Financial Econometrics

Hang Zhou

School of Finance, SUFE

Lecture 3: Multivariate Linear Regression

Roadmap

- In this lecture, we will advance our study to multivariate linear regressions.
- We will use **matrix** notations to write regression model and derive the results.
- First, we introduce the motivation for doing multivariate regressions.

Outline

Omitted Variable Bias

Multivariate Linear Regression

Asymptotic Distribution and Hypothesis Testing

Model Selection

A theoretical motivation

- Suppose that we are interested in the relation between X_1 and Y, say market expected return and individual stock return.
- CAPM relation tells us

$$E(r_i) - r_f = \beta(E(r_m) - r_f)$$

A theoretical motivation

- Suppose that we are interested in the relation between X_1 and Y, say market expected return and individual stock return.
- CAPM relation tells us

$$E(r_i) - r_f = \beta(E(r_m) - r_f)$$

- However, Fama French shows that other factors may also affect stock return. (SMB and HML)

A theoretical motivation

- Suppose that we are interested in the relation between X_1 and Y, say market expected return and individual stock return.
- CAPM relation tells us

$$E(r_i) - r_f = \beta(E(r_m) - r_f)$$

- However, Fama French shows that other factors may also affect stock return. (SMB and HML)
- If we run regression without the two factors, will there be any problem?

A numerical example

- Consider the following population model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i$$

where x_1 and x_2 are 2 variables that both have explanation power to y.

A numerical example

- Consider the following population model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i$$

where x_1 and x_2 are 2 variables that both have explanation power to y.

? What would happen if we **omit** x_2?

Analysis

- Suppose that we ignore the variable x_2, we have

$$y_i = \beta_0 + \beta_1 x_{1,i} + v_i$$

where $v_i = \beta_2 x_{2,i} + u_i$.

Analysis

- Suppose that we ignore the variable x_2, we have

$$y_i = \beta_0 + \beta_1 x_{1,i} + v_i$$

where $v_i = \beta_2 x_{2,i} + u_i$.

- We regress y on x_1 and get the expression of $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\sum(x_{1,i} - \bar{x}_1)(y_i - \bar{y})}{\sum(x_{1,i} - \bar{x}_1)^2}$$

- Plug in the expression for y, we have

$$\hat{\beta}_1 = \beta_1 + \frac{\sum(x_{1,i} - \bar{x}_1)(v_i - \bar{v})}{\sum(x_{1,i} - \bar{x}_1)^2}$$

Omitted Variable Bias

- If the second term is not 0 in expectation, we have a bias due to $E(\hat{\beta}_1) \neq \beta_1$.
- We can proceed by plug in the expression for v_i.

Omitted Variable Bias

- If the second term is not 0 in expectation, we have a bias due to $E(\hat{\beta}_1) \neq \beta_1$.
- We can proceed by plug in the expression for v_i.

$$\hat{\beta}_1 - \beta_1 = \beta_2 \frac{\sum (x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sum (x_{1,i} - \bar{x}_1)^2} + \frac{\sum (x_{1,i} - \bar{x}_1)(u_i - \bar{u})}{\sum (x_{1,i} - \bar{x}_1)^2}$$

- Assuming u_i is the true error term that satisfies **exogeneity** assumption, our bias boils down to

Omitted Variable Bias

- If the second term is not 0 in expectation, we have a bias due to $E(\hat{\beta}_1) \neq \beta_1$.
- We can proceed by plug in the expression for v_i.

$$\hat{\beta}_1 - \beta_1 = \beta_2 \frac{\sum(x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sum(x_{1,i} - \bar{x}_1)^2} + \frac{\sum(x_{1,i} - \bar{x}_1)(u_i - \bar{u})}{\sum(x_{1,i} - \bar{x}_1)^2}$$

- Assuming u_i is the true error term that satisfies **exogeneity** assumption, our bias boils down to

$$\text{Bias} = \underbrace{\beta_2}_{\text{True parameter of } x_2 \text{ on y.}} \times \underbrace{\frac{\sum(x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sum(x_{1,i} - \bar{x}_1)^2}}_{\text{Regression coefficient of } x_2 \text{ on } x_1.}$$

- What are the sufficient conditions for having a non-zero bias?

Omitted Variable Bias

- The sufficient conditions are
  1. $\beta_2 \neq 0$.
  2. $Cov(x_1, x_2) \neq 0$.

