

LECTURE 1 INTRODUCTION

1.1 What is econometrics?

It's a bridge between economic theory and the real world. Use theory from economics, tools from statistics, and data. To test economic hypotheses, to forecast, to answer "how much" question

1.2 Overview of the Econometric Model

- Economic theory – the *average* or *systematic* behaviour, identifies relationship b/w economic variables, make predictions about the direction of outcomes. e.g., $q_d = f(p, p_s, p_c, y)$
- Econometrics – actual behaviours depend upon the sum of a systematic component (economic theory) and a *random* or *unpredictable* component ε . e.g., $q_d = f(p, p_s, p_c, y) + \varepsilon$
 - **Random error ε :**
 - reflects the intrinsic uncertainty in eco activity (unpredictable random behaviour)
 - accounts for factors omitted from the model (any factors other than x that affect y)
 - represents approximation error arising from the assumed linear functional form
 - a systematic component of y is '*explained*' by x ;
 - a random component of y is *not explained* by x , it's called random error ε
 - **Functional form** (only consider linear equation in BE):
 $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon$
 - get a sample of data on $\{y_i, x_i\}$ to learn unknown parameters $\{\beta_0, \beta_1, \dots, \beta_k\}$

1.3 Types of Data

- **time series:** follow a country, region, firm or individual over time
- **cross-sectional:** collects information on several countries, regions, firms or individuals at a single point in time
- **panel:** follows several cross-sectional units over time

LECTURE 2 BASIC LINEAR MODEL: ASSUMPTIONS

2.1 The Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

- **the intercept β_0** represents the average value of y when all the x 's are zero (meaningless if there is no x be zero from sample data)
- **the slope parameter β_j** represents the expected change in y associated with a unit change in x_j , all else constant

$$\beta_j = \left. \frac{E[y|X]}{\Delta X_j} \right|_{\text{all other } X\text{'s constant}} \quad \text{for } j = 1, 2, \dots, K$$

Interpretation: holding all else constant, y changes by ___ unit on average when change x_1 by 1 unit

2.2 Assumptions about the Linear Regression Model

MR1: The correct model is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

- Omitted relevant variables: biased OLS estimators if omitted Z is correlated with original X_1
- Inclusion irrelevant variables: unbiased OLS estimators but larger variance of estimators (not BLUE)

MR2: $E[\varepsilon_i|x_i] = 0 \Rightarrow$ the only assumption to ensure unbiasedness

- the error term has an expected value of 0, given any value of the x 's
- the x 's does not change the expected value of the random error ε
- but doesn't mean zero sample average of the error $E[\varepsilon] \neq 0$

MR3: $Var[\varepsilon_i|x_i] = \sigma^2$ Homoskedasticity \Rightarrow ensure consistency

- The variance of the random errors is constant and independent of the x 's.
- Heteroskedasticity: unbiased, but no longer BLUE (wrong $se(b_j)$ and wrong t-test)

MR4: any pair of random errors are uncorrelated

$$Cov(\varepsilon_i, \varepsilon_j | x_i, x_j) = 0 \quad \text{for all } i, j = 1, 2, \dots, N, i \neq j$$

- Autocorrelation: unbiased, but no longer BLUE

MR5a: x are not-random, x values are "fixed in repeated sampling"

- the values of all x 's are known prior to observing the (realized) values of dependent variable.

MR5b: Non-collinearity

- any one of the x 's is *not an exact linear* function of any of the other x 's.
- none of the x is redundant
- Exact Collinearity: unbiased, but no longer BLUE, the least squares procedure will fail

2.3 The Least Squares Principle

- the Least Squares Principle estimates $\{\beta_0, \beta_1, \dots, \beta_k\}$ such that the squared difference between the fitted line and the observed value of y is minimized.
 - Estimators $\{b_0, b_1, \dots, b_k\}$
 - Fitter line: $\hat{y}_i = b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki}$
 - Least squares residuals: $\hat{e}_i = (y_i - \hat{y}_i) = y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki})$
 - Why min the squared diff $\sum \hat{e}_i^2$ not $\sum \hat{e}_i$ – the minimum value would be at $-\infty$ so the fitted line could be set arbitrarily “high”
 - Why not set the fitted line $\sum \hat{e}_i = 0$ – a large positive value of $\hat{e}_i > 0$ would cancel out a large negative value $\hat{e}_i < 0$. In addition any fitted line passing through the (sample) means of y and x would satisfy this criteria (infinite number of potential fitted line)
- Minimise the sum of squared errors: $S = \sum \hat{e}_i^2$

$$\min_{\{\beta_0, \beta_1, \dots, \beta_k\}} S = \sum_{i=1}^N [y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki})]^2$$

first order conditions, for β_0 :

