Unconditional and conditional expectations

- Definition
  - Discrete random variables: $\mathrm{E}\left(y\right)=\sum_{i=1}^{n}p_{i}y_{i}$
  - Continous random variables: $\mathrm{E}\left(y\right)=\int_{y\in Y}yf\left(y\right)dy$, where Y is the set possible values of y

- Definition
  - Discrete random variables: $\mathrm{E}\left(\left.y\right|x\right)=\sum_{i=1}^{n}p_{i}\left(\left.y\right|x\right)y_{i}$
  - Continous random variables: $\mathrm{E}\left(\left.y\right|x\right)=\int_{x\in X}xf\left(\left.y\right|x\right)dx$

Properties of conditional expectations

- Law of iterated expectation. If $\boldsymbol{x} = f(\boldsymbol{w})$

$$\mathrm{E}(y|\boldsymbol{x}) = \mathrm{E}(\mathrm{E}(y|\boldsymbol{w})|\boldsymbol{x})$$
$$\mathrm{E}(y|\boldsymbol{x}) = \mathrm{E}(\mathrm{E}(y|\boldsymbol{x})|\boldsymbol{w})$$

- The smaller information set always dominates!

- The special case
$$\mathrm{E}(y|\boldsymbol{x}) = \mathrm{E}(\mathrm{E}(y|\boldsymbol{x},\boldsymbol{z})|\boldsymbol{x})$$

- Conditional expectations are linear

- $y, \boldsymbol{x}$ random

- $\boldsymbol{a}(x)$, $\boldsymbol{B}(x)$ functions of $\boldsymbol{x}$

- Function $\boldsymbol{a} + \boldsymbol{B}y$

$$\mathrm{E}\left[\boldsymbol{a}(\boldsymbol{x}) + \boldsymbol{B}(\boldsymbol{x})\, y \,|\boldsymbol{x}\right] = \boldsymbol{a}(\boldsymbol{x}) + \boldsymbol{B}(\boldsymbol{x})\,\mathrm{E}(y|\,\boldsymbol{x})$$

- If $\mathrm{E}\left(y^2\right) < \infty$ and $\boldsymbol{\mu}(\boldsymbol{x}) = \mathrm{E}(y\,|\boldsymbol{x})$, than $\boldsymbol{\mu}(\boldsymbol{x})$ is the solution of

$$\min_{m \in M} \mathrm{E}\left[y - m(\boldsymbol{x})\right]^2$$

- Conditional expectation is the **minimum mean square error predictor** of y based on information contained in $\boldsymbol{x}$

Partial Effects

Example 1. *GUS Data - December labour survey* 2002

- – Question: what is the difference between the wage of women an men.
  – The simplest way of analysis: compare the means

category	mean gross wage
men	2487 zł
women	2042 zł

  Result: difference 445 zł
  – Problem: women work experience is shorter than men's - it is possible that the difference in wages can be explained by the difference in

experience

category	work experience
men	18.49 years
women	18.07 years

– If it is the discrimination on the labour market which is of interest that the proper statement of the research question should be the following:
  * what is the difference between the wages of men and women holding all other factors fixed (**ceteris paribus**)
  * such an effect can be defined as effect of a variable w on $\mathrm{E}\left(y\left|w, \boldsymbol{c}\right.\right)$ holding the **control variables** $\boldsymbol{c}$ constant.
  * For discrete variable w partial effect is defined as

$$\mathrm{E}\left(y\left|w_i, \boldsymbol{c}\right.\right) - \mathrm{E}\left(y\left|w_j, \boldsymbol{c}\right.\right)$$

    where w_i and w_j are possible levels of w
– Usually we infer the causal relationship from the theory rather that from data - the correlation between variables can only prove such

relationships in controlled experiments.

