Methodology of Econometrics

1. Statement of theory or hypothesis
2. Specification of mathematical model of the theory
3. Specification of econometric model of theory
4. Obtaining the data
5. Estimation of the parameters of the econometric model
6. Hypothesis testing
7. Forecasting or prediction
8. Using model for control or policy purposes

Assumptions of Least Squares

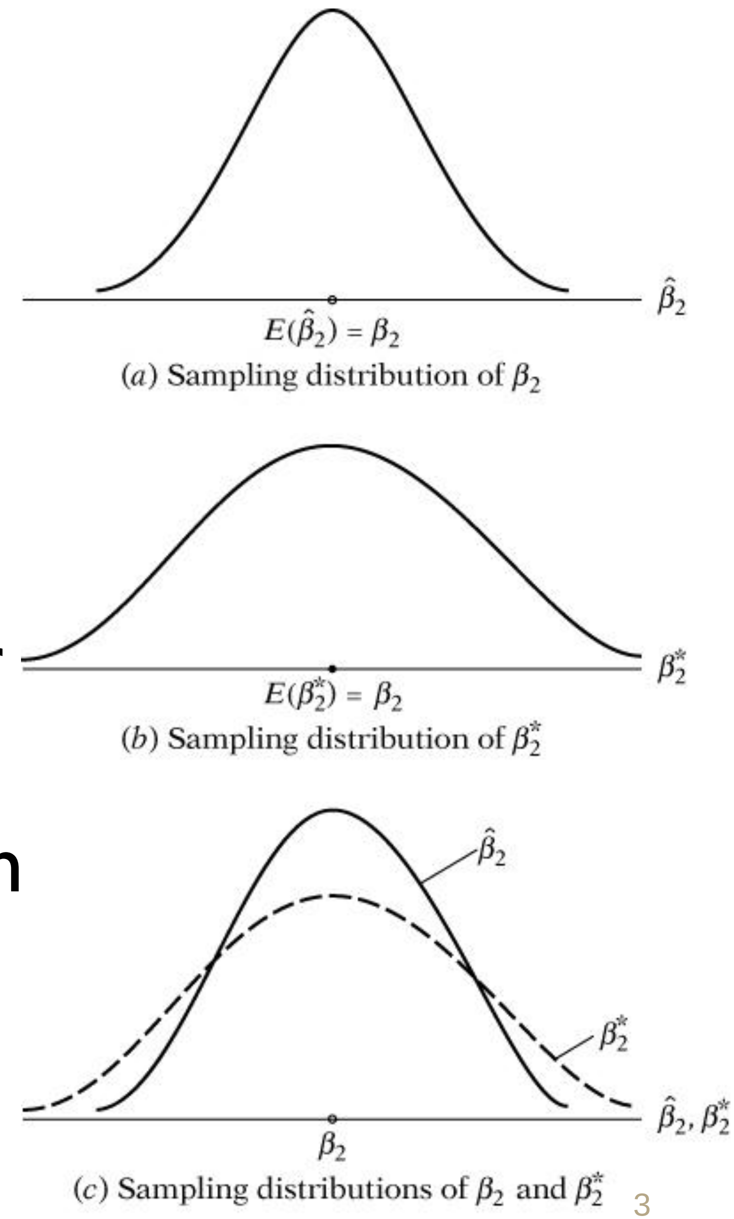
1. Linear Regression Model
2. X_i values are fixed in repeated sampling
3. Zero mean value of disturbance u_i
4. Homoscedasticity or equal variance of u_i
5. No autocorrelation between the disturbance
6. Zero covariance between u_i and X_i
7. Number of observations must be greater than number of parameters to be estimated
8. Variability in X_i values
9. Regression model is correctly specified
10. No perfect multicollinearity

Properties of Least-Squares Estimator

1. Linear
2. Unbiased
3. Efficient estimator

Gauss-Markov Theorem:

Given assumptions of CLRM, the least-squares estimators, in the class of unbiased linear estimators, have minimum variance -- they are Best Linear Unbiased Estimators (BLUE).



Evaluating Estimated Results

1. Sign and meaning of the Coefficients.
   - Whether the estimated results are according to the theory.
2. Overall Test – F-test.
   - Whether all explanatory variables can be used in explaining the dependent variable.
3. R-Squares.
   - How well does the model explain the dependent variable.
4. Individual Test – t-test.
   - Whether each explanatory variables can explain the dependent variable.

Sign of Coefficient – Meaning of Partial Regression Coefficients

Since the regression model derived from the theory, signs of the coefficients should follow what's suggested by the theory.

