The regression model with one stochastic regressor.

3150/4150 Lecture 6

Ragnar Nymoen

Department of Economics, University of Oslo

30 January 2012

- We are now "on" Lecture topic 4
- The main goal in this lecture is to extend the results of the regression model to the case where the explanatory variable is stochastic.
- For that purpose we specificy a second econometric model, Regression model 2, that differs from RM1 in only one respect: the stochastic nature of the regressor.
- References:
  - HGL. Ch 10.1 and B.1.3, B.1.5, B.1.6 and B.1.7 (consider also the same for continous variables)
  - BN. kap 5.6, 5.7. 5.8. Building on kap 4.4

RM2—econometric specification

$$y_i = \beta_1 + \beta_2 x_i + e_i \equiv \alpha + \beta_2(x_i - \bar{x}) + e_i, \quad i = 1, 2, \ldots, n,$$

a. x_i $(i = 1, 2, \ldots, n)$ are stochastic variables (with finite 4th order moments).
b. $E(e_i \mid x_h) = 0$, $\forall$ i and h
c. $var(e_i \mid x_h) = \sigma^2$, $\forall$ i and h
d. $cov(e_i, e_j \mid x_h) = 0$, $\forall$ $i \neq j$, and for all h
e. α, β_0, β_1 and σ^2 are constant parameters

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

f. $e_i \sim N(0, \sigma^2 \mid x_h)$.

Assumptions I

- We assume explicitly that we our model is relevant for n pairs $\{x_i, y_i\}$ of stochastic variables.
- In general all the n pairs can be described by a joint **probability density distribution** (pdf) $f(x_1, y_1, x_2, y_2, ..., y_n, x_n)$.
- In the following we will assume a form of stochastic independence between pairs of variables

$$f(x_1, y_1, x_2, y_2, ..., y_n, x_n) = f(x_1, y_1) \times f(x_1, y_1) \times \ldots \times f(y_n, x_n)$$

- As HGL says, this can be seen as the case of n independent samples.

Assumptions II

- Of course, in economics is it easy to imagine that we have random samples that are **not independent** (time series data for example). In Lecture 7 we will at least indicate how the properties of RM2 needs to be modified to this more general case of stochastic regressor.
- We need to be precise about the notation

$$\mid x_h$$

in assumption **b**.-**f**.

- This means that the assumptions refer to a conditional distribution where x_h is a known parameter.

Assumptions III

- The assumptions therefore refer to the **conditional distribution** for the stochastic variable e_i where x_h is a parameter.
- If we had chosen the notation that upper case letters denote stochastic variables, and lower case letters denote realizations, assumption **b**. would read

$$E(e_i \mid X_h = x_h)$$

which is clearer (but more cumbersome).

- The expressions for the OLS estimators for β_1 and β_2 in RM2 are clearly the same as those we obtained for RM1.
- But:
- Do the *properties* of the OLS estimators, t-tests and confidence intervals
  also hold for RM2?
- If the answer is "Yes"......Why?
- We move in steps:
  - What does Monte Carlo analysis say?
  - Analysis based on probability theory. The keywords are the meaning and implications of **conditioning**

What does Monte Carlo simulation of RM2 show? I

- Consider experiment *Monte Carlo 2* in the note about Monte Carlo simulation
- Compare with the results we had from experiment *Monte Carlo 1*
- The only difference between these two experiments is that x is deterministic in *Monte Carlo 1* and stochastic in *Monte Carlo 2*.
- Since we have learnt about estimation of **standard errors** and **t-test statistics**, we also compare results for them in the two Monte Carlos

Conditional distribution of y given x I

The probability density function (pdf) $f(y_i, x_i)$ can be expressed in terms of a marginal density function and a conditional density function:

$$f(y_i, x_i) = f(y_i \mid x_i) \times f(x_i)$$

which is the density function counterpart to

$$P(A \cap B) = P(A \mid B) \times P(B)$$

where A and B are two events.

- $P(A \cap B)$ is the joint probability,
- $P(A \mid B)$ is the conditional probability for A given B,
- $P(B)$ is the marginal probability of B.

The regression function

We make the following assumption about $f(y_i \mid x_i)$:

1. It is described in terms of two parameters: Variance and expectation.
2. The expectation, μ_i, is a linear function of x_i:

$$\mu_i = E(y_i \mid x_i) = \beta_1 + \beta_2 x_i, \tag{1}$$

- This places the parameters β_1 and $\beta_2$of RM2 in the conditional expectations function (1).
- This function is often called the **regression line** or **the regression function**.

