Basic Econometrics

Christopher Grigoriou

Executive MBA-HEC Lausanne
2007/2008
Overview

Objectives of the day
• Interpreting Econometric Applications
• Understanding how it works

Program of the day
• Introduction (why? how? basic concepts)
• Interpretations
• Case-study

Executive MBA – HEC Lausanne  2007/2008
1. Introduction to Econometrics

=> What for?

- Impact Analysis
- To evaluate the success/failure of a project, reform, law, ...
- To test any economic theory

Executive MBA – HEC Lausanne  2007/2008
1. Introduction to Econometrics

=> How?

- Apply statistical methods to economic data
- Econometric approach:
  - Develop working model from an economic theory
  - Estimate model with real world data.

Real world data is not perfect
Some examples

- Keynesian Consumption Function
- Price and Quantity
- Philips Curve
- Production Function
Keynesian Consumption Function

- Theory: people increase consumption as income increases, but not by as much as the increase in their income.

  - Marginal Propensity to Consume (MPC) is the change in consumption divided by change in income.
Keynesian Consumption Function

- $C = \alpha + \beta I$
  - $C =$ Consumption
  - $\alpha =$ Intercept
  - $I =$ Income
  - $\beta =$ slope (how much $C$ changes for a given change in $I$)

- Not an econometric model
  - Assumes a deterministic relationship
Keynesian Consumption Function

$C = \alpha + \beta I + \varepsilon$

$\varepsilon = \text{error term}$

- Error term captures several factors:
  - omitted variables
  - measurement error in the dependent variable
  - randomness of human behavior

- Expected Results: $\alpha > 0$ and $0 < \beta < 1$
  $\beta$ represents the MPC
Ordinary Least Squares

• How to estimate the model?
  => Fit a line through the data.

• Estimate from the least squares
  - the line of best fit minimizes the sum of the squared deviations of the points on the graph from the points on the straight line.
    - Minimize $\Sigma (CA_i - CP_i)^2$
      - $CA_i =$ Actual Consumption for obs $i$
      - $CP_i =$ Predicted Consumption for obs $i$
Ordinary Least Squares optimization process

\[ y_i = a_0 + a_1 x_i + u_i \]

=> search for \( a_0 \) and \( a_1 \) that minimize the sum of the squared residual (= the global error of the model)

Executive MBA – HEC Lausanne 2007/2008
Suppose we get $C = 1000 + 0.8I$

$\alpha = 1000$

$\beta = 0.8$

Sample income levels

$I = 0$, Consumption = 1000

$I = 1000$, Consumption = 1800

If $I$ increases by 1 dollar, then $C$ increases on average by 0.8 dollars.

These estimates are consistent with theory since $\alpha > 0$ and $0 < \beta < 1$

Suppose $\beta$ was 0.9?

$\beta < 1$, but is it due to the sample? => tests + Confidence Interval
Interpretation (2)

1- Basic specification: \( Y_i = \alpha + \beta X_i + \gamma Z_i + \varepsilon_i \)
   \( \beta = \text{marginal impact:} \)
   \( \Rightarrow \text{an increase of 1 unity in } X \text{ implies ceteris paribus an increase of } \beta \text{ unities in } Y \)

2- Log-log specification: \( \ln Y_i = \alpha + \beta \ln X_i + \gamma \ln Z_i + \varepsilon_i \)
   \( \beta = \text{elasticity:} \)
   \( \Rightarrow \text{an increase of 1 per cent in } X \text{ implies ceteris paribus an increase of } \beta \text{ percent in } Y \)

3- Semi-log: \( \ln Y_i = \alpha + \beta X_i + \gamma Z_i + \varepsilon_i \)
   \( \beta = \text{semi-elasticity:} \)
   \( \Rightarrow \text{an increase of 1 unity in } X \text{ implies ceteris paribus an increase of } \beta \text{ per cent in } Y \)
Other examples (1)

