Seminar 3 Solutions

Omitted Variables, Collinearity, and Heteroskedasticity

Giulio Rossetti*

giuliorossetti94.github.io

February 6, 2026

* email: giulio.rossetti.1@wbs.ac.uk

Roadmap

Part 1: Omitted Variable Bias

- Exercise 1: Omitted Variable Bias
- Exercise 1: Collinearity and Interaction Terms

Part 2: Randomized Experiment

- Exercise 2: Randomized Experiment

Part 3: Heteroskedasticity

- Exercise 3: Heteroskedasticity Consequences

Part 4: Work vs. Sleep

- Exercise 4: Work vs. Sleep

Roadmap

Part 1: Omitted Variable Bias

- Exercise 1: Omitted Variable Bias
- Exercise 1: Collinearity and Interaction Terms

Part 2: Randomized Experiment

- Exercise 2: Randomized Experiment

Part 3: Heteroskedasticity

- Exercise 3: Heteroskedasticity Consequences

Part 4: Work vs. Sleep

- Exercise 4: Work vs. Sleep

Disclaimer

Disclaimer

Full solutions are available on my.wbs. All exercises are examinable material, not just the ones we covered in the seminars.

Exercise 1

Omitted Variable Bias

Model:

$$\log(wage) = \beta_0 + \beta_1 female + \beta_2 train + \beta_3 educ + \beta_4 exper + u$$

- If less able workers are more likely to be selected and ability is *omitted*:

True model:

$$\log(wage) = \beta_0 + \beta_1 female + \beta_2 train + \beta_3 educ + \beta_4 exper + \beta_5 ability + \epsilon$$

- therefore $u = \beta_5 ability + \epsilon$
- What can we say about the bias in the OLS estimate of β_2?

	$Corr(x_2, x_5) > 0$	$Corr(x_2, x_5) < 0$
$\beta_5 > 0$	Positive Bias	Negative Bias
$\beta_5 < 0$	Negative Bias	Positive Bias

Exercise 1

Bias Direction

- Higher worker ability leads to Higher wages: $\beta_5 > 0$.
- Auxiliary model:

$$ability = \delta_0 + \delta_1 train + v$$

- Estimate likely to be $\tilde{\delta}_1 < 0$. i.e. $train$ and $ability$ are negatively correlated (Less able workers are more likely to be selected for training).
- Bias in OLS estimate:

$$\tilde{\beta}_2 = \hat{\beta}_2 + \hat{\beta}_5 \tilde{\delta}_1 < \hat{\beta}_2.$$

- Bias: $\beta_5 > 0$, $Cov(train, ability) < 0$ implies Negative Bias on β_2
- Conclusion: Negative bias in β_2, but the magnitude cannot be exactly quantified.

Exercise 1

Collinearity and Interaction Terms

- Dummy variables and perfect collinearity:
  - By definition, $male = 1 - female$.
  - Including both $male$ and $female$ causes perfect collinearity.
  - If there are N dummy variables, include only $N - 1$ to avoid collinearity.
  - Alternative: exclude the intercept term β_0.
- Interaction term for gender and training program:
  - To test if training effects differ by gender, modify the model:

$$\log(wage) = \beta_0 + \beta_1 female + \beta_2 train + \beta_3 educ + \beta_4 exper + \beta_5 female \times train + u.$$

  - This allows different slopes for $train$ by gender.

Roadmap

Part 1: Omitted Variable Bias

- Exercise 1: Omitted Variable Bias
- Exercise 1: Collinearity and Interaction Terms

Part 2: Randomized Experiment

- Exercise 2: Randomized Experiment

Part 3: Heteroskedasticity

- Exercise 3: Heteroskedasticity Consequences

Part 4: Work vs. Sleep

- Exercise 4: Work vs. Sleep

Exercise 2

Randomized Experiment

- Scholarship *randomly assigned*, independent of other factors.
- OLS is unbiased as long as assumptions hold.
  - No change in OLS mechanics or statistical theory.
  - Interpretation of the coefficient differs.
  - With a single regressor, OLS provides an unbiased estimate as long as SLR.1 through SLR.4 hold.

