Session 3
Quantitative Analysis of Financial Markets

Benjamin Ee

October, 2020

Today

- CAPM
- OLS
- Discrete estimators
- Estimation practicum

1. CAPM

1. INRODUCTION
2. ASSUMPTIONS
3. PORTFOLIO
4. CAPITAL MARKET LINE
5. APPLICATION
6. TAKEAWAYS

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.

Assumptions

- No taxes
- No transaction costs
- All investors are risk-averse.
- All investors know their utility function of terminal wealth.
- All investors can maximize their utility function.
- All investors can choose among portfolios solely on the basis of mean and variance.
- All investors have homogeneous views regarding the parameters of the joint probability distribution of all security returns.
- All investors can borrow and lend at a given risk-less rate of interest.

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

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

Lemma: Variance of $aX + bY$

- Let $\mu_X = \mathrm{E}(X)$, $\mu_Y = \mathrm{E}(Y)$, and $\mu_Z = \mathrm{E}(Z)$.
- Let $Z := aX + bY$.
- $\mu_Z = \mathrm{E}(Z) = a\,\mathrm{E}(X) + b\mathrm{E}(Y) = a\mu_X + b\mu_Y$
- The definition of variance of is

$$\begin{aligned}
\mathbb{V}(Z) &= \mathbb{E}\left((Z-\mu_Z)^2\right) = \mathbb{E}\left((aX - a\mu_X + bY - b\mu_Y)^2\right) \\
&= \mathbb{E}\left((aX-a\mu_X)^2 + (bY-b\mu_Y)^2 + 2(aX-a\mu_X)(bY-b\mu_Y)\right) \\
&= a^2\,\mathbb{E}\left((X-\mu_X)^2\right) + b^2\,\mathbb{E}\left((Y-\mu_Y)^2\right) \\
&\qquad + 2ab\,\mathbb{E}\left((X-\mu_X)(Y-\mu_Y)\right) \\
&= a^2\,\mathbb{V}(X) + b^2\,\mathbb{V}(Y) + 2ab\,\mathbb{C}(X,Y).
\end{aligned}$$

Variance of Portfolio's Return

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

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

- For convenience, we denote

$$\sigma_w^2 := \mathbb{V}\left(r_{w,t}\right), \quad \sigma_i^2 := \mathbb{V}\left(r_{i,t}\right) \quad \text{and} \quad \sigma_m^2 = \mathbb{V}\left(r_{m,t}\right)$$

  Covariance σ_{im} := C ($r_{i,t}$, $r_{m,t}$).

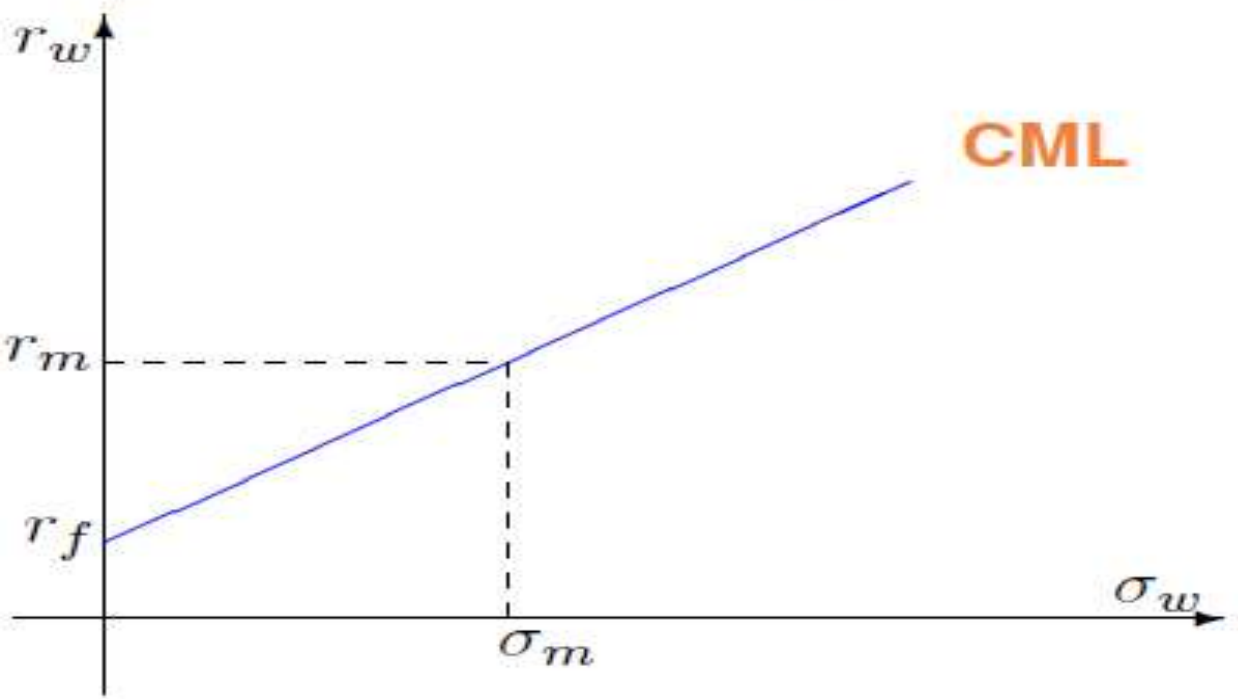
- With these notations, the variance V($r_{w,t}$) simplifies to

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

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 $\frac{dr_w}{d\sigma_w}$ 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 $\frac{dr_w}{dw} = r_i - r_m.$
- 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

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

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

Takeaways

CAPM says that expected excess return $r_i - r_f$ on the portfolio is proportional to the expected excess return $r_m - r_f$ on the market portfolio.