- Price and Quantity

- Demand and elasticities of demand
  - $\ln Q = \alpha + \beta \ln P + \varepsilon$

- Phillips Curve
  - Relationship between change in money wages and unemployment
    - $\Delta w = f (\Delta u)$
Production Function

- Relationship between inputs and outputs.
  - $Y = f (K,L)$
  - Cobb Douglas $Y = AK^aL^\beta$

Wage equation

$$\ln W = \alpha_0 + \alpha_1 \text{EDUC} + \alpha_2 \text{EXP} + \alpha_3 \text{GENDER} + \alpha_4 \text{RACE} + \varepsilon_i$$
General Terminology (1)

- Pooled data: mixture of cross-sectional and time series data
- Panel data: follow a microeconomic unit over time
- Quantitative data: continuous data
- Qualitative data: categorical data
General Terminology (2)

\[ Y_{it} = \alpha + \beta \cdot X_{it} + \varepsilon_{it} \]

- \( Y \): dependent variable
- \( X \): independent or explanatory
- \( \varepsilon \): Error-term
- subscript \( i \): refers to \( i \)th observation
- \( t \): for time series data at time \( t \)
- Cross-sectional data: collected at 1 point in time
- Time series data: collected over a period of time

Executive MBA – HEC Lausanne 2007/2008
What do we know?

✓ From a theoretical economic hypothesis to an econometric validation

✓ Econometric methods = evaluate an average relationship between Y and X

✓ Estimates are done with error

✓ The Ordinary Least Squares: method aiming at assessing the parameters that minimize this error
Statistical Definition Basic Concepts

- Two basic ways to characterize a statistical serie:

  - central parameter => mean, median
    
    \[
    \text{mean} : \quad \bar{X}_i = \frac{1}{n} \sum_{i=1}^{i=n} X_i
    \]

  - dispersion parameter => variance, standard-deviation
    
    \[
    \text{standard-deviation} : \quad \sigma_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^{i=n} (X_i - \bar{X})^2}
    \]
Example: 2 different classrooms

Exam of Statistics...

- 2 groups with on average exactly the same mark... => 11.5

- What information does it provide on your own result?
### Class A

<table>
<thead>
<tr>
<th>Rank</th>
<th>Xi</th>
<th>Mean</th>
<th>Xi-Mean</th>
<th>(Xi-mean)$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>11.5</td>
<td>-9.5</td>
<td>90.25</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>11.5</td>
<td>-8.5</td>
<td>72.25</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>11.5</td>
<td>-5.5</td>
<td>30.25</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>11.5</td>
<td>-4.5</td>
<td>20.25</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>11.5</td>
<td>-2.5</td>
<td>6.25</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>11.5</td>
<td>-1.5</td>
<td>2.25</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>11.5</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>11.5</td>
<td>2.5</td>
<td>6.25</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>11.5</td>
<td>3.5</td>
<td>12.25</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>11.5</td>
<td>4.5</td>
<td>20.25</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>11.5</td>
<td>5.5</td>
<td>30.25</td>
</tr>
<tr>
<td>12</td>
<td>18.5</td>
<td>11.5</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>13</td>
<td>20</td>
<td>11.5</td>
<td>8.5</td>
<td>72.25</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>149.5</strong></td>
<td><strong>Var.</strong></td>
<td><strong>34.3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>11.5</strong></td>
<td><strong>σ</strong></td>
<td><strong>5.9</strong></td>
<td></td>
</tr>
</tbody>
</table>