Partial regression coefficient β_2 measures the change in the mean value of Y or $E(Y|X)$ with respect to X_2, holding other X constant.

Partial regression coefficient β_k measures the change in the mean value of Y or $E(Y|X)$ with respect to X_k, holding other X constant.

Overall Test

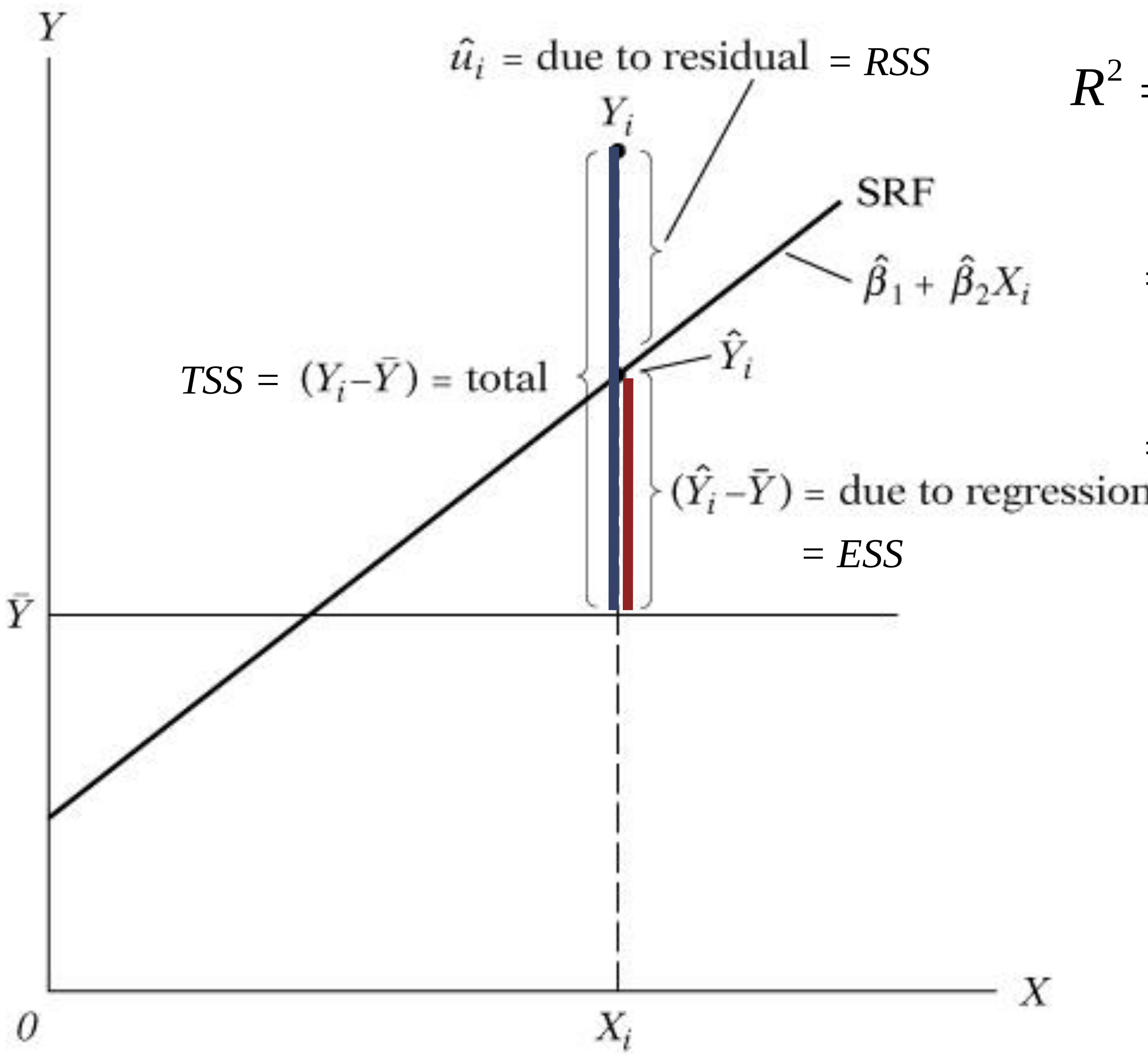
The model: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$