The beta β_i of portfolio i is essentially the covariance between the portfolio and the market, and normalized by the variance of the market portfolio.

Expected excess return $r_m - r_f$ is also known as the market risk premium.

Duality of risk and expected return

RISK FACTOR ←→ RISK PREMIUM

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 multiply 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

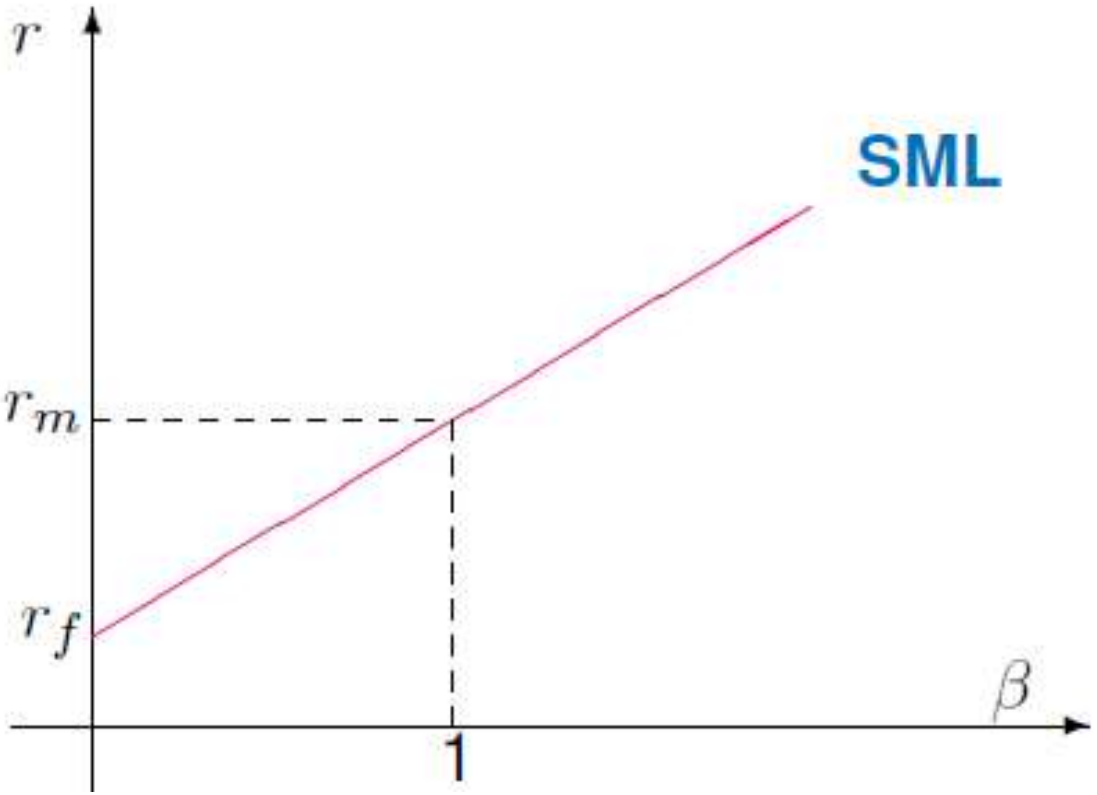
- From (4), we obtain **the security market line**:

$$r_i = r_f + (r_m - r_f)\beta_i. \tag{5}$$

- 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.

Undervalued versus Overvalued

- From (5), we see that the market portfolio has $\beta_m = 1$.
- You can easily draw an SML with two points $(0, r_f)$ and $(1, r_m)$.
- Now you know why the risk-less rate is so important.



Treynor Ratio and Alpha

- All of the portfolios on the SML have the same Treynor ratio as does the market portfolio, i.e.

$$\frac{r_i - r_f}{\beta_i} = \frac{r_m - r_f}{1} = \text{slope of the SML.}$$

- A **stock picking** rule of thumb for assets with positive beta is to buy if the Treynor ratio will be above the SML and sell if it will be below.

- The abnormal extra return above the market's return at a given level of risk is what is called the **alpha**.

2. OLS

INTRODUCTION

SIMPLE OLS

OLS IN MATRIX

HYPOTHESIS TESTS

FORECASTING

CASE STUDY

TAKEAWAYS

Learning Outcomes

- Gain deep insights into **simple** or **univariate OLS**:
  - classical conditions (assumptions) of simple linear regression's FOC (first-order conditions)
  - solutions of two FOC's (OLS estimators)
  - weights of simple OLS
  - distribution of OLS estimators
  - properties of residuals
  - hypothesis testing (significance test) of OLS estimates
- Define BLUE (best linear unbiased estimator).
- Gain deeper insights into asymptotic properties, consistent properties, and coefficient of determination of simple OLS.
- Describe how OLS estimates can be applied to **forecasting**.
- Develop a working knowledge of OLS regression by applying the theory to **hedging an equity portfolio with stock index futures**.