Executive MBA – HEC Lausanne  2007/2008
## Class B

<table>
<thead>
<tr>
<th>Rank</th>
<th>Xi</th>
<th>Mean</th>
<th>Xi-Mean</th>
<th>(Xi-mean)$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>11.5</td>
<td>-1.5</td>
<td>2.25</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>11.5</td>
<td>-1.5</td>
<td>2.25</td>
</tr>
<tr>
<td>3</td>
<td>10.5</td>
<td>11.5</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>11.5</td>
<td>-0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
<td>11.5</td>
<td>-0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>11.5</td>
<td>-0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>7</td>
<td>11.5</td>
<td>11.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>11.5</td>
<td>11.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>11.5</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>10</td>
<td>12.5</td>
<td>11.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>12.5</td>
<td>11.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>11.5</td>
<td>1.5</td>
<td>2.25</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>11.5</td>
<td>1.5</td>
<td>2.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum</th>
<th>149.5</th>
<th>Var.</th>
<th>1.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.5</td>
<td>σ</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Executive MBA – HEC Lausanne 2007/2008
=>...of course, the answer also depends on the dispersion (standard-deviation)

=> To characterize a serie you need

- The mean of the serie (central parameter)
- The standard-deviation (dispersion)
Same thing with a coefficient estimate...

$\Rightarrow$ the coefficient is an \textit{averaged impact}.

$\Rightarrow$ its \textit{significance} depends on its \textit{dispersion}, i.e. the accuracy associated to the estimate.

Don’t forget! the predictions are done with error!

$Y = \alpha + \beta X + \varepsilon$

$\Rightarrow$ Given the error in the estimate or the inaccuracy in the estimate (assessed by the dispersion)

$\Rightarrow$ is $\beta$ \textit{significantly} different from zero?
Estimation of the education return (1)

• One (perhaps you?) wants to know the impact of an additionnal year of education on his wage

• Economic Theory: Mincer’s Equation

• Econometric point: how big is $\beta$?

  => $\ln wage_i = \alpha + \beta . educ_i + \gamma . exper_i + \delta . \text{expersq}_i + \varepsilon_i$
428 observations in the sample

=> R-squared = 15.68%
Our model predicts 15.68% of the fluctuations of the wages

Executive MBA – HEC Lausanne 2007/2008
(average) coefficient

=> 0.1075

=> Any additional year of education implies on average an increase of 0.1075% in the wage

Not only a mean (coefficient), but also a standard deviation (0.0141465)...as any other statistical serie

The standard deviation provides information on the accuracy of the estimate.

Executive MBA – HEC Lausanne 2007/2008
Coefficient estimated with error standard deviation

Confidence Interval :
95% chances for the coefficient to be in the interval
\( (\beta - 2\sigma = 0.07968; \beta + 2\sigma = 0.1352956) \)
Is $\beta$ significantly different from zero?

A test classically used to compare averages: $t$-test.

$$t_\beta = \frac{\hat{\beta} - \overline{\beta}}{\sigma_\hat{\beta}}$$

$\Rightarrow$ Compare the actual coeff. ($\hat{\beta}$) with the restricted coeff. ($\overline{\beta}$) weighted by the dispersion ($= \sigma_\hat{\beta}$ a measure of the accuracy of the estimate)

$\hat{\beta} = 0.1075$; $\overline{\beta} = 0$ !! $\sigma_\hat{\beta} = 0.01414$ $\Rightarrow t_\beta = \frac{0.1075}{0.01414} = 7.60$

So what?

Executive MBA – HEC Lausanne 2007/2008
H₀ : β = 0
Compute a t-statistic
| t-statistic | > 2  => reject H₀
=> β significantly different from zero (what we expected!!)

| t-statistic | < 2  => cannot reject H₀
=> β NOT significantly different from zero (no impact of studying one more year on my wages !!)

