ECON 3150/4150, Spring term 2013. Lecture 4

The regression model with deterministic regressor (part I)

Ragnar Nymoen

University of Oslo

23 January 2013

References to Lecture 4 and 5

- Hill, Griffiths and Lim (**HGL)**
  - Chapter 2 and 3
- Bårdsen and Nymoen (**BN)**
  - Kap 5.1-5.5

Model specification I

We follow convention and formulate our first model as a linear relationship with three parts

1. Dependent variable, Y_i
2. Economic theory (the explanation): $\beta_0 + \beta_1 X_i$
3. A random disturbance term, ε_i

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1, 2, ..., n \quad (1)$$

- $\{Y_i, \varepsilon_i\}$ $i = 1, 2, \ldots, n$ are random variables
- $\{X_i\}$ $i = 1, 2, \ldots, n$ are n values of a non-random variable.
- Since $\varepsilon_i \rightarrow Y_i$, it is clear that the Y_i variables "inherit" their random properties from the disturbance term ε_i.

Model specification II

- The non-random (deterministic variable) X_i only affects the expectation $E(Y_i)$.
- (1) is a generalization of the model we used to make inference about $E(Y_i)$ at the end of Lecture 3:
- In that model, $\beta_1 = 0$ was imposed *a priori*. In (1) we want to estimate β_1 (together with β_0) and make inference about β_1.
- β_0 and β_1 are the unknown **parameters** of the equation.

Model specification III

- As mentioned already in Lecture 2, the interpretation of the *slope coefficient* β_1 in particular depends on how Y and X are measured:
  - If, for example, Y is expenditure on a certain good in kroner and X is total consumption expenditure in kroner, then β_1 is the **derivative** of Y with respect to X
  - If Y and X are variables that have been transformed to the natural logarithms of the corresponding kroner expenditures, then the interpretation of β_1 changes to **elasticity.**
- HGL cover many of these issues in Ch 4, including R^2, and the spilt of total sum of squares into explained sum of squares and residual sum of squares, that we have already covered, since their role is to characterize the regression line's fit to a given sample.

Re-parameterisation of the equation I

- The following *re-parameterisation* of (1) is often useful:

$$Y_i = \alpha + \beta_1(X_i - \bar{X}) + \varepsilon_i, \; i = 1, 2, ..., n \tag{2}$$

  where

$$\alpha \equiv \beta_0 + \beta_1 \bar{X}, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- The "trick" is the same as the one we used for in Lecture 1 when Y_i was simply given numbers.
- It is a valid operation also in the context that we are in now, where Y_i is a random variable which is function of the random variable ε_i.

Re-parameterisation of the equation II

- The point is that the disturbance ε_i is unaffected. It is only the parameters of the equation that is affected (and only the constant term in this case).
- Therefore, (2) is a re-parameterisation of (1).

RM1—econometric specification

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \equiv \alpha + \beta_1 (X_i - \bar{X}) + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

a. X_i are fixed numbers, $(i = 1, 2, \ldots, n)$
b. $E(\varepsilon_i) = 0$, $\forall\ i$, ("for all i")
c. $Var(\varepsilon_i) = \sigma^2$, $\forall\ i$
d. $Cov(\varepsilon_i, \varepsilon_j) = 0$, $\forall\ i \neq j$
e. α, β_0, β_1 and σ^2 are constant parameters

For the purpose of statistical inference we will often assume normally distributed disturbances:

f. $\varepsilon_i \sim N(0, \sigma^2)$.

Comments to the econometric specification I

b., **c.** and **d.** are often referred to as the **Classical assumptions** about the regression disturbance. We will also follow that convention.

b. $E(\varepsilon_i) = 0$, Note that if $E(\varepsilon_i) = b \neq 0$, then the model can be re-stated as

$$Y_i = \underbrace{\beta_0 + b}_{\beta_1'} + \beta_1 X_i + \underbrace{\varepsilon_i - b}_{\varepsilon_i'}$$

with assumptions b. - f. holding for ε_i'.

Despite the warning at the bottom of page 46 in HGL, assumption b. $E(\varepsilon_i) = 0$ therefore seems to be innocuous, as long as we are not interested in the intercept β_0 per se.

Comments to the econometric specification II

c. $Var\left(\varepsilon_i\right) = \sigma^2, \forall\ i$

This assumption is called **Homoskedasticity.**

$$Var\left(\varepsilon_i\right) \neq \sigma^2, \forall i$$

is called **Heteroskedasticity.**

Heteroskedasticity is **not** an innocuous assumption!

To understand why, we need to develop the theory of estimators for RM1.

Comments to the econometric specification III

d. $Cov\left(\varepsilon_i, \varepsilon_j\right) = 0, \forall\ i \neq j$
Failure to meet this assumption about uncorrelated disturbances also has serious consequences.
For cross-section data, $Cov\left(\varepsilon_i, \varepsilon_j\right) \neq 0$ may be called "cross-section dependence".
For time series data, the case of

$$Cov(\varepsilon_t, \varepsilon_{t-s}) \neq 0 \text{ for } \ s = \pm 1, \pm 2, ...$$

is called **serial correlated errors** or **autocorrelated errors**.