$H_0: \beta_2 = \beta_3 = \cdots = \beta_k = 0$ and H_a: *Otherwise*

Source of Variation	*SS*	*df*	*MSS*
Due to Regression *(ESS)*	$\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$	$k-1$	$\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 \Big/ k-1$
Due to Residuals *(RSS)*	$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{n}\hat{u}_i^2$	$n-k$	$\sum_{i=1}^{n}\hat{u}_i^2 \Big/ n-k$
TSS	$\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$	$n-1$	

$$F = \frac{ESS/k-1}{RSS/n-k}$$

Coefficient of Determination R^2



$$R^2 = \frac{\sum\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum\left(Y_i - \bar{Y}\right)^2} = \frac{ESS}{TSS}$$

$$= 1 - \frac{RSS}{TSS}$$

$$= 1 - \frac{\sum \hat{u}_i^2}{\sum\left(Y_i - \bar{Y}\right)^2}$$

Coefficient of Determination R^2

Nonnegative quantity

R^2 lies between 0 and 1

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

R^2 is nondecreasing function of the number of explanatory variables.

As number of explanatory variables increases, R^2 almost invariably increases and never decreases.

R^2 and Adjusted R^2

To compare R^2, we should adjust the number of explanatory variables or degree of freedom.

Adjusted R^2 $\bar{R}^2 = 1 - \left(1 - R^2\right)\dfrac{n-1}{n-k}$

1. For $k>1$, $\bar{R}^2 < R^2$

2. $\bar{R}^2$ can be negative while R^2 is nonnegative

Comparing Two R^2 Values

R^2 can be compared only when the sample size n and the dependent variable must be the same.

Explanatory variables may take any form.

R^2 and Adjusted R^2

Example:

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (1)$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \qquad (2)$$

R^2 of these two models are not comparable.

Eq. (1) R^2 measures proportion of the variation in $ln\ Y_i$ explained by X_2 and X_3

Eq. (2) R^2 measures proportion of the variation in Y_i explained by X_2 and X_3

Individual Test – t-test

Since residuals are assumed to be normally distributed, Y as linear function of residuals will also be normally distributed.

$\hat{\beta}$ computed from Y should also be normally distributed. However, during computational process, $\hat{\beta}$ loss degree of freedom, thus, $\hat{\beta}$ are t-distributed.

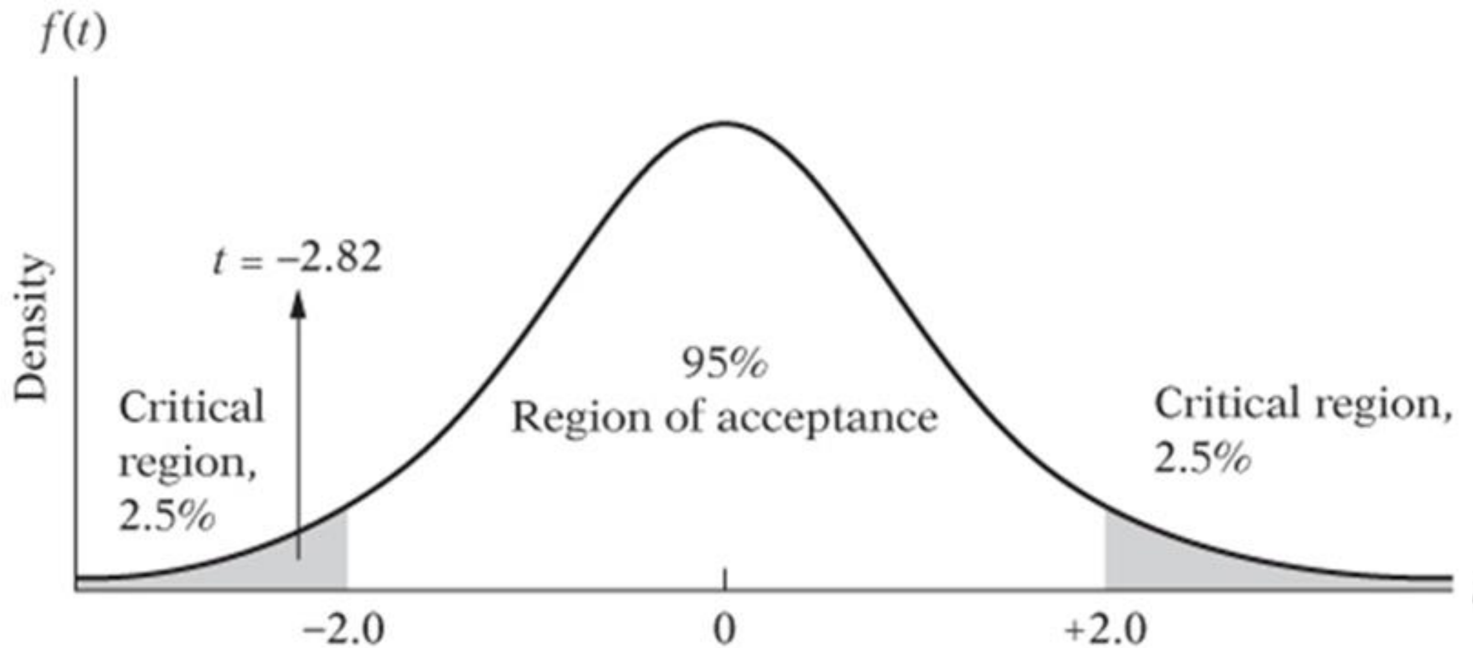
$$H_0 : \beta_1 = 0 \qquad H_0 : \beta_2 = 0 \qquad H_0 : \beta_k = 0$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{se\left(\hat{\beta}_1\right)} \qquad t = \frac{\hat{\beta}_2 - \beta_2}{se\left(\hat{\beta}_2\right)} \qquad t = \frac{\hat{\beta}_k - \beta_k}{se\left(\hat{\beta}_k\right)}$$

