Properties of Regression Coefficients and Hypothesis Testing

Parthasarathi Edupally[1]

[1]DataConscientious LLP, Mumbai

Elements of Econometrics,
Russell Square International College, Mumbai

Outline

Outline

Outline

Outline

Different Types of Data

- Cross-sectional Data
  - observations on individuals, households, enterprises etc at one moment in time
  - reasonable assumption of random sampling can be made

Different Types of Data

- Cross-sectional Data
  - observations on individuals, households, enterprises etc at one moment in time
  - reasonable assumption of random sampling can be made
- Time Series Data
  - Observations on income, consumption etc over number of periods in time (years, quarters, months...)
  - samples are correlated in time, they are not independent

Different Types of Data

- Cross-sectional Data
  - observations on individuals, households, enterprises etc at one moment in time
  - reasonable assumption of random sampling can be made
- Time Series Data
  - Observations on income, consumption etc over number of periods in time (years, quarters, months...)
  - samples are correlated in time, they are not independent
- Panel Data
  - Observations on one cross-section of individuals, enterprises etc over number of time periods
  - can assume random sampling across cross-section but not across time dimension

Types of Models

- Classical Linear Regression Model
  - regressors are assumed to be fixed or non-random, only error term is random
  - rules out the possibility of interaction between error and regressors

Types of Models

- Classical Linear Regression Model
  - regressors are assumed to be fixed or non-random, only error term is random
  - rules out the possibility of interaction between error and regressors
- Linear Regression Model
  - regressors are now assumed to be stochastic and random sampling assumption holds
  - can explicitly talk about interaction of error term with explainatory variables
  - it is easier to derive asymptotic properties of regression coefficients

Types of Models

- Classical Linear Regression Model
  - regressors are assumed to be fixed or non-random, only error term is random
  - rules out the possibility of interaction between error and regressors
- Linear Regression Model
  - regressors are now assumed to be stochastic and random sampling assumption holds
  - can explicitly talk about interaction of error term with explanatory variables
  - it is easier to derive asymptotic properties of regression coefficients
- Time Series Model
  - complex relationship between explanatory variables across time is modeled
  - many standard assumptions of linear regression model do not hold

Assumptions of CLRM

- Assumption.1
  - The model is correctly specified and is linear in parameters

$$Y = \beta_1 + \beta_2 X + u, \quad \text{It is correct model!}$$

Assumptions of CLRM

- Assumption.1
  - The model is correctly specified and is linear in parameters

$$Y = \beta_1 + \beta_2 X + u, \quad \text{It is correct model!}$$

- Assumption.2
  - Their is some variance in the regressors in the sample
  - We are not looking at constant values

Assumptions of CLRM

- Assumption.1
  - The model is correctly specified and is linear in parameters

$$Y = \beta_1 + \beta_2 X + u, \quad \text{It is correct model!}$$

- Assumption.2
  - Their is some variance in the regressors in the sample
  - We are not looking at constant values
- Assumption.3
  - The disturbance term has zero expectation
  - This is reasonable assumption when intercept term is used
  - It means inherent error in the model for generating Y is on an average zero

Assumptions of CLRM

- Assumption.4
    - The disturbance term is Homoscedastic
    - Error in generating Y doesnt depend on X value

Assumptions of CLRM

- Assumption.4
  - The disturbance term is Homoscedastic
  - Error in generating Y doesnt depend on X value
- Assumption.5
  - Disturbance terms have independent distributions
  - Error terms are independent across X values

Assumptions of CLRM

- Assumption.4
  - The disturbance term is Homoscedastic
  - Error in generating Y doesnt depend on X value
- Assumption.5
  - Disturbance terms have independent distributions
  - Error terms are independent across X values
- Assumption.6
  - Disturbance term has normal distribution
  - Effect of all factors other than X can be modeled as following normal distribution

Unbiased Coefficients

- We can see how assumptions of CLRM lead to unbiased regression coefficients

Unbiased Coefficients

- We can see how assumptions of CLRM lead to unbiased regression coefficients

$$\text{Fitted Model:} \quad \hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

$$\hat{\beta}_2 = \beta_2 + \frac{\sum (X_i - \overline{X}) u_i}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$$

$$E(\hat{\beta}_2) = \beta_2$$

- Also OLS estimator is efficient of all the estimators, though with same bias (zero)

Monte Carlo Experiment

- Monte Carlo experiment is generally used to derive distributions
- We can use same method to see that the error term is infact following normal distribution as specified in our assumption

Monte Carlo Experiment

- Monte Carlo experiment is generally used to derive distributions
- We can use same method to see that the error term is infact following normal distribution as specified in our assumption

