}(X = x)$.

First compute a new table of conditional probabilities using the Bayes law: $\mathbb{P}(Y | X) = \frac{\mathbb{P}(Y \cap X)}{\mathbb{P}(X)}$. Then compute the conditional expectations $\mathbb{E}(Y|X)$: first line is $(1 * 0.2) + (2 * 0.2) + (3 * 0.4) + (4 * 0.2) = 2.6$.

X	Y	1	2	3	4		$\mathbb{E}(Y X)$	$\mathbb{P}_{x.}$	$\mathbb{P}_{x.} \mathbb{E}(Y X)$
1		0.2	0.2	0.4	0.2	1	2.6	0.5	1.3
2		0	0.5	0.5	0	1	2.5	0.2	0.5
3		1/3	0	0	2/3	1	3	0.3	0.9

- We obtain that

$$\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y) = 2.7$$

This is called the **iterated law of expectation**: $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$

- Also

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X) = 1.8$$

The linear regression: model and assumptions

The model specification

We assume that each observation in a sample $(y_i, x_{i1}, x_{i2}, \dots, x_{iK})$, $i = 1, \dots, n$, is generated by an underlying process described by the model:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + u_i$$

The observed value of y_i is the sum of two parts: a **deterministic part**, x_{ik} and the **random part**, u_i (disturbance or error term).

Aim: use data to estimate the unknown parameters β_k . How we proceed from here depends crucially on what we assume about the stochastic process that has led to observations of the data in hand.

The linear regression: model and assumptions

Data matrix

Real Investment (Y)	Constant (1)	Trend (T)	Real GNP (G)	Interest Rate (R)	Inflation Rate (P)
0.161	1	1	1.058	5.16	4.40
0.172	1	2	1.088	5.87	5.15
0.158	1	3	1.086	5.95	5.37
0.173	1	4	1.122	4.88	4.99
0.195	1	5	1.186	4.50	4.16
0.217	1	6	1.254	6.44	5.75
0.199	1	7	1.246	7.83	8.82
y = 0.163	X = 1	8	1.232	6.25	9.31
0.195	1	9	1.298	5.50	5.21
0.231	1	10	1.370	5.46	5.83
0.257	1	11	1.439	7.46	7.40
0.259	1	12	1.479	10.28	8.64
0.225	1	13	1.474	11.77	9.31
0.241	1	14	1.503	13.42	9.44
0.204	1	15	1.475	11.02	5.99

Model in matrix notation

$$y = \mathbf{X}\beta + u$$

The linear regression: model and assumptions

$$y = \mathbf{X}\beta + u$$

Role of the disturbance or error term u

- **Capture omitted factors** in the explanation of y : we cannot hope to capture every influence on an economic variable in a model, no matter how elaborate.
- **Measurement error**: difficulty to obtain accurate measurement of some variables.

Example: measurements of innovation are proxy

The linear regression: model and assumptions

Assumptions of the model (con'd)

- **A1. Linearity:** $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + u_i$
The model specifies a linear relationship between y and x_1, \dots, x_K .

$$\text{Assumption: } y = \mathbf{X}\beta + \mathbf{u}$$

- **A2. Full rank:** There is no exact linear relationship among any of the independent variables in the model. This assumption is known as an **identification condition**.

$$\text{Assumption: } \mathbf{X} \text{ is an } n \times K \text{ matrix with rank } K$$

The linear regression: model and assumptions

Assumptions of the model (con'd)

- **A3. Exogeneity of independent variables**

The expected value of the disturbance at observation i in the sample is not a function of the independent variables observed at any observation, including this one. This means that the independent variables will not carry useful information for prediction of u_i

$$\text{Assumption: } \mathbb{E}[\mathbf{u} | \mathbf{X}] = \begin{bmatrix} \mathbb{E}[u_1 | \mathbf{X}] \\ \mathbb{E}[u_2 | \mathbf{X}] \\ \vdots \\ \mathbb{E}[u_n | \mathbf{X}] \end{bmatrix} = \mathbf{0}$$

- **A4. Distribution of u :** The disturbances are i.i.d.

$$\text{Assumption: } \mathbf{u} | \mathbf{X} \sim N[\mathbf{0}, \sigma^2 \mathbf{I}]$$

