Basic Econometric Revision

Topic 1 Basic Linear Model (Lecture 2-7)

1) Introduction

- Economic theory describes average behaviour of many individuals: identifies relationships between economic variables; make predictions about direction of outcomes when a variable is altered
- Dependent variable: y
- Explanatory variables: X=x1,x2,….,xk
- Unknown parameters: β_i
- Error term: ε_i (any factors other than X that affect y and are not included in the model: i.e. assumed linear function form; unpredictable random behaviour)
- Linear equation: $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$

2) Economic model

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

- Intercept: β_0 → average value of y when all the X's are zero
- Slope parameters: β_j → expected change in y associated with a unit change in X_j, all else constant
- **Assumptions:**
- (1) Correct model is $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$
- (2) $E[\varepsilon_i|x_i] = 0$: error term has an expected value of 0, given any value of X's
- → $E[y_i|x_i] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + 0$
- → $y_i = E[y_i|x_i] + \varepsilon_i$; systematic component of y "explained" by X; a random component of y "not explained" by X
- (3) VAR$[\varepsilon_i|x_i] = \sigma^2$: variance of random errors is constant and independent of the X's → **Homoskedasticity**
- (4) COV $[\varepsilon_i, \varepsilon_j|x_i, x_j] = 0$ for all I,j=1,2,…N, i$\neq j$: any pair of random errors are uncorrelated
- (5a) The explanatory variables are not-random (values of all X's are known prior to observe the values of the dependent variable)
- (5b) any one of the X's is not exact linear function of any other X's

3) Lease Squares Principle

- estimates $(\beta_{0,}\beta_1, \ldots, \beta_k)$ such that the **squared difference** btw fitted value and observed value of y is minimised → why "Squared" ? so positive diff won't cancel out negative diff.
- $(b_{0,}b_1, \ldots, b_k)$ are estimators, random variables; values for $(b_{0,}b_1, \ldots, b_k)$ are least squares estimates
- Fitted line: $\hat{y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki}$
- Least squares residuals: $\hat{e}_i = (y_i - \hat{y}_i) = y_i - (b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki})$
- Sum of Squared residual (RSS): $\sum_{i=1}^{N} \hat{e}_i^{\,2}$
-
- Note: since $\sum_{i=1}^{N} \hat{e}_i = 0$ & $\sum_{i=1}^{N} \hat{e}_i X_{1i} = 0$…..$\sum_{i=1}^{N} \hat{e}_i X_{Ki} = 0$
- Implies $\sum_{i=1}^{N} \hat{e}_i \hat{y}_i = \sum_{i=1}^{N} \hat{e}_i [b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki}] = 0$

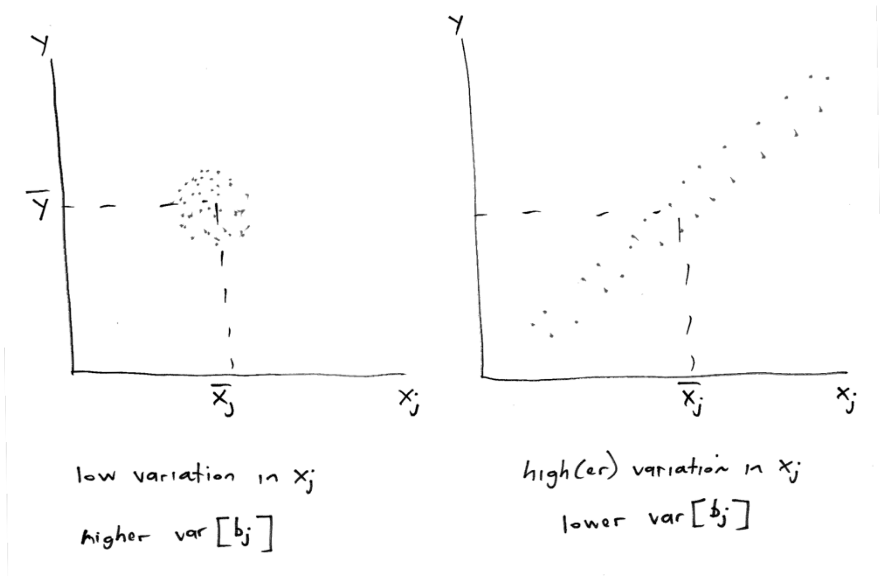
- → sum of “product btw fixed value and residue” is 0

