Quantitative Analysis of Financial Markets

Session 4: Cross Sectional Estimation Frameworks and CAPM

Benjamin Ee

Week 4

SMU

SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

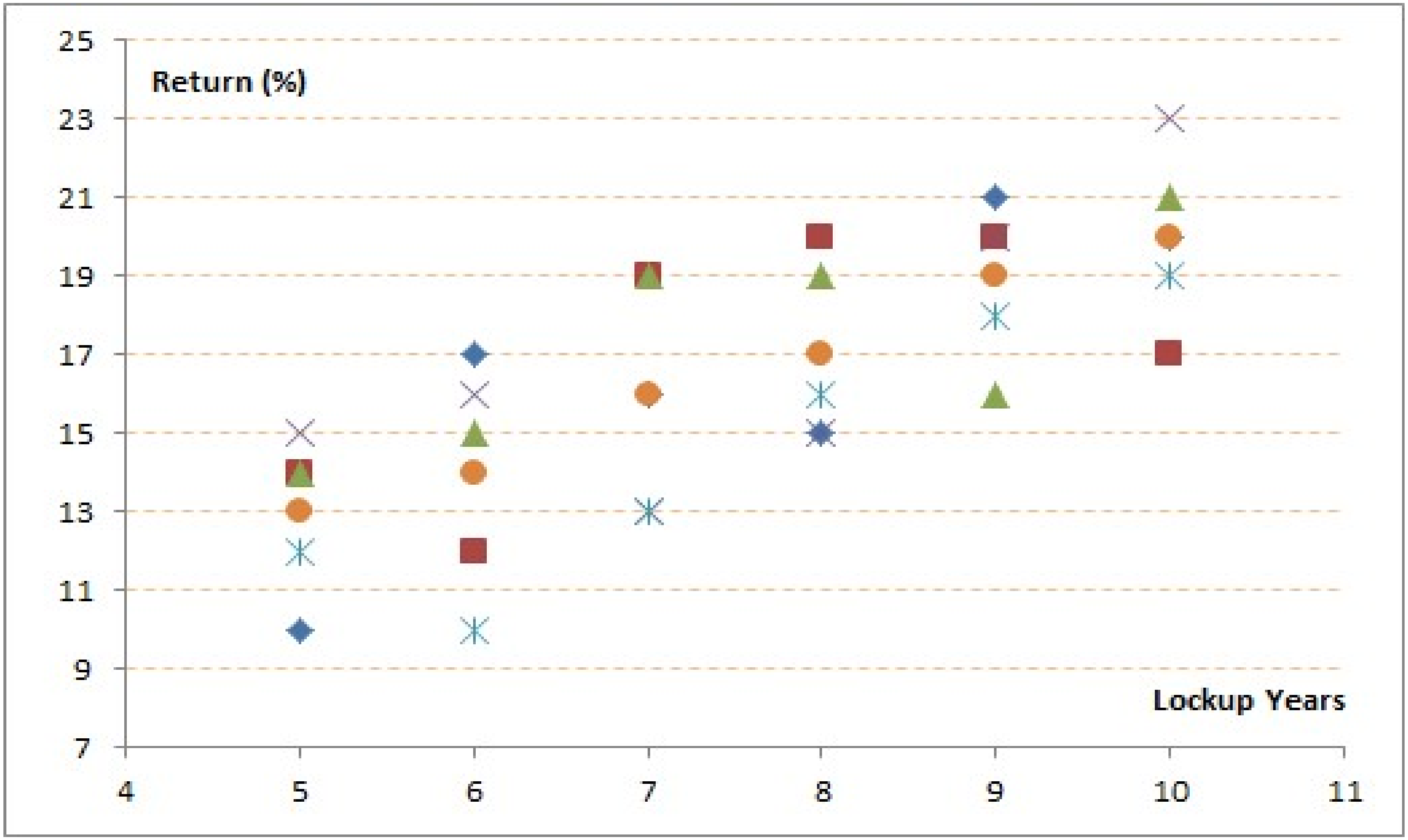
Week 4

Pre-class Prep 1/10

Today

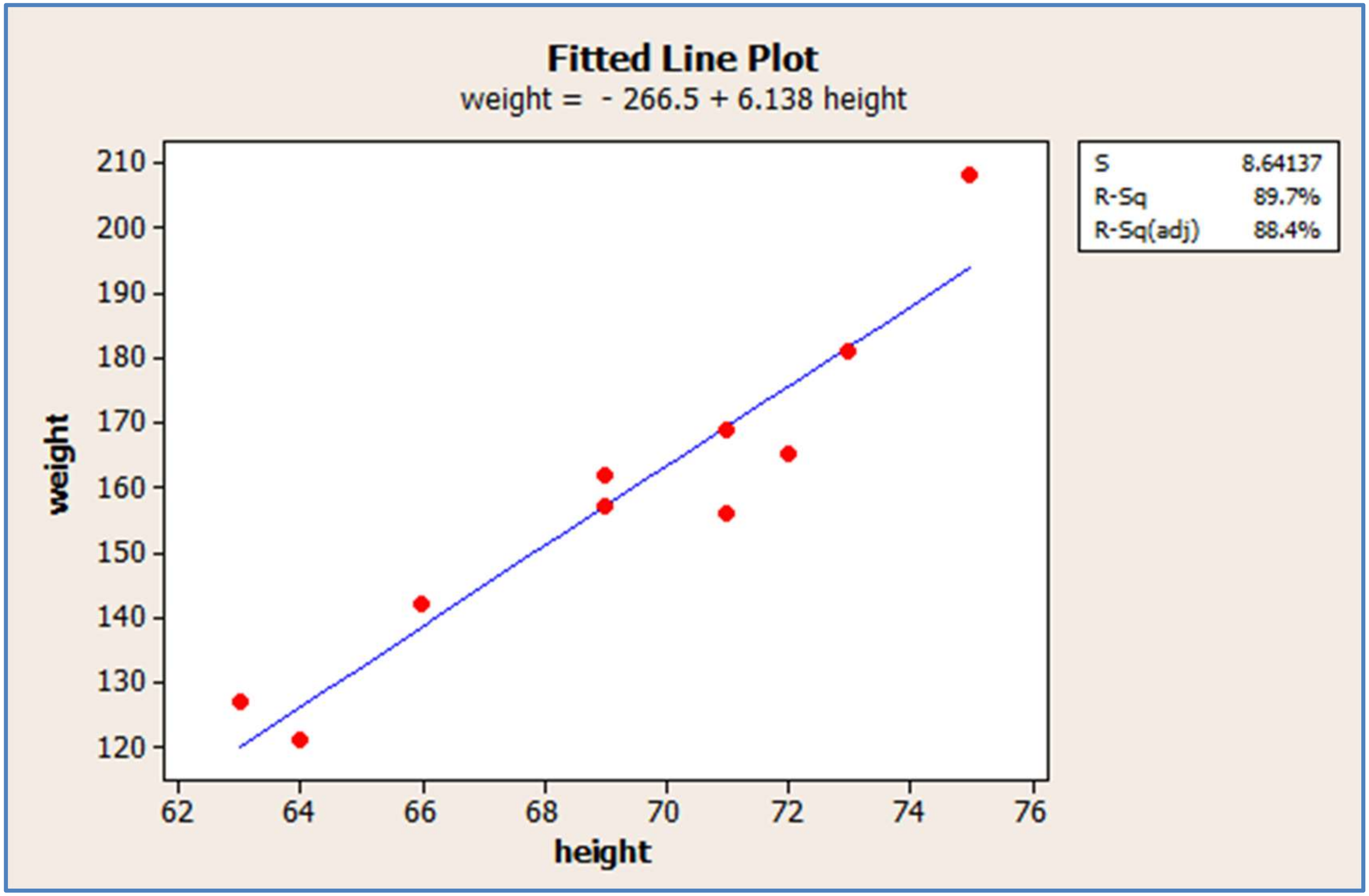
1. **Ordinary Least Squares Estimators (OLS)**
2. **Capital Asset Pricing Model (CAPM)**
3. **Tests of CAPM**