$\rightarrow \beta_2$ ability

$$score = \beta_0 + \beta_1 scholarship + u$$

ass $E[u|x] = 0 \Rightarrow corr(sch, u) = 0$

Exercise 2

OLS and Dummy Variables

- Should we add additional controls? Do we have an OVB?

Exercise 2

OLS and Dummy Variables

- Should we add additional controls? Do we have an OVB?
- MLR4 Zero Conditional Mean Assumption:

$$\mathbb{E}[u_i \mid x_i] = 0 \quad (1)$$

$$\mathbb{E}[u_i \mid scholarship] = 0 \quad (2)$$

- Is MLR4 satisfied? If not, we have an OVB.
- OVB vs better model fit

β_1 is unbiased

Roadmap

Part 1: Omitted Variable Bias

- Exercise 1: Omitted Variable Bias
- Exercise 1: Collinearity and Interaction Terms

Part 2: Randomized Experiment

- Exercise 2: Randomized Experiment

Part 3: Heteroskedasticity

- Exercise 3: Heteroskedasticity Consequences

Part 4: Work vs. Sleep

- Exercise 4: Work vs. Sleep

Exercise 3

Heteroskedasticity Consequences

homosk : $var(u|X) = \sigma^2 I_n$

$\rightarrow var(\hat{\beta}|X) = \sigma^2 (X'X)^{-1}$

Which of the following are consequences of heteroskedasticity?

1. The OLS estimator, $\hat{\beta}_j$, is biased.
2. The OLS estimator is no longer BLUE.
3. The usual t-statistic no longer has a t distribution.

Roadmap

Part 1: Omitted Variable Bias

- Exercise 1: Omitted Variable Bias
- Exercise 1: Collinearity and Interaction Terms

Part 2: Randomized Experiment

- Exercise 2: Randomized Experiment

Part 3: Heteroskedasticity

- Exercise 3: Heteroskedasticity Consequences

Part 4: Work vs. Sleep

- Exercise 4: Work vs. Sleep

Exercise 4

Work vs. Sleep – Regression Output

$H_0: \beta_4 = 0$

$H_1: \beta_4 > 0$

$t_{\hat{\beta}_4} = \frac{\hat{\beta}_4 - \beta_{H_0}}{se(\hat{\beta}_4)}$

Model: $sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 male + u$

sleep	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
totwrk	-.1657914	.0179622	-9.23	0.000	-.2010576	-.1305253
educ	-11.75612	5.866382	-2.00	0.045	-23.27391	-.2383405
age	1.964277	1.442942	1.36	0.174	-.8687296	4.797283
male	87.99325	34.32329	2.56	0.011	20.6045	155.382
_cons	3642.467	111.8443	32.57	0.000	3422.877	3862.056

Model: $sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 male + \beta_5 male \times totwork + u$

sleep	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
totwrk	-.1438338	.026148	-5.50	0.000	-.1951717	-.0924959
educ	-11.78482	5.865035	-2.01	0.045	-23.29998	-.2696511
age	1.723503	1.457574	1.18	0.237	-1.138238	4.585244
male	174.457	82.333	2.12	0.034	12.80782	336.1062
male_totwrk	-.0419258	.0362901	-1.16	0.248	-.1131762	.0293246
_cons	3614.41	114.4244	31.59	0.000	3389.754	3839.066

Exercise 4

Work vs. Sleep – Interpretation

- Do men sleep more than women?
  - Male tend to sleep more than females $\hat{\beta}_4 = 87.99$, $(p = 0.011)$
  - At which confidence level can we reject the *null hypothesis* $H_0 : \beta_4 = 0$? 5%
- Trade-off between work and sleep:
  - Statistically significant tradeoff: $\hat{\beta}_1 = -0.166$
  - Strong significance: $t_{\hat{\beta}_1} = -9.23$, $p < 0.001$
  - Intuition: The more you work, the less you sleep.
- Being male and working hard:
  - No effect($\hat{\beta}_5 = -0.042$, $t_{\hat{\beta}_5} = -1.16$, $p = 0.248$).
  - Hardworking men still tend to sleep more than females.
  - The interaction term does not significantly affect the impact of being male on sleep time.