Applied Econometrics (MSc.)
Lecture 1: Course Introduction
The OLS Model

Yonas Alem

Department of Economics
University of Gothenburg

December 3, 2014

Table of Contents

Course Objectives

What is Econometrics?

- In Economics, we are interested in relationships between different variables, for e.g., "What is the impact of an additional years of schooling on earnings?"
- In Econometrics, we learn the skill how to quantify these relationships on the basis of available data, use statistical techniques, and interpret, use or exploit the resulting outcomes appropriately.
- Econometrics therefore is the interaction of economic theory, observed data and statistical methods.
- The interaction of these three makes econometrics interesting, challenging, and perhaps, difficult!
- Traditionally econometrics has focused upon aggregate economic relationships.

Course Objectives
What is Econometrics? Cont.

- Since the 1970s econometric methods have increasingly been employed in micro-economic models (Focusing on the analysis of individual, household, or firm behavior)
- This was driven by:
  - The development of appropriate econometric models and estimators that take into account problems like discrete dependent variable and sample selection,
  - Availability of large survey data sets
  - Increasing computational possibilities.
- More recently, the empirical analysis of financial markets has required and stimulated many theoretical developments in econometrics.

Course Objectives

What is Econometrics?

- Econometrics, currently, plays a major role in empirical work in all fields of economics, almost without exception, and in most cases it is no longer sufficient ot be able to run a few regressions and interpret the results.
- As a result, introductory econometrics textbooks, usually provide insufficient coverage for applied researchers
- On the other hand, the more advanced econometrics textbooks are often too technical or too detailed for the average economist to grasp the essential ideas and to extract the information that is needed
- Applied econometrics fills this gap! Discusses the recent and relatively more advanced developments

Course Objectives

- Econometrics is different from statistical analysis: guided by economic theories
- This course to a great extent takes this approach & emphasizes:
    - Application of a broad range of econometric techniques to address research questions
    - Effective communication of regression results
    - Very little in derivation of the different estimator and proving their properties
    - You are assumed to have covered this in the first year econometrics course

Course Objectives Cont.

- In this course, you will study application and implementation of core econometric concepts which you will be using to write your MSc. thesis
- So the most important objective of the course is to give you the skill and confidence to undertake rigorous empirical analysis
- A variety of published articles using the econometric methods we will be discussing will be used in the application section
- This would give you the opportunity to identify what you are interested in for your MSc. thesis.
- As a result, at the end of the course, you will be in a position to identify a topic and write a reasonably good MSc. thesis proposal
- The course is ideal for pursuing a career as a researcher or continue PhD in economics as well

Table: Course Outline

Topic
Lec. 1, Course Introduction, the OLS Model
Lec. 2, OLS - Application, IV Estimators
Lec. 3, IV Estimators - Application
Lec. 4, Binary Choice Models
Lab 1, OLS & IV Methods
Lec. 5, Models with timeseries data
Lec. 6 Models with Non-stationary Timeseries Data
Lec. 7 Models with Cointegrated timeseries data
Lab 2, Models with time-series Data
Lec. 8, Binary Choice Models - Application
Lab 3, Binary Choice Models
Lec. 9, Linear Panel Data Models
Lec. 10, LPM Applications & Nonlinear Panel Data Models
Lab 4, Panel Data Models
Presentations

Labs 25, 15, 15, 25 Points

- There will be four labs aimed to give you the opportunity to implement the different estimators in Stata and interpret your results
- Two of the lab reports should be written in a journal article format (8-12) pages containing the relevant sections of a standard article
  - I will give you clear guidelines on how to write an article-type paper
  - I will also post a couple of model articles which you can learn from on how to structure a paper
- Reports should be submitted before the respective deadlines (which would be 7-10 days after the lab session).
- If you submit 2-5 days late, you will lose half of the points
- If you submit 5 days later, you won't receive any points

Presentations 20 Points

- The last two days of the course are allocated for presentation of articles
- You will present an article of your choice in 15-20 minutes
- You will learn how to prepare good power point slides and make presentation of a research article. The objective of the presentation sessions is to give you the opportunity to practice presentation and learn from each other.