Recap: Scatter Plots can highlight linear relationships that we can consider modelling with Ordinary Least Squares (OLS)



- We hypothesize a positive relationship between hedge fund returns and the lockup period. How can we quantify this positive relationship?

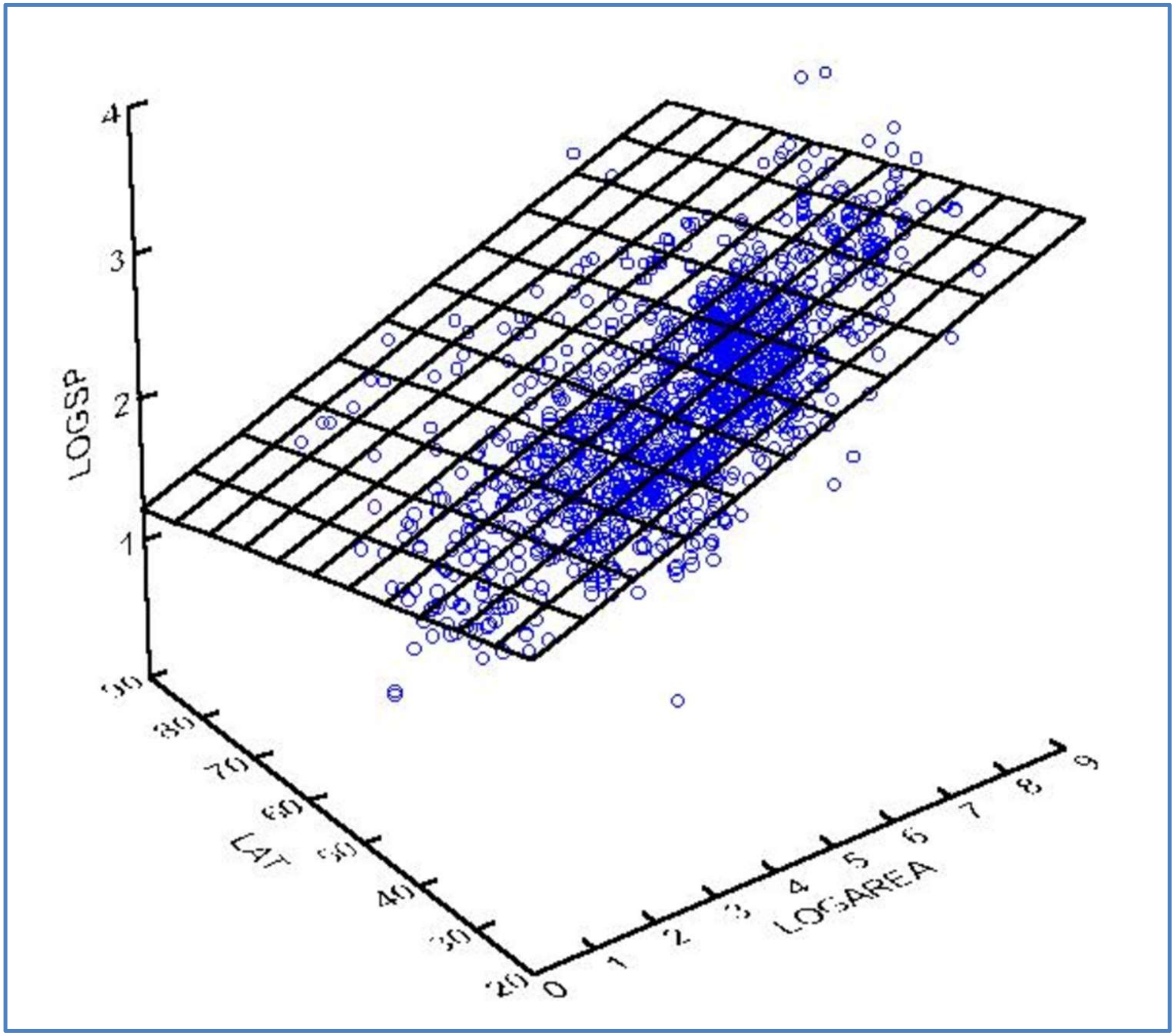
With these assumptions, **we generate a linear equation for the variable of interest in terms of predictors**



Source: https://online.stat.psu.edu/stat415/book/export/html/880

- A 'linear equation' is fully defined by a y-intercept and a gradient. How do we quantify these?

We can also extend this to multiple predictor variables



$$Y_i = a + bX_1 + cX_2 + e_i$$

http://ordination.okstate.edu/MULTIPLE.htm

Recap: OLS Assumptions

- Given n pairs of observations on explanatory variable X_i and dependent variable Y_i, we can have **a linear model** postulating that

$$Y_i = a + bX_i + e_i, \qquad i = 1, 2, \ldots, n,$$

where e_i is the noise.

