LECTURE #5

Econometrics I

MULTIPLE REGRESSION ANALYSIS INFERENCE

Jiri Kukacka, Ph.D.
Institute of Economic Studies, Faculty of Social Sciences, Charles University

Summer semester 2024, March 19

In the previous lecture #4

- We derived the OLS estimator for multiple regression models:

$$\hat{\beta}^{OLS} = (X^T X)^{-1} X^T Y.$$

- 'Partial' interpretation: $\hat{\beta}_j = \frac{\Delta \hat{y}}{\Delta x_j}$, holding all other $x_{\neq j}$ fixed.
- We listed **four MLR assumptions** $\Rightarrow \mathbb{E}(\hat{\beta}) = \beta$.
- We discussed model overspecification (**irrelevant** variables) and underspecification (**omitted** variables).
- **MLR.5 Homoskedasticity:** $\text{Var}(u|x_1, \ldots, x_k) = \sigma^2 \mathbb{I} \quad \Rightarrow$

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)} \quad \text{or} \quad \text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}.$$

- We finally estimated $\quad \hat{\sigma}^2 = \frac{1}{n-k-1} \sum_{i=1}^n \hat{u}_i^2 \Rightarrow se(\hat{\beta}_j)$.
- Readings for lecture #5:
  - Chapter 3: 3.5, Chapter 4: 4.1–4.4

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter

- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter

- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

Gauss-Markov theorem

Under MLR.1 through MLR.5, the OLS estimator is the **best linear unbiased estimator (BLUE)**.

- We know OLS is an 'E'stimator.
- We know it is 'U'nbiased.
- 'Linear' estimator means that it can be expressed as a linear function of the data on the dependent variable:

$$\hat{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i,$$

  where w_{ij} can be a function of the sample values of all independent variables.
- 'Best' here refers to the one with the smallest variance (within this specific family of estimators), i.e., the efficient one.

Linear estimator

- We show that OLS is a linear estimator for the simple regression case.
- It means that the estimator can be written as $\hat{\beta}_j = \sum w_{ij} y_i$, i.e., it is a linear combination of y_i, $i = 1, \ldots, n$.
- To show this, we rewrite the OLS estimator as

$$\hat{\beta}_1 = \frac{\sum y_i(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = y_1 \frac{x_1 - \bar{x}}{\sum (x_i - \bar{x})^2} + \ldots + y_n \frac{x_n - \bar{x}}{\sum (x_i - \bar{x})^2}.$$

- We thus have weights $w_{11} = \frac{x_1 - \bar{x}}{\sum (x_i - \bar{x})^2}, \ldots, w_{n1} = \frac{x_n - \bar{x}}{\sum (x_i - \bar{x})^2}$ and OLS is a linear estimator.
- Keep in mind that, in general, a specific weight w_{ij} can be zero; it can also be a constant or the same value across more i as long as it does not contain the dependent variable.

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter
- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

Normality assumption

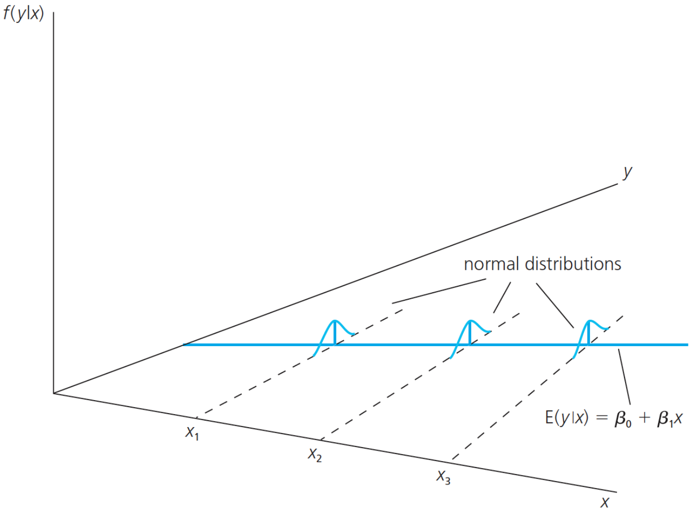
- We know the expected value and variance of the OLS estimator.
- However, the shape of the distribution has not been defined or inferred yet.
- As we condition on $x_1, \ldots, x_k$ (i.e., they are not treated as random variables), the distribution of u becomes crucial.
- **MLR.6 Normality:** The population error u is **independent** of the explanatory variables $x_1, \ldots, x_k$ and is **normally** distributed with zero mean and variance σ^2, i.e., $u \sim N(0, \sigma^2)$.
- **MLR.6** covers **MLR.4** and **MLR.5**.
- Assumptions **MLR.1** through **MLR.6** are called the **classical linear model (CLM) assumptions**.