Presentations 20 Points

- I will give you the freedom to choose whatever article you would like to present. I recommend an article related to what you would like to write-up your thesis on.
- The article should be published in a good international field or general economics journal and should have an empirical strategy section applying any of the models we covered in class
- Please e-mail me the choice of your article before Jan. 12, 2015.

Attendance

- The course relies On *Verbeek, M. 2012. A Guide to Modern Econometrics, Fourth Edition, John Wiley & Sons, Ltd.* for presentation of the main econometric topics
- Discussion of the different topics will be followed by presentation of a published article using the respective methodology
- This implies that you will have to come to lectures to make the best out of the course
- I won't penalize you for not coming to lectures, but I will use attendance in "Borderline Cases" while grading
- Please try to read the required articles before every lecture.

Course Materials on the Course Home Page

- Lecture Plan
- Lecture Notes
- Computer Labs & Data
- List of Required Readings

OLS - Rationale

- One is interested in the best linear combination of $x_2, ..., x_K$ and a constant gives a good approximation of the dependent variable y
- Write an arbitrary linear combination:

$$\tilde{\beta}_1 + \tilde{\beta}_2 x_2 + ... + \tilde{\beta}_K x_K, \tag{1}$$

- Where $\tilde{\beta}_1, ..., \tilde{\beta}_K$ are constants to be chosen
- Index the observations by $i, i = 1, ..., N$
- The difference between the observed value of y and its linear approximation is given by

$$y_i - [\tilde{\beta}_1 + \tilde{\beta}_2 x_{i2} + ... + \tilde{\beta}_K x_{iK}]. \tag{2}$$

OLS Cont.

- Let $x_i = (1 \; x_{i2} \; x_{i3} \; ... \; x_{iK})'$
- Collecting the $\tilde{\beta}$ coefficients in a K-dimensional vector $\tilde{\beta} = (\tilde{\beta}_1 \; ... \; \tilde{\beta}_K)'$, we can write (2) as

$$y_i - x_i'\tilde{\beta}. \tag{3}$$

- The objective is to choose values for $\tilde{\beta}_1, ..., \tilde{\beta}_K$ such that these differences are small
- OLS: choose $\tilde{\beta}$ such that the sum of squared differences is as small as possible. i.e.,

$$S(\tilde{\beta}) \equiv \sum_{i=1}^{N} (y_i - x_i'\tilde{\beta})^2 \tag{4}$$

- What do we achieve by taking squares, instead of just differences?

OLS Cont.

- FOC: Differentiate $S(\tilde{\beta})$ w.r.t the vector $\tilde{\beta}$, which gives the following K conditions, which are also called the **Normal Equations**

$$-2\sum_{i=1}^{N} x_i(y_i - x_i'\tilde{\beta}) = 0 \tag{5}$$

or

$$(\sum_{i=1}^{N} x_i x_i')\tilde{\beta} = \sum_{i=1}^{N} x_i y_i. \tag{6}$$

- K unknowns and, hence unique solution is for $\tilde{\beta}$ provided that $(\sum_{i=1}^{N} x_i x_i')$ is invertible (what does this mean?).

OLS Cont.

- The solution to the minimization problem, for b is given by

$$b = (\sum_{i=1}^{N} x_i x_i')^{-1} \sum_{i=1}^{N} x_i y_i \qquad (7)$$

- One can confirm from the SOCs that b indeed corresponds to a minimum. Thus,

$$\hat{y} = x_i' b \qquad (8)$$

- Giving the **best linear approximation** of y from $x_2, ..., x_K$
- Why do we call it best?
- They only assumption one needs to estimate this model is **no-multicollinearity**
- But, consistency of the estimator requires more assumptions!

OLS Cont.

- Let the residual $e_i = y_i - \hat{y}_i = y_i - x_i' b$
- Thus,

$$y_i = \hat{y}_i + e_i = x_i' b + e_i \tag{9}$$

- The minimum value for the objective function (which is referred to as the **residual sum of squares**could thus be written as

$$S(b) = \sum_{i=1}^{N} e_i^2 \tag{10}$$

- The approximated value $x_i' b$ and the residual e_i satisfy certain properties by construction

OLS Cont.