- Assumptions:

(A1) $\mathbb{E}(e_i) = 0$ for every i

(A2) $\mathbb{E}(e_i^2) = \sigma_e^2$

(A3) $\mathbb{E}(e_i\, e_j) = 0$ for every i, j

(A4) X_i, e_j are independent for each i, j

(A5) $e_i \overset{d}{\sim} N(0, \sigma_e^2)$

We will reference specific assumptions at different stages of our derivation

SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

Week 4

Pre-class Prep 2/10

Goal: Solving for OLS ‘coefficients’

1. In the general case, we may have n observations, **each observation is (i) variable that we are modelling, and (ii) multiple predictor (independent) variables**

$$Y_i = a + X_i b + e_i \,, \qquad i = 1, 2, \ldots, n,$$

2. Here, Y, the variable we are modelling, is **n rows by 1 column matrix (or vector)**

3. ‘a’ is scalar (single number). We call this **the ‘constant’**

4. X is a **n rows by m columns matrix**
    a. 1 row for each observation, **each row is composed of m data points**

    b. For instance, a single observation at **t = 31-12-2021** may have 3 datapoints: **{income = 100, revenue = 200, debt = 10}**

    c. In this example, if we have observations across 10 time periods, **then X is a 10 rows by 3 columns matrix**

5. ‘b’ is **m rows by 1 column** coefficient vector

Example of OLS coefficients setup

$$\begin{bmatrix} 1 \\ 3 \\ 9 \\ -10 \\ 20 \end{bmatrix} = a + \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \\ 9 & 10 \end{bmatrix} \begin{matrix} b_1 \\ b_2 \end{matrix} + e_i$$

Dependent variable

Independent (predictor) variables

Goal: Solving for OLS ‘coefficients’

1. To solve for OLS ‘coefficients’ **means to look for numerical values of the constant ‘a’ and the coefficient vector ‘b’ in terms of the raw data**

$$\begin{bmatrix} 1 \\ 3 \\ 9 \\ -10 \\ 20 \end{bmatrix} = a + \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix} b + e_i$$

Simplified case: only 1 predictor/independent variable

Optimization goal: minimize sum of squared errors

1. Lets say that we have estimated specific values of a and b [**using simplified 1 independent variable case for now**]

2. For each independent variable obs X_i in input data set, we now have a 'fitted' value, which is
   a. $\mathbf{Y_{\_cap} = a + X_i b}$

3. Generally, there will be a **difference between Y_i (actual value of Y) and $Y_{\_cap}$ (fitted value)**

4. Difference is called the **residual, e_i**

5. We will look at deriving equations for the constant a and slope coefficient b in terms of the data values, assuming that our objective is to minimize the **sum of squared residuals**

First-Order Conditions of Least Squares

- Least Squares: **Minimizing the sum of squared errors**:

$$\min_{\widehat{a}\,,\,\widehat{b}} \sum_{i=1}^{n} e_i^2$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \widehat{a}} = -2\sum_{i=1}^{n} \left(Y_i - \widehat{a} - \widehat{b}X_i\right) = 0$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \widehat{b}} = -2\sum_{i=1}^{n} X_i\left(Y_i - \widehat{a} - \widehat{b}X_i\right) = 0$$

Do not need to memorize

- Similar to differentiating "x"

First-Order Conditions of Least Squares

- Least Squares: **Minimizing the sum of squared errors**:

$$\min_{\widehat{a},\widehat{b}} \sum_{i=1}^{n} e$$

$$\min_{\widehat{a},\widehat{b}} \sum_{i=1}^{n} e_i^2$$

$+ e_i$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \widehat{a}} = -2\sum_{i=1}^{n}\left(Y_i - \widehat{a} - \widehat{b}X_i\right) = 0$$

$e_1^2 + e_2^2$

$= (1\text{-}a\text{-}b)$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \widehat{b}} = -2\sum_{i=1}^{n} X_i\left(Y_i - \widehat{a} - \widehat{b}X_i\right) = 0$$

$– 9b)^2$

To chose a to minimize:

$f(a,b) = (1\text{-}a\text{-}b)^2 + (3 – a – 3b)^2 + (9\text{-}a\text{-}5b)^2 + (\text{-}10 – a\text{-}5b)^2 + (20 - a – 9b)^2$

Set partial derivative of f(x) w.r.t a to 0

Ordinary Least Squares Solutions

- Solution of first FOC

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \widehat{a} + \sum_{i=1}^{n} \widehat{b} X_i$$

$$\Longrightarrow \quad n\overline{Y} = n\widehat{a} + n\widehat{b}\overline{X}$$

$$\Longrightarrow \quad \overline{Y} = \widehat{a} + \widehat{b}\overline{X}$$

$$\Longrightarrow \quad \widehat{a} = \overline{Y} - \widehat{b}\overline{X}$$

Do not need to memorize

- Solution of second FOC

$$\sum_{i=1}^{n} X_i Y_i = \sum_{i=1}^{n} X_i \widehat{a} + \sum_{i=1}^{n} \widehat{b} X_i^2$$

$$\Longrightarrow \quad \sum_{i=1}^{n} X_i Y_i = \sum_{i=1}^{n} X_i \widehat{a} + \widehat{b} \sum_{i=1}^{n} X_i^2$$

$$\Longrightarrow \quad \sum_{i=1}^{n} X_i Y_i = \sum_{i=1}^{n} X_i (\overline{Y} - \widehat{b}\overline{X}) + \widehat{b} \sum_{i=1}^{n} X_i^2$$

$$\Longrightarrow \quad \sum_{i=1}^{n} X_i (Y_i - \overline{Y}) = \widehat{b} \sum_{i=1}^{n} X_i (X_i - \overline{X})$$

$$\Longrightarrow \quad \widehat{b} = \frac{\sum_{i=1}^{n} X_i (Y_i - \overline{Y})}{\sum_{i=1}^{n} X_i (X_i - \overline{X})}$$

SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

Week 4

Pre-class Prep 3/10

How do we know that our original objective function is appropriate?

1. So far, starting with objective of minimizing sum of squared errors, we have:

2. **Derived expressions for intercept and slope estimators** in terms of the ‘data’

3. **Todo**: How do we know our **original goal of minimizing sum squared errors is appropriate**?

4. **Todo**: Show that **estimators are unbiased**

5. **Todo**: Evaluate quality of estimates by **hypotheses testing and other diagnostics**

Gauss-Markov Theorem

- Gauss-Markov Theorem states that among all linear and unbiased estimators, the OLS estimators $(\widehat{a})$ and $(\widehat{b})$ have the minimum variances, i.e., $\mathbb{V}(\widehat{a})$ and $\mathbb{V}(\widehat{b})$ are the smallest possible and thus the OLS estimators are efficient (**estimation efficiency**).

- OLS estimators under the classical conditions are BLUE, i.e., Best Linear Unbiased Estimators for the linear regression model:

We derived expressions for this before

$$Y_i = a + bX_i + e_i, \qquad i = 1, 2, \ldots, n,$$

which can be written in the vector-matrix form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}, \; \boldsymbol{Y} := \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \; \boldsymbol{X} := \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}, \; \boldsymbol{\beta} := \begin{pmatrix} a \\ b \end{pmatrix}, \; \boldsymbol{e} := \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Simple OLS Estimators in Vector-Matrix Form

- Multiply from the left the matrix $\boldsymbol{X}'$ to both sides of $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$ to obtain

$$X'y = X'X\widehat{\beta} + X'e.$$

- By the classical assumption (A4), $\boldsymbol{X}'\boldsymbol{e} = \mathbf{0}$.
- Suppose $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists.

This is merely another way of expressing 'a' and 'b' in terms of the data; **an alternative to the equations from Clip 2 but means the same thing. Do not need to memorize**

- Multiply $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ to both sides to obtain

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

which is

$$\widehat{\beta} = (X'X)^{-1}X'y.$$

OLS Estimators Are Unbiased

Proposition **1**

Given the data matrix X, estimator is unbiased.