BUE and beyond

- Under **MLR.1** through **MLR.6**, the OLS estimator is **BUE**, i.e., the best unbiased estimator.
- Conditioning on X and assuming normality of u implies

$$y|X \sim N(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2).$$

BUE and beyond: Example



Source: Wooldridge (2012)

BUE and beyond

- CLT: because the error term u is the sum of many different unobserved factors affecting y, we can conclude that it has an approximate normal distribution.
- Assuming normality can be troublesome:
  - 'CLT argument' for u has some weaknesses: independence/uncorrelatedness, different/various distributions, additivity, etc.,
  - truncated distributions/variables,
  - data (functional) transformations sometimes help: in economics and finance, the logarithmic transformation is the most popular one.

Normal sampling distribution

- Normality of u translates into normal sampling distribution of the OLS estimator.
- Under the **CLM assumptions MLR.1** through **MLR.6**, conditional on the sample of the independent variables,

$$\hat{\beta}_j \sim N\big(\beta_j, \mathrm{Var}(\hat{\beta}_j)\big).$$

- Therefore,

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0,1).$$

- However, this holds only if we know σ^2, which is not the case in most cases.

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter
- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter
- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

t distribution for the standardized estimators

- Under the CLM assumptions MLR.1 through MLR.6,

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} = t_{df}, \tag{1}$$

  where $k+1$ is the number of unknown parameters in the population model (including the intercept) and $n-k-1$ is the degrees of freedom (df).
- Note the difference between sd and se.

t ratio

- Most commonly tested null hypothesis

$$H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0.$$

- **Important:** we are testing the value of the population parameter β_j (without 'hat'), not of its estimate!
- Under $H_0 : \beta_j = 0$, equation (1) gives us the t **ratio**

$$\boxed{t_{\hat{\beta}_j}} \equiv \frac{\hat{\beta}_j - 0}{se(\hat{\beta}_j)} = \boxed{\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}}.$$

- t **ratio** identifies the statistically significant independent variables, i.e., the ones whose partial effect (after controlling for all other independent variables) is statistically significantly different from zero.

Hypothesis testing using the t ratio: $\underline{t \text{ test}}$

t **ratio** construction gives us the standard testing framework you know from Statistics, i.e., it can be used for testing:

- One-sided alternatives, e.g.: $H_0 : \beta_j = / \leq 0 \quad \text{vs.} \quad H_1 : \beta_j > 0$.
- Two-sided alternative: $H_0 : \beta_j = 0 \quad \text{vs.} \quad H_1 : \beta_j \neq 0$.
- Through the **critical value(s)**:
  significance level $\alpha \longrightarrow$ t **ratio** vs. $c \ldots$ critical value(s) under H_0 according to df and the H_1 type $\longrightarrow$ reject/not reject H_0.
- Through the p-**value**:
  (significance level $\alpha \longrightarrow$) t **ratio** $\longrightarrow$ p-value for the given H_0 according to df and the H_1 type ($\longrightarrow$ reject/not reject H_0).
- Keep in mind the difference between **'not rejecting'** the null hypothesis and 'accepting' it (incorrect)!
- For $n - k - 1 > 100$, it is often assumed $t_{n-k-1} \approx N(0, 1)$, which gives us the 'rule of 2 (sigma)' for $\alpha = 5\%$ for the **two-tailed test**.

t test using the t statistic

- t **statistic** is a generalized version of the t ratio following (1) under a generalized $H_0 : \beta_j = a_j$

$$\boxed{t_{\hat{\beta}_j} \equiv \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)}}.$$

- t **statistic** identifies independent variables statistically significantly different from a_j.
- For example, $H_0 : \beta_j = 1$ is very useful for constant elasticity models.
- Steps of the hypotheses testing are the same as for the t ratio above.