4) Statistical properties

- Sampling distribution of the OLS estimators: Mean and variances of $(b_{0,}b_1, \ldots, b_k)$
- **Mean**; If:
- $E[b_j] = \beta_j$, for j=1,2,….,K with the assumption $E[\varepsilon_i|x_i] = 0$ for all i=1,2,….,N hold
- $E[b_0] = \beta_0$
- Then:
- the **estimator** is said to be **unbiased**

- **Variance**: $VAR[b_j]$ & $COV[b_j, b_k]$ with the following assumptions hold:
- $E[\varepsilon_i|x_i] = 0$ for all i=1,2,….,N
- $\text{VAR}[\varepsilon_i|x_i] = \sigma^2$
- $\text{COV}\left[\varepsilon_i, \varepsilon_j|x_i, x_j\right] = 0$
- X’s are not random

- unbiased estimator that has a higher prob. of getting an estimate “close” to β_j
- The lower the variance of an estimator, the greater the sampling precision of the estimator
- Factors affecting Variance of OLS estimators:
- (1) a larger σ^2 raises $\text{VAR}[b_j]$
- (2) greater dispersion in values of X measured by term $\sum(X_{ji} - \bar{X}_j)^2$ lower the variance of $\text{VAR}[b_j]$



- (3) Larger sample size lower VAR$[b_j]$
- (4) Larger correlation raises VAR$[b_j]$

5) Gauss-Markov Theorem

- Under the assumptions of the linear regression model (1--5b), the OLS estimators $(b_{0,}b_1, \ldots, b_k)$ have the smallest variance of all linear and unbiased estimators of $(\beta_{0,}\beta_1, \ldots, \beta_k)$ OR the Best linear unbiased estimators of $(\beta_{0,}\beta_1, \ldots, \beta_k)$
    - the assumptions must be true for Gauss-Markov holds

- Unbiased Estimator of the error variance: $\widehat{\sigma^2} = \frac{\sum_{i=1}^{N} \hat{e_i}^2}{(N-K-1)} = \frac{RSS}{(N-K-1)}$ where K+1 =no. of parameters being estimated

- Note: In Eviews: → S.E of regression = $\hat{\sigma}$ → $\widehat{\sigma^2} = \hat{\sigma}$^2
- → Sum Squared resid (RSS) = $\sum_{i=1}^{N} \hat{e_i}^2$ → $\widehat{\sigma^2} = \frac{RSS}{(N-K-1)}$

- RSS: residual sum of squares
- TSS: total sum of squares
- $R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS}$, the variation in the dependent variable y about its mean that is explained by the regression model (how well the model fits the data)
- R^2 also measures the degree of linear association btw the values of y_i and the fitted values $\hat{y}_i$ → $R^2 = [\widehat{CORR}(y, \hat{y})]$^2
- $0 \leq R^2 \leq 1,$
- *Interpretation*: e.g. 21% of the variation in y is explained by variation in X1 and X2.
- *where* $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$
- **problem**: R^2may be made bigger by including irrelevant X variables (no significantly related to y) ; Note: intuitively R^2 cannot decrease as RSS cannot increase by adding more X variables (As RSS will decrease by adding more Xs → so R^2 will increase.)
- Solution: measure the cost of imposing irrelevant explanatory variables

6) Unrestricted and restricted model

- Restricted model: restrict β_k=0 → one less X variable than the unrestricted model
- Minimisation problem: minimise the sum of squared errors (RSS is the minimised value of the objective function evaluated at the solution $b_{0,}b_1, \ldots, b_k$
- Thus, $RSS_R \geq RSS_{UR}$ must hold: an extra factor might explain the model better so error decreases.
- → adding one more regressor decreases RSS and thus increases R^2

$$RSS_R \geq RSS_{UR} \rightarrow R^2_{UR} \geq R^2_R$$

- Adjusted$\bar{R}^2$: this measure does not always rise with additional X's due to "degree of freedom" correction (N-K-1) → as more X's are added, $\sum \hat{e_i}^2$decreases, but (N-K-1) also decreases.