The linear regression: model and assumptions

Assumptions of the model (end)

- **A5. Homoscedasticity and non-autocorrelation:** Each disturbance, u_i has the same finite variance, σ^2 and is uncorrelated with every other disturbance, u_j

$$\mathbb{V}[u_i | \mathbf{X}] = \sigma^2, \quad \text{for all } i = 1, \dots, N,$$

and

$$\text{Cov}[u_i, u_j | \mathbf{X}] = 0, \quad \text{for all } i \neq j.$$

The two assumptions imply that

$$\text{Assumption : } \mathbb{V}[\mathbf{u} | \mathbf{X}] = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}$$

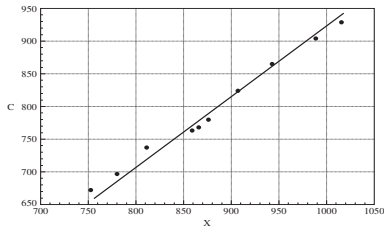
The Ordinary Least Squares Estimator (OLS): easy!

Consider the linear regression model with one regressor (univariate):

$$y_i = \alpha + \beta x_i + u_i, \quad i = 1, \dots, N$$

1. Adjust a linear draw within the cloud of points given by pair observations (y_i, x_i)

Here the estimated coefficient $\hat{\beta}$ is positive



2. Simple formulae to compute the adjusted coefficients

$$\hat{\beta} = \frac{\text{cov}(x_i, y_i)}{\text{var}(x_i)} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad \text{and } \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

The Ordinary Least Squares Estimator (OLS): easy!

Consider the linear regression model: $y = \mathbf{X}\beta + \mathbf{u}$

OLS: computation

Premultiply the model equation above by \mathbf{X}' , and take expectations:

$$\mathbb{E}(\mathbf{X}'y) = \mathbb{E}(\mathbf{X}'\mathbf{X})\beta + \mathbb{E}(\mathbf{X}'\mathbf{u})$$

By A.3, $\mathbb{E}(\mathbf{X}'\mathbf{u}) = 0$. Solving for β gives and using A.2 (Full rank matrix):

$$\beta = [\mathbb{E}(\mathbf{X}'\mathbf{X})]^{-1} \mathbb{E}(\mathbf{X}'y)$$

Replace the population moments with the sample analogues:

$$\mathbb{E}(\mathbf{X}'\mathbf{X}) = \left(\frac{1}{N} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1}; \quad \mathbb{E}(\mathbf{X}'y) = \left(\frac{1}{N} \sum_{i=1}^N \mathbf{x}'_i y_i \right)$$

yields the OLS estimator: Formula: $\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$

The Ordinary Least Squares Estimator (OLS): easy!

OLS: basic properties

- $\hat{\beta}$ is **consistent** for β
- $\hat{\beta}$ is **normally distributed**, with

$$\mathbb{V}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}, \text{ with } \hat{\sigma}^2 = \frac{1}{N-K} \|\hat{u}'\hat{u}\|, \text{ and } \hat{u} = y - \mathbf{X}\hat{\beta} \text{ (residuals)}$$

OLS: inference

- **Test of parameters significancy**, $H_0 : \hat{\beta}_k = 0$ Compute the Student statistic:

$$t_{\hat{\beta}_k} = \frac{\hat{\beta}_k}{\text{diag} \sqrt{\mathbb{V}(\hat{\beta}_k)}}$$

- **Goodness of fit**: coefficient of determination $R^2 \in (0, 1)$. It measures the proportion of the total variance in y that is accounted for by variations in the regressors.

The Ordinary Least Squares Estimator (OLS): easy!

OLS: Relaxing some assumptions

- Relaxing A.5 (**Homoskedasticity**) for **Heteroskedasticity** of unknown form:

$$\mathbb{V}[u_i | \mathbf{X}] = \sigma_i^2, \quad \text{for all } i = 1, \dots, N,$$

and

$$\text{Cov}[u_i, u_j | \mathbf{X}] = 0, \quad \text{for all } i \neq j.$$

yields the Huber, Halbert robust estimation to **Heteroskedasticity** of unknown form (see STATA output!)