Motivation

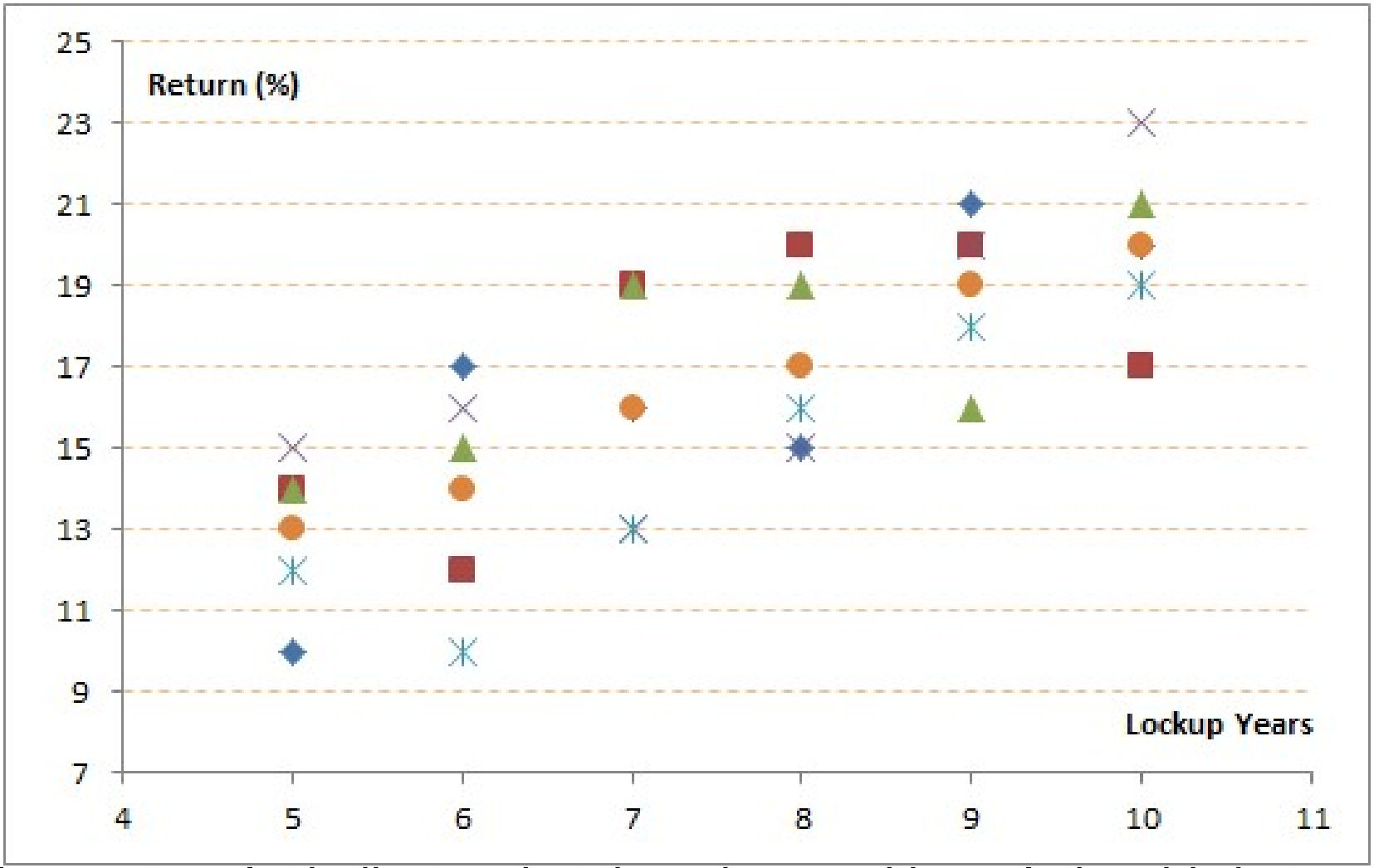
- So far we look at one time series or one set of data X.
- What about two sets of data X and Y?

Example: Annual Returns of 30 Hedge Funds

The population consists of 30 hedge funds that follow the same strategy, but of different length of the lockup period (minimum number of years an investor must keep funds invested).

Lockup (years)	Return (% per year)					Average Return
5	10	14	14	15	12	13
6	17	12	15	16	10	14
7	16	19	19	13	13	16
8	15	20	19	15	16	17
9	21	20	16	20	18	19
10	20	17	21	23	19	20

Scatter Plot



- The scatter plot indicates that there is a positive relationship between the hedge fund returns and the lockup period.

Model 1 and Classical Conditions

- Model 0 is $Y_i = a + e_i$.
- But given n pairs of observations on explanatory variable X_i and dependent variable Y_i, we can have **Model 1** by 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) $e_i \overset{d}{\sim} N(0, \sigma_e^2)$

First-Order Conditions of Least Squares

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

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

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

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

- These least squares minimization conditions are “ordinary”.

