ECON 3150/4150, Spring term 2014. Lecture 4

The regression model (part I)

Ragnar Nymoen

University of Oslo

23 January 2014

References to Lecture 4

- **SW**
  - Ch. 4 (plus CH. 17, references are given in the slide set)
- Bårdsen and Nymoen (**BN)**
  - Kap 5.1-5.8

Looking back—and ahead I

- In Lecture 2 we reviewed the *statistical model*

$$Y_i = \mu_Y + \varepsilon_i,\ i = 1, 2, \ldots, n$$

where $\varepsilon_i \sim i.i.d(0, \sigma^2)\ \forall i$ or $\varepsilon_i \sim IIN(0, \sigma^2)\ \forall i$.

- Reminded ourselves that $\hat{\mu}_Y = \bar{Y}$ is a good estimator of μ_Y that can be used to test hypotheses like

$$H_0\colon\ \mu_Y = \mu_Y^0 \quad \text{against} \quad H_1\colon\ \mu_Y > \mu_Y^0.$$

or

$$H_0\colon\ \mu_Y = \mu_Y^0 \quad \text{against} \quad H_1\colon\ \mu_Y = \mu_Y^0.$$

for example.

- We now want to extend the statistical model to include an economic explanatory variable X to obtain an *econometric model.*

Modelling concepts and terminology I

- We formulate our first econometric model as a linear (in parameters) relationship between the *regressand,* Y, and the *regressor* X.
- The relationship, often called the population regression line, holds "on average" for n variables $\{Y_i, X_i\}$ $i = 1, 2, \ldots, n$:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, ..., n \tag{1}$$

- β_0 and $\beta_1$are *parameters* (non random numbers) to be estimated using a sample of n observations of $\{Y_i, X_i\}$.
- ε_i is the *disturbance term* (u_i in SW)
- Y_i and ε_i are always random variables.
- X_i can be either random or deterministic.

Modelling concepts and terminology II

- The model formulation in SW (Key Concept 4.3) appears to rule out the case of deterministic regressor. This is hardly the intention, since it would crash with the efficient way of estimating the *Difference-of-Means* model in Ch. 3.5.
  - Cf. Lecture 3: Using an indicator variable (dummy) variable to estimate the treatment effect in a natural experiment.
- In this lecture, we therefore make a distinction between the regression model with deterministic X (call it *RM1*) and the regression model with stochastic regressor (*RM2*).
- The parameter β_0 is called the *constant term*, or the *intercept coefficient*.
- The parameter β_1 is referred to as the *regression coefficient*, the *slope coefficient* or the *derivative coefficient.*

Modelling concepts and terminology III

- The economic interpretation of the *slope coefficient* β_1 depends on how Y and X are measured:
  - If, 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*.
  - If X is an indicator variable (dummy) or step-variable (composite dummy) the interpretation is different, as we shall see.

Re-parameterisation of the regression equation I

- Before we specify the models, we note a useful way of re-writing equation (1), namely

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

  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 3 for the equation for the straight line in the scatter plot of data point.
- One important point in the present setting is that the random disturbance ε_i is unaffected.

Re-parameterisation of the regression equation II

- It is only the parameters of the equation that are changed (in fact, only the constant term in this case).
- We say that (2) is a *re-parameterisation* of (1) that leave the stochastic properties of the model unchanged.

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. $\{X_i\}$ $i = 1, 2, \ldots, n$ are fixed numbers, $\sum_{i=1}^{2} (X_i - \bar{X})^2 > 0$
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.i $\varepsilon_i \sim IIN(0, \sigma^2)$.

With reference to asymptotic theory (Lecture 1 and 2), the normality assumption can be replaced by the weaker assumption:

f.ii $\{\varepsilon_i\}$ $i = 1, 2, \ldots, n$. are $i.i.d.$ distributed with finite (but non-zero) fourth order moments—no excess kurtosis in large samples.

Comments to the econometric specification, RM1 I

- Since X_i is deterministic we can set

$$\mu_{Y_i} = \beta_1 + \beta_1 X_i$$

and write assumption b.-d. as:

$$\begin{aligned} \mu_{Y_i} \equiv E(Y_i) &= \beta_1 + \beta_1 X_i \\ var\left(Y_i\right) &= \sigma^2, \, \forall i \\ cov\left(Y_i, Y_j\right) &= 0, \forall i \neq j \end{aligned}$$

Comments to the econometric specification, RM1 II

- **b.**, **c.** and **d.** are often referred to as the *Classical assumptions* about the regression disturbance.
- $var(\varepsilon_i) = \sigma^2, \forall\ i$
  This assumption is called **Homoskedasticity,** while

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

  is called **Heteroskedasticity.**
- $Cov(\varepsilon_i, \varepsilon_j) = 0, \forall\ i \neq j$
  For cross-section data, $Cov(\varepsilon_i, \varepsilon_j) \neq 0$ 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**.

