Model Specification

1. Types of Specification Errors
2. Consequences of Model Specification Errors
3. Tests of Specification Errors
4. Errors of Measurement
5. Incorrect Specification of Stochastic Error Term
6. Nested versus Non-nested Models
7. Model Selection Criteria

Types of Specification Errors

1. Omitting Relevant Variable(s)
2. Including Unnecessary or Irrelevant Variable(s)
3. Adopting Wrong Functional Form
4. Errors of Measurement
5. Incorrect Specification of Stochastic Error Term

Omitting Relevant Variable(s)

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{1i}$$

Incorrect Specified Model

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_{2i}$$

Specified Error – Omitting a Relevant Variable

$$u_{2i} = u_{1i} + \beta_4 X_{4i}$$

Including Irrelevant Variable(s)

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{1i}$$

Incorrect Specified Model

$$Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \lambda_4 X_{4i} + \lambda_5 X_{5i} + u_{3i}$$

Specified Error – Including a Irrelevant Variable

$$u_{3i} = u_{1i} - \lambda_5 X_{5i}$$

Adopting Wrong Functional Form

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{1i}$$

Incorrect Specified Model

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{1i}$$

Specified Error – Wrong Functional Form

Y should be linear but log-linear.

Errors of Measurement

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i}$$

Incorrect Specified Model

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + \beta_3^* X_i^{*2} + \beta_4^* X_i^{*3} + u_i^*$$

Specified Error – Errors of Measurement

$$Y_i^* = Y_i + \varepsilon_i \quad \text{and} \quad X_i^* = X_i + w_i$$

ε_i and w_i are errors of measurement bias.

Incorrect Specification of Stochastic Error Term

Correct Specified Model

$$Y_i = \beta X_i u_i$$

Incorrect Specified Model

$$Y_i = \alpha X_i + u_i$$

Specified Error – Stochastic Error

u_i is log-normally distributed,
not normally distributed.

Consequences of Model Specification Errors

1. Underfitting a Model
   (Omitting a Relevant Variable)
2. Overfitting a Model
   (Inclusion of an Irrelevant Variable)

Underfitting a Model (Omitting a Relevant Variable)

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{1i}$$

Incorrect Specified Model

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_{2i}$$

Consequences:

1. $\hat{\alpha}_1, \hat{\alpha}_2, and\ \hat{\alpha}_3$ are biased and inconsistent.
2. σ^2 is incorrectly estimated.
3. S.E. of $\hat{\alpha}_2\ and\ \hat{\alpha}_3$ are biased.

Overfitting a Model (Inclusion of an Irrelevant Variable)

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{1i}$$

Incorrect Specified Model

$$Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \lambda_4 X_{4i} + \lambda_5 X_{5i} + u_{3i}$$

Consequences:

1. $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, and\ \hat{\lambda}_4$ are unbiased and consistent.
2. σ^2 is correctly estimated.
3. Variance of $\hat{\lambda}_2, \hat{\lambda}_3, and\ \hat{\lambda}_4$ are larger.

Nested versus Non-nested Models

Nested Models

Model A: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$

Model B: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

Non-nested Models

Model C: $Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$

Model D: $Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$

Functional Form

Model E: $Y_i = \beta_1 + \beta_2 \ln Z_{2i} + \beta_3 \ln Z_{3i} + w_i$

Tests of Specification Errors

1. Detecting the Presence of Unnecessary Variables (Overfitting a Model)
   - F-test – Redundant Variables
2. Tests for Omitted Variables and Incorrect Functional Form
   - Examination of Residuals
   - LM Test – Restricted Least-Squares (Omitted Variables)

Tests of Non-nested Hypotheses

1. Discrimination Approach
   - Goodness-of-fit Criterion – R^2 or Adjusted R^2
   - Akaike's Information Criterion (AIC)
   - Schwarz's Information Criterion (SIC)
2. Discerning Approach
   - Non-nested F Test
   - Davidson-MacKinnon J Test

Model Selection Criteria

1. The R^2 Criterion
2. Adjusted R^2
3. Akaike Information Criterion (AIC)
4. Schwarz Information Criterion (SIC) or Bayesian Information Criterion (BIC)

$$AIC = e^{2k/n} \frac{\sum \hat{u}_i^2}{n}$$

$$SIC = n^{k/n} \frac{\sum \hat{u}_i^2}{n}$$

Errors of Measurement

Correct Specified Model

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i}$$

Incorrect Specified Model

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + \beta_3^* X_i^{*2} + \beta_4^* X_i^{*3} + u_i^*$$

Specified Error – Errors of Measurement

$$Y_i^* = Y_i + \varepsilon_i \quad \text{and} \quad X_i^* = X_i + w_i$$

ε_i and w_i are errors of measurement bias.

Errors of Measurement in Dependent Variable Y

The Model

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + u_i$$

Y_i^* is incorrectly measured $\qquad Y_i = Y_i^* + \varepsilon_i$

$$Y_i = (\beta_1 + \beta_2 X_{2i} + u_i) + \varepsilon_i$$
$$= \beta_1 + \beta_2 X_{2i} + (u_i + \varepsilon_i)$$

Consequences:

Although estimators are unbiased, variance of $\hat{\beta}_2$ are larger.

Errors of Measurement in Explanatory Variable X

The Model

$$Y_i = \beta_1 + \beta_2 X_i^* + u_i$$

X_i^* is incorrectly measured $\qquad X_i = X_i^* + w_i$

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 (X_{2i} - w_i) + u_i \\ &= \beta_1 + \beta_2 X_{2i} + (u_i - \beta_2 w_i) \end{aligned}$$

Consequences:

Estimators are biased and inconsistent.

Incorrect Specification of Stochastic Error Term

(1) $Y_i = \beta_1 X_i^{\beta_2} u_i \longrightarrow \ln Y_i = \alpha + \beta_2 \ln X_i + \ln u_i$

(2) $Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \longrightarrow \ln Y_i = \alpha + \beta_2 \ln X_i + u_i$

(3) $Y_i = \beta_1 X_i^{\beta_2} + u_i \longrightarrow \ln Y_i = \ln\left(\beta_1 X_i^{\beta_2} + u_i\right)$

where $\alpha = \ln \beta_1$

Equation (1) and (2) are linear-in-parameter.

Equation (3) is nonlinear-in-parameter.

(1) $\ln u_i \sim N\left(0, \sigma^2\right)$ Log-normal Distribution

(2) $u_i \sim N\left(0, \sigma^2\right)$ Normal Distribution

Incorrect Specification of Stochastic Error Term

Correct Specified Model $Y_i = \beta_1 X_i^{\beta_2} u_i$

$$\ln Y_i = \alpha + \beta_2 \ln X_i + \ln u_i$$

Incorrect Specified Model

$$\ln Y_i = \alpha + \beta_3 \ln X_i + u_i$$

Specified Error – Stochastic Error

u_i is log-normally distributed, not normally distributed.

<u>Consequences</u>: $E(\hat{\beta}_3) = \beta_2 e^{\sigma^2/2}$

$\hat{\beta}_3$ is biased estimator.