Ordinary Least Squares Solutions

- Solution of first FOC

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

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

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

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

- Solution of second FOC

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

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

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

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

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

OLS with Centered Regressor

- More convenient to start with the centralized linear model

$$Y_i = a^* + b(X_i - \overline{X}) + e_i\,, \quad a^* = a + b\overline{X}$$

- OLS

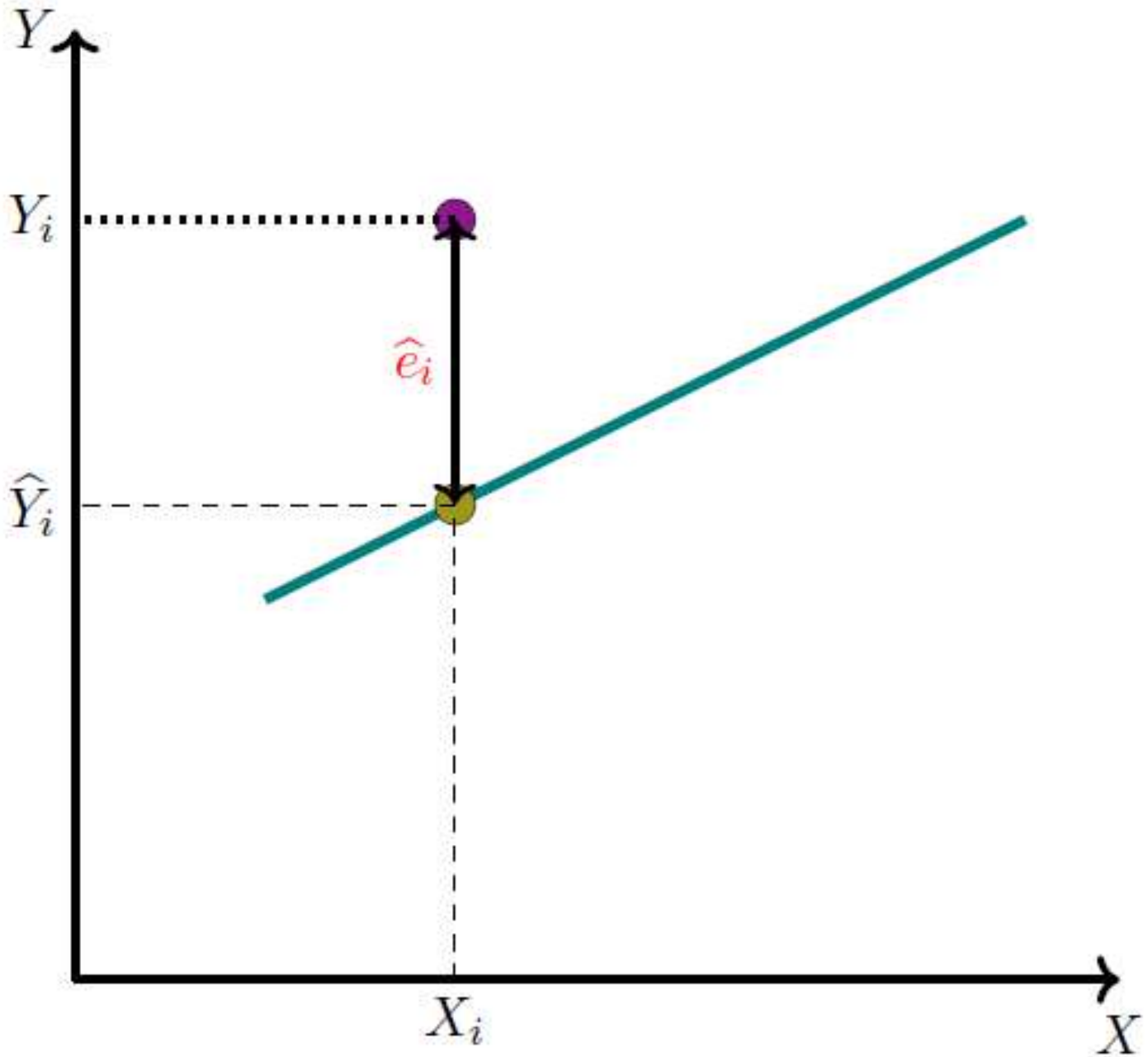
$$\min_{\widehat{a^*}\,,\widehat{b}} \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \widehat{a^*} - \widehat{b}(X_i - \overline{X})\right)^2$$

- FOC

$$\sum_{i=1}^{n} \left(Y_i - \widehat{a^*} - \widehat{b}(X_i - \overline{X})\right) = 0$$

$$\sum_{i=1}^{n} (X_i - \overline{X}) \left(Y_i - \widehat{a^*} - \widehat{b}(X_i - \overline{X})\right) = 0$$

Residual is the Vertical Length



Linear Estimators

Solution of FOC's

$$\begin{aligned} \widehat{a^*} &= \overline{Y} \\ \widehat{b} &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})} \end{aligned}$$

Define the weights

$$v_i := \frac{1}{n} - \frac{(X_i - \overline{X})\overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2}; \qquad w_i := \frac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Hence, linear combinations:

$$\widehat{a} = \sum_{i=1}^{n} v_i Y_i; \qquad \widehat{b} = \sum_{i=1}^{n} w_i Y_i$$

Remark:

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = (n-1)\widehat{\sigma}_X^2$$

Properties of Weights and OLS Estimators

- Properties of v_i

$$\sum_{i=1}^{n} v_i = 1\,, \quad \sum_{i=1}^{n} v_i^2 = \frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}\,, \quad \sum_{i=1}^{n} v_i X_i = 0$$