- Proof:

$$\begin{aligned}
\widehat{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\beta + \boldsymbol{e}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e} \\
&= \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e} \qquad (1)
\end{aligned}$$

- It follows that

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}}\left(\widehat{\beta}\right) &= \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}_{\boldsymbol{X}}(e) \\
&= \beta
\end{aligned}$$

Variance of $\widehat{\boldsymbol{\beta}}$

Proposition 2

The variance-covariance matrix of the OLS estimator is

$$\mathbb{V}_{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2(\boldsymbol{X}\boldsymbol{X}')^{-1}.$$

Proof: First we note from (1) that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$.

$$\begin{aligned}
\mathbb{V}_{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}\right) &= \mathbb{E}_{\boldsymbol{X}}((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})') \\
&= \mathbb{E}_{\boldsymbol{X}}\Big(((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{e})((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{e})'\Big) \\
&= \mathbb{E}_{\boldsymbol{X}}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}\boldsymbol{e}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathbb{E}_{\boldsymbol{X}}(\boldsymbol{e}\boldsymbol{e}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}$$

Proof of Gauss-Markov Theorem

* Note that $\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is a linear combination of $\boldsymbol{y}$.
* Let $\tilde{\boldsymbol{\beta}} = \boldsymbol{C}\boldsymbol{y}$ be another linear estimator of $\boldsymbol{\beta}$ with

$$\boldsymbol{C} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{D},$$

where $\boldsymbol{D}$ is a $2 \times n$ non-zero matrix.

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{X}}\left(\widetilde{\beta}\right) &= \mathbb{E}_{\boldsymbol{X}}(\boldsymbol{C}\boldsymbol{y}) \\
&= \mathbb{E}_{\boldsymbol{X}}\Big(\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{D}\right)(\boldsymbol{X}\beta + \boldsymbol{e})\Big) \\
&= \left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{D}\right)\boldsymbol{X}\beta + \left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{D}\right)\mathbb{E}_{\boldsymbol{X}}(\boldsymbol{e}) \\
&= \left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{D}\right)\boldsymbol{X}\beta \qquad \because \mathbb{E}_{\boldsymbol{X}}(\boldsymbol{e}) = \boldsymbol{0} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta + \boldsymbol{D}\boldsymbol{X}\beta \\
&= \beta + \boldsymbol{D}\boldsymbol{X}\beta.
\end{aligned}$$

Proof of Gauss-Markov Theorem (cont'd)

- Therefore, is unbiased if and only if ***DX*** =**0**. Then

$$
\begin{aligned}
\mathbb{V}_{\boldsymbol{X}}\left(\widetilde{\beta}\right) &= \mathbb{V}_{\boldsymbol{X}}(\boldsymbol{C}\boldsymbol{y}) = \boldsymbol{C}\,\mathbb{V}_{\boldsymbol{X}}(\boldsymbol{y})\boldsymbol{C}' = \sigma_e^2\boldsymbol{C}\boldsymbol{C}' \\
&= \sigma_e^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{D}\right)\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{D}'\right) \\
&= \sigma_e^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right. \\
&\qquad \left. + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{D}\boldsymbol{D}'\right) \\
&= \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{D}\boldsymbol{X})' + \sigma_e^2\boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma_e^2\boldsymbol{D}\boldsymbol{D}' \\
&= \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma_e^2\boldsymbol{D}\boldsymbol{D}' \qquad\qquad \because \boldsymbol{D}\boldsymbol{X} = \boldsymbol{0} \\
&= \mathbb{V}_{\boldsymbol{X}}\left(\widehat{\beta}\right) + \sigma_e^2\boldsymbol{D}\boldsymbol{D}' \qquad\qquad \because \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \mathbb{V}_{\boldsymbol{X}}\left(\widehat{\beta}\right)
\end{aligned}
$$

- Since $\boldsymbol{D}\boldsymbol{D}'$ is a positive semidefinite matrix, $\mathbb{V}_{\boldsymbol{X}}\left(\widetilde{\beta}\right)$ exceeds $\mathbb{V}_{\boldsymbol{X}}\left(\widehat{\beta}\right)$.

(here ends proof of GM theorem)

Recap [1/3]

1. In this clip, we:

2. Showed that OLS estimators are unbiased

3. Derived an expression for the variance of the OLS estimators which we can use for hypotheses testing

4. Showed that among all possible unbiased and linear estimators, OLS estimators have the lowest variance

Recap [2/3]

1. Also, in Proposition 2, we showed $\mathbb{V}_{\boldsymbol{X}}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$

2. Recall also that the coefficient vector is a 2 by 1 vector composed of the y-intercept and slope coefficient. Hence, equation above gives variance of both y-intercept and slope coefficient

3. Expanding this matrix (stating without derivation):

   a) Var $\widehat{a} = \sigma_e^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2} \right)$

   b) Var $\widehat{b} = \sigma_e^2 \left(\frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2} \right)$

Do not need to memorize

Recap [3/3]

1. What is the distribution of each of these estimators?

a) $$\text{Var}\ \widehat{a} = \boxed{\sigma_e^2}\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$$

b) $$\text{Var}\ \widehat{b} = \sigma_e^2\left(\frac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$$

Both highlighted expressions are constants!

Do not need to memorize

In practice, this is unobservable, so we approximate with sample variance estimator, which will be a sum of squared normal variables. Overall, this will be chi-squared distributed.

What is the degrees of freedom of the chi-square variable? It will be N-2, because there are 2 parameters, which are a and b

SMU

SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

Week 4

Pre-class Prep 4/10

Extension to multiple independent variables

1. In our derivation of OLS estimators [whether in equation or matrix form], we have so far worked with **simple case of just 1 independent variable**

2. In this clip, we will work to extend OLS framework to **multiple independent variables**

3. As number of variables becomes larger, it becomes increasingly **more convenient to use matrix notation form for discussion**

4. Therefore, using matrix notation form, multiple linear regression has the same form, **except *X matrix* now has more than two columns**

5. We assume **all columns in X matrix are linearly independent**

Sum of Squared Errors (SSE) and OLS

1. In model $y = X\beta + \boldsymbol{\varepsilon}$, OLS-estimate $\hat{\beta}$ minimizes:

$$\mathrm{SSE} = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta.$$

2. $X'X$ is a square symmetric matrix. Hence, first-order condition is:

$$\frac{\partial \mathrm{SSE}}{\partial \beta'} = -2y'X - 2\hat{\beta}'X'X = \mathbf{0}.$$

3. Applying the transpose operation, we obtain $X'X\hat{\beta} = Xy$

4. Due to our assumption that all columns of X are linearly independent, $X'X$ has an inverse and one can premultiply both sides by $(X'X)^{-1}$ to obtain $\hat{\beta} = (X'X)^{-1}Xy$.