Omitted Variable Bias

- The sufficient conditions are
  1. $\beta_2 \neq 0$.
  2. $Cov(x_1, x_2) \neq 0$.
- The bias is positive (upward bias) if and only if

$$\text{sgn}(\beta_2) = \text{sgn}(Cov(x_1, x_2)),$$

  negative (downward bias) if and only if

$$\text{sgn}(\beta_2) \neq \text{sgn}(Cov(x_1, x_2)).$$

- Let's take a look at a numerical example.

Numerical example

- Consider the following true model

$$y_i = 1.5 + 2 \times x_{1,i} + 3 \times x_{2,i} + u_i$$

- And we run regression

$$y_i = \beta_0 + \beta_1 x_{1,i} + v_i$$

- We set the correlation coefficient $Corr(x_1, x_2) = 0.7$.

Numerical example

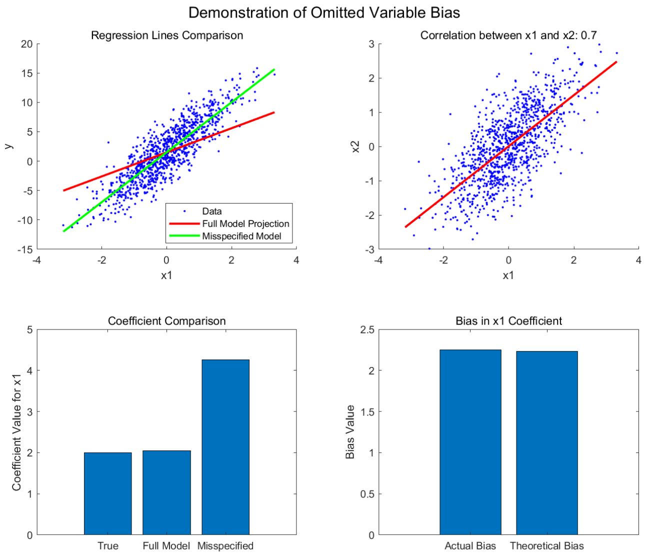
- Consider the following true model

$$y_i = 1.5 + 2 \times x_{1,i} + 3 \times x_{2,i} + u_i$$

- And we run regression

$$y_i = \beta_0 + \beta_1 x_{1,i} + v_i$$

- We set the correlation coefficient $Corr(x_1, x_2) = 0.7$.
- In theory, we shall have an upward bias.



Demonstration of Omitted Variable Bias
Regression Lines Comparison
Data
Full Model Projection
Misspecified Model
y
x1
Correlation between x1 and x2: 0.7
x2
x1
Coefficient Comparison
Coefficient Value for x1
True
Full Model
Misspecified
Bias in x1 Coefficient
Bias Value
Actual Bias
Theoretical Bias

Back to CAPM

- In theory, if we assume the true asset pricing model has more factors than market risk, we always have **misspecified** model.
- Portfolio choices based on CAPM will be misleading.
- How should we fix this?

Back to CAPM

- In theory, if we assume the true asset pricing model has more factors than market risk, we always have **misspecified** model.
- Portfolio choices based on CAPM will be misleading.
- How should we fix this?

* We can use **multivariate linear regression** to include more factors in the model.

Outline

Omitted Variable Bias

Multivariate Linear Regression

Asymptotic Distribution and Hypothesis Testing

Model Selection

Introduction

- Having more than one X variable does not complicate things too much. Suppose that we have m factors to explain the return of a stock.

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} \cdots + \beta_m x_{m,i} + u_i$$

- We can still perform OLS to get estimates for the $\boldsymbol{\beta}$s.
- We call each factor $\boldsymbol{x_m}$ a **regressor** in the regression model.
- It is more convenient to write in **matrix** form.

Matrix notations

- We make the following definition.
- $\mathbf{1} = (1, 1, ..., 1) \in R^n$ is the n-dim row vector of ones.
- Let $\boldsymbol{x}_m = (x_{1,m}, x_{2,m}, \cdots x_{n,m}) \in R^n$ to denote the m-th regressor.
- Let $\boldsymbol{u} = (u_1, u_2, \cdots, u_n) \in R^n$ to denote the error terms.