Simple Linear Regression

- For e.g., if one substitutes for $\tilde{\beta}$ by b in equation (5),

$$\sum_{i=1}^{N} x_i(y_i - x_i'b) = \sum_{i=1}^{N} x_i e_i = 0 \tag{11}$$

- I.e, the vector $e = (e_1, ..., e_N)'$ is orthogonal to each vector of observations on an x-variable.

OLS Cont.
Simple Linear Regression

- In this case, $K = 2$
- We follow the same procedure of minimizing the RSS:

$$S(\tilde{\beta}_1, \tilde{\beta}_2) = \sum_{i=1}^{N} (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2 \tag{12}$$

- FOCs:

$$\partial S(\tilde{\beta}_1, \tilde{\beta}_2)/\partial \tilde{\beta}_1 = -2 \sum_{i=1}^{N} (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i) = 0, \tag{13}$$

$$\partial S(\tilde{\beta}_1, \tilde{\beta}_2)/\partial \tilde{\beta}_2 = -2 \sum_{i=1}^{N} x_i (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i) = 0, \tag{14}$$

OLS Cont.

Simple Linear Regression Cont.

- From (13):

$$b_1 = 1/N * \sum_{i=1}^{N} y_i - b_2 * 1/N \sum_{i=1}^{N} x_i = \overline{y} - b_2\overline{x}, \tag{15}$$

- Rewriting (14),

$$\sum_{i=1}^{N} x_i y_i - b_1 \sum_{i=1}^{N} x_i - (\sum_{i=1}^{N} x_i^2) b_2 = 0 \tag{16}$$

- And substituting in (15) in (16) gives:

$$\sum_{i=1}^{N} x_i y_i - N\overline{xy} - (\sum_{i=1}^{N} x_i^2 - N\overline{x}^2) b_2 = 0 \tag{17}$$

OLS Cont.

Simple Linear Regression Cont.

- One can thus solve (17) for b_2 as follows

$$b_2 = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{18}$$

- Divide (18) by $N-1$ to show that the OLS solution for b_2 is the ratio of the sample covariance between $x\&y$ and the sample variance of x!
- The OLS estimator in matrix notation is given by:

$$b = (X'X)^{-1}X'y \tag{19}$$

- How do we derive (19)?

OLS Cont.
Matrix Notation

- $$X = \begin{pmatrix} 1 & x_{i2} & \dots & x_{iK} \\ . & . & . & . \\ 1 & x_{N2} & \dots & x_{NK} \end{pmatrix}$$
$$= \begin{pmatrix} x_1' \\ . \\ x_N' \end{pmatrix}$$
$$y = \begin{pmatrix} y_1 \\ . \\ y_N \end{pmatrix}$$

OLS Cont.
Matrix Notation

- The minimization criterion stated in (4) can be rewritten as

$$S(\tilde{\beta}) = (y - X\tilde{\beta})'(y - X\tilde{\beta}) = y'y - 2y'X\tilde{\beta} + \tilde{\beta}'X'X\tilde{\beta} \quad (20)$$

- Differentiating (20) w.r.t $\tilde{\beta}$ and setting the result equal to zero would give

$$\frac{\partial S(\tilde{\beta}}{\partial \tilde{\beta}} = -2(X'y - X'X\tilde{\beta}) = 0 \quad (21)$$

- Solving 21 gives the OLS solution

$$b = (X'X)^{-1}X'y \quad (22)$$

- Exactly the same with (7) but now expressed in Matrix Notation.

Small Sample Properties of OLS
The Gauss-Markov Assumptions

- Correct approximation (consistency) of the OLS estimator for b depends crucially upon the assumptions that are made about the distribution of ε_i and its relation to x_i
- The Gauss-Markov conditions for the linear regression model given by

$$y_i = x_i'\beta + \varepsilon_i, \tag{23}$$

are given by:

A1 $E\{\varepsilon_i\} = 0, i = 1, ..., N$

A2 $\{\varepsilon_1, ..., \varepsilon_N\}$ & $\{x_1, ..., x_N\}$ are independent

A3 $V\{\varepsilon_i\} = \sigma^2,\ i = 1, ..., N$

A4 $cov\{\varepsilon_i, \varepsilon_j\} = 0,\ i, j = 1, ..., N, i \neq j$

Small Sample Properties of OLS

The Gauss-Markov Assumptions Cont.