P.S: It does not matter how many columns there are in X

OLS Algorithm for Model M

1. Estimate the model by OLS: $\hat{\beta}$= $(X'X)^{-1}Xy$
2. Compute the fitted values of y : $\widehat{\mathbf{y}}\,\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$

3.Compute the residuals or "surprise" : $\widehat{\boldsymbol{u}} = y - \widehat{\mathbf{y}}$

4. Previously with 1 independent variable (and a constant term), variance of estimators are chi-squared with n-2 degrees of freedom

$$\text{Var} \quad \widehat{a} = \sigma_e^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2} \right)$$

$$\text{Var} \quad \widehat{b} = \sigma_e^2 \left(\frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2} \right)$$

5. In the case of K variables (including constant term), it would be **chi-squared with n - K** degrees of freedom

$$\widehat{\sigma}_u^2 = \frac{1}{n-K} \widehat{\boldsymbol{u}}' \widehat{\boldsymbol{u}}$$

Summary of OLS Regression

1. The estimates are unbiased, i.e., $\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$.

2. Efficiency: According to the Gauss-Markov theorem (whether we have multiple independent variables or not), among classical linear regression models, OLS estimator is the linear unbiased estimator of $\boldsymbol{\beta}$ with the minimum variance

3. As derivation of Gauss-Markov theorem relied on matrix expression of least squares estimators, it is indifferent to the number of columns in the X matrix (i.e. number of independent variables does not matter, Gauss-Markov still holds)

SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

Week 4

Pre-class Prep 5/10

How do we evaluate ‘quality’ of the OLS estimation?

We have:

1. **Derived expressions for coefficient estimates** (i.e. constant and slope(s)) in terms of the data. Using matrix notation, we can use the same expressions regardless of number of independent variables

2. Shown **coefficient estimates are unbiased**

3. **Derived expressions for variance of coefficient estimates**, **and their distributions**

4. Shown that OLS estimators have **lowest variance of all unbiased linear estimators**

5. Closing out our discussion of OLS, **how can evaluate ‘quality’ of our OLS estimates**?

Decomposition

Do not need to memorize

- Consider

$$\widehat{Y}_i = \widehat{a} + \widehat{b}\,X_i$$
$$\widehat{e}_i = Y_i - \widehat{a} - \widehat{b}\,X_i = Y_i - \widehat{Y}_i$$

- **TSS = ESS + RSS**

$$\underbrace{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}_{\text{Total Sum of Squares}} = \underbrace{\sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2}_{\text{Explained Sum of Squares}} + \underbrace{\sum_{i=1}^{n} \widehat{e}_i^{\;2}}_{\text{Residual Sum of Squares}}$$

- ESS can be expressed as

$$\text{ESS} = \sum_{i=1}^{n} \left(\widehat{a} + \widehat{b}\,X_i - \widehat{a} - \widehat{b}\,\overline{X}\right)^2 = \widehat{b}^{\,2} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

Coefficient of Determination, R^2

1. Coefficient of determination R^2 is defined as

$$R^2 := \frac{\text{ESS}}{\text{TSS}}$$

"% of variation explained"

2. One feature of R^2 is it will always increase as we add more independent variables

3. For instance, if we have N observations (total) and we wanted to use n independent variables, we can 'explain' the data perfectly.

4. R^2 in this situation would be 100%

5. However, that is not a good measure of how the model will perform out of sample for actual prediction

Adjusted R^2

1. We can correct for flaws in R^2 **by using adjusted R^2** instead.

2. Adjusted R^2 **penalizes additional independent variables which do not explain 'enough' of variation in dependent variable**

3. **Adjusted R^2** = $1 - \frac{(1-R^2)(n-1)}{(n-p)}$

4. For the above formula, **p is number of variables in the model**, n is sample size, and R^2 follows definition from previous slide

Do not need to memorize

Hypothesis Testing (2 variable example)

- Series of residuals

$$\widehat{e}_i = Y_i - \widehat{a} - \widehat{b}\, X_i\,, \quad i = 1, 2, \ldots, n$$

- Unbiased estimator of residual variance

$$\widehat{\sigma}_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} \widehat{e}_i^{\,2}$$

- Testing null hypothesis $H_0: \; b = \; \beta$ (e.g. $\beta = 0$)

$$t_{n-2} = \frac{\widehat{b} - \beta}{\widehat{\sigma}_e \sqrt{\frac{1}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}}}$$

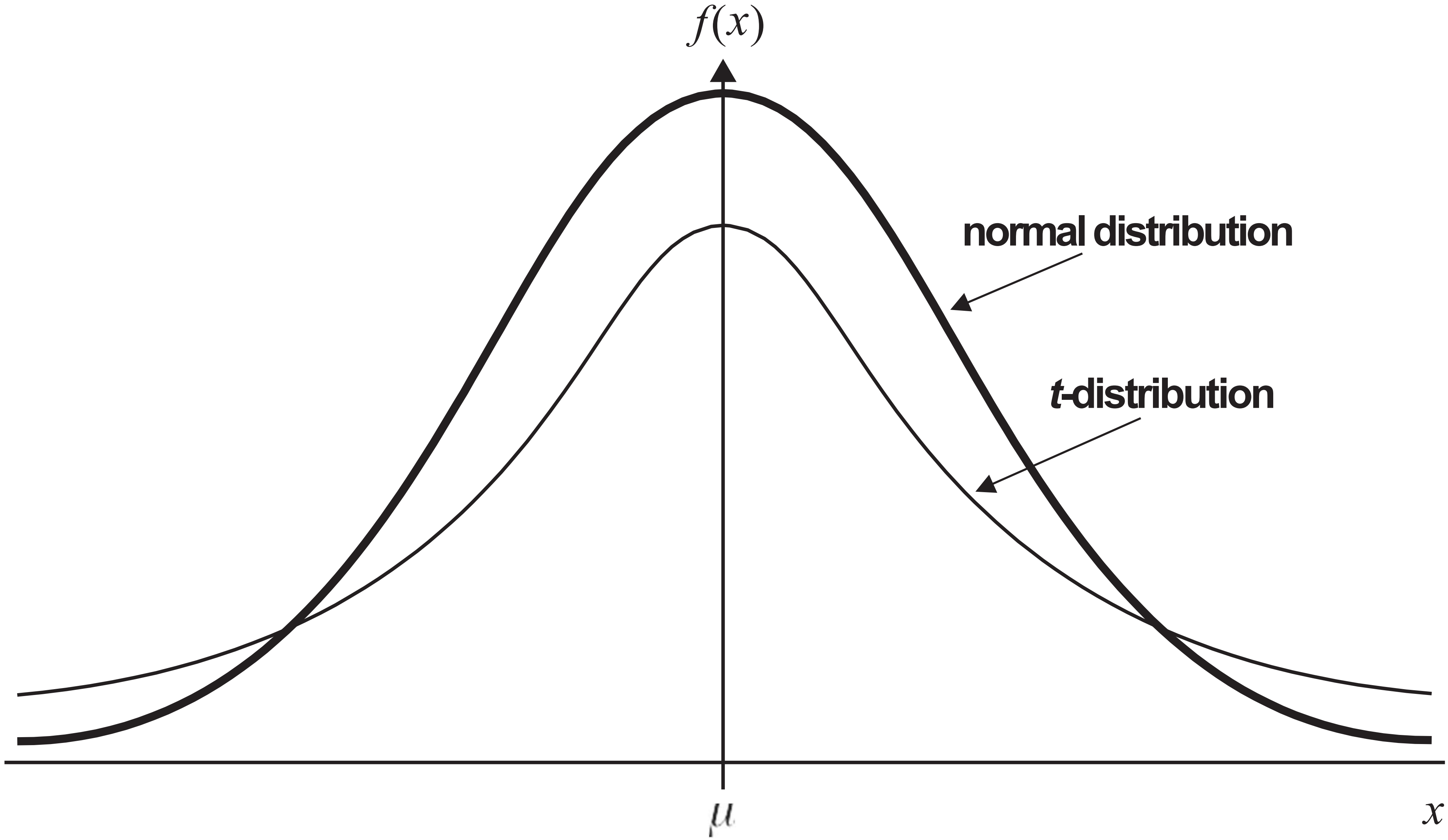
- Testing null hypothesis $H_0: \; a = \; \alpha$ (e.g. $\alpha = 0$)