Matrix notations

- We make the following definition.
- $\mathbf{1} = (1, 1, ..., 1) \in R^n$ is the n-dim row vector of ones.
- Let $\boldsymbol{x}_m = (x_{1,m}, x_{2,m}, \cdots x_{n,m}) \in R^n$ to denote the m-th regressor.
- Let $\boldsymbol{u} = (u_1, u_2, \cdots, u_n) \in R^n$ to denote the error terms.
- Stack the coefficient β and y as a column vector.

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} \in R^{m+1}; \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in R^n$$

Matrix notations

- Stack the regressors $\boldsymbol{x}_1, ...\boldsymbol{x}_m$ as $\mathbf{X}$, we have

$$\mathbf{X} = [1', x_1', x_2', \cdots x_m']$$

which is equivalent to

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix}_{n\times(m+1)}$$

Matrix notations

- Stack the regressors $\boldsymbol{x}_1, ...\boldsymbol{x}_m$ as $\mathbf{X}$, we have

$$\mathbf{X} = [1', x_1', x_2', \cdots x_m']$$

which is equivalent to

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix}_{n\times(m+1)}$$

- In matrix notation, we can write the population model as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}.$$

The dimension is *n by 1; n by (m+1); (m+1) by 1; n by 1*, respectively.

Number of Regressors and Observations

- In multivariate regression, we have m regressors and n observations.
- Usually, we have $n \gg m + 1$.

Number of Regressors and Observations

- In multivariate regression, we have m regressors and n observations.
- Usually, we have $n \gg m + 1$.
- If m is close to n, we call this **high-dimensional** regression, which requires **machine learning** techniques.

Number of Regressors and Observations

- In multivariate regression, we have m regressors and n observations.
- Usually, we have $n \gg m + 1$.
- If m is close to n, we call this **high-dimensional** regression, which requires **machine learning** techniques.
- In high-dimensional regression, OLS will fail because $\mathbf{X}'\mathbf{X}$ is not invertible. (There exists infinitely many solutions to $\mathbf{X}\beta = \mathbf{y}$)

Number of Regressors and Observations

- In multivariate regression, we have m regressors and n observations.
- Usually, we have $n \gg m + 1$.
- If m is close to n, we call this **high-dimensional** regression, which requires **machine learning** techniques.
- In high-dimensional regression, OLS will fail because $\mathbf{X}'\mathbf{X}$ is not invertible. (There exists infinitely many solutions to $\mathbf{X}\beta = \mathbf{y}$)
- We need to use other techniques like Lasso or Ridge regression.

OLS in matrix form

- In OLS, we minimizes the residual sum square:

$$\beta \in \arg\min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

OLS in matrix form

- In OLS, we minimizes the residual sum square:

$$\beta \in \arg\min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

- The first order condition in matrix form is

$$D_{\beta} RSS = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta = 0$$

- Which solves the OLS estimator $\hat{\beta}$ as

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

with one assumption that $\mathbf{X}'\mathbf{X}$ is invertible.

Assumptions

- As we can see, the invertibility of $\mathbf{X}'\mathbf{X}$ is critical for the OLS estimator. This leads to our new assumption.

Assumptions

- As we can see, the invertibility of $\mathbf{X}'\mathbf{X}$ is critical for the OLS estimator. This leads to our new assumption.

6 **No perfect collinearity**: the data set of regressors $\mathbf{X}$ has full rank.

Assumptions

- As we can see, the invertibility of $\mathbf{X}'\mathbf{X}$ is critical for the OLS estimator. This leads to our new assumption.

6 **No perfect collinearity**: the data set of regressors $\mathbf{X}$ has full rank. Normally we have perfect collinearity because:

* Two or more regressors are linear functions of each other.

Assumptions

- As we can see, the invertibility of $\mathbf{X}'\mathbf{X}$ is critical for the OLS estimator. This leads to our new assumption.

6 **No perfect collinearity**: the data set of regressors $\mathbf{X}$ has full rank. Normally we have perfect collinearity because:

* Two or more regressors are linear functions of each other. For instance, if your data on asset pricing is

$$\mathbf{X} = (\mathbf{1}, R_m, SMB, 2 * SMB, HML),$$

there is a clear collinearity for SMB and $2 * SMB$.

Assumptions

- As we can see, the invertibility of $\mathbf{X}'\mathbf{X}$ is critical for the OLS estimator. This leads to our new assumption.

