ECON4150 - Introductory Econometrics

Lecture 4: Linear Regression with One Regressor

Monique de Haan
(moniquened@econ.uio.no)

Stock and Watson Chapter 4
Lecture outline

- The OLS estimators
  - The effect of class size on test scores
- The Least Squares Assumptions
  - $E(u_i | X_i) = 0$
  - $(X_i, Y_i)$ are i.i.d
  - Large outliers are unlikely
- Properties of the OLS estimators
  - unbiasedness
  - consistency
  - large sample distribution
- The compulsory term paper
Question of interest: What is the effect of a change in $X_i$ on $Y_i$?

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

Last week we derived the OLS estimators of $\beta_0$ and $\beta_1$:

\[
\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}
\]

\[
\hat{\beta}_1 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{s_{xy}}{s_x^2}
\]
OLS estimates: The effect of class size on test scores

Question of interest: What is the effect of a change in class size on test scores?

\[ TestScore_i = \beta_0 + \beta_1 ClassSize_i + u_i \]

. regress test_score class_size, robust

Linear regression
Number of obs = 420
F(1, 418) = 19.26
Prob > F = 0.0000
R-squared = 0.0512
Root MSE = 18.581

|            | Coef.   | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|------------|---------|-----------|-------|------|---------------------|
| test_score |         |           |       |      |                     |
| class_size | -2.279808 | 0.5194892 | -4.39 | 0.000 | -3.300945 to -1.258671 |
| _cons      | 698.933 | 10.36436  | 67.44 | 0.000 | 678.5602 to 719.3057  |

\[ TestScore_i = 698.93 - 2.28 \cdot ClassSize_i \]
The Least Squares assumptions

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

Under what assumptions does the method of ordinary least squares provide appropriate estimators of \( \beta_0 \) and \( \beta_0 \)?

Under what assumptions does the method of ordinary least squares provide an appropriate estimator of the effect of class size on test scores?

The Least Squares assumptions:

**Assumption 1:** The conditional mean of \( u_i \) given \( X_i \) is zero

\[ E(u_i|X_i) = 0 \]

**Assumption 2:** \( (Y_i, X_i) \) for \( i = 1, \ldots, n \) are independently and identically distributed \((i.i.d)\)

**Assumption 3:** Large outliers are unlikely

\[ 0 < E\left(X_i^4\right) < \infty \quad \& \quad 0 < E\left(Y_i^4\right) < \infty \]
The Least Squares assumptions: Assumption 1

\[ E(u_i|X_i) = 0 \]

The first OLS assumption states that:

All other factors that affect the dependent variable \( Y_i \) (contained in \( u_i \)) are unrelated to \( X_i \) in the sense that, given a value of \( X_i \), the mean of these other factors equals zero.

In the class size example:

All the other factors affecting test scores should be unrelated to class size in the sense that, given a value of class size, the mean of these other factors equals zero.
The least squares assumptions: Assumption 1

The first OLS assumption can also be written as:

\[ E(Y_i | X_i) = E(\beta_0 + \beta_1 X_i + u_i | X_i) \]

**Expectation rules**

\[ = \beta_0 + \beta_1 E(X_i | X_i) + E(u_i | X_i) \]

**ASS#1:** \( E(u_i | X_i) = 0 \)

\[ = \beta_0 + \beta_1 X_i \]
The Least Squares assumptions: Assumption 1

\[ E(Y_i|X_i) = \beta_0 + \beta_1 \]
Example of a violation of assumption 1:

Suppose that

- districts which wealthy inhabitants have small classes and good teachers
  - these districts have a lot of money which they can use to hire more and better teachers
- districts with poor inhabitants have large classes and bad teachers.
  - These districts have little money and can hire only few and not very good teachers

In this case class size is related to teacher quality.

Since teacher quality likely affects test scores it is contained in $u_i$.