$$t_{n-2} = \frac{\widehat{a} - \alpha}{\widehat{\sigma}_e \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}}}$$

Normal / Chi-sq

= t-distributed

Recap: *t*-Distribution



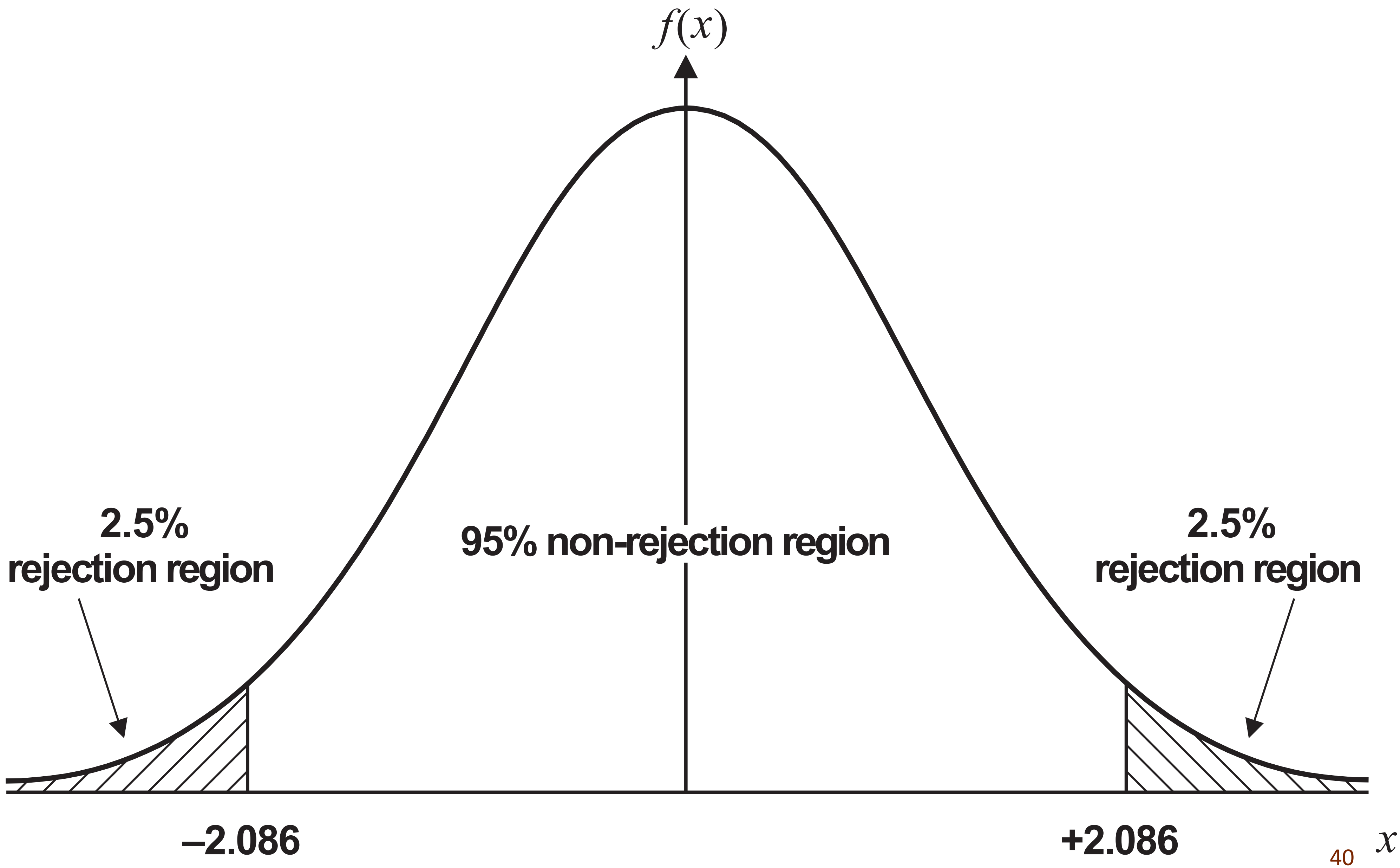
Recap: *t* versus Normal Distributions

- A *t*-distribution with an infinite number of degrees of freedom is a standard normal, i.e. $t_\infty \overset{d}{\sim} N(0,1)$.
- Examples

Significance level	t_∞	t_{40}	t_4
50%	0	0	0
5%	1.64	1.68	2.13
2.5%	1.96	2.02	2.78
0.5%	2.57	2.70	4.60

- The reason for using the *t*-distribution rather than the standard normal is that we need to estimate σ_e^2, the variance of the disturbances (aka noise or errors).

Rejection Regions for Two-Tailed Test



SMU
SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

Week 4

Pre-class Prep 6/10

Question 1: How can we compare two otherwise 'valid' OLS models?

1. In this setting, we have **two OLS models, M1 and M2**

2. M1 is a superset of M2. i.e. **M1 has all the parameters that M2 has, plus additional parameters**

3. In this situation, M1 will have a higher R^2 than M2. We can ask if **difference in R^2 between both models is statistically significant**

4. Apart from comparing OLS model performances, we can also use this approach to compare a model's performance against a situation without any predictors (i.e. **M2 only has constant** and no other variables)

Question 2: Testing Multiple Hypotheses

1. We have used the t-test to test single hypotheses, i.e. hypotheses involving only one coefficient. But what if we want to test more than one coefficient simultaneously?