OLS estimates I

- In lecture 3 we derived the OLS estimates $\hat{\beta}_0$ (alternatively $\hat{\alpha}$) and $\hat{\beta}_1$.
- These estimates are sample specific numbers.
- 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 II

- Apply the least-squares principle again!
- And again, for a third and fourth realization of the random variables!

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

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

where

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

- One interpretation that sometimes is helpful is:

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

- $\hat{\alpha}$ 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).

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 write the empirical variance of X as:

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

(instead of dividing 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, σ^2, increases $var(\hat{\beta}_1)$ and therefore estimation uncertainty
2. Large variability in the explanatory variable ($\hat{\sigma}_X^2$) reduces $var(\hat{\beta}_1)$
3. More observations (n) reduces $var(\hat{\beta}_1)$

Intercept estimator properties I

- You can show that

$$E(\hat{\alpha}) = \alpha \tag{7}$$
$$E(\hat{\beta}_1) = \beta_1 \tag{8}$$

and

$$\begin{aligned} var(\hat{\alpha}) &= \frac{\sigma^2}{n} \\ var(\hat{\beta}_0) &= var(\hat{\alpha}) + \bar{X}^2 var(\hat{\beta}_1) - 2\bar{X} cov(\hat{\alpha}, \hat{\beta}_1) \\ &= \frac{\sigma^2}{n}(1 + \bar{X}^2 \frac{1}{\hat{\sigma}_x^2}) \end{aligned} \tag{9}$$

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).
- As another DIY exercise: Derive an 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.
- **Consistency** of $\hat{\beta}_1$ and $\hat{\beta}_0$ (and $\hat{\alpha}$) follows from unbiasedness, and $var(\hat{\beta}_1) \rightarrow 0$ when $n \rightarrow \infty$ (under mild assumptions: $\sum_{i=1}^{\infty}(X_i - \bar{X})^2\} > 0$)

$$plim(\hat{\beta}_1) = \beta_1, plim(\hat{\beta}_0) = \beta_1,\ plim(\hat{\alpha}) = \alpha \qquad (11)$$

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:

Gauss-Markov theorem II

- SW, p. 110 for OLS estimation of constant and appendix 5.2 (not Ch 4!) for regression case. BN: kap 5.3.4

- 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.

Gauss-Markov theorem III

We require

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

which implies the following for δ_i:

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

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

which allows us to write

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

Gauss-Markov theorem IV

so that

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

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 3—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

- With reference to page 28 in the Lecture 1, if $\varepsilon_i \sim IIN(0, \sigma^2)$ we have that

$$\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\sigma^2} \sim \chi^2(n-2) \tag{12}$$

where the d.f is $n-2$ instead of $n-1$ because we now have *two restrictions* between the n random variables:

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \tag{13}$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i (X_i - \overline{X}) = 0. \tag{14}$$

when the model is $Y_i = \beta_0 + \varepsilon_t$ there are $n-1$ independent variables.

Estimating the variance of the disturbance III

- Because of the $\chi^2(n-2)$ distribution in (12) it follows that

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-2}\left[\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{\sigma^2}\right] = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-2} \qquad (15)$$

  is an unbiased and consistent estimator of $\hat{\sigma}^2$.
- Show!
- What is the expression for $var(\hat{\sigma}^2)$?
- If we relax the normality assumption and use $\varepsilon_i \sim i.i.d(0, \sigma^2)$ instead, we still have consistency of $\hat{\sigma}^2$ given by (15).
- Notation: In SW:

$$SER \equiv \hat{\sigma}^2$$

  see page 163.

Distribution function of OLS estimators I

- With the $\varepsilon_i \sim IIN\left(0, \sigma^2\right)$assumption, $\hat{\beta}_1$ is itself a normally distributes variable, for every sample size n.

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2})$$

- In the case of $i.i.d.$ we can refer to the CLT to understand that

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} \text{ normal distribution,}$$

since $\hat{\beta}_1 - \beta_1$ is a weighted sum of $i.i.d.$ random variables:

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

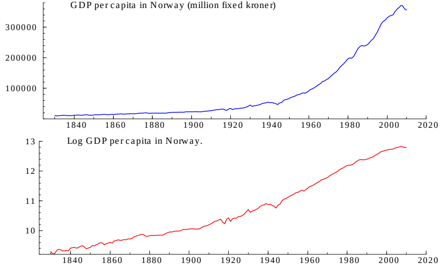
Gender gap in US earnings

- See Table 3.1 in SW.
- Let Y_i denote average hourly earnings (of working college graduates) in USA.
- Let X_i be 1 if the average hourly earnings is for women, and 0 if it is for men.
- When we use OLS on the $n = 10$ observations, the result is:

$$Y_i = \underset{(0.4627)}{24.1460} - \underset{(0.6543)}{3.81400} X_i + \hat{\varepsilon}_i$$

where the numbers below the estimates are the estimated standard errors $\sqrt{var(\hat{\beta}_0)}$ and $\sqrt{var(\hat{\beta}_1)}$.