Hypotheses testing: Example

Dependent Variable: log(*salary*)			
Independent Variables	**(1)**	**(2)**	**(3)**
log(*sales*)	.224 (.027)	.158 (.040)	.188 (.040)
log(*mktval*)	——	.112 (.050)	.100 (.049)
profmarg	——	−.0023 (.0022)	−.0022 (.0021)
ceoten	——	——	.0171 (.0055)
comten	——	——	−.0092 (.0033)
intercept	4.94 (0.20)	4.62 (0.25)	4.57 (0.25)
Observations	177	177	177
R-squared	.281	.304	.353

© Cengage Learning, 2013

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter
- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

Economic vs. statistical significance

- Check the statistical significance first: if significant, assess its economic/practical importance.
- Even if not statistically significant, we might still care about the expected effect of x on y, especially if it seems practically large: check the actual p-value whether it is not close to being significant at the given α.
- Also, think carefully about a potential bias!
- It is much more troublesome to have a significant variable with an unexpected sign and a practically large effect than an insignificant variable we thought would play a role.

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter

- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

Confidence interval

- Under the CLM assumptions MLR.1 through MLR.6, we can easily construct a **confidence interval (CI)** for the population parameter β_j.
- Point vs. interval estimates.
- Using the distribution of $\hat{\beta}_j$: $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-k-1}$, we compute a $1-\alpha$ **confidence interval** as

$$\hat{\beta}_j \pm t_{n-k-1,1-\alpha/2} se(\hat{\beta}_j).$$

- For $n-k-1 > 100$, the 'rule of 2 (sigma)' for $\alpha = 5\%$ can be again used for a rough idea.

Outline

Efficiency of the OLS estimator

Sampling distributions of the OLS estimator

Testing hypotheses about a single population parameter

- t test
- Economic vs. statistical significance
- Confidence intervals

Testing hypotheses about a single linear combination of parameters

A specific type of the null hypothesis

- We might be interested in testing $H_0 : \beta_1 = \beta_2$ rather than the commonly used $H_0 : \beta_1 = 0$.
- We need to transform/rewrite the null hypothesis into a 'testable' version, i.e., a linear combination $H_0 : \beta_1 - \beta_2 = 0$.
- This allows us to write the testing t statistic as

$$\boxed{t} \equiv \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{se(\hat{\beta}_1 - \hat{\beta}_2)} = \boxed{\frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}}.$$

- What is the catch here?

Rewriting the model

- This issue can be overcome by using the rewritten null hypothesis $H_0 : \theta \equiv \beta_1 - \beta_2 = 0$ and substituting into the original model using $\beta_1 = \theta + \beta_2$.
- Example:

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u = \\
&= \beta_0 + (\theta + \beta_2) x_1 + \beta_2 x_2 + u = \\
&= \beta_0 + \theta x_1 + \beta_2 (x_1 + x_2) + u.
\end{aligned}
$$

- We can now estimate the model and test $H_0 : \theta = 0$.
- And what is the catch here now?

Seminars and the next lecture

- Seminars:
  - hypotheses testing: single parameter (t test, p-value)
  - confidence intervals
  - hypotheses testing: single linear combination
  - discussing statistical vs. economic significance
  - two computer exercises: hypotheses testing in R
- Next lecture #6:
  - testing multiple linear restrictions: F test
  - OLS asymptotics:
    - consistency
    - asymptotic normality and efficiency
    - large sample inference
- Readings for lecture #6:
  - Chapter 4: 4.5–6, Chapter 5