2. Answer: F-test, which involves estimating two regressions:

3. The unrestricted regression is the one in which the coefficients are freely determined by the data

4. The restricted regression is the one in which the coefficients are restricted, i.e. the restrictions are imposed on a subset of the parameter vector $\boldsymbol{\beta}$

5. This situation also helps to answer Question 1, because we can restrict coefficient values to be 0 if we do not 'want' them

A F-test Example

* The general regression is

$$y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + u_t.$$

* The unrestricted regression is as above, but what is the restricted regression?

* We want to test the restriction that $\beta_3 + \beta_4 = 1$ as we have some theory suggesting this relationship.

$$y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + u_t \qquad \text{s.t. } \beta_3 + \beta_4 = 1.$$

* We substitute the restriction ($\beta_3 + \beta_4 = 1$) into the regression so that it is automatically imposed on the data.

$$\beta_3 + \beta_4 = 1 \Rightarrow \beta_4 = 1 - \beta_3$$

The F-test: Forming the Restricted Regression

* Hence, the restricted regression is

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + (1 - \beta_3) x_{4,t} + u_t \\
\Longrightarrow y_t &= \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + x_{4,t} - \beta_3 x_{4,t} + u_t \\
\Longrightarrow (y_t - x_{4,t}) &= \beta_1 + \beta_2 x_{2,t} + \beta_3 (x_{3,t} - x_{4,t}) + u_t
\end{aligned}
$$

* For estimation, we create two new variables, P_t and Q_t.

$$
\begin{aligned}
P_t &= y_t - x_{4t} \\
Q_t &= x_{3t} - x_{4t}
\end{aligned}
$$

* So

$$P_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 Q_t + u_t$$

is the restricted regression we actually estimate.

Calculating the F-Test Statistic

* The test statistic is given by

$$\text{test statistic} = \frac{\text{RRSS} - \text{URSS}}{\text{URSS}} \times \frac{T-K}{m}$$

where URSS	=	RSS from unrestricted regression
RRSS	=	RSS from restricted regression
m	=	number of restrictions
T	=	number of observations
K	=	total number of parameters to be estimated

The F-Distribution

* The test statistic follows the F-distribution, which has a pair of degree-of-freedom parameters.

* The value of the degrees of freedom parameters are m and $T - K$, respectively.

* Note that the order of the d.f. parameters is important.

* The F-distribution has only positive values and is not symmetrical.

* We therefore only reject the null if the test statistic > critical F-value.

Determining the Number of Restrictions

* Examples :

H_0 : hypothesis	Number of restrictions, m
$\beta_1 + \beta_2 = 2$	1
$\beta_2 = 1$ and $\beta_3 = -1$	2
$\beta_2 = 0, \beta_3 = 0$ and $\beta_4 = 0$	3

* If the model is

$$y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_{4,t} x_{4,t} + u_t,$$

then the null hypothesis is

$$H_0\text{: } \beta_2 = 0, \text{ and } \beta_3 = 0 \text{ and } \beta_4 = 0,$$

* Alternative hypothesis

$$H_1 : \beta_2 \neq 0, \text{ or } \beta_3 \neq 0 \text{ or } \beta_4 \neq 0$$

t and the F-Distributions

* Any hypothesis which could be tested with a t-test could have been tested using an F-test, but not the other way around.
* For example, consider the hypothesis

$$H_0 : \beta_2 = 0.5$$
$$H_1 : \beta_2 \neq 0.5$$

We could have tested this using the usual t-test:

test stat$= \dfrac{\hat{\beta}_2 - 0.5}{\mathrm{SE}(\hat{\beta}_2)}$.

* Or, it could be tested in the framework above for the F-test.
* Note that the two tests always give the same result since the t-distribution is just a special case of the F-distribution.

* For example, if we have some random variable Z, and $Z \sim t_{T-K}$ then also $Z^2 \sim F(1, T-K)$.

SMU

SINGAPORE MANAGEMENT UNIVERSITY

Masters of Quantitative Finance

QF 603

Week 4

Pre-class Prep 7/10

Background

- Markowitz (1959) model suggests that investors choose a portfolio that will minimize the variance of portfolio return, given a specific level of expected return, or maximize expected return, given a specific level of variance.

- Sharpe (1964) and Lintner (1965) introduce two (hypothetical) constructs:
    - **risk-less instrument**
    - **risky market portfolio**

Expected rate of excess return is proportional to the market risk premium, without having to do mean-variance optimization.

Portfolio Construction, Expected Return

- Consider a portfolio with w portion invested in an asset I of expected return $r_i := \mathrm{E}(r_{i,t})$ and $1 - w$ portion invested in the market portfolio of expected return $r_m := \mathrm{E}(r_{m,t})$

- The return of this portfolio, denoted by $r_{w,t}$, is a weighted average of $r_{i,t}$ and $r_{m,t}$.

$$r_{w,t} = wr_{i,t} + (1 - w)r_{m,t}. \tag{1}$$

- By the linear property of the expectation operator $\mathrm{E}(\,.\,)$, the expected return of this portfolio is

$$r_w = wr_i + (1 - w)r_m. \tag{2}$$

Variance of Portfolio's Return

- To (1), apply the variance operator(.). Using the lemma, we get

$$\mathbb{V}\left(r_{w,t}\right) = w^2\,\mathbb{V}\left(r_{i,t}\right) + (1-w)^2\,\mathbb{V}\left(r_{m,t}\right) + 2w(1-w)\,\mathbb{C}\left(r_{i,t}, r_{m,t}\right).$$

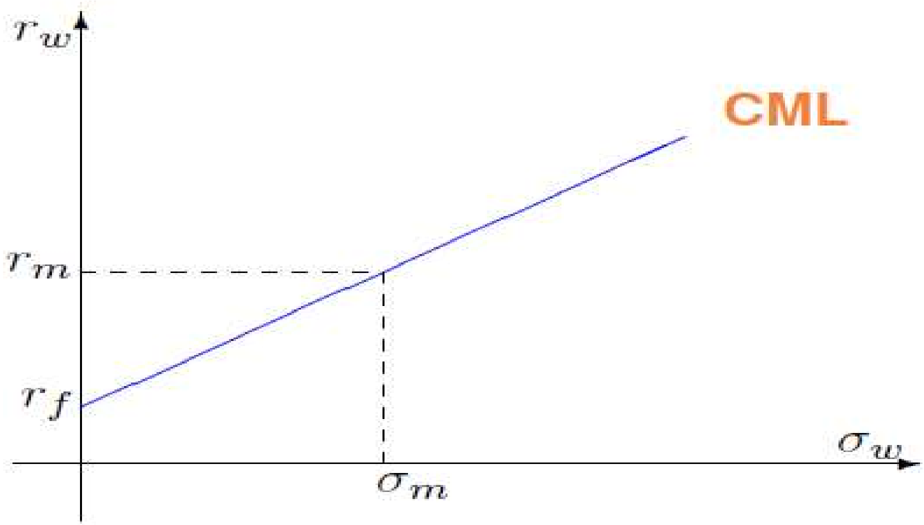
- Changing notation ***only***, with no other alterations:

$$\sigma^2 = w^2\sigma_i^2 + 2w(1-w)\sigma_{im} + (1-w)^2\sigma_m^2 \qquad (3)$$

Slope of Capital Market Line

- The slope of the CML is the Sharpe ratio. At $w = 0$, or for σ_m, we have

$$\frac{r_m - r_f}{\sigma_m} = \left.\frac{dr_w}{d\sigma_w}\right|_{w=0}.$$



Derivation of Slope

- It is tedious to compute $\dfrac{dr_w}{d\sigma_w}$ directly.