Comments to the econometric specification IV

DIY exercise 1:

1. Show that when assumption **a.** is true, then **d**. can alternatively be written as:

$$E(\varepsilon_i \varepsilon_j) = 0, \forall \ i \neq j$$

2. Show that the model specification implies

$$E(Y_i) = \alpha + \beta_1(X_i - \bar{X})$$
$$Var(Y_i) = Var(\varepsilon_i) = \sigma^2, \forall i$$
$$Cov(Y_i, Y_j) = 0, \ \forall i \neq j$$

OLS estimates I

- A given data consists of one realization (value) of each of the n random variables Y_i , $i = 1, 2, \ldots, n$ which can write $y_1, y_2, \ldots, y_n$ and the n fixed values x_1,x_2,$\ldots$, x_n of the explanatory variable.Use of OLS estimation on a data set (Lecture 2) result in the OLS estimates:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad (\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0)$$

$$\hat{\alpha} = \bar{y}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

- These estimates are sample specific numbers.

OLS estimates II

- We see that the only difference from Lecture 2 is that we have used lower case letters to represent values of the variables.
- However, we can imagine that we get access to a second sample, with another realization of the n stochastic variables Y_i.
- What would you do in terms of estimation?

OLS estimates III

- Apply OLS again!
- And again for a third and fourth realization of the random variables!

- Hence we can define a stochastic variable $\hat{\beta}_1$ which is a function of the random variables Y_i, $i = 1, 2, \dots, n$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sum_{i=1}^{n} w_i Y_i \qquad (3)$$

where

$$w_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (4)$$

- The interpretation is:

$$\underset{\text{random}}{\varepsilon_i} \underset{(2)}{\rightarrow} Y_i \rightarrow \sum_{i=1}^{n} w_i Y_i = \frac{\sum_{i=1}^{n}(X_i - \bar{X})Y_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \rightarrow \underset{\text{random}}{\hat{\beta}_1}$$

- $\hat{\alpha}$ and $\hat{\beta}_0$ are also reinterpreted as random variables:

$$\hat{\alpha} = \bar{Y}, \tag{5}$$
$$\hat{\beta}_0 = \hat{\alpha} - \hat{\beta}_1 \overline{X} \tag{6}$$

- We see that $\hat{\alpha}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ take a *double-meaning*, as estimates and estimators (random variables).
- It will be clear form the context which interpretation we have in mind.
- For the same reason, Y_i from now on takes the same double meaning as a random variable and a realization of that variable

Expectation and bias I

- We are interested in $E(\hat{\beta}_1)$ since we want to evaluate the bias $E(\hat{\beta}_1 - \beta_1)$
- $\hat{\beta}_1$ is a linear function of the Y_i variables. The OLS estimator is a *linear estimator*.
- Can therefore find $E(\hat{\beta}_1)$ by use of the rules for expectation.

Re-write the estimator as:

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i(\beta_0 + \beta_1 X_i + \varepsilon_i) = \beta_1 + \sum_{i=1}^{n} w_i \varepsilon_i$$

Expectation and bias II

using

$$\sum_{i=1}^{n} w_i = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \sum_{i=1}^{n}(X_i - \bar{X})_i = 0$$

$$\sum_{i=1}^{n} w_i X_i = \frac{1}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \sum_{i=1}^{n}(X_i - \bar{X})_i X_i = 1$$

Take the expectation through;

$$E(\hat{\beta}_1 - \beta_1) = E\left(\sum_{i=1}^{n} w_i \varepsilon_i\right) = \sum_{i=1}^{n} w_i E(\varepsilon_i) = 0$$

Hence

$E(\hat{\beta}_1 - \beta_1) = 0$, **unbiasedness** of $\hat{\beta}_1$

Variance I

$$Var(\hat{\beta}_1) = Var(\beta_1 + \sum_{i=1}^{n} w_i \varepsilon_i) = \sigma^2 \sum_{i=1}^{n} w_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

If we re-introduce the empirical variance of the deterministic X:

$$\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

(alternatively divide by $(n-1)$) we get the compact expression

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{n \hat{\sigma}_X^2}$$

Variance II

1. Larger disturbances variance increases $Var(\hat{\beta}_1)$ and therefore estimation uncertainty
2. Large variability in the explanatory variable reduces $Var(\hat{\beta}_1)$
3. More observations reduce $Var(\hat{\beta}_1)$

Intercept estimator properties I

- You can show that

$$E(\hat{\alpha}) = \alpha \quad (7)$$

$$E(\hat{\beta}_1) = \beta_1 \quad (8)$$

and

$$\begin{aligned} Var(\hat{\alpha}) &= \frac{\sigma^2}{n} \\ Var(\hat{\beta}_0) &= Var(\hat{\alpha}) + \bar{X}^2 Var(\hat{\beta}_0) - 2\bar{X} Cov(\hat{\alpha}, \hat{\beta}_0) \\ &= \frac{\sigma^2}{n}(1 + \bar{X}^2 \frac{1}{\hat{\sigma}_x^2}) \quad (9) \end{aligned}$$

Intercept estimator properties II

$Var(\hat{\beta}_0)$ makes use of

$$Cov(\hat{\alpha}, \hat{\beta}_1) = 0 \tag{10}$$

- Why is (10) true? See BN Appendix 5.A for a proof (English translation on the web-page).
- In the exercises to *Seminar 2* you are asked to find the expression for $Cov(\hat{\beta}_0, \hat{\beta}_1)$.