- Properties of w_i

$$\sum_{i=1}^{n} w_i = 0\,, \quad \sum_{i=1}^{n} w_i^2 = \frac{1}{\sum_{i=1}^{n} (X_i - \overline{X})^2}\,, \quad \sum_{i=1}^{n} w_i X_i = 1$$

- Finite sample properties of OLS estimators:

$$\widehat{a} = \sum_{i=1}^{n} v_i (a + bX_i + e_i) = a + \sum_{i=1}^{n} v_i e_i \quad \Longrightarrow \quad \mathbb{E}(\widehat{a}) = a$$

$$\widehat{b} = \sum_{i=1}^{n} w_i (a + bX_i + e_i) = b + \sum_{i=1}^{n} w_i e_i \quad \Longrightarrow \quad \mathbb{E}(\widehat{b}) = b$$

Variance and Covariance of OLS Estimators

$$\begin{aligned}
\mathbb{V}(\widehat{a}) &= \mathbb{E}\Big((\widehat{a}-a)^2\Big) = \mathbb{E}\left(\left(\sum_{i=1}^{n} v_i\, e_i\right)^2\right) = \sum_{i=1}^{n} \mathbb{E}(v_i^2)\,\mathbb{E}(e_i^2) \\
&= \sigma_e^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right) \\
\mathbb{V}\Big(\widehat{b}\Big) &= \mathbb{E}\Big((\widehat{b}-b)^2\Big) = \mathbb{E}\left(\left(\sum_{i=1}^{n} w_i\, e_i\right)^2\right) = \sum_{i=1}^{n} \mathbb{E}(w_i^2)\mathbb{E}(e_i^2) \\
&= \sigma_e^2\left(\frac{1}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right) \\
\mathbb{C}\Big(\widehat{a},\widehat{b}\Big) &= \mathbb{E}\Big((\widehat{a}-a)(\widehat{b}-b)\Big) = \mathbb{E}\left(\left(\sum_{i=1}^{n} v_i\, e_i\right)\left(\sum_{j=1}^{n} w_j\, e_j\right)\right) = \sigma_e^2 \sum_{i=1}^{n} v_i\, w_i \\
&= -\sigma_e^2\left(\frac{\overline{X}}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\right)
\end{aligned}$$

Distribution of OLS Estimators

- Slope estimator

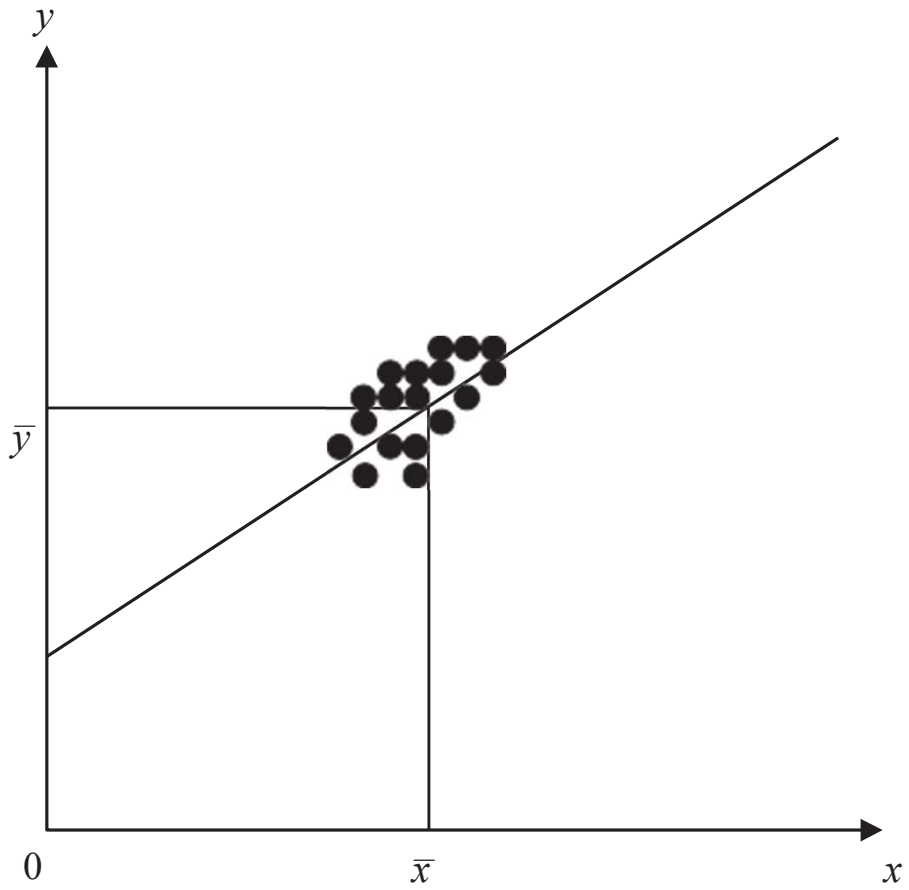
$$\widehat{b} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2}; \qquad \widehat{b} \overset{d}{\sim} N\left(b, \sigma_e^2 \left(\frac{1}{\sum_{i=1}^{n} (X_i - \overline{X})^2}\right)\right)$$