- Instead, we have, by chain rule,

$$\frac{dr_w}{d\sigma_w} = \frac{\frac{dr_w}{dw}}{\frac{d\sigma_w}{dw}}.$$

- From (2), we obtain $\dfrac{dr_w}{dw} = r_i - r_m.$

- From (3), we obtain

$$2\sigma_w \frac{d\sigma_w}{dw} = 2w\sigma_i^2 + 2(1-2w)\sigma_{im}, -2(1-w)\sigma_m^2,$$

equivalently,

$$\frac{d\sigma_w}{dw} = \frac{w\sigma_i^2 + (1-2w)\sigma_{im} - (1-w)\sigma_m^2}{\sigma_w}.$$

Slope at w = 0

- Putting everything together,

$$\frac{dr_w}{d\sigma_w} = \frac{\frac{dr_w}{dw}}{\frac{d\sigma_w}{dw}} = \frac{r_i - r_m}{\frac{w\sigma_i^2 + (1-2w)\sigma_{im} - (1-w)\sigma_m^2}{\sigma_w}}.$$

- At $w = 0$, $\sigma_w = \sigma_m$. Moreover, given that the slope is the Sharpe ratio, we have

$$\frac{r_m - r_f}{\sigma_m} = \frac{r_i - r_m}{\left(\frac{\sigma_{im} - \sigma_m^2}{\sigma_m}\right)}$$

$$r_m - r_f = \frac{r_i - r_m}{\left(\frac{\sigma_{im} - \sigma_m^2}{\sigma_m^2}\right)} = \frac{r_i - r_m}{\left(\frac{\sigma_{im}}{\sigma_m^2} - 1\right)}.$$

Slope at $w = 0$ (Cont'd)

- For any asset i that is not a market portfolio, $\frac{\sigma_{im}}{\sigma_m^2} - 1 \neq 0$. So we multiple it to both sides to obtain

$$(r_m - r_f)\left(\frac{\sigma_{im}}{\sigma_m^2} - 1\right) = r_i - r_m,$$

$$\Longrightarrow \qquad \frac{\sigma_{im}}{\sigma_m^2}(r_m - r_f) - (r_m - r_f) = r_i - r_m.$$

- Knowing that $\frac{\sigma_{im}}{\sigma_m^2} = \beta_i$, we write,

$$\beta_i(r_m - r_f) = (r_m - r_f) + r_i - r_m = r_i - r_f.$$

- Hence CAPM ensues:

$$r_i - r_f = \beta_i(r_m - r_f). \tag{4}$$

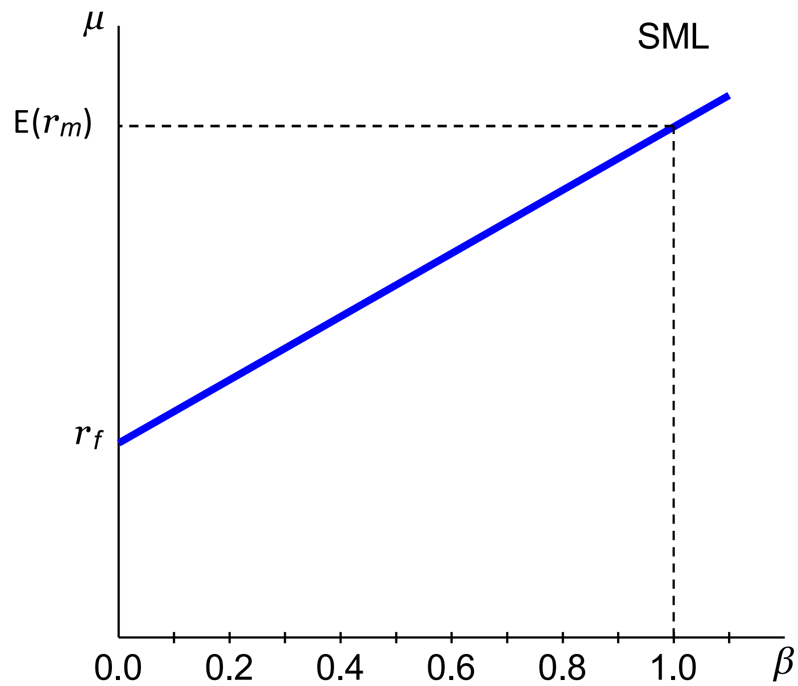
Security Market Line

- From (4), we obtain **the security market line**:

$$r_i = r_f + (r_m - r_f)\beta_i.$$

- It shows you the relationship between β_i and the required return r_i
- It is a useful tool in determining if an asset being considered for a portfolio offers a reasonable expected return for risk.
- If the security's (β_i, r_i) is plotted above (below) the SML, it is considered undervalued (overvalued), and the security gives a greater (smaller) return against its inherent risk.

Illustration of SML



- A line in a graph of expected return versus beta:

$$\mu := \mathbb{E}(r_j) = r_f + \beta_j\big(\mathbb{E}(r_m) - r_f\big)$$

or

$$\mu - r_f = \frac{\mathbb{C}(r_j\,, r_m)}{\mathbb{V}(r_m)}\big(\mathbb{E}(r_m) - r_f\big)$$

- The market portfolio has a beta of 1.

Estimating Beta: Capital Asset Pricing Model

- Imposing theoretical CAPM structure, OLS regression of market model becomes

$$r_{it} = r_{ft} + b(r_{mt} - r_{ft}) + e_{it}.$$

- Capital Asset Pricing Model

$$\mathbb{E}(r_{it} - r_{ft}) = b_i \, \mathbb{E}(r_{mt} - r_{ft}).$$

- Regression specification

$$r_{it} - r_{ft} = a_i + b_i(r_{mt} - r_{ft}) + e_{it}.$$

Example

FJ Linear regression model: $r_{it} - r_{ft} = a_i + b_i \left(r_{mt} - r_{ft} \right) + e_{it}$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003119	0.002653	1.175459	0.2426
MKT EXC RET	0.624225	0.080491	7.755199	0.0000
R-squared	0.375559	Mean dependent var.		0.000904
Adjusted R-squared	0.369314	S.D. dependent var.		0.033544
S.E. of regression	0.026639	Akaike info criterion		-4.393447
Sum squared resid	0.070965	Schwarz criterion		-4.341977
Log likelihood	226.0658	F-statistic		60.14311
Durbin-Watson stat	2.513910	Prob(F-statistic)		0.000000