GDP pr capita growth



- 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 regression with a continuous deterministic X

Model specification with random X (RM2) I

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \equiv \alpha + \beta_1 (X_i - \bar{X}) \; i = 1, 2, \ldots, n \qquad (16)$$

Assumptions:

a. $\{X_i, Y_i\}$ $(i = 1, 2, \ldots, n)$ are IID pairs, with $cov(X_i, Y_i) \geq 0$, $var(X_i) = \sigma_X^2 > 0$, $var(Y_i) = \sigma_Y^2 > 0 \; \forall \; i$

b. $E\left(\varepsilon_i \mid X_h\right) = 0$, $\forall \; i$ and h

c. $var\left(\varepsilon_i \mid X_h\right) = \sigma^2$, $\forall \; i$ and h and finite 4th moments (no excess kurtosis)

d. $cov\left(\varepsilon_i, \varepsilon_j \mid X_h\right) = 0$, $\forall \; i \neq j$, and for all h

e. β_0, β_1 and σ^2 are constant parameters

Model specification with random X (RM2) II

For the purpose of statistical inference in small samples we may invoke normally distributed disturbances:

f. $\varepsilon_i \sim IIN\left(0, \sigma^2 \mid X_h\right)$.

- This specification is (more or less) the same as Key Concept 4.3 and 17.1 in SW
- Remark about the IID assumption:
  - We have seen that if

$$E(Y_i \mid X_i) \equiv \beta_0 + \beta_1 X_i$$

    it is true that

$$E\left(\varepsilon_i \mid X_i\right) = 0 \;\; \forall i$$

    also for non $i.i.d.$ random variables.

Model specification with random X (RM2) III

- Hence, if we set $h = i$, assumption *b.* is really a consequence of the linearity of $E(Y_i \mid X_h)$
- Conversely, *b.*, *c*, and *d* in the list of assumptions can be seen as a consequence of linearity of $E(Y_i \mid X_h)$ **and** the i.i.d. assumption.

- Note that in RM2 because of the conditioning we have the reasonable implication that $var\left(\varepsilon_i \mid X_i\right) = \sigma^2 \leq \sigma_Y^2$

Properties of estimators, RM2

- RM2 is really RM1 re-expressed in terms of the *conditional moments* of ε_i
- Therefore, **all the properties that we have shown for $\hat{\beta}_1$ also hold for RM2**.
- In particular we have, for *RM2*

$$E(\hat{\beta}_1 \mid X) = \beta_1 \quad (17)$$

where $\mid X$ means conditional on given values of $X_1, X_2, \ldots, X_n$. The proof is exactly the same calculation as for *RM1* since all the X's are fixed numbers.
- Next, use the *Law of iterated expectations*

$$E(\hat{\beta}_1) = E(E(\hat{\beta}_1 \mid X)) = \beta_1 \quad (18)$$

to show that the OLS estimator of $\hat{\beta}_1$ is also unconditionally unbiased.
- The same is true for the OLS estimators of the intercept ($\hat{\alpha}$ and $\hat{\beta}_0$)
- The variance expression is also the same

$$var(\hat{\beta}_1) = var(\hat{\beta}_1 \mid X) = \frac{\sigma^2}{(n-1)\hat{\sigma}_X^2}$$

and the large sample distribution of $\hat{\beta}_1$ becomes

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \frac{\sigma^2}{\sigma_X^2}) \quad (19)$$

- References for self-study:
  - SW, p 170-170 and Appendix 4.3 and 4.4 omits some details that are filled in if you solve exercise 17.3
  - BN, a complete proof of (19) is in kap. 5.8.3
- The consistency property in (11) also holds for the random regressor model *RM2*.
  - Kap 5.8.2 has a complete proof for $plim(\hat{\beta}_1) = \beta_1$.

Summary—and looking ahead

- We have formulated two *classical regression models* (the random X version is known as the i.i.d. model for reasons that should by now be clear)
- We have seen that the properties of the OLS estimators are the same in the two models (including the BLUE property)
- The difference between RM1 and RM2 therefore lies in the interpretation:
- In RM1 the parameter β_1 shows how the unconditional expectation of Y varies, as a function of the deterministic variable t.
- That variation can be continuous, in which case β_1 is the derivative coefficient, or a step-function
- In RM2, β_1 is the slope coefficient in the conditional expectations function of Y given X.
- In the following we will have random X as of reference case.
- Can we have both deterministic and random regressors in one and the same regression model?
- Yes! A model like

$$Y_i = \beta_1 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

  where X_{1i} is random and X_{2i} is deterministic will be the typical case in applied work, and if the classical assumption of ε_i conditional on X_{1i} holds, the OLS estimators are BLUE in this model as well.
- Technically, this is an example of multivariate regression that we will return to later.
- Next: How can we use the regression model for statistical inference?
- And: How are the properties of the OLS estimators affected if one or more of the classical assumptions do not hold?