- Assumptions $[A1][A2]\&[A4] \Longrightarrow$

$$E\{\varepsilon\} = 0 \quad \& \quad V\{\varepsilon\} = \sigma^2 I_N \tag{24}$$

- If the Gauss-Markov assumptions are satisfied, OLS is **BLUE!**, i.e

$$E\{b\} = E\{(X'X)^{-1}X'y\} = E\{\beta + (X'X)^{-1}X'\varepsilon\} \tag{25}$$

$$\beta + E\{(X'X)^{-1}X'\varepsilon\} = \beta \tag{26}$$

$$Var\{b|X\} = \sigma^2(X'X)^{-1} = \sigma^2(\sum_{i=1}^{N} x_i x_i')^{-1} \tag{27}$$

Small Sample Properties of OLS

The Gauss-Markov Assumptions Cont.

- An unbiased estimator of σ^2 is given by:

$$s^2 = \frac{1}{N-K} \sum_{i=1}^{N} e_i^2 \tag{28}$$

Where N refers to the number of observations, and K, the number of coefficients to be estimated.

- Apart from the Gauss-Markov assumptions, exact statistical inference needs assumption on the distribution of ε. The most common assumption:

$$[A5] \quad \varepsilon \sim (0, \sigma^2 I_N) \tag{29}$$

Or

$$\varepsilon \sim NID(0, \sigma^2 I_N) \tag{30}$$

OLS
Goodness-of-fit

- The goodness of fit measure for OLS is the R^2
- It answers the question "How well does the estimated regression line fit the observations?"
- Is the proportion of the (sample) variance of y that is explained by the model:

$$R^2 = \frac{\hat{V}\{\hat{y}_i\}}{\hat{V}\{y_i\}} = \frac{1/(N-1)\sum_{i=1}^{N}(\hat{y}-\bar{y})^2}{1/(N-1)\sum_{i=1}^{N}(y-\bar{y})^2} \tag{31}$$

- Where $\hat{y}_i = x_i'b$ and $\bar{y} = (1/N)\sum_i y_i$ represents the sample mean of y_i

OLS
Goodness-of-fit Cont.

- From the FOCs, it follows that

$$\sum_{i=1}^{N} e_i x_{ik} = 0, k = 1, ..., K. \tag{32}$$

- One can therefore write $y_i = \hat{y}_i + e_i$, where $\sum_i e_i \hat{y}_i = 0$
- When the model contains an intercept term, we have

$$\hat{V}\{y_i\} = \hat{V}\{\hat{y}_i\} + \hat{V}\{e_i\} \tag{33}$$

- Where $\hat{V}\{e_i\} = \tilde{s}^2$

OLS

Goodness-of-fit Cont.

- Using the above, one can rewrite the R^2

$$R^2 = 1 - \frac{\hat{V}\{e_i\}}{\hat{V}\{y_i\}} = 1 - \frac{1/(N-1)\sum_{i=1}^{N} e_i^2}{1/(N-1)\sum_{i=1}^{N}(y_i - \bar{y})^2} \qquad (34)$$

- Eq. [33] shows how the sample variance of y_i can be decomposed into the sum of the sample variances of two orthogonal components: $\hat{y}_i$ and e_i
- The R^2 thus indicates which proportion of the sample variation in y_i is explained by the model.

31 & [34] are equivalent if the model contains an intercept term.

- In this case, it is also true that $0 \leq R^2 \leq 1$
- When exactly is it true that $R^2 = 0, \& 1 respectively?$

OLS

Hypothesis testing

- A simple t-test: $\frac{b_k - \beta_k^p}{se(b_k)}$
- Testing one linear restriction: e.g., $\beta_2 + \beta_3 + ... + \beta_K = 1$
- A joint test of significance of regression coefficients: all coefficients, except the intercept β_1, are equal to zero.

OLS

Asymptotic Properties of OLS

- Consistency: $plimb = \beta$
- Asymptotic normality: OLS is shown to be asymptotically normally distributed