Executive MBA – HEC Lausanne  2007/2008
Whatever the econometric results & the purpose of the study

1- Coefficient = the impact studied
2- Either the standard deviation or the t-statistic or the p-value (critical probability i.e. the type 1 error)
3- T-test => \( \beta = 0 \) for each coefficient
4- P-value associated = Type 1 error.
| lwage  | Coef.      | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|--------|------------|-----------|-------|------|---------------------|
| educ   | 0.1074896  | 0.0141465 | 7.60  | 0.000| 0.0796837 – 0.1352956|
| exper  | 0.0415665  | 0.0131752 | 3.15  | 0.002| 0.0156697 – 0.0674633|
| expersq| -0.0008112 | 0.0003932 | -2.06 | 0.040| -0.0015841 – -0.000382|
| _cons  | -0.5220407 | 0.1986321 | -2.63 | 0.009| -0.9124669 – -0.1316145|

- Comparing the t-statistic to 2 = a 5% type-1 error
  « A type I error = the probability to reject Ho while it’s true... »
  => i.e. 5% chances to be wrong when rejecting $H_0$

- A more accurate way: the p-value = the exact type 1 error:

- Less than 1% chance to be wrong when rejecting $H_0$
  ($H_0 = « the coefficient is not significantly different from zero »$)

Executive MBA – HEC Lausanne 2007/2008
Case-Study: Feldstein and Horioka (1980)

- From the liberalisation of the capital flights
  => how did the capital really move?

- The impact of economic policy depends on the degree of mobility of the capital

- Feldstein and Horioka (1980): correlation between savings and investments
Correlation between Savings and Investment and the openness of the economy

1- Correlation between Savings and Investments close to 1:
   Closed economy
   any increase in national savings induces an identical increase in investments
   => low degree of capital mobility

2- Correlation between Savings and Investments close to 0:
   Opened (integrated) economy:
   National savings respond to investment opportunities on the world market/
national investment is financed by savings from the rest of the world
   => high degree of capital mobility
Econometric model (Feildstein and Horioka)

- Testing the correlation between Investments and Savings
  \[ \frac{I_i}{Y_i} = \alpha + \beta \frac{S_i}{Y_i} + \epsilon_i \]

- T-test on the \( \beta \)
  (i) \( H_0 : \beta = 0 \)
  (ii) \( H_0 : \beta = 1 \)

- The sample: 19 countries of the OECD
  (a) long-term effect (1970-1998)
  (b) short-term (three 10 year periods)
### Saving and Investment: long period (1970-98)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Saving and Investment:</strong></td>
<td>1970-1998</td>
</tr>
<tr>
<td><strong>$s_i$</strong></td>
<td>0.62</td>
</tr>
<tr>
<td><strong>(standard-error)</strong></td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.08</td>
</tr>
<tr>
<td><strong>(standard-error)</strong></td>
<td>0.03</td>
</tr>
<tr>
<td><strong>R- Squared</strong></td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>19</td>
</tr>
</tbody>
</table>

1- Is $b$ significantly different from zero? $H_0 : \beta = 0$

⇒ t-test (to compare means)

$$t_\beta = \frac{\hat{\beta} - \bar{\beta}}{\sigma_\beta} = \frac{0.62 - 0}{0.13} = 4.85$$

⇒ $|4.85| > 2$

⇒ At the 5% level $H_0$ is rejected

⇒ Note the p-value (computed by the software) is inferior to 1%

⇒ In the long-term we cannot conclude to a perfect degree of capital mobility
Saving and Investment: long period (1970-98)

1- Is $\beta$ significantly different from one? $H_0: \beta = 1$

$t-test$ (to compare means)

$t_\beta = \frac{\hat{\beta} - \bar{\beta}}{\sigma_\beta} = \frac{0.62 - 1}{0.13} = 2.94$

$\Rightarrow |2.94| > 2$

$\Rightarrow$ At the 5% level $H_0$ is rejected

$\Rightarrow$ Note the p-value (computed by the software) is inferior to 1%

$\Rightarrow$ In the long-term we cannot conclude to closed economies

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i$</td>
<td>0.62</td>
</tr>
<tr>
<td>(standard-error)</td>
<td>0.13</td>
</tr>
<tr>
<td>Constant</td>
<td>0.08</td>
</tr>
<tr>
<td>(standard-error)</td>
<td>0.03</td>
</tr>
<tr>
<td>R- Squared</td>
<td>0.58</td>
</tr>
<tr>
<td>Observations</td>
<td>19</td>
</tr>
</tbody>
</table>
**Saving and Investment: short-term**