6 **No perfect collinearity**: the data set of regressors $\mathbf{X}$ has full rank. Normally we have perfect collinearity because:

* Two or more regressors are linear functions of each other. For instance, if your data on asset pricing is

$$\mathbf{X} = (\mathbf{1}, R_m, SMB, 2 * SMB, HML),$$

there is a clear collinearity for SMB and $2 * SMB$.

* Rank($\mathbf{X}$) = Rank($\mathbf{X}'\mathbf{X}$) = m+1.

* Two or more observations for different firms have the exact same value. For instance, if you mistakenly input the data for the same firm twice, you will have collinearity.

Interpretation of the coefficients

- Through multivariate regression, we get a vector of estimator $(\hat{\beta}_0, \hat{\beta}_1, ...\hat{\beta}_m)$.
- Each $\hat{\beta}_s$ represents the partial effect of x_s on y.

$$\Delta y = \beta_s \Delta x_s$$

* Holding others constant, if x_s increase by 1 unit, y will respond by β_s unit.

Logarithm Transformation

- Our previous model

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots \beta_m x_{i,m} + u_i$$

studies the effect of x_s on y.

- For instance, if y is stock price and x_s is the total amount of debt. In this case, β_s is interpreted as once the firm's debt increase by 1 dollar, its value responds by β_s dollar.

Logarithm Transformation

- Our previous model

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots \beta_m x_{i,m} + u_i$$

  studies the effect of x_s on y.

- For instance, if y is stock price and x_s is the total amount of debt. In this case, β_s is interpreted as once the firm's debt increase by 1 dollar, its value responds by β_s dollar.
- However, sometimes we care about the **percentage change**, not the actual value.

Logarithm Transformation

- Our previous model

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots \beta_m x_{i,m} + u_i$$

  studies the effect of x_s on y.

- For instance, if y is stock price and x_s is the total amount of debt. In this case, β_s is interpreted as once the firm's debt increase by 1 dollar, its value responds by β_s dollar.

- However, sometimes we care about the **percentage change**, not the actual value.

- In this case, we do the logarithmic transformation of the data.

$$\ln(y_i) = \tilde{\beta}_0 + \tilde{\beta}_1 \ln(x_{i,1}) + \tilde{\beta}_2 \ln(x_{i,2}) + \cdots + \tilde{\beta}_m \ln(x_{i,m}) + \tilde{u}_i$$

Logarithm Transformation and Elasticity

- Consider the log-transformed regression, we have

$$\ln(y_i) = \tilde{\beta}_0 + \tilde{\beta}_1 \ln(x_{i,1}) + \tilde{\beta}_2 \ln(x_{i,2}) + \cdots + \tilde{\beta}_m \ln(x_{i,m}) + \tilde{u}_i.$$

- Take $\tilde{\beta}_1$ as an example. We have

$$\tilde{\beta}_1 = \frac{\partial \ln(y)}{\partial \ln(x_1)} = \frac{dy/y}{dx/x} = \text{Elasticity of x on y}$$

- $\tilde{\beta}_1$ is interpreted as: **if x increase by 1 percentage**, then **y will respond by $\tilde{\beta}_1$ percentage.**

Demo of log-log transformation

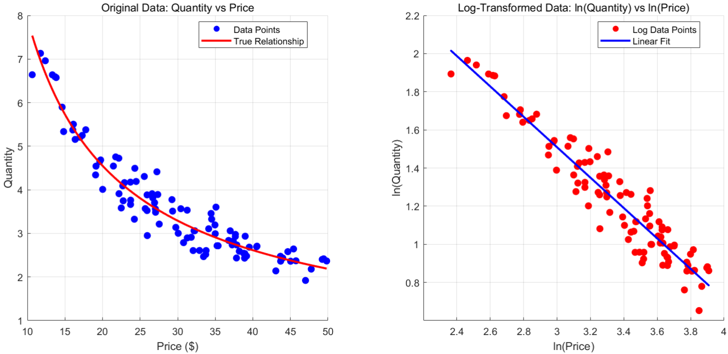


Figure 1: Price-Quantity Relation

Other log transformations

- Depending on the data and the underlying economic model, you could do the following transformation:
- $\log(y) =$ linear x beta
- $y = \log(x)$ beta
- $\log(y) = \log(x)$ beta

Other log transformations

- Depending on the data and the underlying economic model, you could do the following transformation:
- log(y) = linear x beta
- y = log(x) beta
- log(y) = log(x) beta

* These transformation are still linear in β. We can still do OLS. We will discuss non-linear functions of β later.