The disturbance in RM 2 I

We can now *define* the disturbance e_i in *RM2*

$$e_i \equiv y_i - E\left(y_i \mid x_i\right). \quad (2)$$

i.e., as the stochastic variable y_i minus the conditional expectation of y_i. e_i has cond. expectation:

$$\begin{aligned} E\left(e_i \mid x_i\right) &= E\left(y_i \mid x_i\right) - E\left[E\left(y_i \mid x_i\right)\right] \\ &= E\left(y_i \mid x_i\right) - E\left(y_i \mid x_i\right) = 0 \end{aligned} \quad (3)$$

We have **only** used

1. The understanding that $\mid x_i$ means conditioning on a fixed x_i so that x_i is a parameter

The disturbance in RM 2 II

2. The basic property of the expectations operator that the expectation of a stochastic variable minus a constant parameter is the expectation of the variable minus the constant

Implications for OLS estimator properties in RM2 I

The other assumptions of the model specification

c. $var\left(e_i \mid x_h\right) = \sigma^2$, $\forall$ i and h

d. $cov\left(e_i, e_j \mid x_h\right) = 0$, $\forall$ $i \neq j$, and for all h

are also assumptions about the conditional distribution of e_i given x_i.

- When we compare the specification of *RM1* and *RM2*, the difference is that assumptions **c.–d**. are for the **conditional distribution** of e_i given x_i in *RM2* and for the **marginal (or unconditional) distribution** in *RM1*.
- Since there are no principal differences between conditional distributions and other probability distributions, we can regard *RM2* as a special case of *RM1*.

Implications for OLS estimator properties in RM2 II

Therefore, **all the statistical properties** that we have shown for the estimators of *RM1*, will also hold for the conditional model *RM2*.

Specifically, if we take on board the assumption about conditional normality

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

$\hat{\beta}_2$ has a normal distribution, conditional on x_h:

$$\hat{\beta}_2 \sim N\left(\beta_2, \frac{\sigma^2}{n\hat{\sigma}_x^2} \mid x_h\right) \ \forall \ h$$

$$\hat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Implications for OLS estimator properties in RM2 III

This means in particular that we can show unbiasedness, e.g., for $\hat{\beta}_2$.

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})e_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The conditional expectation of $\hat{\beta}_2$:

$$\begin{aligned} E(\hat{\beta}_2 \mid x_h) &= E(\beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})e_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \mid x_h) \\ &= \beta_2 + E(\frac{\sum_{i=1}^{n}(x_i - \bar{x})e_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \mid x_h) \\ &= \beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})E(e_i \mid x_h)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}) = \beta_2 \end{aligned}$$

Unconditional properties I

- That all properties hold also for $RM2$, only conditional on x_h is important
- But still somewhat unsatisfactory since we want our conclusions to have generality (hold "in the population").
- Hence the conclusion should not depend in any important sense on the values that we happened to obtain for the n stochastic x variables.
- Intuitively, this cannot be a very large problem, since the t-distribution that we use for hypothesis testing does not depend on x.

Unconditional properties II

In particular, the distribution of the test statistic

$$\frac{\hat{\beta}_2 - \beta_2^0}{\widehat{se}\left(\hat{\beta}_2\right)} \sim t\left(n-2\right)$$

depends on the number of degrees of freedom, not on x_h.
Furthermore, we can show that the unconditional (marginal) expectation of $\hat{\beta}_2$is

$$E\left(\hat{\beta}_2\right) = \beta_2 \tag{4}$$

We then need to use the **law of iterated expectations**, which we will pay attention to below, since it plays an important in econometric theory.

Unconditional properties III

Note that the Monte Carlo simulations for RM2 showed (or suggests depending on your attitude) that the unconditional result (4) holds!

Summing up so far I

- In all cases where the parameters of interest are in the linear conditional expectation function

$$E(y_i \mid x_i) = \beta_1 + \beta_2 x_i,$$

  it plays no role for the properties of the OLS estimators whether the x_i $(i = 1, 2, ..., n)$ are deterministic or stochastic variables.
- The same conclusion is true for the t-statistics
- Extremely good news, also because in practice when we estimate multivariate model, we will typically have deterministic and stochastic x's in one and the same model

Summing up so far II

- What **is** important for the properties of the estimators is that the assumptions about
  - homoskedasticity
  - cross-section independence (no autocorrelation), and
  - exogeneity

  are valid for the model that we estimate.
- In Lecture 7, we will have time to talk briefly about exogeneity, and later lectures will mention the consequences "other departures from the model assumptions".
- But this type of model specification is a large field, that will re-appear in later courses in econometrics.

An example of a simultaneous distribution I

Consider the simultaneous discrete distribution function:

		x			
		-8	0	8	$f_y(y_i)$
	-2	0.1	0.5	0.1	0.7
y	6	0	0.2	0.1	0.3
	$f_x(x_i)$	0.1	0.7	0.2	

If you solve the Exercises to Seminar 4 you will show that the conditional distribution for y given x is

An example of a conditional distribution I

		x		
		-8	0	8
y	-2	$\frac{0.1}{0.1}$	$\frac{0.5}{0.7}$	$\frac{0.1}{0.2}$
	6	$\frac{0}{0.1}$	$\frac{0.2}{0.7}$	$\frac{0.1}{0.2}$

- For $x = -8$ the probabilities are $f(y = -2 \mid x = -8) = 1$ and $f(y = 6 \mid x = -8) = 0$
- For $x = 0$: $f(y = -2 \mid x = 0) = 0.5/0.7$ and $f(y = 6 \mid x = 0) = 0.2/0.7$

Conditional expectation function—example I

In the same seminar exercise you will show that

$$E(y \mid x = -8) = (-2) * 1 + 6 * 0 = -2$$
$$E(y \mid x = 0) = \frac{2}{7}$$
$$E(y \mid x = 8) = 2$$

which shows that the conditional expectations function for y is a deterministic function.