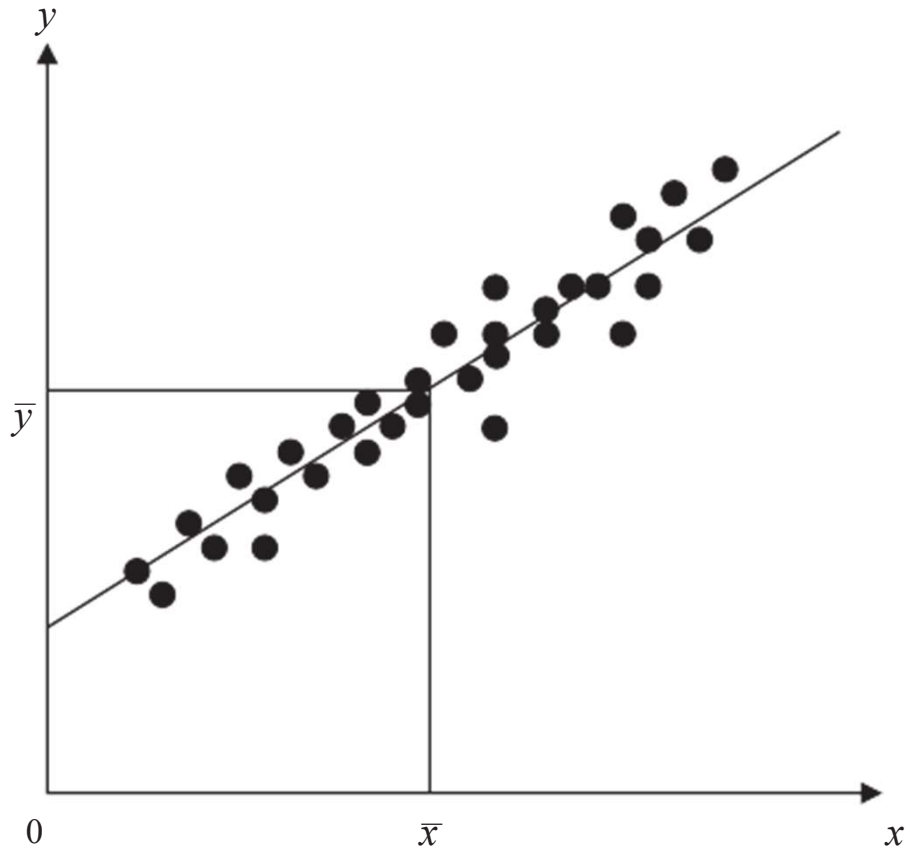
- Intercept estimator

$$\widehat{a} = \overline{Y} - \widehat{b}\,\overline{X}; \qquad \widehat{a} \overset{d}{\sim} N\left(a, \sigma_e^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}\right)\right)$$

Effect of the Variance of *X*

- What happens if $\sum_{i=1}^{n}(X_i - \overline{X})^2 = n\widehat{\sigma}_X^2$ is big or small?