This implies a violation of assumption 1:

$$E(u_i | \text{ClassSize}_i = \text{small}) \neq E(u_i | \text{ClassSize}_i = \text{large}) \neq 0$$
The Least Squares assumptions: Assumption 2

\[(Y_i, X_i) \text{ for } i = 1, \ldots, n \text{ are } i.i.d\]

- If the sample is drawn by simple random sampling assumption 2 will hold.

Example: What is effect of mother’s education \((X_i)\) on child’s education \((Y_i)\)?

Example of simple random sampling:

- randomly draw sample of mother’s with information on her education and the education of \textit{one randomly selected} child.
- \((Y_i, X_i) \text{ for } i = 1, \ldots, n \text{ are } i.i.d\)

Example of a violation of simple random sampling:

- randomly draw sample of mothers with information on her education and the education of \textit{all} of her children.
- \((Y_i, X_i) \text{ for } i = 1, \ldots, n \text{ are NOT } i.i.d\)
- Observations on children from the same mother are not \textit{independent}!
The Least Squares assumptions: Assumption 3

Large outliers are unlikely

$$0 < E \left( X_i^4 \right) < \infty \quad \& \quad 0 < E \left( Y_i^4 \right) < \infty$$

- Outliers are observations that have values far outside the usual range of the data
- Large outliers can make OLS regression results misleading
- Another way to state assumption is that $X$ and $Y$ have finite kurtosis.
- Assumption is necessary to justify the large sample approximation to the sampling distribution of the OLS estimators
The Least Squares assumptions: Assumption 3
Use of the Least Squares assumptions

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]

Assumption 1: \( E(u_i | X_i) = 0 \)

Assumption 2: \((Y_i, X_i)\) for \(i = 1, \ldots, n\) are i.i.d

Assumption 3: Large outliers are unlikely

If the 3 least squares assumptions hold the OLS estimators \(\hat{\beta}_0\) and \(\hat{\beta}_1\)

- Are unbiased estimators of \(\beta_0\) and \(\beta_1\)
- Are consistent estimators of \(\beta_0\) and \(\beta_1\)
- Have a jointly normal sampling distribution
Properties of the OLS estimator: unbiasedness

\[ Y_i = \beta_0 + \beta_1 X_i + u_i \]
\[ \bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{u} \]

\[
E \left[ \hat{\beta}_1 \right] = E \left[ \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right]
\]

Substitute for \( Y_i, \bar{Y} \)

\[
= E \left[ \frac{\sum_{i=1}^{n} (X_i - \bar{X})(\beta_0 + \beta_1 X_i + u_i - (\beta_0 + \beta_1 \bar{X} + \bar{u}))}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right]
\]

Rewrite (\( \beta_0 \) drops out)

\[
= E \left[ \frac{\sum_{i=1}^{n} (X_i - \bar{X})(\beta_1(X_i - \bar{X}) + (u_i - \bar{u}))}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right]
\]

Rewrite & use expectation rules

\[
= E \left[ \frac{\beta_1 \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right] + E \left[ \frac{\sum_{i=1}^{n} (X_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right]
\]
Properties of the OLS estimator: unbiasedness

\[ E\left[ \hat{\beta}_1 \right] = E \left[ \frac{\beta_1 \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})} \right] + E \left[ \frac{\sum_{i=1}^{n} (x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})} \right] \]

*take \( \beta_1 \) out of 1st expectation*

*Algebra trick*

\[ = \beta_1 + E \left[ \frac{\sum_{i=1}^{n} (x_i - \bar{x})u_i}{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})} \right] \]