Individual Test – t-test

$$H_0: \quad \beta_2 = 0 \quad \text{and} \quad H_a: \quad \beta_2 \neq 0$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{se\left(\hat{\beta}_2\right)} = \frac{\hat{\beta}_2}{se\left(\hat{\beta}_2\right)}$$



Specific Test – Restricted Test

Unrestricted Models

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$$

$H_0 : \beta_3 = \beta_4 = 0$ and H_a: *Otherwise*

Restricted Models

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_5 X_{5i} + \varepsilon_i$$

Specific Test – Restricted Test

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i}$$

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

Constant Return to Scale

$$H_0: \quad \beta_2 + \beta_3 = 1 \quad or \quad \beta_2 = 1 - \beta_3$$

$$\ln Y_i = \ln \beta_1 + \boxed{(1 - \beta_3)} \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$\ln Y_i = \ln \beta_1 + \boxed{\ln X_{2i}} + \beta_3 \boxed{(\ln X_{3i} - \ln X_{2i})} + u_i$$

$$\boxed{\ln \left(Y_i / X_{2i} \right)} = \ln \beta_1 + \beta_3 \boxed{(\ln X_{3i} / X_{2i})} + u_i$$

Specific Test – Restricted Test

$$F = \frac{\left(RSS_R - RSS_{UR}\right)/m}{RSS_{UR}/(n-k)}$$

$$= \frac{\left(R^2_{UR} - R^2_R\right)/m}{\left(1 - R^2_{UR}\right)/(n-k)}$$

UR = Unrestricted regression

R = Restricted regression

m = Number of linear restrictions

k = Number of parameters in unrestricted regression

n = Number of observation

Specific Test – Restricted Test

Unrestricted Models

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i$$

$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ and H_a: *Otherwise*

Restricted Models

$$Y_i = \beta_1 + \varepsilon_i$$

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} = \frac{ESS/k-1}{RSS/(n-k)}$$

Dummy Variable

In the case that we have special event that cannot be quantify (i.e. economics crisis), dummy variable can be used.

Dummy variable is a discrete and binary-choice variable (0 or 1) that used to quantify qualitative independent variable.

Dummy variable can only be 0 or 1. It cannot be 1 or 2.

In case that qualitative variable has more than 2 choices, the model will have more than 1 dummy variable.

(Two choice – 1 dummy), (Three choice – 2 dummy), (Four choice – 3 dummy).

Intercept Dummy Variable

General model $$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

Model with Intercept Dummy Variable

$$Y_t = \beta_0 + \gamma_0 D_t + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

where: $D_t = 0$ before crisis
$= 1$ after crisis.

This model can be interpreted as:

Before Crisis: $$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

After Crisis: $$Y_t = (\beta_0 + \gamma_0) + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$$

Intercept & Slope Dummy Variables

Model with Intercept and Slope Dummy Variable

$$Y_t = \beta_0 + \gamma_0 D_t + \beta_1 X_{1t} + \gamma_1 D_t X_{1t} + \beta_2 X_{2t} + \gamma_2 D_t X_{2t} + u_t$$

where: $D_t = 0$ before crisis
$= 1$ after crisis.