Example 2. *cont.*

- The partial effect of the sex on wage:

$$\mathrm{E}\left(y\,|\textit{man,experience}\right)-\mathrm{E}\left(y\,|\textit{woman,experience}\right)$$

  – Result of estimation (OLS)

partial effect	estimate
sex	-435.4

  – But working women are batter educated than man.
  – Partial effect:

$$\mathrm{E}\left(y\,|\textit{man,experience,education}\right)-\mathrm{E}\left(y\,|\textit{woman,experience,education}\right)$$

– Estimate (OLS):

partial effect	estimate
sex	-747.9

Assumptions

- Sample
  - We will adopt random sampling assumption:
    - population model (e.g. for people *in* the workforce)
    - independent identically distributed (iid) draws from the sample
  - for most of the lecture we will analyse the survey data
  - generalizations will be required of iid asummption are necessary to cover cluster sampling, stratified sampling or panels (these samples are not iid)

- Regressors
  - in Classical Regression Model we have

$$y_i = \boldsymbol{\beta}_0 + \boldsymbol{x}_i \boldsymbol{\beta} + u_i$$

- we assume that regressors $\boldsymbol{x}_i$ are fixed (nonrandom)
- for u_i we assume that u_i are iid with $\mathrm{E}\left(u_i\right) = 0$ and $\mathrm{Var}\left(u_i\right) = \boldsymbol{\sigma}^2$
- if sample is random u_i are iid by definition!
- what is interesting in nonexperimental framework is the relation between $\boldsymbol{x}_i$ and u_i
- if $\boldsymbol{x}_i$ is nonrandom than $\boldsymbol{x}_i$ and u_i are independent by definition!
- does it make sense to assume that regressors are nonrandom if we do not control them?
- for nonexperimental data fixed regressors assumption is nonrealistic
- **Conclusion**: regressors should be treated as random variables
- **Consequence:** most of the results derived will only be asymptotic

Conditional Expectations

- Most applied econometrics deals with estimation of an effect on
  - *expectation* of explained variable (dependent variable, regressand, response variable)
  - *conditional* on a set of explanatory variables (independent, regressors, control variables, covariates)

- If $\mathrm{E}\left(|y|\right) < \infty$ than we can define

$$\mathrm{E}\left(y\,|x_1,\ldots,x_K\right) = \boldsymbol{\mu}\left(x_1,\ldots,x_K\right)$$

- Conditional expectation is a function determining what will be the expected (average) value of y for given $x_1,\ldots,x_K$

- If $x_1, \ldots, x_K$ are random than conditional expectation $\mathrm{E}\left(y\,|x_1, \ldots, x_K\right)$ is also random

- Simple properties of conditional expectations

  – if we know x, than the expected value of x is equal to x

$$\mathrm{E}\left(x|\,x\right) = x$$

  – conditional expected value (average) of weighted sum of variables is equal to weighted sum of conditional expected values (averages) of the variables

$$\mathrm{E}\left(a_1 x_1 + \ldots + a_k x_k|\,\boldsymbol{z}\right) = a_1\,\mathrm{E}\left(x_1|\,\boldsymbol{z}\right) + \ldots + a_K\,\mathrm{E}\left(x_K|\,\boldsymbol{z}\right)$$

- Model in error form

$$y = \mathrm{E}\left(y\,|\boldsymbol{x}\right) + u$$

$$\mathrm{E}(u|\boldsymbol{x}) = 0$$

Example 3. *Typical models*

$$\mathrm{E}(y|x_1, x_2) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2 \tag{1}$$

$$\mathrm{E}(y|x_1, x_2) = \exp(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \log(x_1) + \boldsymbol{\beta}_2 x_2) \tag{2}$$

- Define $u = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2$, then $\mathrm{E}(u|\, x_1, x_2) = 0$ and model (1) can be reformulated as follows

$$y = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2 + u$$

as

$$
\begin{aligned}
\mathrm{E}\left(y\left|x_1, x_2\right.\right) &= \mathrm{E}\left(\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2 + u\left|x_1, x_2\right.\right) \\
&= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2 + \mathrm{E}\left(u\right| x_1, x_2) = 0 \\
&= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2
\end{aligned}
$$

Partial effects

- continuous variable

$$\Delta \, \mathrm{E}\left(y \,|\boldsymbol{x}\right) \approx \frac{\partial \boldsymbol{\mu}\left(\boldsymbol{x}\right)}{\partial x_j} \Delta x_j \text{ for fixed } x_1, \ldots, x_{j-1}, x_{j+1}, x_K$$