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^N [y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki})] = 0$$

for $\beta_j, j = 1, 2, \dots, k$

$$\frac{\partial S}{\partial \beta_j} = -2 \sum_{i=1}^N [y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki})] x_{ji} = 0$$

2.4 Properties of the OLS Residuals

OLS (Ordinary Least Squares)

- **Estimators** $\{b_0, b_1, \dots, b_k\}$
- **Random variables** $\{b_0, b_1, \dots, b_k\}$: values depend on the sample data y and x
- **Least squares estimates**: when the sample data are substituted into the formulas we obtain numbers that are the observed values of random variables

P1: when there is an intercept term β_0 , $\sum \hat{e}_i = 0$

P2: for each $\beta_j, j = 1, 2, \dots, k$

$$\sum \hat{e}_i x_{1i} = 0 \text{ and } \sum \hat{e}_i x_{2i} = 0 \text{ and } \dots \sum \hat{e}_i x_{ki} = 0$$

P3: these two properties imply:

$$\begin{aligned} \sum \hat{e}_i \hat{y}_i &= \sum \hat{e}_i [b_0 + b_1x_{1i} + b_2x_{2i} + \dots + b_kx_{ki}] \\ &= b_0 \sum \hat{e}_i + b_1 \sum \hat{e}_i x_{1i} + \dots + b_k \sum \hat{e}_i x_{ki} = 0 \end{aligned}$$

LECTURE 3 BASIC LINEAR MODEL: STATISTICAL PROPERTIES

3.1 The Sampling Properties of the OLS Estimators

- The means and variances of the estimators $\{b_0, b_1, \dots, b_k\}$ provide *location* and *dispersion* of the probability distribution of $\{b_0, b_1, \dots, b_k\}$.

- OLS Estimator is an Unbiased Estimator**

$$E(b_j) = \beta_j \text{ for } j=1, 2, \dots, k$$

$$E(b_0) = \beta_0$$

- Assumption $E(\varepsilon_i | x_i) = 0$ for all i (*conditional expectation*: the average of ε_i over all outcomes in x_i)
- When the *expected value* of any estimator of a population parameter is *equal to* the *true value* of that population parameter, the estimator is said to be *unbiased*
- Intuition**: if say 10,000 sample of size N were collected and b_j were calculated for each of these 10,000 samples, the average value of these estimates would be equal to β_j
- Unbiasedness does *not* mean a specific estimate of b_j is "close" to the true population parameter β_j , we never know how close they are, it depends on variance.
- We say an estimator is unbiased *not* an estimate

- Variability of the OLS estimators:**

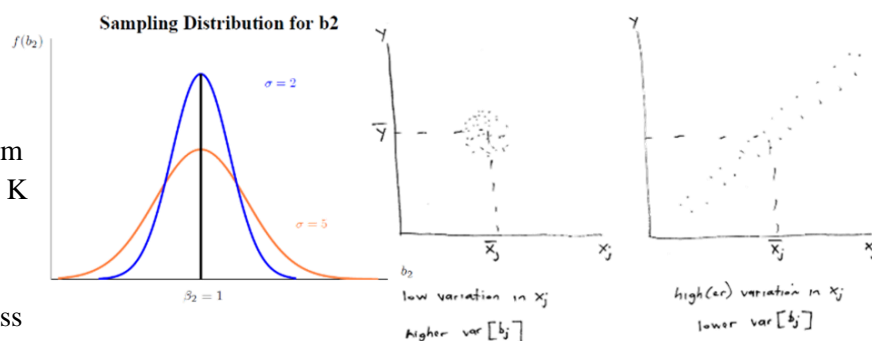
- the *lower* the variance of an estimator, the greater the sampling precision of the estimator
- if any of these OLS assumptions don't hold, the expressions for $Var(b_1), Var(b_2) \dots Var(b_k), Cov(b_j, b_k)$ will be *wrong*. Not the $Min Var(b_j)$

- MR2: $E(\varepsilon_i | x_i) = 0$
- MR3: $Var(\varepsilon_i | x_i) = \sigma^2$
- MR4: $Cov(\varepsilon_i, \varepsilon_j | x_i, x_j) = 0$
- MR5a: The X 's are not-random

- Factors affect $Var(b_j)$ for $j = 1, 2, \dots, K$

$$Var(b_j) = \frac{\sigma^2}{\sum (X_{ji} - \bar{X}_j)^2}$$

- $\uparrow \sigma^2 \Rightarrow \uparrow Var(b_j) \Rightarrow$ randomness
- $\uparrow \sum (X_{ji} - \bar{X}_j)^2 \Rightarrow \downarrow Var(b_j) \Rightarrow$ dispersion in the values of X
- \uparrow sample size $\Rightarrow \downarrow Var(b_j) \Rightarrow$ normality, greater job in closing to the sample mean
- $\uparrow corr(X_j, X_i) \Rightarrow \uparrow Var(b_j)$ for $i \neq j \Rightarrow$ collinearity, higher difficulty disentangling separate effects