Please refer Dougherty Slides on Monte Carlo experiment

Precision of Regression estimates

- OLS regression estimators being RVs have variance given by,

$$Var\hat{\beta}_2 = \sigma^2_{\hat{\beta}_2} = \frac{\sigma_u^2}{nMSD(X)}$$

$$\text{where} \quad MSD(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

- Clearly, variance of coefficient depends on following:

Precision of Regression estimates

- OLS regression estimators being RVs have variance given by,

$$Var\hat{\beta}_2 = \sigma^2_{\hat{\beta}_2} = \frac{\sigma_u^2}{nMSD(X)}$$

$$\text{where} \quad MSD(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

- Clearly, variance of coefficient depends on following:
  - MSD, mean square distance of x, which is a measure of dispersion of x values
  - Variance of error term itself, which is just inherent noise in the model

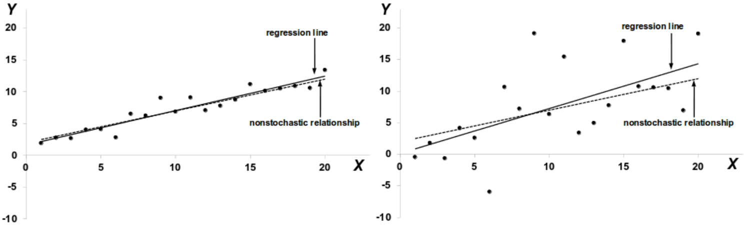
Precision of Regression estimates

- OLS regression estimators being RVs have variance given by,

$$Var\hat{\beta}_2 = \sigma^2_{\hat{\beta}_2} = \frac{\sigma_u^2}{nMSD(X)}$$

$$\text{where} \quad MSD(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

- Clearly, variance of coefficient depends on following:



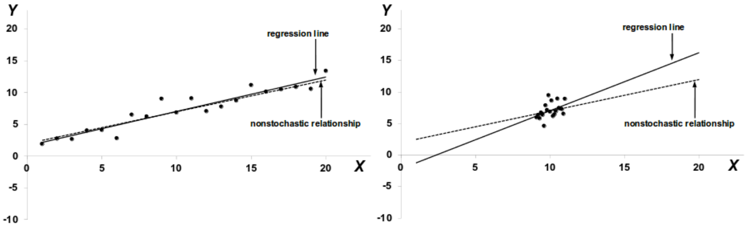
Precision of Regression estimates

- OLS regression estimators being RVs have variance given by,

$$Var\hat{\beta}_2 = \sigma^2_{\hat{\beta}_2} = \frac{\sigma_u^2}{nMSD(X)}$$

$$\text{where} \quad MSD(X) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}$$

- Clearly, variance of coefficient depends on following:



Standard Errors of Coefficients

- Variance of error term is not available,

Standard Errors of Coefficients

- Variance of error term is not available,
  - we can use residual errors as reasonable estimator for error variance

$$\hat{\sigma_u}^2 = \frac{\sum_{i=i}^{n} \hat{u}^2}{n-2}$$

Standard Errors of Coefficients

- Variance of error term is not available,
  - we can use residual errors as reasonable estimator for error variance

$$\hat{\sigma_u}^2 = \frac{\sum_{i=i}^{n} \hat{u}^2}{n-2}$$

- Thus standard error (s.e) of regression coefficients is given by,

$$s.e(\beta_1) = \hat{\sigma_u}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

Standard Errors of Coefficients

- Variance of error term is not available,
  - we can use residual errors as reasonable estimator for error variance

$$\hat{\sigma_u}^2 = \frac{\sum_{i=i}^{n} \hat{u}^2}{n-2}$$

- Thus standard error (s.e) of regression coefficients is given by,

$$s.e(\beta_1) = \hat{\sigma_u}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

$$s.e(\beta_2) = \sqrt{\frac{\hat{\sigma_u}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

Standard Errors of Coefficients

- Variance of error term is not available,
  - we can use residual errors as reasonable estimator for error variance

$$\hat{\sigma_u}^2 = \frac{\sum_{i=i}^{n} \hat{u}^2}{n-2}$$

- Thus standard error (s.e) of regression coefficients is given by,

$$s.e(\beta_1) = \hat{\sigma_u}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

$$s.e(\beta_2) = \sqrt{\frac{\hat{\sigma_u}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

- Lower the standard error, more precise the regression line approximation

Hypothesis Testing

Please use Dougherty Slides from Study Guide

Confidence Intervals

Please use Dougherty Slides from Study Guide

F tests for Goodness of Fit

Please use Dougherty Slides from Study Guide

"In God we trust, all others bring data."

William Edwards Deming (1900 - 1993).