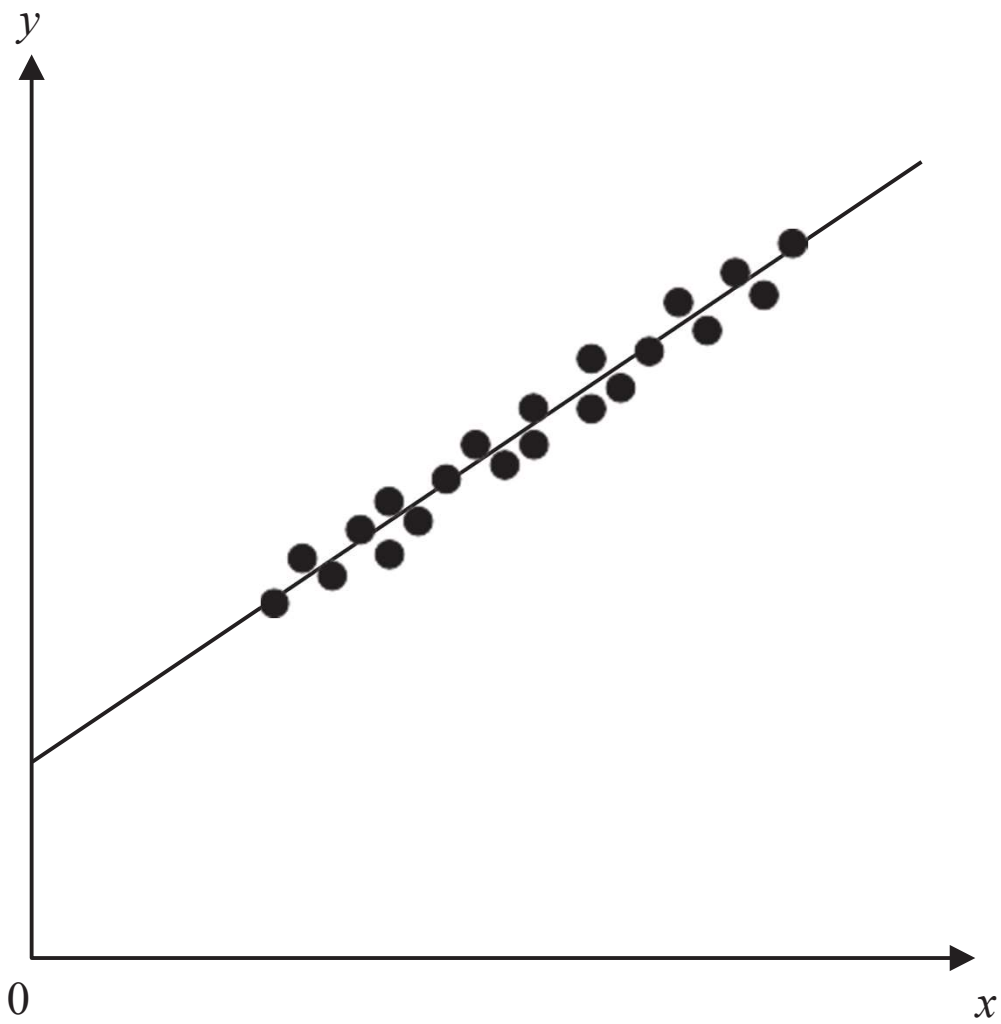




- The larger the sample size, n, the smaller will be the coefficient variances.

Accuracy of Intercept Estimate

- Care needs to be exercised when considering the intercept estimate, particularly if there are no or few observations close to the y-axis:



Distribution of OLS Estimators in Matrix Form

- To incorporate

$$\mathbb{C}\left(\widehat{a}\,,\,\widehat{b}\right) = -\sigma_e^2 \left(\frac{\overline{X}}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right).$$

- Normal distribution

$$\begin{pmatrix} \widehat{a} \\ \widehat{b} \end{pmatrix} \overset{d}{\sim} N\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} \sigma_e^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right) & -\sigma_e^2 \left(\frac{\overline{X}}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right) \\ -\sigma_e^2 \left(\frac{\overline{X}}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right) & \sigma_e^2 \left(\frac{1}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2} \right) \end{pmatrix} \right)$$

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:

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

which can be written in the vector-matrix form:

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

Simple OLS Estimators in Vector-Matrix Form

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

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

- By the classical assumption (A4), $X'e = \mathbf{0}$.
- Note that $X'X$ is a_____×_____matrix.
- Suppose $(X'X)^{-1}$ exists.

- Multiply $(X\ X)$ to both sides of $X\ Y = X\ X\beta$ to obtain

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

which is

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

OLS Esimators Are Unbiased

Proposition **1**

Given the data matrix X, the estimator β is unbiased.

- Proof:

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

- It follows that

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

Conditional Variance of y

Proposition 2

Given the data matrix X, the variance of y is the variance of the error σ_e^2.

- Proof:

$$\begin{aligned}\mathbb{V}_{\boldsymbol{X}}(\boldsymbol{y}) &= \mathbb{V}_{\boldsymbol{X}}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}) \\ &= \mathbb{V}_{\boldsymbol{X}}(\boldsymbol{X}\boldsymbol{\beta}) + \mathbb{V}_{\boldsymbol{X}}(\boldsymbol{e}) + 2\,\mathbb{C}_{\boldsymbol{X}}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{e}) \\ &= 0 + \sigma_e^2 + 0 \\ &= \sigma_e^2.\end{aligned}$$

Variance of $\widehat{\beta}$

Proposition 3

The variance-covariance matrix of the OLS estimator is

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

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

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

Proof of Gauss-Markov Theorem

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

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

Proof of Gauss-Markov Theorem (cont'd)

- Therefore, is unbiased if and only if $\boldsymbol{DX} = \mathbf{0}$. Then

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

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

Properties of Residuals

- Once the estimates and are obtained, we can compute the **residuals**:

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

- The variance of residual $\widehat{e}_i, i = 1, 2, \ldots, n$ is estimated as

$$\widehat{\sigma_e}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \widehat{e}_i^2.$$

- Mean and variance conditional on X_i

$$\mathbb{E}_{X_i}(\widehat{e}_i) = \mathbb{E}_{X_i}(Y_i) - a - b\,X_i;$$

$$\mathbb{V}_{X_i}(\widehat{e}_i) = \sigma_e^2\left(1 + \frac{1}{n} + \frac{(X_i - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right).$$

Hypothesis Testing

- 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 $H_0: a = a$ (e.g. $\alpha = 0$)

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

Lockup and Hedge Fund Return

- Does the number of lockup years "explain" hedge fund return?

	Lockup	Return (%)				
	X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$\widehat{\sigma}_{XY}$	$\widehat{\sigma}_X^2$
	5	10	-2.5	-6	15	6.25
	6	12	-1.5	-4	6	2.25
	7	19	-0.5	3	-1.5	0.25
	8	16	0.5	0	0	0.25
	9	18	1.5	2	3	2.25
	10	21	2.5	5	12.5	6.25
Sum	45	96	0	0	35	17.5
Average	7.5	16				

- The OLS estimates are $\widehat{b} = \dfrac{35}{17.5} = 2$, and $\widehat{a} = 16 - 2 \times 7.5 = 1.$

Standard Errors

- First, compute the fitted value: $\widehat{Y}_i = \widehat{a} + \widehat{b}X_i$

$$\widehat{Y}_i : \ 11, 13, 15, 17, 19, 21$$

- Next compute the residuals: $\widehat{e}_i = Y_i - \widehat{Y}_i$

$$\widehat{e}_i : \ -1, -1, 4, -1, -1, 0$$

- Sum of squared residuals $\sum_{i=1}^{6} \widehat{e}_i^2 = 20 \implies \widehat{\sigma}_e^2 = 20/(6-2) = 5.$