- $\bar{R}^2 = 1 - \frac{\frac{\sum \hat{e}_i^2}{N-K-1}}{\frac{\sum (y_i - \bar{y})^2}{N-1}} = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma_y}^2}$
- $= 1 - \left\{(1 - R^2)\frac{N-1}{N-K-1}\right\}$
- The effect on $\bar{R}^2$ depends on the reduction in $\sum \hat{e}_i^2$relative to (N-K-1).
- In terms of R^2: $\bar{R}^2 = 1 - \left\{(1 - R^2)\frac{N-1}{N-K-1}\right\}$
- When N is sufficiently small and K sufficiently large, the $\bar{R}^2$ might be actually negative → BUT! R^2 cannot be negative when intercept is included in the model

7) Hypothesis testing I

(a) Adding normality: assumption of normality makes statistical inference much easier

- assume $\varepsilon_i \sim N(0, \sigma^2)$ then: $y_i \sim N(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots \beta_K X_{Ki}, \sigma^2)$

- If **errors** are normally distributed → **y'**s also be normally distributed (y's contains weighted sum of OLS estimators)

- so OLS estimators are weighted sums of normal variables for j=1,2,.....K

$$b_0 \sim N(\beta_0, VAR[b_0]$$

$$b_j \sim N(\beta_j, VAR[b_j]$$

- OLS estimators will have normal distribution if N is sufficiently large

(b)Steps for Hypothesis testing

- formulate H_0 and H_A specify a test

(null hypothesis is usually stated in terms of the magnitude or sign of β_j that we **do not expect**

(based on economic theory)

$$\boldsymbol{H_0: \beta_j = c, \quad H_A: \beta_j \neq c}$$

- test statistic (a r.v.)and its distribution when H_0 is true

$$t = \frac{b_j - \beta_j}{se(b_j)} \sim t(N - K - 1)$$

- choose a level of significance and determine the rejection region

rejection region for 2 sided test: $t > t_c or t < -t_c \ where \ P[t \geq t_c] = P[t \leq -t_c] = \frac{\alpha}{2}$

$[H_A: \beta_j \neq c]$

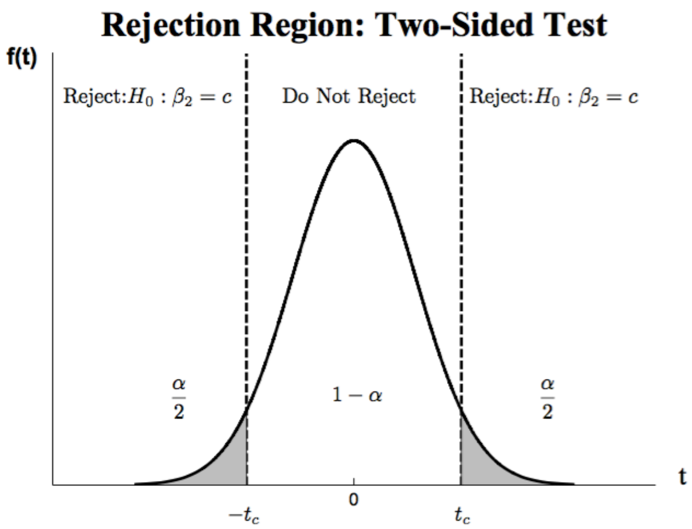
rejection region for 1 sided test: $t > t_c \ P[t \geq t_c] = \alpha$ $\quad [\ H_A: \beta_j > c\ ]$

rejection region for 1 sided test: $t < -t_c \ P[t \leq -t_c] = \alpha$ $\quad [\ H_A: \beta_j < c\ ]$

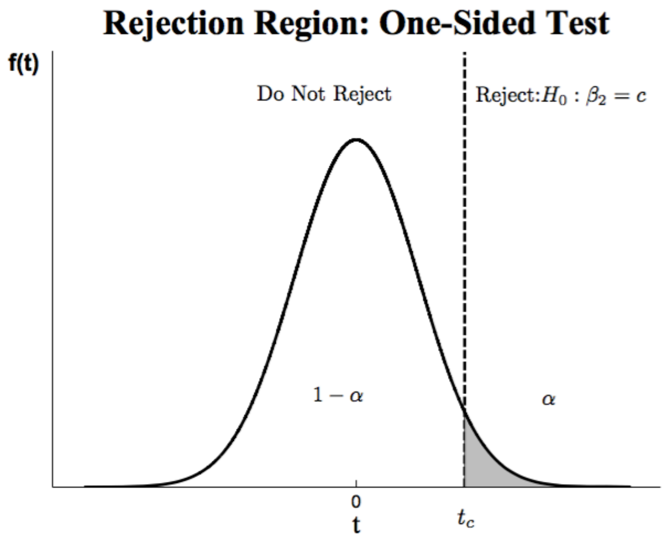
- obtain the sample estimates for b_j and se(b_j) apply the decision rule

