LECTURE #6

Econometrics I

TESTING MULTIPLE LINEAR RESTRICTIONS & OLS ASYMPTOTICS

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

Summer semester 2024, March 26

In the previous lecture #5

- We discussed the **Gauss-Markov theorem:** OLS is **BLUE**.
- We added **MLR.6 Normality:** OLS is **BUE** and

$$
\begin{aligned}
y|X &\sim N(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2), \\
\hat{\beta}_j &\sim N\big(\beta_j, \mathrm{Var}(\hat{\beta}_j)\big).
\end{aligned}
$$

- We introduced the t **test:** under MLR.1–6 and $H_0 : \beta_j = a_j$,

$$
t_{\hat{\beta}_j} = \frac{\hat{\beta}_j - a_j}{se(\hat{\beta}_j)} \sim t_{n-k-1} = t_{df}.
$$

- $1 - \alpha$ **confidence interval** for β_j: $\quad \hat{\beta}_j \pm t_{n-k-1,1-\alpha/2} se(\hat{\beta}_j)$.
- Readings for lecture #6:
  - Chapter 4: 4.5–6, Chapter 5

Next week: Midterm test

- No seminars
- Midterm test instead of lecture:
  - 65-minute, written, closed-book; 0–30 points, results via SIS,
  - **April 2**, **11:00**, lecture halls: **MM 109 (A–P), 206 (R–Z)**,
  - no calculators, formula sheets, own blank papers, statistical tables, or vocabularies can be used; a pen fully suffices,
  - please bring your ID card (ISIC, national ID, passport).
- Midterm structure:
  1. multiple choice questions,
  2. empirical exercise: interpretation of regression results, functional forms (logs, quadratic), data scaling, inference (t test, F test), expected bias/misspecification analysis,
  3. theoretical derivations: OLS algebraic properties, 'sums of squares', unbiasedness (SLR, MLR), variance (SLR, MLR), omitted variable bias, linear estimator.
- HA #1 reminder:
  - deadline on March 28, 23:59:59,
  - delivered via the Study group roster (Lecture JEB109) in SIS.

Outline

Outline

Testing multiple linear restrictions

Consistency

Asymptotic normality, efficiency, and large sample inference

- Asymptotic normality and efficiency of OLS
- Lagrange multiplier test

Testing exclusion restrictions

- Up until now, we have been testing only simple hypotheses.
- However, we might be interested in testing a set of hypotheses/restrictions.
- Null hypothesis can then have the following form:

$$H_0 : \beta_1 = 0, \beta_2 = 0$$

against the alternative

$$H_1 : H_0 \text{ does not hold.}$$

- H_1 thus holds if at least one of β_1 and β_2 is non-zero (i.e., it complements the null).
- t statistic is not appropriate here, as we need to test the restrictions **jointly** (vs. multiple **separate** restrictions).

Unrestricted and restricted models

- Let us have an original model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

- As there are no restrictions, the model can be referred to as the **unrestricted model**.
- Using the restrictions from H_0 on the previous slide, we obtain the **restricted model**

$$y = \beta_0 + \beta_3 x_3 + u.$$

- For each, we can obtain SSR or the coefficient of determination R^2 to judge their performance, i.e., how well the models explain the variation in the dependent variable y.
- Recall that SSR of the unrestricted model SSR_U will **always be lower** than SSR of the restricted model SSR_R. Why?

F test

- Has SSR_R increased enough to allow us to reject the null hypothesis?
- In other words, has the model changed so much that the two models are statistically (based on SSR or R^2) distinguishable?
- To answer this question, we use the F **statistic** defined as

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n-k-1)} = \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/(n-k-1)},$$

  where $n-k-1$ is the degrees of freedom of the original unrestricted model, and q is the number of restrictions.
- $F \geq 0$ (**always** non-negative!)
- Under MLR.1 through MLR.6, if the null hypothesis holds,

$$F \sim F_{q,n-k-1},$$

  where q and $n-k-1$ are the two degrees of freedom.
- Why is F distributed this way?