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i$</td>
<td>0.82</td>
<td>0.65</td>
<td>0.40</td>
</tr>
<tr>
<td>(standard-error)</td>
<td>0.14</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>Constant</td>
<td>0.08</td>
<td>0.08</td>
<td>0.11</td>
</tr>
<tr>
<td>(standard-error)</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>R- Squared</td>
<td>0.61</td>
<td>0.55</td>
<td>0.30</td>
</tr>
<tr>
<td>Observations</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

=> What can you say about openness of OECD countries over each 10 year periods?

Executive MBA – HEC Lausanne 2007/2008
Exercise:
Saving and Investment: short-term

⇒ What can you say about openness of OECD countries for each of the three periods?
(for each period: perfect capital mobility? Closed economies? Etc...)

⇒ What would you say regarding the evolution of the capital mobility over the whole period?
Exercise:
Saving and Investment: short-term

⇒ Results:

⇒ Conclusion:

Executive MBA – HEC Lausanne 2007/2008
Exercise:
Saving and Investment: short-term

⇒ Results:
Period 1: $\hat{\beta} = 0.82$
   (i) $H_0 : \beta = 0 \Rightarrow t = 5.93 \Rightarrow |t| > 2 \Rightarrow$ rejection of $H_0$
   (ii) $H_0 : \beta = 1 \Rightarrow t = -1.32 \Rightarrow |t| < 2 \Rightarrow$ non-rejection of $H_0$
Period 2: $\hat{\beta} = 0.65$
   (i) $H_0 : \beta = 0 \Rightarrow t = 4.62 \Rightarrow |t| > 2 \Rightarrow$ rejection of $H_0$
   (ii) $H_0 : \beta = 1 \Rightarrow t = -2.46 \Rightarrow |t| > 2 \Rightarrow$ rejection of $H_0$
Period 3: $\hat{\beta} = 0.40$
   (i) $H_0 : \beta = 0 \Rightarrow t = 2.67 \Rightarrow |t| > 2 \Rightarrow$ rejection of $H_0$
   (ii) $H_0 : \beta = 1 \Rightarrow t = -4.04 \Rightarrow |t| > 2 \Rightarrow$ rejection of $H_0$

⇒ Conclusion:
- None of the three periods with a perfect capital mobility
- Even a behaviour of closed economies over the first period
⇒ A change toward openness over the 2 last periods?
A change toward openness over the 2 last periods?

In other words...

=> are the two coefficients significantly different?

=> \( H_0 : \beta_{8089} = \beta_{9098} \)?

=> not exactly the same t-statistic as usual because the both terms are estimated (...with error)

\[
t^* = \left| \frac{\hat{\beta}_{8089} - \hat{\beta}_{9098}}{\sqrt{\sigma_{\beta_{8089}}^2 + \sigma_{\beta_{9098}}^2}} \right| = \frac{0.40 - 0.65}{\sqrt{0.14^2 + 0.15^2}} = 1.2419 < 2
\]

=> No significant decrease in \( \beta \) over the two last periods, we cannot conclude to an increased liberalisation of the capital market for this sample of countries and these periods.

Executive MBA – HEC Lausanne  2007/2008
Conclusion - What have we learnt?
Conclusion - What have we learnt?

1- Basic methodology regarding econometrics
   - Economic problem => data => econometric validation

2- Characterizing a statistical serie
   - Central parameter, dispersion character...

3- The most common econometric estimator
   - Ordinary Least Squares, concept of error-term

4- Reading/interpreting econometric results
   - R-squared, Marginal impact, elasticity, semi-elasticity, confidence interval, p-value,...

5- Statistical test of the coefficients
   - t-test (student test): against a constant, against another estimate

Executive MBA – HEC Lausanne  2007/2008