*Law of iterated expectations*

\[ = \beta_1 + E \left[ \frac{\sum_{i=1}^{n} (x_i - \bar{x})E[u_i|X_i]}{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})} \right] \]

\[ E\left[ \hat{\beta}_1 \right] = \beta_1 \quad \text{if} \quad E[u_i|X_i] = 0 \]
Algebra trick

\[ \sum_{i=1}^{n} \left( X_i - \bar{X} \right) (u_i - \bar{u}) \]

\[ = \sum_{i=1}^{n} X_i u_i - \sum_{i=1}^{n} X_i \bar{u} - \sum_{i=1}^{n} \bar{X} u_i + \sum_{i=1}^{n} \bar{X} \bar{u} \]

\[ = \sum_{i=1}^{n} X_i u_i - n \cdot \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) \bar{u} - \sum_{i=1}^{n} \bar{X} u_i + n\bar{X} \bar{u} \]

\[ = \sum_{i=1}^{n} X_i u_i - n\bar{X} \bar{u} + \sum_{i=1}^{n} \bar{X} u_i + n\bar{X} \bar{u} \]

\[ = \sum_{i=1}^{n} X_i u_i - \sum_{i=1}^{n} \bar{X} u_i \]

\[ = \sum_{i=1}^{n} \left( X_i - \bar{X} \right) u_i \]
Consistency: $\hat{\beta}_1 \xrightarrow{p} \beta_1$ or $\text{plim} \hat{\beta}_1 = \beta_1$

\[
\text{Plim} \hat{\beta}_1 = \text{plim} \left( \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right)
\]

\[
= \text{Plim} \left( \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})} \right) = \frac{s_{XY}}{s_X^2}
\]

law of large numbers
OLS assumptions 2 and 3

\[
= \frac{\text{Cov}(X_i, Y_i)}{\text{Var}(X_i)}
\]

substitute for $Y_i$

\[
= \frac{\text{Cov}(X_i, \beta_0 + \beta_1 X_i + u_i)}{\text{Var}(X_i)}
\]

see Key Concept 2.3

\[
= \frac{\beta_1 \text{Var}(X_i) + \text{Cov}(X_i, u_i)}{\text{Var}(X_i)}
\]
Consistency

\[ \text{Plim } \hat{\beta}_1 = \frac{\beta_1 \text{Var}(X_i) + \text{Cov}(X_i, u_i)}{\text{Var}(X_i)} = \beta_1 \frac{\text{Var}(X_i)}{\text{Var}(X_i)} + \frac{\text{Cov}(X_i, u_i)}{\text{Var}(X_i)} \]

Substitute covariance expression

\[ = \beta_1 + \frac{E[(X_i - \mu_x)(u_i - \mu_u)]}{\text{Var}(X_i)} \]

Algebra trick

\[ = \beta_1 + \frac{E[(X_i - \mu_x)u_i]}{\text{Var}(X_i)} \]

Law of iterated expectations

\[ = \beta_1 + \frac{E[(X_i - \mu_x)E[u_i|X_i]]}{\text{Var}(X_i)} \]

So

\[ \text{Plim } \hat{\beta}_1 = \beta_1 \text{ if } E[u_i|X_i] = 0 \]
Unbiasedness vs Consistency

- Unbiasedness & consistency both rely on $E [u_i | X_i] = 0$

- Unbiasedness implies that $E \left[ \hat{\beta}_1 \right] = \beta_1$ for a given sample size $n$

- Consistency implies that the sampling distribution becomes more and more tightly distributed around $\beta_1$ if the sample size $n$ becomes larger and larger.
Consistency: A simulation example

- Let's create a data set with 100 observations
- $X_i \sim N(0, 1)$
- $u_i \sim N(0, 1)$
- We define Y to depend on X as: $Y_i = 1 + 2X_i + u_i$

```
set obs 1000
gen x=invnorm(uniform())
gen y=1+2*x+invnorm(uniform())
```

```
. sum y x

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>100</td>
<td>.6123606</td>
<td>2.211365</td>
<td>-5.05828</td>
<td>5.462746</td>
</tr>
<tr>
<td>x</td>
<td>100</td>
<td>-.1479108</td>
<td>.9928607</td>
<td>-2.633841</td>
<td>1.80305</td>
</tr>
</tbody>
</table>
```
A simulation example