- Relaxing A3. (**Exogeneity** of independent variables) for possible **endogeneity**

$$\mathbb{E}[\mathbf{u} | \mathbf{X}] \neq 0$$

– MOVE TO INSTRUMENTAL VARIABLE ESTIMATION –

The Instrumental Variable estimation (IV)

1. **Definition:** An explanatory variable x_k is said to be **endogenous** in the linear regression model if it is correlated with the error term u :

$$\mathbb{E}[u | \mathbf{X}] \neq 0$$

[back to the role of A3. (exogeneity) in computing the OLS estimator]

2. Sources of endogeneity

- ▶ **Omitted variables:** unavailability of data prevents from the use of other determinants as additional regressors. $\mathbb{E}[u | \mathbf{X}, q] \neq \mathbb{E}[u | \mathbf{X}]$ when q and \mathbf{X} are correlated (self-selection problem).
- ▶ **Measurement error:** imperfect measure of variables (innovation in developing countries).
- ▶ **Simultaneity:** at least one the regressor x_k is determined simultaneously along with y (feedback effect).

3. Consequence of ignoring endogeneity: OLS non consistent!

IV provides a general solution to the issue of endogenous regressors.

The IV Method

Consider the linear regression model: $y = \mathbf{X}\beta + \mathbf{u}$

The implementation of the IV method requires what is called **instruments** or **instrumental variables**

What is instrument?

- **Definition:** An instrument (or a set of instruments) is an observable variable, say, \mathbf{Z} , not in equation studied that satisfies two conditions:

- ▶ **Condition 1:** \mathbf{Z} must be uncorrelated with \mathbf{u} , that is

$$\mathbb{E}[\mathbf{u} | \mathbf{Z}] = 0, \quad \text{which implies that } \mathbb{E}[\mathbf{Z}'\mathbf{u}] = 0$$

In others words, \mathbf{Z} is exogenous!

- ▶ **Condition 2:** \mathbf{Z} must be correlated with \mathbf{X}

The IV Method

Consider the linear regression: $y = \mathbf{X}\beta + \mathbf{u}$. Assume that we have instrumental variables \mathbf{Z} that satisfy conditions 1 & 2. Let H be the number of instruments, and K the number of regressors.

IV: computation (case $H = K$)

Premultiply the model equation above by \mathbf{Z}' , and take expectations:

$$\mathbb{E}(\mathbf{Z}'y) = \mathbb{E}(\mathbf{Z}'\mathbf{X})\beta + \mathbb{E}(\mathbf{Z}'\mathbf{u})$$

By condition 1, $\mathbb{E}(\mathbf{Z}'\mathbf{u}) = 0$. Solving for β yields:

$$\beta_{IV} = [\mathbb{E}(\mathbf{Z}'\mathbf{X})]^{-1} \mathbb{E}(\mathbf{Z}'y)$$

Replacing the population moments with the sample analogues yields the IV estimator:

Formula: $\hat{\beta}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'y$

Other cases: i) $H < K$: system failed, ii) $H > K$: select optimal instruments $\mathbf{Z}^* = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ yielding 2SLS procedure (see STATA!)

The IV Method

IV: How to find suitable instruments?

- Test the two conditions defining an instrument:
 - ▶ Possible to test $\mathbb{E}[\mathbf{Z}'\mathbf{x}] = 0$ as \mathbf{Z} and \mathbf{x} are observable
 - ▶ Impossible to test $\mathbb{E}[\mathbf{Z}'\mathbf{u}] = 0$ as \mathbf{u} is not observable
 - ▶ Testing for endogeneity (**Hausman test**: see STATA!)

- Some examples of convincing instruments in the literature:
 - ▶ **Dynamic models**: use the lag of variable as instruments
 - ▶ **Program evaluation** (where individuals are randomly selected to be eligible for a program). Actual participation maybe endogenous because it can depend on unobserved factors that affect the response. However, eligibility can be assumed exogenous. Because participation and eligibility are correlated, the latter can be used as instrument.
 - ▶ **Generated instruments** $\hat{\mathbf{z}}_i \equiv g(w, \hat{\lambda})$. In the case of **panel data**, instruments can be obtained within the dataset.

Thanking you !
Enjoy your STATA