

Multiple Regression Analysis: Problem of Estimation

1. Three-Variable Model: Notation and Assumptions
2. Interpretation of Multiple Regression Equation
3. Meaning of Partial Regression Coefficients
4. OLS and ML Estimation of Partial Regression Coefficients
5. Multiple Coefficient of Determination R^2
6. R^2 and Adjusted R^2

Three-Variable Model: Notation and Assumptions

In practice, the two-variable regression model is inappropriate.

The three-variable regression model:

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

Where Y_i is dependent variable

X_2 and X_3 are explanatory variable

β_2 and β_3 are called the partial regression coefficient

Three-Variable Model: Notation and Assumptions

Assumptions of the model:

Zero mean $E(u_i | X_{2i}, X_{3i}) = 0$ for each i

No autocorrelation $\text{cov}(u_i | u_j) = 0$ for $i \neq j$

Homoscedasticity $\text{var}(u_i) = \sigma^2$

Zero covariance $\text{cov}(u_i, X_{2i}) = \text{cov}(u_i, X_{3i}) = 0$

No Specification Bias

No Exact Linear Relationship between X_2 and X_3

Three-Variable Model: Notation and Assumptions

No Multicollinearity (or Linearly Independent)

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0 \text{ and } \lambda_2 = \lambda_3 = 0$$

Linear Dependent $\lambda_2 \neq 0$ or $\lambda_3 \neq 0$

$$\lambda_2 X_{2i} + 4X_{3i} = 0 \text{ or } X_{2i} = -4X_{3i}$$

Nonlinear Relationship will show *no exact linear relationships.*

$$\lambda_2 X_{3i} = X_{2i}^2$$

Three-Variable Model: Notation and Assumptions

Example of Multicollinearity

If $X_{3i} = 2X_{2i}$

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

Then
$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 (2X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 2\beta_3) X_{2i} + u_i \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned}$$

where $\alpha = (\beta_2 + 2\beta_3)$

Interpretation of Multiple Regression Equation

Conditional mean or expected value of Y conditional upon the given or fixed values of the variables X_2 and X_3

$$E(Y_i | X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

Multiple regression analysis reveals the average or mean value of Y or mean response of Y for the fixed valued of the X variables.

Meaning of Partial Regression Coefficients

Partial regression coefficient β_2 measures the change in the mean value of Y or $E(Y/X_2, X_3)$ with respect to X_2 , holding X_3 constant.

Partial regression coefficient β_3 measures the change in the mean value of Y or $E(Y/X_2, X_3)$ with respect to X_3 , holding X_2 constant.

In calculus meaning, β_2 and β_3 are partial derivatives of $E(Y/X_2, X_3)$ with respect to X_2 and X_3

OLS Estimation of Partial Regression Coefficients

OLS Estimators

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

OLS minimize the residual sum of squares

$$\min \sum \hat{u}_i^2 = \sum \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} \right)^2$$

OLS Estimation of Partial Regression Coefficients

To get minimization, take partial derivative of the residual sum of squares function with respect to β_1 , β_2 , and β_3 and set them equal to zero.

Normal equations:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2$$

OLS Estimation of Partial Regression Coefficients

Derive normal equations and get:

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

OLS Estimation of Partial Regression Coefficients

Variations & Standard Errors of OLS Estimators

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$$

$$\text{var}(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$$

$$\text{var}(\hat{\beta}_3) = \frac{\hat{\sigma}^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$$

OLS Estimation of Partial Regression Coefficients

Properties of OLS Estimators

1. In k -variable linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

$$\hat{\beta}_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \bar{X}_3 - \cdots - \beta_k \bar{X}_k$$

2. Mean value of the estimated Y_i is equal to the mean value of the actual Y_i

3. $\sum \hat{u}_i = \bar{\hat{u}} = 0$

4. $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = 0$

OLS Estimation of Partial Regression Coefficients

Properties of OLS Estimators

5. $\sum \hat{u}_i \hat{Y}_i = 0$

6. When $r_{23} = 1$ (i.e. perfect multicollinearity), the variances of estimators become infinite.

7. The greater the variation in the sample values of X_i , the smaller the variance of β

8. Given the assumptions of the classical linear regression model, the OLS estimators are BLUE.

Multiple Coefficient of Determination R^2

R^2 lies between 0 and 1

$$R^2 = \frac{ESS}{TSS} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

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.

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

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:

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

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

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

R^2 and Adjusted R^2

“Game” of Maximizing Adjusted R^2

Sometimes researchers play the game of maximizing *adjusted R^2* , that is, choosing the model that gives the highest *adjusted R^2* .

However, the objective of regression analysis is not to obtain the highest *adjusted R^2* model, but rather to obtain dependable estimates of the true population regression coefficients and draw statistical inferences about them.