Goodness of fit revised

- We introduce the R^2 as the measure of goodness of fit.
- However, with more regressors, we need to update our definition because: adding more regressors will not decrease the model fit.

$$\hat{\beta} \in \arg\min \sum u_i^2$$

Goodness of fit revised

- We introduce the R^2 as the measure of goodness of fit.
- However, with more regressors, we need to update our definition because: adding more regressors will not decrease the model fit.

$$\hat{\beta} \in \arg\min \sum u_i^2$$

- In a multivariate regression, we define the standard error of the regression as

$$SER = \underbrace{\frac{1}{n-m-1}}_{\text{Degree of freedom adjustment}} \sum_{i=1}^{n} \hat{u}_i^2$$

The adjusted R^2

- We modified the definition of R^2 as

$$\bar{R}^2 = 1 - \frac{RSS/_{n-m-1}}{TSS/_{n-1}}$$

The adjusted R^2

- We modified the definition of R^2 as

$$\bar{R}^2 = 1 - \frac{RSS/_{n-m-1}}{TSS/_{n-1}}$$

- Intuitively, to estimate RSS, we need to impose $m+1$ equality in the first order conditions.
- However, to get TSS, we simply regress y on a constant $\mathbf{1}$ and this consumes 1 df.

The adjusted R^2

- We modified the definition of R^2 as

$$\bar{R}^2 = 1 - \frac{RSS/_{n-m-1}}{TSS/_{n-1}}$$

- Intuitively, to estimate RSS, we need to impose $m+1$ equality in the first order conditions.
- However, to get TSS, we simply regress y on a constant $\mathbf{1}$ and this consumes 1 df.

* $\bar{R}^2$ is not non-decreasing w.r.t the number of regressors m.

Outline

Omitted Variable Bias

Multivariate Linear Regression

Asymptotic Distribution and Hypothesis Testing

Model Selection

Asymptotic distribution

- Recall that in matrix form we have

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

Asymptotic distribution

- Recall that in matrix form we have

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \boldsymbol{u})$$

  for the population model.

- Simplify the equation:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}$$

Asymptotic distribution

- Recall that in matrix form we have

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \boldsymbol{u})$$

  for the population model.

- Simplify the equation:

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}$$

- Assuming that all regressors are **exogenous**, we shall have

$$E(\hat{\beta} - \beta \mid \boldsymbol{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{u} \mid \boldsymbol{X}) = 0$$

  which shows the unbiasness.

- To compute the variance, we know that $E(\hat{\beta} - \beta) = 0$. Its variance-covariance matrix is simply

$$\begin{aligned}
\Sigma_{\hat{\beta}} &= E((\hat{\beta} - \beta)(\hat{\beta} - \beta)') \\
&= E\left[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u})'\right] \\
&= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]
\end{aligned}$$

- The key insight is that $\mathbf{X}'\mathbf{X}$ is symmetric so is its inverse.

- To compute the variance, we know that $E(\hat{\beta}-\beta)=0$. Its variance-covariance matrix is simply

$$
\begin{aligned}
\Sigma_{\hat{\beta}} &= E((\hat{\beta}-\beta)(\hat{\beta}-\beta)') \\
&= E\left[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u})((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u})'\right] \\
&= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right]
\end{aligned}
$$

- The key insight is that $\mathbf{X}'\mathbf{X}$ is symmetric so is its inverse.
- The formula is called the sandwich form.

Further analysis

- The key in the variance-covariance matrix is the variance-covariance of $\boldsymbol{u}$.

$$\boldsymbol{u}\boldsymbol{u}' = \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots u_1 u_n \\ u_1 u_2 & u_2^2 & \cdots u_2 u_n \\ \vdots & \vdots & \vdots \\ u_1 u_n & u_2 u_n & \cdots u_n^2 \end{bmatrix}$$

Further analysis

- The key in the variance-covariance matrix is the variance-covariance of $\boldsymbol{u}$.

$$\boldsymbol{u}\boldsymbol{u}' = \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots u_1 u_n \\ u_1 u_2 & u_2^2 & \cdots u_2 u_n \\ \vdots & \vdots & \vdots \\ u_1 u_n & u_2 u_n & \cdots u_n^2 \end{bmatrix}$$

- Assume that i.i.d sample, the error term across different observations should be 0.