F test: SSR vs. R^2 form of the statistic

- Using the definition of $R^2 = 1 - \frac{SSR}{SST}$, we can rearrange to get $SSR = (1 - R^2)SST$.
- This can be used both for the unrestricted and restricted model.
- Keeping in mind that SST is the same for both models, as the dependent variable remains the same, we have

$$
\begin{aligned}
F &= \frac{SSR_R - SSR_U}{SSR_U} \frac{n-k-1}{q} = \\
&= \frac{(1 - R_R^2 - 1 + R_U^2)SST}{(1 - R_U^2)SST} \frac{n-k-1}{q} = \\
&= \frac{R_U^2 - R_R^2}{1 - R_U^2} \frac{n-k-1}{q} = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(n-k-1)}.
\end{aligned}
$$

F test

- Recall how to test a hypothesis utilizing the F distribution.
- How do we obtain the p-value for the test?
- If the null hypothesis is rejected, we say that the variables of interest are **jointly statistically significant**, ceteris paribus.
- If the null hypothesis is not rejected, we say that the variables of interest are **jointly insignificant**.
- In addition, F test can be used as one of the variable selection/elimination criteria to drop insignificant explanatory variables from the model.

F statistic for testing overall significance of a regression

- Econometric software usually reports the **overall F statistic** as a complement to the coefficient of determination R^2 as a measure of the estimated model's overall explanatory power.
- Statistical significance of R^2 is usually not provided.
- Such F statistic is used to test a joint hypothesis that **all model parameters but intercept** are equal to zero, i.e.,

$$H_0 : \beta_1 = 0, \ldots, \beta_k = 0.$$

- I.e., we test the **overall significance of a regression**.
- As the R^2 of the restricted model is 0 (no variation in y is explained as there are no explanatory variables) and $q = k$ in this specific case, the F test statistic can be slightly simplified.
- Passing the F test for overall significance is considered the very minimum for the model quality judgment.

Testing general linear restrictions

- Frequently, we are interested in a more complicated joint hypothesis containing general linear restrictions, i.e., not only the ‘zero excluding restrictions’.
- The procedure remains the same but often requires some creativity.
- Let us have the **unrestricted** model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u.$$

- A set of linear restrictions $H_0 : \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$ gives the **restricted** model

$$\begin{aligned} y &= \beta_0 + x_1 + u, \\ y - x_1 &= \beta_0 + u. \end{aligned}$$

- We can now only use the *SSR* form of the *F* statistic! Why?

The CAPM model

- An interesting example: capital asset pricing model (CAPM)
- The model is specified as

$$r_{i,t} - r_{rf,t} = \alpha_i + \beta_i(r_{M,t} - r_{rf,t}) + u_{i,t}$$

  with an interesting null hypothesis of $H_0 : \alpha_i = 0, \beta_i = 1$ of an efficient market hypothesis-following asset.
- If we substitute for H_0, we obtain

$$r_{i,t} - r_{rf,t} = (r_{M,t} - r_{rf,t}) + u_{i,t},$$

  so there is nothing to estimate!
- We thus take that as if it was estimated:

$$SSR_R = \sum_{t=1}^{T} \hat{u}_{i,t}^2 = \sum_{t=1}^{T} (r_{i,t} - r_{M,t})^2.$$

Outline

Consistency

- Unbiasedness cannot always be achieved.
- Consistency is thus considered a minimal requirement for an estimator.
- In layman's terms, an estimator is considered consistent if it gets **closer** to and **more tightly** distributed around the true population parameter with increasing sample size n.
- For $n \to +\infty$, a consistent estimator collapses to a single point β_j, i.e., to the true value of the population parameter.
- Under assumptions MLR.1 through MLR.4, the **OLS estimator** $\hat{\beta}_j$ **is consistent** for all $j = 0, \ldots, k$.
- This can be shown utilizing LLN (in seminar #6 for $\hat{\beta}_0$).
- Consistency vs. unbiasedness

Consistency

- MLR.5 and MLR.6 are thus not needed for consistency, in the same way as for unbiasedness.
- Regarding MLR.4, we only need zero correlation (not mean independence) between explanatory variables and the error term for consistency.
- **MLR.4′ Zero mean and zero correlation:** $\mathbb{E}(u) = 0$ and $\text{Cov}(x_j, u) = 0$ for $j = 1, 2, \ldots, k$.
- However, the 'original' MLR.4 is still needed for unbiasedness.
- Inconsistency of $\hat{\beta}_1$ for the simple regression case is given by

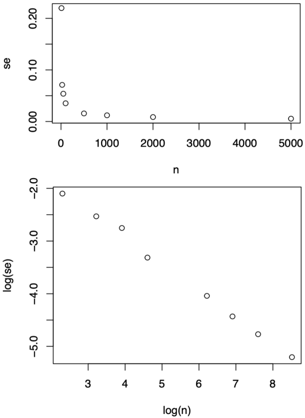
$$\text{plim}\hat{\beta}_1 - \beta_1 = \frac{\text{Cov}(x_1, u)}{\text{Var}(x_1)},$$

which is parallel to the derivation of the bias, only the inconsistency is expressed in population terms.