For (1)

$$\frac{\partial \, \mathrm{E}\left(y \,|\boldsymbol{x}\right)}{\partial x_1} = \boldsymbol{\beta}_1, \frac{\partial \, \mathrm{E}\left(y \,|\boldsymbol{x}\right)}{\partial x_2} = \boldsymbol{\beta}_2$$

For (2)

$$\frac{\partial \, \mathrm{E}\left(y \,|\boldsymbol{x}\right)}{\partial x_1} = \exp\left(\cdot\right) \frac{\boldsymbol{\beta}_1}{x_1}, \frac{\partial \, \mathrm{E}\left(y \,|\boldsymbol{x}\right)}{\partial x_2} = \exp\left(\cdot\right) \boldsymbol{\beta}_2$$

Notice that for model (2) partial effects are not constant across values of x_1, x_2

- Partial elasticity

$$\frac{\partial\,\mathrm{E}\left(y\,|\boldsymbol{x}\right)}{\partial x_j}\frac{x_j}{\mathrm{E}\left(y\,|\boldsymbol{x}\right)}=\frac{\partial\boldsymbol{\mu}\left(\boldsymbol{x}\right)}{\partial x_j}\frac{x_j}{\boldsymbol{\mu}\left(\boldsymbol{x}\right)}$$

- For $\mathrm{E}\left(y\,|\boldsymbol{x}\right)>0$ and $x_j>0$ it can also be defined as

$$\frac{\partial\log\mathrm{E}\left(y\,|\boldsymbol{x}\right)}{\partial\log x_j} \tag{3}$$

Example 4. *Partial elasticity with respect to x_1 in model (2) is equal to:*

$$\frac{\partial\log\mathrm{E}\left(y\,|x_1,x_2\right)}{\partial\log x_1}=\frac{\partial\left(\boldsymbol{\beta}_0+\boldsymbol{\beta}_1\log\left(x_1\right)+\boldsymbol{\beta}_2x_2\right)}{\partial\log\left(x_1\right)}=\boldsymbol{\beta}_1$$

$\boldsymbol{\beta}_1$

- Sometimes we also use the definition

$$\frac{\partial \operatorname{E}(\log y \,|\boldsymbol{x})}{\partial \log x_1} \tag{4}$$

Example 5. *We often define similar (but not identical) model to model (2)*

$$\operatorname{E}(\log y \,|x_1, x_2) = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \log(x_1) + \boldsymbol{\beta}_2 x_2$$

in this case it is not possible to use directly the definition (3) as $\log \operatorname{E}(y \,|\boldsymbol{x}) \neq \operatorname{E}(\log y \,|\boldsymbol{x})$ *but it easy to use definition (4)*

$$\frac{\partial \operatorname{E}(\log y \,|x_1, x_2)}{\partial \log x_1} = \frac{\partial (\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \log(x_1) + \boldsymbol{\beta}_2 x_2)}{\partial \log x_1} = \boldsymbol{\beta}_1$$

- Definitions (3) and (4) equivalent only if $\boldsymbol{x}$ independent of u.

Example 6. *Partial elasticity y in model (2) with respect to x_1 is equal to $\boldsymbol{\beta}_1$*

- Partial semielasticity

$$100 \times \frac{\partial \, \mathrm{E}\left(y \,|\boldsymbol{x}\right)}{\partial x_1} \frac{1}{\mathrm{E}\left(y \,|\boldsymbol{x}\right)}$$

- Interpretation of partial elasticity: expected percentage change of y if x change by 1 percent

- Interpretation of partial semielasticity: expected percentage change of y if x change by 1

- Sometimes y is influenced by some unobserved variable q (unobserved heterogeneity)

- The partial effect calculated from the model not including q is equal to the average over q of partial offect: average partial effect (APE)

$$\boldsymbol{\delta}_j\left(\boldsymbol{x}^0\right) = \mathrm{E}_q\left[\mathrm{E}\left(y\,|\boldsymbol{x},q\right)\right]$$