Example

Cobb-Douglas Production Function

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

where

Y = Output

X_2 = Labor Input

X_3 = Capital Input

u = Stochastic disturbance term

e = Base of natural logarithm

Linearize the model by taking logarithm:

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

Polynomial Regression Models

Polynomial regression models have been extensively used in econometric research, such as, MC.

Second-degree polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

k^{th} -degree polynomial regression:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i$$

Multiple Regression Analysis: Problem of Inference

1. Normality Assumption
2. Hypothesis Testing about Individual Partial Regression Coefficients
3. Testing Overall Significance of Sample Regression
4. Testing Equality of Two Regression Coefficients
5. Restricted Least Squares
6. Comparing Two Regressions: Testing for Structural Stability of Regression Models

Normality Assumption

Since we also need to test hypothesis of the model, we have to make assumption of the distribution of the disturbances u_i .

$$u_i \sim N(0, \sigma^2)$$

With normality assumption, OLS estimators of partial regression coefficient are BLUE.

Normality Assumption

Estimators of partial regression coefficients are t -distributed with $n-3$ degree of freedom.

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)}$$

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

$$t = \frac{\hat{\beta}_3 - \beta_3}{se(\hat{\beta}_3)}$$

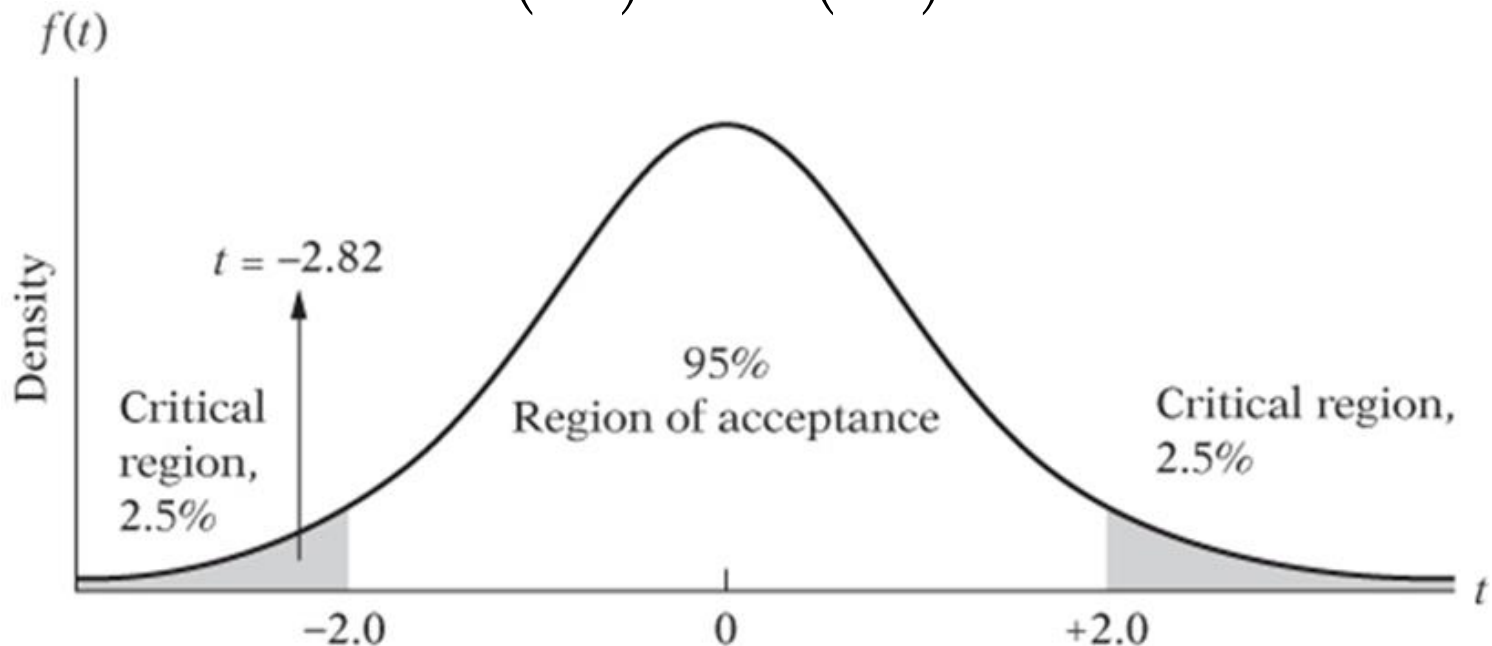
Hypothesis Testing in Multiple Regression

1. Hypothesis testing about individual partial regression coefficients.
2. Testing the overall significance of the sample regression.
3. Testing the equality of two regression coefficients.
4. Testing linear equality restriction.
5. Testing for structural stability of regression models.
6. Testing the functional form of regression.