Further analysis

- The key in the variance-covariance matrix is the variance-covariance of $\boldsymbol{u}$.

$$\boldsymbol{u}\boldsymbol{u}' = \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots u_1 u_n \\ u_1 u_2 & u_2^2 & \cdots u_2 u_n \\ \vdots & \vdots & \vdots \\ u_1 u_n & u_2 u_n & \cdots u_n^2 \end{bmatrix}$$

- Assume that i.i.d sample, the error term across different observations should be 0. Imposing this and taking expectation, we simplify the matrix to

$$E(\boldsymbol{u}\boldsymbol{u}' \mid \mathbf{X}) = \begin{bmatrix} \sigma_{u_1}^2 & 0 & \cdots 0 \\ 0 & \sigma_{u_2}^2 & \cdots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & \cdots \sigma_{u_n}^2 \end{bmatrix}$$

Homoscedasticity

- Previously, we assume homoscedastic variance that

$$Var(u_i \mid X) = \sigma_u^2, \;\; \forall i$$

- If this is still true, we simplify the matrix as

$$E(\boldsymbol{uu}' \mid \mathbf{X}) = \sigma_u^2 \cdot \mathrm{I}$$

Homoscedasticity

- Previously, we assume homoscedastic variance that

$$Var(u_i \mid X) = \sigma_u^2, \ \ \forall i$$

- If this is still true, we simplify the matrix as

$$E(\boldsymbol{uu}' \mid \mathbf{X}) = \sigma_u^2 \cdot \mathrm{I}$$

- Then, our sandwich form reduces to

$$\Sigma_{\hat{\beta}} = \sigma_u^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- To estimate the $\Sigma_{\hat{\beta}}$, we simple use

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n - m - 1}$$

Heteroscedasticity

- However, if you data appears to be heteroscedastic, i.e,

$$Var(u_i) = f(\boldsymbol{x}_i) \neq \text{a constant.}$$

Heterocedasticity

- However, if you data appears to be heteroscedastic, i.e,

$$Var(u_i) = f(\boldsymbol{x}_i) \neq \text{a constant.}$$

Then we call this heteroscedasticity.

- For HSK robust standard error estimation, we use the following sandwich formula

$$\widehat{\Sigma_{\hat{\beta}}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n} \hat{u}_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- This is the White robust variance estimator. (Named after Halbert White)

Visualization of HSK

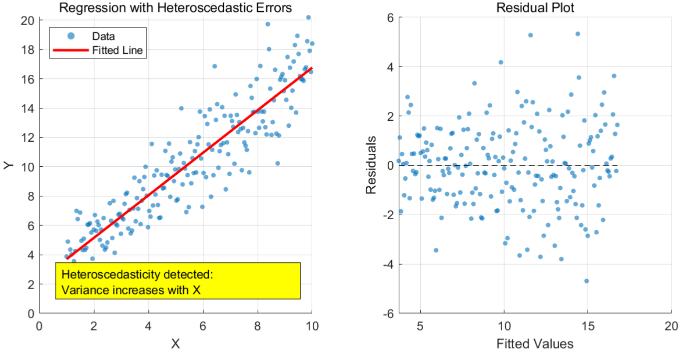


Figure 2: illustration of HSK

HSK and HMK variance estimator

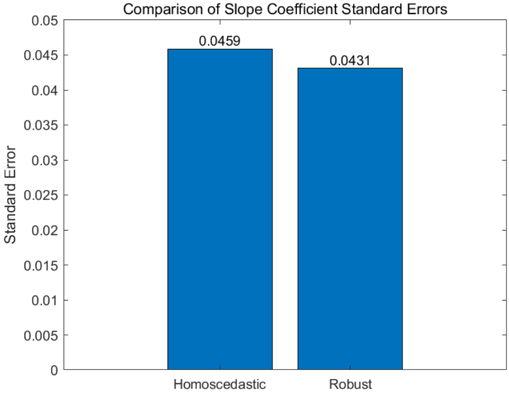


Figure 3: Comparison of two variance estimator

Hypothesis testing

- Given the assumption: (1) zero conditional mean; (2) i.i.d sample; (3) No outliers; (4) No perfect collinearity,

Hypothesis testing

- Given the assumption: (1) zero conditional mean; (2) i.i.d sample; (3) No outliers; (4) No perfect collinearity, we can show that $\hat{\beta} \xrightarrow{p} \hat{\beta}$ and $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\beta})$.
- Because $\hat{\beta}$ is jointly normal distributed for large sample, we can test multiple constraints.
- Usually, for multiple linear constraints, we do F test.