- OLS estimator is a Linear Estimator** (weighted sum of y 's)

$$b_2 = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \frac{\sum_{i=1}^N y_i (x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} = \sum w_i y_i$$

$$\sum_{i=1}^N \bar{y} (x_i - \bar{x}) = \bar{y} \sum_{i=1}^N (x_i - \bar{x}) = \bar{y}(0) = 0$$

$$w_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

3.2 The Gauss-Markov Theorem

- Under the assumptions of the linear regression model (MR1–MR5), the OLS estimators $\{b_0, b_1, \dots, b_k\}$ have the of **Best Linear Unbiased Estimators (BLUE)** $\{\beta_0, \beta_1, \dots, \beta_k\}$
- smallest variance* of all linear and unbiased estimators
- We can't say it as the best estimator of all possible estimators, but in this course, we don't consider non-linear estimators, just remember these two factors
- The Gauss-Markov theorem *does not* depend on the assumption of *normality*
- The Gauss-Markov theorem applies to OLS estimators – *not* an OLS estimate from a single sample
- BLUE is not one of MLMR assumptions, Gauss-Markov sets the BLUE conclusion if all assumptions held.

3.3 Estimator of the Error Variance

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \{b_0 + b_1 x_{1i} + \dots + b_k x_{ki}\}$$

Use the unbiased estimator of the error variance:

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{(N - K - 1)} \text{ or } \frac{\sum \hat{\varepsilon}_i^2}{(N - K)}$$

Both are correct: sample size takes away the number of coefficients including intercepts.

where $k + 1$ is the number of parameters being estimated. The simple regression model $k + 1 = 2$

- **EViews get $\hat{\sigma}^2$:**

- S. E of regression = standard error of regression gives $\hat{\sigma}$
- Sum Squared resid = SSR sum of squared residue gives $\sum \hat{e}_i^2$

$$RSS = \sum (y_i - \hat{y})^2 = \sum \hat{e}_i^2 = \hat{\sigma}^2 * (N - K - 1)$$

- S. D dependent var = standard deviation of dependent variable s or $\hat{\sigma}_y$

$$TSS = \sum (y_i - \bar{y})^2 = s^2 * (N - 1)$$

- **EViews get covariance matrix:**

➤ After OLS estimation: view – covariance matrix

	c	cigs	income	
c	0.004298	-0.000121	-0.000100	$\begin{bmatrix} \text{VAR}[b_1] & \text{COV}[b_1, b_2] & \text{COV}[b_1, b_3] \\ \text{COV}[b_2, b_1] & \text{VAR}[b_2] & \text{COV}[b_2, b_3] \\ \text{COV}[b_3, b_1] & \text{COV}[b_3, b_2] & \text{VAR}[b_3] \end{bmatrix}$
cigs	-0.000121	3.28e-05	1.81e-06	
income	-0.000100	1.81e-06	3.33e-06	

- Use the entries on the **main diagonal** to get the estimated variances of the estimates
- Use the entries on the **off-diagonal** to get the estimated covariances of the estimates

3.4 Measuring Goodness of Fit in the MLRM (multiple linear regression model)

- **RSS:** Residual Sum of Squares (= SSR)

TSS: Total Sum of Squares

R^2 shows the variation in the dependent variable y about its mean that is **explained** by the expression model (i.e. explained by **all of** the explanatory variables)

$$R^2 = 1 - \frac{\sum \hat{e}_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS}$$

R^2 also measures the degree of linear association b/w the values of y_i and the fitted values \hat{y}_i

$$R^2 = [\widehat{\text{corr}}(y, \hat{y})]^2$$

- **Problem of R-square**

R^2 may too **larger** by including irrelevant x (in the extreme case, $R^2 = 1$ by including $(N - 1)$ X variables)

- For Unrestricted and Restricted Model, it must be, $RSS_R \geq RSS_{UR}$ or $RSS_{UR} \leq RSS_R$, so $R^2_{UR} \geq R^2_R$

- **Solution of R-square**

- Use the **adjusted R^2** symbolized as \bar{R}^2

$$\bar{R}^2 = 1 - \frac{\sum \hat{e}_i^2 / N - K - 1}{\sum (y_i - \bar{y})^2 / N - 1} = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_y^2}$$

- \bar{R}^2 does not always rise with additional X 's due to the 'degrees of freedom' correction $(N - K - 1)$

$$\bar{R}^2 = 1 - \{(1 - R^2) \frac{(N - 1)}{(N - K - 1)}\}$$

- \bar{R}^2 can be **negative**, if N is sufficiently small and K sufficiently large (R^2 cannot be negative)

$$\bar{R}^2 < R^2$$

- \bar{R}^2 no longer measures the percent of variation in the dependent variable explained by the model. How to interpretation?