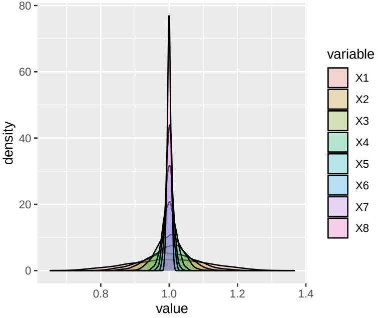
Consistency: R example

- $n \in \{10, 25, 50, 100, 500, 1000, 2000, 5000, (10k, 30k, 100k)\}$
- $y = 1 + 1x + u$

(a) convergence of *se*



(b) approximated $\hat{\beta}_1$ densities



Outline

Outline

Asymptotic normality of OLS

- Normality of u is a very strong assumption, but it is needed for normality of the OLS estimator and, in turn, also for t-distributed t statistics and F-distributed F statistics.
- Fortunately, CLT comes to the rescue when MLR.6 is violated.

Asymptotic normality of OLS

Under the Gauss-Markov assumptions MLR.1 through MLR.5:

- $\sqrt{n}(\hat{\beta}_j - \beta_j) \stackrel{a}{\sim} N(0, \textit{asymptotic}\ \mathrm{Var}_j)$, i.e., $\hat{\beta}_j$ is **asymptotically normally distributed**.
- $\hat{\sigma}^2$ is a consistent estimator of $\sigma^2 = \mathrm{Var}(u)$.
- For each j,

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

  and

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

Asymptotic normality and efficiency of OLS

- That means, if the sample is large enough and the Gauss-Markov assumptions MLR.1 through MLR.5 are met, **we do not need MLR.6** for statistical inference based on t and F statistics and the confidence intervals.
- We still need MLR.5 Homoskedasticity!
- To make sure there is a difference in notation, these are sometimes referred to as the **asymptotic** t **statistics**, etc.
- In addition, under the Gauss-Markov assumptions MLR.1 through MLR.5, the OLS estimators are **asymptotically efficient**, i.e., the estimators with the lowest variance.
- $\widehat{\text{Var}}(\hat{\beta}_j)$ shrinks to zero at a rate of $1/n \Rightarrow se(\hat{\beta}_j)$ at $1/\sqrt{n}$.
- How many observations do we need for the asymptotics?

Outline

Lagrange multiplier tests (for q exclusion restrictions)

- **LM statistic/test** is an alternative approach for testing multiple exclusion restrictions without assuming MLR.6.
- *df* play no role bcs. of the asymptotic nature of the statistic.
- Basic notion: if the 'restricted' independent variables are truly irrelevant, then these should not be correlated with residuals of the restricted model in the sample.
- This recalls MLR.4 as the 'specification assumption'.
- LM test is constructed in the following steps:
  1. H_0 and H_1 are the same as for the respective F test, e.g.:
  $$H_0 : \beta_1 = 0, \beta_2 = 0 \quad vs. \quad H_1 : H_0 \text{ does not hold},$$
  2. estimate the **restricted model** and save the residuals $\tilde{u}$,
  3. run an **auxiliary regression:** regress $\tilde{u}$ on **all independent variables** and obtain R^2, i.e., $R^2_{\tilde{u}}$ (if H_0 is true, $R^2_{\tilde{u}}$ is 'close' to zero),
  4. compute $\boxed{LM = nR^2_{\tilde{u}}}$,
  5. under the null hypothesis, $LM \overset{a}{\sim} \chi^2_q$,
  6. if $LM > c$, we reject H_0 at the given significance level α.

Seminars and the next lecture

- Seminars:
  - testing multiple linear restrictions: F test
  - consistency derivation for $\hat{\beta}_0$
  - normality of residuals and asymptotic variance in practice
  - Lagrange multiplier test
- Next lecture #7 (in two weeks):
  - midterm summary
  - effects of data scaling on OLS statistics
  - more on functional forms: quadratic, logarithmic, interactions
  - more on goodness-of-fit: adjusted R-squared
  - selection of explanatory variables
- Readings for lecture #7:
  - Chapter 6: 6.1–6.3