This model can be interpreted as:

Before Crisis: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$

After Crisis: $Y_t = (\beta_0 + \gamma_0) + (\beta_1 + \gamma_1) X_{1t} + (\beta_2 + \gamma_2) X_{2t} + u_t$

Test Structural Change – Dummy Variable Test vs Chow Test

Dummy Variable Technique

All $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$

Before $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$

After $Y_t = (\beta_0 + \gamma_0) + (\beta_1 + \gamma_1) X_{1t} + (\beta_2 + \gamma_2) X_{2t} + u_t$

Chow Test

1 All $Y_t = \lambda_0 + \lambda_1 X_{1t} + \lambda_2 X_{2t} + u_t$ *for* $t = 1, 2, \ldots, t_1 + t_2$

2 Before $Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + u_{1t}$ *for* $t = 1, 2, \ldots, t_1$

3 After $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_{2t}$ *for* $t = n_1+1, n_1+2, \ldots, t_1+t_2$

Dummy Variable vs Chow Test

Chow Test

$$H_0: \quad \alpha_0 = \beta_0 = \lambda_0$$
$$\text{and} \quad \alpha_1 = \beta_1 = \lambda_1$$
$$\text{and} \quad \alpha_2 = \beta_2 = \lambda_2$$
$$H_a: \quad Otherwise$$

$$F = \frac{\left(RSS_1 - RSS_2 - RSS_3\right)/k}{\left(RSS_2 + RSS_3\right)/\left(n_1 + n_2 - 2k\right)}$$

Dummy Variable

$$H_0: \quad \gamma_0 = \gamma_1 = \gamma_2 = 0$$
$$H_a: \quad Otherwise$$

$$F = \frac{\left(RSS_R - RSS_{UR}\right)/k}{RSS_{UR}/\left(n-k\right)}$$

Comparing the Two Methods

Chow test is to test whether all estimators are the same or not.

Dummy variables technique test whether intercept dummy coefficient and all slope dummy coefficients are all equal to zero or not.

Dummy Variable Alternative to Chow Test

Chow Test

$$
\begin{aligned}
H_0: \quad & \alpha_0 = \beta_0 = \lambda_0 \\
\text{and} \quad & \alpha_1 = \beta_1 = \lambda_1 \\
\text{and} \quad & \alpha_2 = \beta_2 = \lambda_2 \\
H_a: \quad & Otherwise
\end{aligned}
$$

Dummy Variable Technique

$$
H_{01}: \gamma_0 = 0 \quad H_{02}: \gamma_1 = 0 \quad H_{03}: \gamma_2 = 0
$$

$$
H_{a1}: \gamma_0 \neq 0 \quad H_{a2}: \gamma_1 \neq 0 \quad H_{a3}: \gamma_2 \neq 0
$$

<u>Distinction between Two Methods</u>

Chow test is to test whether all estimators are the same or not.

Dummy variables technique is to separately test whether each estimator is the same or not. Some estimators may be different while some may not be.

Interaction Effects Using Dummy Variables

Model with Intercept and Interaction Dummy Variables

$$Y_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{1t} D_{2t} + \beta X_t + u_t$$

where: $D_{1t} = 0$ male or $= 1$ female.
$D_{2t} = 0$ Thai or $= 1$ foreigner.

The coefficients of the dummy variables can be interpreted as:

α_1 = Differential effect of being a female.
α_2 = Differential effect of being a foreigner.
α_3 = Differential effect of being a female foreigner.

Violation of OLS Assumptions

1. Multicollinearity
2. Specification Error
3. Autocorrelation
4. Heteroscedasticity

Issues of Concern

1. Cause of the Problem
2. Properties of OLS in Presence of Problem
3. Detection of Problem
4. Remedy of the Problem

Multicollinearity

1. Cause of the Problem

- Linear correlation among independent variables – Perfect Multicollinearity.
- High linear correlation among independent variables – Imperfect (High or Serious) Multicollinearity.

Multicollinearity

2. Properties of OLS in Presence of Problem

Perfect Multicollinearity – Can't be estimated.

Imperfect (High or Serious) Multicollinearity

1. OLS estimated coefficients are BLUE but variances and covariances are large.
2. Wide confidence intervals of estimated regression coefficients.
3. t-ratio tends to be statistically insignificant.
4. R^2 can be very high.
5. OLS estimators and their standard errors can be sensitive to small changes in the data

Multicollinearity

3. Detection of the Problem

1. High R^2 but few significant t ratios.
2. High pair-wise correlations among regressors.
3. Examination of partial correlations.
4. Auxiliary Regressions.
5. Tolerance and Variance Inflation Factor (VIF).
6. Eigenvalues and Condition Index (CI).

Multicollinearity

4. Remedy of the Problem

1. Adding More Data.
2. Transformation of Variables.
3. Dropping Variable(s) and Specification Bias.
4. Other methods--e.g. Factor Analysis, Principal Component, or Ridge Regression.

Model Specification

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