Testing for Joint Significance

- Suppose that we've run multivariate regression and get the estimated coefficients $\hat{\beta}$.
- We are interested in the following test

$$
\begin{aligned}
H_0 &: \quad \beta_1 = \beta_2 = \cdots = \beta_m = 0 \\
H_a &: \quad \text{Not } H_0
\end{aligned}
$$

- We do the F test.

Testing for Joint Significance

- Suppose that we've run multivariate regression and get the estimated coefficients $\hat{\beta}$.
- We are interested in the following test

$$
\begin{aligned}
H_0 : &\quad \beta_1 = \beta_2 = \cdots = \beta_m = 0 \\
H_a : &\quad \text{Not } H_0
\end{aligned}
$$

- We do the F test. However, it only works for HMK.

Testing Procedure

1. Run the unrestricted model that

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... \beta_m x_{i,m} + u_i$$

to get the unrestricted RSS, denote as RSS_u.

2. Run the restricted regression

$$y_i = \beta_0 + v_i$$

to get the restricted RSS, denote as RSS_r.

- If the H_0 is true, then we should have

$$F = \frac{(RSS_r - RSS_u)/m}{RSS_u/n-m-1} \sim F_{m,n-m-1}$$

- Using R^2 to represent the test statistic, we have

$$F = \frac{(R_u^2 - R_r^2)/m}{(1 - R_u^2)/n-m-1} \sim F_{m,n-m-1}$$

which is equivalent to the previous result.

- In general, if you have some null hypothesis involves linear constraints, you can always do F test.
- For example

$$\begin{aligned} H_0 : &\quad \beta_1 = \beta_2 = \beta_3 \\ H_a : &\quad \text{Not } H_0 \end{aligned}$$

- There are 2 " $=$ " in the null hypothesis. So the number of constraints is 2.
- The testing statistic is

$$F = \frac{(RSS_r - RSS_u)/k}{RSS_u/n-k-1} \sim F_{k,n-k-1}$$

where k is the number of " $=$ " in the H_0.

F-test Demo

- Consider the following **data generating process**:

$$y = 1.00 + 2.00 * x1 + 1.50 * x2 + u$$

- We can do F test based on the simulated data.

F-test Demo Results

- Here are the results from our regression analysis:

Parameter	Estimate	Std. Error	t-value	p-value
Intercept	0.8327	0.2341	3.5563	0.0006
β_1	1.9233	0.3293	5.8404	0.0000
β_2	1.8584	0.2487	7.4736	0.0000

Table 1: Full Model Regression Results

F-test Results

- **F-test for Joint Significance of Slope Coefficients:**
- $H_0 : \beta_1 = \beta_2 = 0$ (Restricted model: $y = \beta_0 + v$)
- H_1 : At least one slope coefficient is non-zero (Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$)

F-test Results

- **F-test for Joint Significance of Slope Coefficients:**
- $H_0 : \beta_1 = \beta_2 = 0$ (Restricted model: $y = \beta_0 + v$)
- H_1 : At least one slope coefficient is non-zero (Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$)
- **Results:**

$$\begin{aligned} \text{F-statistic} &= 77.2445 \\ \text{Critical F-value (95\%)} &= 3.0902 \\ \text{p-value} &= 0.000000 \\ R^2 &= 0.6221 \\ \text{Adjusted } R^2 &= 0.6143 \end{aligned}$$

- **Conclusion:** We reject H_0 at the 5% significance level. At least one of the slope coefficients is statistically significant. The model as a whole is statistically significant.