$$t = \frac{b_j - c}{se(b_j)} \sim t(N - K - 1)$$

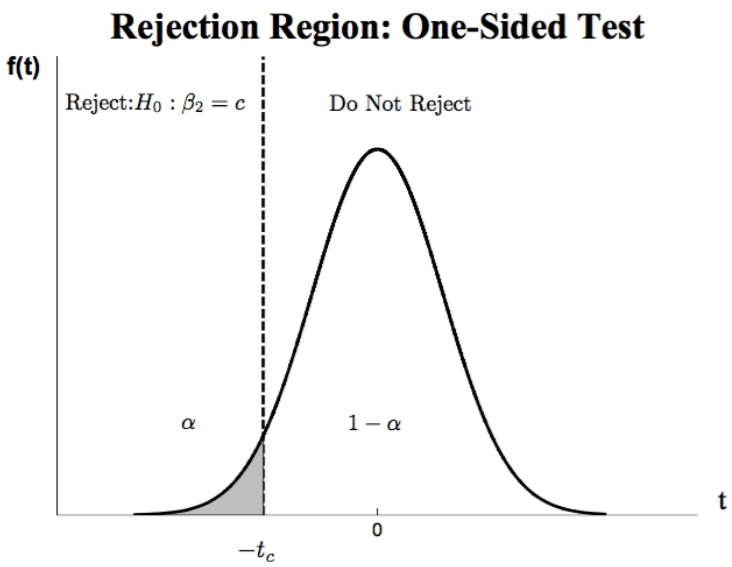
- state your conclusion



$H_0 : \beta_2 = c$ and $H_A : \beta_2 > c$



$H_0 : \beta_2 = c$ and $H_A : \beta_2 < c$



- rejection of null $H_0 : \beta_j = 0$ implies there is a statistically significant relationship between X_j and y.

Summary (example from A1)

Step 1: identify the null hypothesis and alternative hypothesis

$$\boldsymbol{H_0: \beta_1 = 1}$$

$$\boldsymbol{H_A: \beta_1 \neq 1}$$

Step 2: specify a test statistic and its distribution when H_0 is true

- If H_0 is true, the probability distribution of the test statistic is t-distribution.

$$t = \frac{b_1 - \beta_1}{se(b_1)} \sim t(N - K - 1) \; j = 0,1,2 \ldots k$$

- where the number of parameters estimated $K + 1 = 5$, the sample size $N = 79$, the degree of freedom $d.f. = N - K - 1 = 79 - 5 = 74$

Step 3: choose a level of significance α and determine the rejection region

- The assumed level of significance $\alpha = 0.05$ (two-tails test)
- The critical value $t_{0.025,74} = \text{approx.}\, t_{0.025,70} = 1.9944$
- So reject H_0 if $t \geq 1.9944$ or if $t \leq -1.9944$

Step 4: obtain the sample estimates for b_j and $se(b_j)$

$$t = \frac{b_1 - \beta_1}{se(b_1)} = \frac{0.813821 - 1}{0.013969} = -13.328$$

Step 5: apply the decision rule

$$-13.328 \leq -1.9944$$

Step 6: State the conclusion

- There is sufficient evidence to reject the null hypothesis. We have 95% confidence to conclude that the production technology will not exhibit

(c) Type I and Type II errors

	H_0 is true	H_0 is false
reject H_0	**Type I Error**	**Correct Decision**
not reject H_0	**Correct Decision**	**Type II Error**

Type I errors

- P[reject H_0 | H_0 true]= α
- P[not reject H_0 | H_0 true]=$1-\alpha$
    - ✓ we can control the prob. of Type I error since we control α (if rejecting a true H_0 is "costly", we should set α to be small)

Type II errors

- probability of a Type II error is not under our control and we cannot determine this probability without knowing the true value of the unknown population parameter
- the probability of a Type I error and the probability of a Type II error are **inversely related**—so if we make α smaller, the probability of a Type II error will increase

  NOTES: both the probability of a Type I and II error will be lower for a larger sample size (imagine a bigger pie)

(d) P-value

- p value of a hypothesis test is the probability that the t-distribution takes on a value at least as large (in absolute value) as the sample value of the t-statistic
- $0 \leq p \leq 1$
- p-value < α → reject null