. regress y x

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>385.987671</td>
<td>1</td>
<td>385.987671</td>
<td>F( 1, 98) = 385.45</td>
</tr>
<tr>
<td>Residual</td>
<td>98.1357149</td>
<td>98</td>
<td>1.00138485</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>484.123386</td>
<td>99</td>
<td>4.89013521</td>
<td>R-squared = 0.7973</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.7952</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 1.0007</td>
</tr>
</tbody>
</table>

| y         | Coef.       | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-----------|-------------|-----------|-------|------|---------------------|
| x         | 1.988753    | .1012965  | 19.63 | 0.000| 1.787733 2.189772   |
| _cons     | .9065187    | .1011847  | 8.96  | 0.000| .705721 1.107316    |
A simulation example n=100

We can create 999 of these data sets with 100 observations and use OLS to estimate

\[ Y_i = \beta_0 + \beta_1 + u_i \]

1. program define ols, rclass
   1.   drop _all
   2.   set obs 100
   3.   gen x=invnorm(uniform())
   4.   gen y=1+2*x+invnorm(uniform())
   5.   regress y x
   6.   end

2.
3. simulate _b, reps(999) nodots : ols

   command:  ols

4. sum

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>_b_x</td>
<td>999</td>
<td>1.997521</td>
<td>.1018595</td>
<td>1.67569</td>
<td>2.308795</td>
</tr>
<tr>
<td>_b_cons</td>
<td>999</td>
<td>1.003246</td>
<td>.1019056</td>
<td>.6844429</td>
<td>1.285363</td>
</tr>
</tbody>
</table>
A simulation example $n=100$

OLS estimates of $B_1$ in 999 samples with $n=100$
A simulation example n=1000

1. program define ols, rclass
   1.  drop _all
   2.  set obs 1000
   3.  gen x=invnorm(uniform())
   4.  gen y=1+2*x+invnorm(uniform())
   5.  regress y x
   6.  end

2.
3.  simulate _b, reps(999) nodots : ols

    command:  ols

4.  sum

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>_b_x</td>
<td>999</td>
<td>2.000035</td>
<td>.030417</td>
<td>1.908725</td>
<td>2.112585</td>
</tr>
<tr>
<td>_b_cons</td>
<td>999</td>
<td>1.000791</td>
<td>.0311526</td>
<td>.8970624</td>
<td>1.088724</td>
</tr>
</tbody>
</table>
A simulation example $n=1000$

OLS estimates of $B_1$ in 999 samples with $n=1000$
A simulation example n=10000

1. program define ols, rclass
   1.   drop _all
   2.   set obs 10000
   3.   gen x=invnorm(uniform())
   4.   gen y=1+2*x+invnorm(uniform())
   5.   regress y x
   6.   end

2.
3.   simulate _b, reps(999) nodots: ols
    command:  ols

4.   sum

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>_b_x</td>
<td>999</td>
<td>1.999748</td>
<td>.0099715</td>
<td>1.969678</td>
<td>2.034566</td>
</tr>
<tr>
<td>_b_cons</td>
<td>999</td>
<td>1.000391</td>
<td>.0100135</td>
<td>.9699681</td>
<td>1.033458</td>
</tr>
</tbody>
</table>
A simulation example $n=10000$

OLS estimates of $B_1$ in 999 samples with $n=10000$
Consistency of the OLS estimator of $\hat{\beta}_1$

**True model**: $Y_i = 1 + 2X_i + u_i$,  
**Estimated model**: $Y_i = \beta_0 + \beta_1 X_i + u_i$

OLS estimates of $\beta_1$ in 999 samples with $n=100$; $n=1000$ and $n=10000$
We discussed the sampling distribution of the sample average $\bar{Y}$:

