Exam 1 Review

Chapters 4, 5, 6, 7

Gauss-Markov Theorem

Measures of fit

Omitted variable bias

Joint tests

Practice

CH4 Learning objectives

- Set up appropriate equations to estimate relationship between two variables using OLS
- Interpret intercept and slope coefficients for simple linear regression
- Define and calculate residuals
- Calculate measures of fit, including R^2, ESS, TSS, SSR, and SER
- Understand underlying assumptions for estimation of β_0 and β_1

CH5 Learning objectives

- Create hypotheses about slope coefficients and test them using $\hat{\beta}_1$ and its standard error.
- Correctly interpret the results of hypothesis tests
- Calculate confidence intervals for β_1
- Take binary regressors in stride (and interpret them correctly)
- Understand the implications of heteroskedasticity and correct your standard errors
- Know and apply the Gauss-Markov theorem to understand the circumstances under which OLS is BLUE.

CH6 Learning objectives

- Just go to town on some multiple linear regression - implementing and interpreting
- Deepen our understanding of omitted variable bias
- Calculate and interpret a new measure of fit, the adjusted R^2
- Update our knowledge of least square assumption and the sampling distribution of the OLS estimator in the case of multiple independent variables

CH7 Learning objectives

- Construct and interpret tests of joint hypotheses
- Construct and test hypothesis test involving one restriction and multiple coefficients

Today

- CH4/5/6 Know and apply the Gauss-Markov theorem to understand the circumstances under which OLS is BLUE.
- CH4/6 Calculate measures of fit, including R^2, ESS, TSS, SSR, and SER
- CH6 Deepen our understanding of omitted variable bias
- CH7 Complicated hypothesis testing
- CH6 Just go to town on some multiple linear regression - implementing and interpreting

Gauss-Markov Theorem

Ordinary Least Squares Assumptions

We worked in three stages: (Chapter 4): Consider our three LS assumptions (needed for unbiasedness):

1. $E(u|X = x) = 0$ (zero conditional mean)
2. $(X_i, Y_i), i = 1, , n$ are i.i.d.
3. Large outliers are rare

(Chapter 5) Plus, one more!

4. u is homoskedastic

(Chapter 6) JK, one more (*but not part of GM theorem*)

5. No multicollinearity

Gauss-Markov Theorem

Under these **four** extended LS assumptions, $\hat{\beta}_1$ has the smallest variance among *all linear estimators* (estimators that are linear functions of $Y_1, ..., Y_n$).

This is the **Gauss-Markov theorem**

Under the GM theory, OLS estimators are **BLUE**:

- Best
- Linear
- Unbiased
- Estimators

Common violations

- Violation of zero conditional mean: omitted variable bias
- $(X_i, Y_i), i = 1, , n$ are i.i.d.: panel data, time-series data
- u is homoskedastic, $Var(u|X_i = x) = \sigma$ (constant): if variance depends on X (happens a lot!)

What happens when we violate these assumptions

- No homoskedasticity: $\hat{\beta}$ remains unbiased. OLS no longer BLUE. If you do not adjust standard errors, they will be wrong
- Violation of other assumptions: $\hat{\beta}$ biased

Measures of fit

Goodness-of-fit

We define the total sum of squares, estimated sum of squares, and residual sum of squares:

$$y_i = \hat{y}_i + \hat{u}_i$$

$$\begin{aligned} TSS &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\ ESS &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \\ SSR &= \sum_{i=1}^{n} \hat{u}_i^2 \end{aligned}$$

Properties of OLS on any Sample of Data

- Assuming $TSS > 0$, we can define the fraction of the total variation in y_i that is explained by x_i (or the OLS regression line) as

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

- Called the **R-squared** of the regression.

$$0 \leq R^2 \leq 1$$

Do not fixate on R^2. Having a " high" R-squared is neither necessary nor sufficient to infer causality.

Standard error of the regression (SER)

We can estimate the variance of the regression

$$\boxed{\hat{\sigma}^2 = s_e^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = \frac{SSR}{n-k-1}}$$

- Divide by $n-2$ in simple linear regression because we've used up two d.f: one on $\hat{\beta}_0$ and one on $\hat{\beta}_1$.
- We call $s_e = \sqrt{s_e^2}$ the **standard error of the regression** (SER)

Omitted variable bias

Omitted variable bias

Population model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i$$

Estimated model

$$y_i = \widetilde{\beta_0} + \widetilde{\beta}_1 x_{1,i} + +u_i$$

Three cases:

1. $cov(y, x_2) = 0$
2. $cov(y, x_2) \neq 0$ and $cov(x_1, x_2) = 0$
3. $cov(y, x_2) \neq 0$ and $cov(x_1, x_2) = 0$

Signing the direction of the bias

- With one omitted variable, we can sign the bias if we know the direction of β_2 and δ_1
- Conditional on x_1 and x_2, we can compute $E[\widetilde{\beta}_1]$

$$E[\widetilde{\beta}_1] = \beta_1 + \beta_2 \widetilde{\delta}_1 \qquad (1)$$

- Note that the sign of $\widetilde{\delta}_1$ is the same as the sign of $Cov(x_{i1}, x_{i2})$.

	$corr(x_1, x_2) > 0$	$corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

Joint tests

Three types of tests

1. Hypothesis tests with one restriction, one coefficient
    - Example: $H_0 : \beta_j = \beta_{j,0}$ vs. $H_a : \beta_j \neq \beta_{j,0}$
2. Hypothesis tests with one restriction, multiple coefficients
    - General: $H_0 : \beta_j = \beta_m$
    - Example: $H_0 : \beta_1 = 0$
3. Hypothesis tests involving a multiple tests at once (joint hypothesis tests)
    - General: $H_0 : \beta_j = \beta_{j,0}, \beta_m = \beta_{m,0}, \ldots$
    - Example: $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$
    - Special case - test of *all* regressors

Practice

Practice Question 1

Consider the following population model:

$$WorkHrs_i = \beta_0 + \beta_1 age_i + \beta_2 educ_i + \beta_3 married_i + u_i$$

where $WorkHrs_i$ is the usual weekly work hours, age_i is a person's age, $educ_i$ is the number of years of education completed, and $married_i$ is a binary variable equal to 1 if the person is married, and 0 otherwise.

Practice Question 1

Using OLS, you estimate the following equation:

$$WorkHrs_i = 33.23 + 0.11age_i + 0.11educ_i + 3.58married_i$$

What is the interpretation on the coefficient on age_i, $\widehat{\beta}_1$?

Practice Question 1

Using OLS, you estimate the following equation:

$$WorkHrs_i = 33.23 + 0.11age_i + 0.11educ_i + 3.58married_i$$

Consider Boris. Boris is 56 years old, has completed 16 years of education, and is married. He works 35 hours per week. What is his residual?

Practice Question 2

Consider the following population model:

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

Suppose that $E[u|x_1, x_2, x_3] = 0$ and $Var(u|x_1, x_2, x_3) = \sigma^2 \sqrt{x_2}$

1. Is the OLS estimate of β_2 likely to be BLUE?
2. List each assumption that needs to hold for $\widehat{\beta}_2$ to be BLUE. For each one, explain whether it is likely to be violated or note

Conclusion

Gauss-Markov Theorem

Measures of fit

Omitted variable bias

Joint tests

Practice