Interpretation of Results

- All coefficients are statistically significant at the 1% level:
  - The intercept ($\beta_0 = 0.8327$) is significant with $p = 0.0006$
  - $\beta_1 = 1.9233$ is highly significant with $p < 0.0001$
  - $\beta_2 = 1.8584$ is highly significant with $p < 0.0001$
- The F-test strongly rejects the null hypothesis that both slope coefficients are zero:
  - F-statistic (77.2445) far exceeds the critical value (3.0902)
  - Extremely small p-value ($p < 0.000001$)
- The model explains approximately 62.21% of the variation in the dependent variable ($R^2 = 0.6221$)
- The $\bar{R}^2$ (0.6143) remains high, indicating that the explanatory power is not artificially inflated by the number of predictors

Graphic Illustration

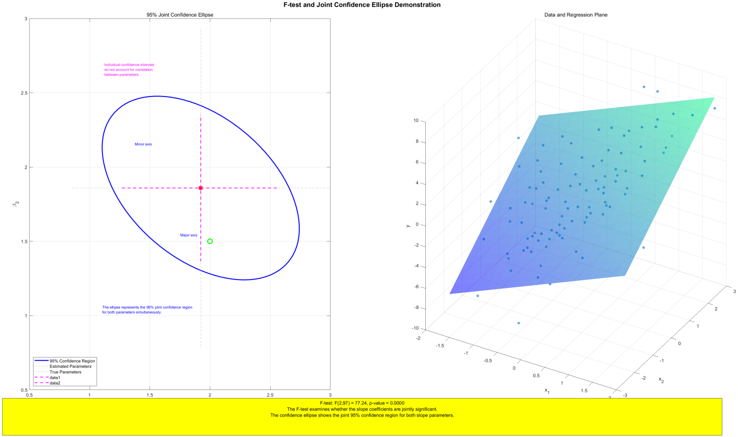


Figure 4: 95% Confidence Sets

Outline

Omitted Variable Bias

Multivariate Linear Regression

Asymptotic Distribution and Hypothesis Testing

Model Selection

Model Selection

- We are in the era of big data. Essentially, you might have large datasets containing numerous X variables.
- How should we select the correct model?

Model Selection

- We are in the era of big data. Essentially, you might have large datasets containing numerous X variables.
- How should we select the correct model?
- Do not simply rely on R^2 or $\bar{R}^2$!
- Use your economic intuition!

Control Variables

- In multi-variate regression models, you may still have **omitted variable bias** if your error term is correlated with some of the X variables.
- We introduce the idea of control variable.

Control Variable

A control variable Z is not the variable of interest, i.e., you don't care about the coefficients of Z.

Control Variables

- In multi-variate regression models, you may still have **omitted variable bias** if your error term is correlated with some of the X variables.
- We introduce the idea of control variable.

Control Variable

A control variable Z is not the variable of interest, i.e., you don't care about the coefficients of Z. However, you need the control variable for zero-conditional mean to hold. If

$$E(u \mid X, Z) = E(u \mid Z) \quad \text{(Conditional Mean Independence)},$$

then after adding Z to our regression, it seems that the variable of interest X is randomly assign to individuals.

Control Variables

- In multi-variate regression models, you may still have **omitted variable bias** if your error term is correlated with some of the X variables.
- We introduce the idea of control variable.

Control Variable

A control variable Z is not the variable of interest, i.e., you don't care about the coefficients of Z. However, you need the control variable for zero-conditional mean to hold. If

$$E(u \mid X, Z) = E(u \mid Z) \quad \text{(Conditional Mean Independence)},$$

then after adding Z to our regression, it seems that the variable of interest X is randomly assign to individuals.

- After controlling for Z, our estimate of $\hat{\beta}_X$ is unbiased and consistent.

Role of R^2 and $\bar{R}^2$

Key Role:

- R^2 and $\bar{R}^2$ indicate how well regressors predict the dependent variable.
- Close to 1: Good prediction (small OLS residual variance relative to dependent variable variance).
- Near 0: Poor prediction power.

Role of R^2 and $\bar{R}^2$

Key Role:

- R^2 and $\bar{R}^2$ indicate how well regressors predict the dependent variable.
- Close to 1: Good prediction (small OLS residual variance relative to dependent variable variance).
- Near 0: Poor prediction power.

Limitations (What They Don't Show):

- Statistical significance of included variables.
- Regressors being true causes of dependent variable movements.
- Presence of omitted variable bias.
- Whether the regressor set is most appropriate.

Summary

- In this lecture, we introduce multivariate linear regressions.
- We derive the formula for multivariate OLS model and its asymptotic distributions.
- We also study how to perform joint test and model selection based on R^2 and $\bar{R}^2$.