- sampling distribution is complicated for small $n$, but if $Y_1, \ldots, Y_n$ are i.i.d. we know that
  \[ E(\bar{Y}) = \mu_Y \]

- By the Central Limit theorem the large sample distribution can be approximated by the normal distribution:
  \[ \bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{n}\right) \]

If the 3 least squares assumptions hold we can make similar statements about the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$
Large-sample distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

- Technically the Central Limit theorem concerns the large sample distribution of averages (like $\bar{Y}$)

- Examining the formulas of the OLS estimators shows that these are functions of sample averages:

  \[
  \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \\
  \hat{\beta}_1 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y}) \cdot \frac{1}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})(x_i - \bar{X})}
  \]

- It turns out that the Central Limit theorem also applies to these functions of sample averages.
Sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

If the first least squares assumption holds:

- The OLS estimators are unbiased which implies that (for any sample size $n$)
  \[ E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad E(\hat{\beta}_1) = \beta_1 \]

In addition, if all 3 least squares assumptions hold:

- The Central Limit theorem implies that $\hat{\beta}_0$ and $\hat{\beta}_1$ are approximately jointly normally distributed in large samples:
  \[ \hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0}) \]
  \[ \hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \]
In large samples

\[\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0})\]

\[\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})\]

where it can be shown that

\[\sigma^2_{\hat{\beta}_0} = \frac{1}{n} \frac{\text{Var}(H_i u_i)}{\left[\text{E}(H_i^2)\right]^2} \quad \text{with} \quad H_i = 1 - \left[\frac{\mu_X}{\text{E}(X_i^2)}\right] X_i\]

\[\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \frac{\text{Var}[(X_i - \mu_X) u_i]}{[\text{Var}(X_i)]^2}\]

Expression for \(\sigma^2_{\hat{\beta}_1}\) shows that the larger the variation in the regressor \(X_i\) the smaller the variance of \(\hat{\beta}_1\)
Large-sample distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

- When $\text{Var}(X_i)$ is low, it is difficult to obtain an accurate estimate of the effect of $X$ on $Y$ which implies that $\text{Var}(\hat{\beta}_1) = \sigma_{\hat{\beta}_1}^2$ is high.
- If there is more variation in $X$, then there is more information in the data that you can use to fit the regression line.
• Traffic fatalities are the leading cause of death for Americans between the ages of 5 and 32.

• The government wants to decrease the number of traffic fatalities by increasing seat belt usage.

• If many people wear seat belts the chance that people die in a car crash is likely smaller.

• People who wear seat belts might however be more careful drivers.

• Regions with many seat belt users might have fewer traffic fatalities not because of the seat belt usage but because the drivers are more careful.
In the term paper you are going to investigate the following research question.

What is the causal effect of seat belt usage on traffic fatalities?

This research question can be addressed by using the data set seatbelts.dta.

Data consists of a panel of 50 U.S. States, plus the District of Columbia, for the years 1983-1997.

The data sets can be downloaded from the course website site.

In analyzing this data you may consider the use of panel data methods on top of a pure cross-section analysis.
The term paper should consist of the following sections:

- Introduction
- Empirical approach
- Data
- Results
- Conclusion
- References
- Appendix with Stata code & output

The term paper should be at most 10 pages including tables and figures (but excluding the stata code and output).

The quality (and not the quantity) of the content of the term paper will determine your grade.
Compulsory term paper

You are expected to work in a group of two students.

- You can form a group of two students yourself
- Register this group before 29 January 2017 00:00, by using link in email you will receive today.
- If you are unable to form a group, please let me know before 29 January 2017.
  - you will be randomly assigned to another student.

Important dates:

- 25 January 2017 – Hand-out of term paper
- 22 March 2017 – Hand-in of term paper on Fronter
- 11 April 2017 – Notification of grade (pass/fail)
- 21 April 2017 – Hand-in of improved term paper for those who failed
- 4 May 2017 – Everyone is informed about final grade for term paper