Conditional expectation function—definition I

Let y be a discrete variable with conditional distribution function $f_{y|x}(y_i \mid x_l)$. The conditional mean of y is then:

$$E(y \mid x_h) = \sum_{i=1}^{n} y_i f_{y|x}(y_i \mid x_h)$$

If y is a continuous variable, the conditional mean is

$$E(y \mid x) = \int_{-\infty}^{\infty} y f_{y|x}(y \mid x)\, dy$$

where $f_{y|x}(y \mid x)$ is the relevant probability density function (pdf)

- For a given value of x, $E(y \mid x)$ is a (deterministic) number.

Conditional expectation function—definition II

- When we calculate $E(y \mid x)$ for different given values of x, we get the deterministic function exemplified above
- Generally we can let $E(y \mid x)$ represent both the expectation of y for **any given** x, and as a stochastic variable that is a function of the stochastic variable x.
- We use the term **conditional expectations function** for $E(y \mid x)$ in both meanings.
  - When we looked at the confidence interval for μ_i in RM1, lecture 5, $\mu_i \equiv E(y_i \mid x_i) = \beta_1 + \beta_2 x_i$ was deterministic.

Conditional expectation function—definition III

- In the interpretation where $E(y_i \mid x_i)$ is a function of the stochastic variable x, and **the function is linear** we get the regression function in (1) in $RM2$:

$$E(y_i \mid x_i) = \beta_1 + \beta_2 x_i, \qquad (5)$$

- If the regression function is non-linear we need another estimation principle than OLS
  - Non Linear OLS for example, but that's for another course
  - May also need to make other changes, because the non-linearity will also possibly affect the disturbances

Conditional expectation function—definition IV

- But already this is enough to start understanding that **the choice of functional form** of the conditional expectations function is one of the most important decisions that we make in econometric modelling.
- The focus on variable transformations, prior to OLS estimation, in our course should be understood in this perspective as well.

The law of iterated expectations I

Let z_1 and z_2 be two stochastic variables. The law of iterated expectations, also called the law of double expectations states that

$$E\left[E\left(z_1 \mid z_2\right)\right] = E\left(z_1\right).$$

The law of iterated expectations says that if we take the expectation with respect to all the values of the variables that we have the regression function is conditioned on, we obtain the unconditional expectation.

The intuition is that if consider the probabilities for all values of z_2, then those values no longer condition the mean of z_1.

If you solve the exercises to seminar 4 you will prove this important theorem

Independence when the conditional expectation is zero I

An important consequence of the above, and the linearity property of conditional expectations, is that if

$$E\left(z_1 \mid z_2\right) = c$$

where c is a constant, then the two stochastic variables z_1 and z_2 are uncorrelated:

$$cov(z_1, z_2) = 0$$

We can write

$$y_i = \beta_1 + \beta_2 x_i + e_i$$

in terms of the conditional expectation and the disturbance e_i

$$y_i = E(y_i \mid x_i) + e_i \tag{6}$$

where

$$E(y_i \mid x_i) = \beta_1 + \beta_2 x_i \tag{7}$$

Consider the distrubance:

$$e_i = y_i - \beta_1 + \beta_2 x_i$$

We have shown, in (3), that

$$E(e_i \mid x_i) = 0$$

Since 0 is a constant, it is then true that in $RM2$

$$cov(e_i, x_i) = 0 \tag{8}$$

The disturbance is uncorrelated with the stochastic variable x_i. Moreover, since by definition

$$cov(e_i, x_i) = E(e_i x_i) - E(e_i)E(x_i)$$

and by the use of the law of iterated expectations

$$E\left[E(e_i \mid x_i)\right] = E(e_i) = 0$$

then we also have

$$E(e_i x_i) = 0 \tag{9}$$

in RM2.

From before we have conditional unbiasedness:

$$E(\hat{\beta}_2 \mid x_i) = \beta_2$$

and with reference to Monte Carlo simulation we have claimed that $\hat{\beta}_2$ that is unconditionally unbiased, see (4).
Now we can prove this claim.
Start with writing the OLS estimator in RM2 in terms of β_2 and $\frac{\sum_{i=1}^{n}(x_i - \bar{x})e_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$:

$$\hat{\beta}_2 = \beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})e_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The apply the rule of double expectations.

$$E(\hat{\beta}_2 - \beta_2) = E\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})e_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})E(e_i \mid x_i)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = 0$$