

## 1 Model Specification

Here we discuss whether the model is “correct” and address issues such as...

- omission of a relevant variable (underfitting a model),
- inclusion of an irrelevant variable (overfitting a model),
- measurement error,
- specification tests, and
- model selection criteria.

### Underfitting a Model

- True model:  $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$
- Estimated model:  $Y_i = \alpha_1 + \alpha_2 X_{2i} + u_i$
- OLS estimate  $\hat{\alpha}_2$  is biased if  $\text{corr}(X_{2i}, X_{3i}) \neq 0$
- OLS  $se(\hat{\alpha}_2)$  is biased too
- Hypothesis testing is invalid

### Overfitting a Model

- True model:  $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$
- Estimated model:  $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$
- OLS estimates are unbiased:  $E(\hat{\alpha}_2) = \beta_2$  and  $E(\hat{\alpha}_3) = \beta_3 = 0$
- OLS estimates are inefficient
- Hypothesis testing is valid

### Measurement Error

- Dependent Variable:
  - True model:  $Y_i^* = \beta_1 + \beta_2 X_{2i} + u_i$
  - $Y_i^*$  is unobserved

- $Y_i = Y_i^* + \epsilon_i$  is observed
- Estimated model:  $Y_i = \beta_1 + \beta_2 X_i + (u_i + \epsilon_i)$
- $\hat{\beta}_2$  is unbiased

- Independent Variable:

- True model:  $Y_i = \beta_1 + \beta_2 X_i^* + u_i$
- $X_i^*$  is unobserved
- $X_i = X_i^* + \epsilon_i$  is observed
- Estimated model:  $Y_i = \beta_1 + \beta_2 X_i + (u_i - \beta_2 \epsilon_i)$
- $\hat{\beta}_2$  is biased because  $cov(X_i, u_i - \beta_2 \epsilon_i) = -\beta_2 \sigma_\epsilon^2 \neq 0$
- $\hat{\beta}_2$  exhibits attenuation bias (i.e., biased towards zero)

### Model Specification Tests and Selection Criteria

- Graphical test of residuals
- Overfitting test:  $t$  and  $F$  tests
- Ramsey RESET test
- Model selection criteria:  $\bar{R}^2$ ,  $AIC$ ,  $SIC$

### Application: Phillips Curve