Summing up so far

- For RM1, and before invoking the assumption about normality of ε_i, we have that the OLS estimators for β_0,α and β_1 are:
- **Unbiased** (On average $\hat{\beta}_1 - \beta_1$ is zero, for example)
- And have **well defined variances and covariances** that depend on σ^2, the sample size n, and how much variation there is in X.

Gauss-Markov theorem I

As noted, the OLS estimator $\hat{\beta}_2$ is a **linear** estimator

$$\hat{\beta}_1 = \sum_{i=1}^{n} w_i Y_i, \text{ with } w_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

which is unbiased.

- The **Gauss-Markov theorem** says that there is no other estimator for the parameter β_1 in RM1 that is linear and unbiased and that has lower variance than $\hat{\beta}_1$ for a given sample size n
- The same is true for $\hat{\beta}_0$ (and $\hat{\alpha}$). We say that for RM1, the OLS estimators are **best linear unbiased estimators** (**BLUE**)
    - There are proofs is in both books,:
    - HGL appendix 2.F, BN: kap 5.3.4

Gauss-Markov theorem II

- so we only outline the argument here, and leave the details for self study.

That other estimator for β_1 takes the form

$$\hat{\beta}_1' = \sum_{i=1}^{n} c_i Y_i, \text{ with fixed weights } c_i$$

We can define δ_i

$$\delta_i = c_i - w_i,\ i = 1, 2, \ldots, n$$

as a measure of the difference between the two set of weights.
We require

$$E(\hat{\beta}_1') = \beta_1$$

Gauss-Markov theorem III

which implies the following for δ_i:

$$\sum_{i=1}^{n} \delta_i = 0$$

$$\sum_{i=1}^{n} \delta_i Y_i = 0$$

which allows us to write

$$Var(\hat{\beta}_1') = \sigma^2 \left[\sum_{i=1}^{n} w_i^2 + \sum_{i=1}^{n} \delta_i^2 \right]$$

so that

$$Var(\hat{\beta}_1') > Var(\hat{\beta}_1) \text{ unless } \delta_i = 0$$

Gauss-Markov theorem IV

and in that case

$$\hat{\beta}_1' \equiv \hat{\beta}_1.$$

Estimating the variance of the disturbance I

- The OLS principle itself—the normal equation (1ocs) from Lecture 2—does not give an estimator for σ^2.
- But it is natural to use the sum of squares of the OLS residuals, i.e.,

$$\sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

with $\hat{\varepsilon}_i$ interpreted as random variable.

$$\hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta}_1(X_i - \bar{X})$$

Estimating the variance of the disturbance II

- It is possible to show that

$$\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sigma^2} \sim \chi^2(n-2) \tag{11}$$

where the loss of one degree of freedom compared to the case in Lecture 3 where the model was

$$Y_i = \beta_0 + \varepsilon_i$$

Estimating the variance of the disturbance III

has to do with the fact we now have two restrictions between the n random variables in the form of the two normal equations:

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \tag{12}$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i (X_i - \overline{X}) = 0. \tag{13}$$

Estimating the variance of the disturbance IV

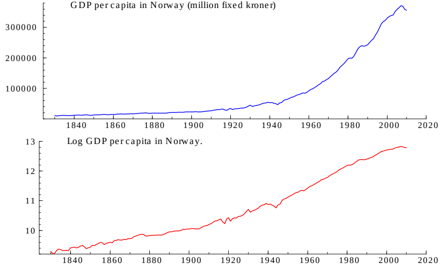
- Because of the $\chi^2(n-2)$ distribution in (11) we have

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-2}\left[\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sigma^2}\right] = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-2}$$

  is an unbiased estimator of $\hat{\sigma}^2$ given the regression model that we have formulated.
- Show!
- What is the expression for $Var(\hat{\sigma}^2)$?

Checking the results by simulation

- We can use Monte Carlo simulation to “check” the theory that we have developed
- See Appendix 2G in HGL (the explanation of the methodology) or Kap 5.3.1 in BN
- A note about Monte Carlo simulations on the course web page [POSTPONED TO LATER]
- Two small Monte Carlo’s in the seminars!



- Blue graph: GDP per capita Y against time, t
- t is deterministic
- Approx non-linear $Y(t)$ by $Y = Ae^{g_Y t + \varepsilon_t}$
- ε_t is a random error
- Red graph shows $\ln Y$ against time $\ln Y_t = \ln A + g_y t + \varepsilon_t$ is an example of RM1.