Hypothesis Testing about Individual Partial Regression Coefficients

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

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



Testing Overall Significance of Sample Regression

$H_0: \beta_2 = \beta_3 = 0$ and $H_a: \text{Otherwise}$

$$F = \frac{ESS/df}{RSS/df}$$

Testing Equality of Two Regression Coefficients

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

$$H_0 : \beta_3 = -\beta_4 \text{ or } (\beta_3 + \beta_4) = 0$$

$$H_a : \beta_3 \neq -\beta_4 \text{ or } (\beta_3 + \beta_4) \neq 0$$

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{se(\hat{\beta}_3 - \hat{\beta}_4)}$$

$$se(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}$$

Testing Linear Equality Restrictions

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

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

Constant Return to Scale

$$H_0 : \beta_2 + \beta_3 = 1$$

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{se(\hat{\beta}_2 + \hat{\beta}_3)}$$

$$se(\hat{\beta}_2 + \hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2\text{cov}(\hat{\beta}_2, \hat{\beta}_3)}$$

Restricted Least Squares: F -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

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

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

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

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

Restricted Least Squares: F -test

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$$
$$= \frac{(R_{UR}^2 - R_R^2)/m}{(1 - R_{UR}^2)/(n-k)}$$

UR = Unrestricted regression

R = Restricted regression

m = Number of linear restrictions

k = Number of parameters in unrestricted regression

n = Number of observation

Comparing Two Regressions

Testing for Structural Stability of Regression Models

Whole Period $Y_t = \lambda_1 + \lambda_2 X_{2t} + \lambda_3 X_{3t} + u_{3t}$
(RSS = S_1) $t = 1, 2, \dots, n_1 + n_2$

Before Liberalization $Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_{1t}$
(RSS = S_2) $t = 1, 2, \dots, n_1$

After Liberalization $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_{2t}$
(RSS = S_3) $t = n_1 + 1, n_1 + 2, \dots, n_1 + n_2$

Comparing Two Regressions

Testing for Structural Stability of Regression Models

$$H_0 : \quad \alpha_1 = \beta_1 = \lambda_1$$

and $\alpha_2 = \beta_2 = \lambda_2$

and $\alpha_3 = \beta_3 = \lambda_3$

$$H_a : \quad \textit{Otherwise}$$

$$\textit{Chow Test} : \quad F = \frac{(S_1 - S_2 - S_3)/k}{(S_2 + S_3)/(n_1 + n_2 - 2k)}$$

Evaluating Estimated Results

1. Sign 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.

Dummy Variable Regression Models

1. Nature of Dummy Variables
2. ANOVA Models
3. Regression with a Mixture of Quantitative and Qualitative Regressors: ANCOVA Models
4. Dummy Variable Alternative to Chow Test
5. Interaction Effects Using Dummy Variables
6. Some Applications of Dummy Variables

Nature of Dummy Variables

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).

ANOVA Models

The model includes dependent variable and independent variables which are all dummy variables.

$$Y_t = \beta_1 + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \beta_5 D_{5t} + u_t$$

Caution!

In the case of 5 choices (Monday, Tuesday, Wednesday, Thursday, Friday), only 4 dummy variables can be included in the model with intercept term.

If we would like to use 5 dummy variables, the model must be model without intercept term.

Regression with a Mixture of Quantitative and Qualitative Regressors: ANCOVA Models

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

Regression with a Mixture of Quantitative and Qualitative Regressors: ANCOVA Models

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

Dummy Variable Alternative to Chow Test

Chow Test

$$\text{All} \quad Y_t = \lambda_0 + \lambda_1 X_{1t} + \lambda_2 X_{2t} + u_t \quad \text{for } t = 1, 2, \dots, n_1 + n_2$$

$$\text{Before} \quad Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + u_{1t} \quad \text{for } t = 1, 2, \dots, n_1$$

$$\text{After} \quad Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_{2t} \quad \text{for } t = n_1 + 1, n_1 + 2, \dots, n_1 + n_2$$

Dummy Variable Technique

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

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

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

Dummy Variable Alternative to Chow Test

Chow Test

$$H_0: \alpha_0 = \beta_0 = \lambda_0$$

$$\text{and } \alpha_1 = \beta_1 = \lambda_1$$

$$\text{and } \alpha_2 = \beta_2 = \lambda_2$$

$$H_a: \textit{Otherwise}$$

Dummy Variable Technique

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

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

Distinction between Two Methods

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.

Some Applications of Dummy Variables

- Seasonal Analysis
- Spline Regression Model
- Fixed-effect Panel Data Model
- Switching Regression Model
- Disequilibrium Model