- Compute the standard error of $\widehat{b}$:

$$\mathsf{SE}(\widehat{b}) := \widehat{\sigma}_e \sqrt{\frac{1}{\sum_{i=1}^{n} (X_i - \overline{X})^2}} = \sqrt{\frac{5}{17.5}} = 0.5345.$$

- Compute the standard error of

$$\mathsf{SE}(\widehat{a}) := \widehat{\sigma}_e \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}} = \sqrt{\frac{5}{6} + \frac{5 \times 7.5^2}{17.5}} = 4.1115.$$

t Statistics

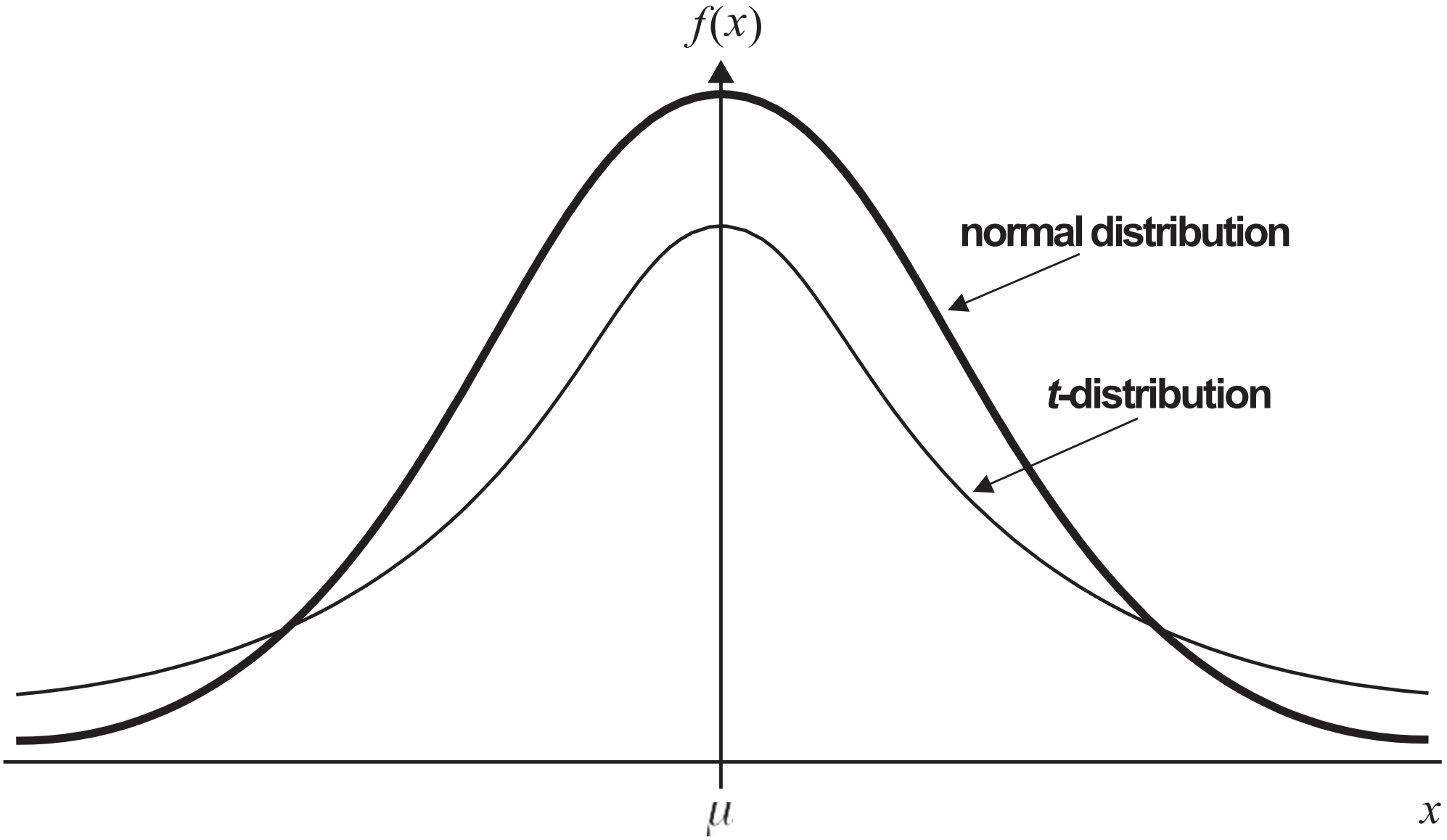
- To test the null hypothesis $H_0: b = 0$,

$$t_4(\widehat{b}) = \frac{\widehat{b} - 0}{\mathrm{SE}(\widehat{b})} = \frac{2}{0.5345} = 3.74$$

- To test the null hypothesis $H_0: a = 0$,

$$t_4(\widehat{a}) = \frac{\widehat{a} - 0}{\mathrm{SE}(\widehat{a})} = \frac{1}{4.1115} = 0.24.$$

What Does the *t*-Distribution Look Like?



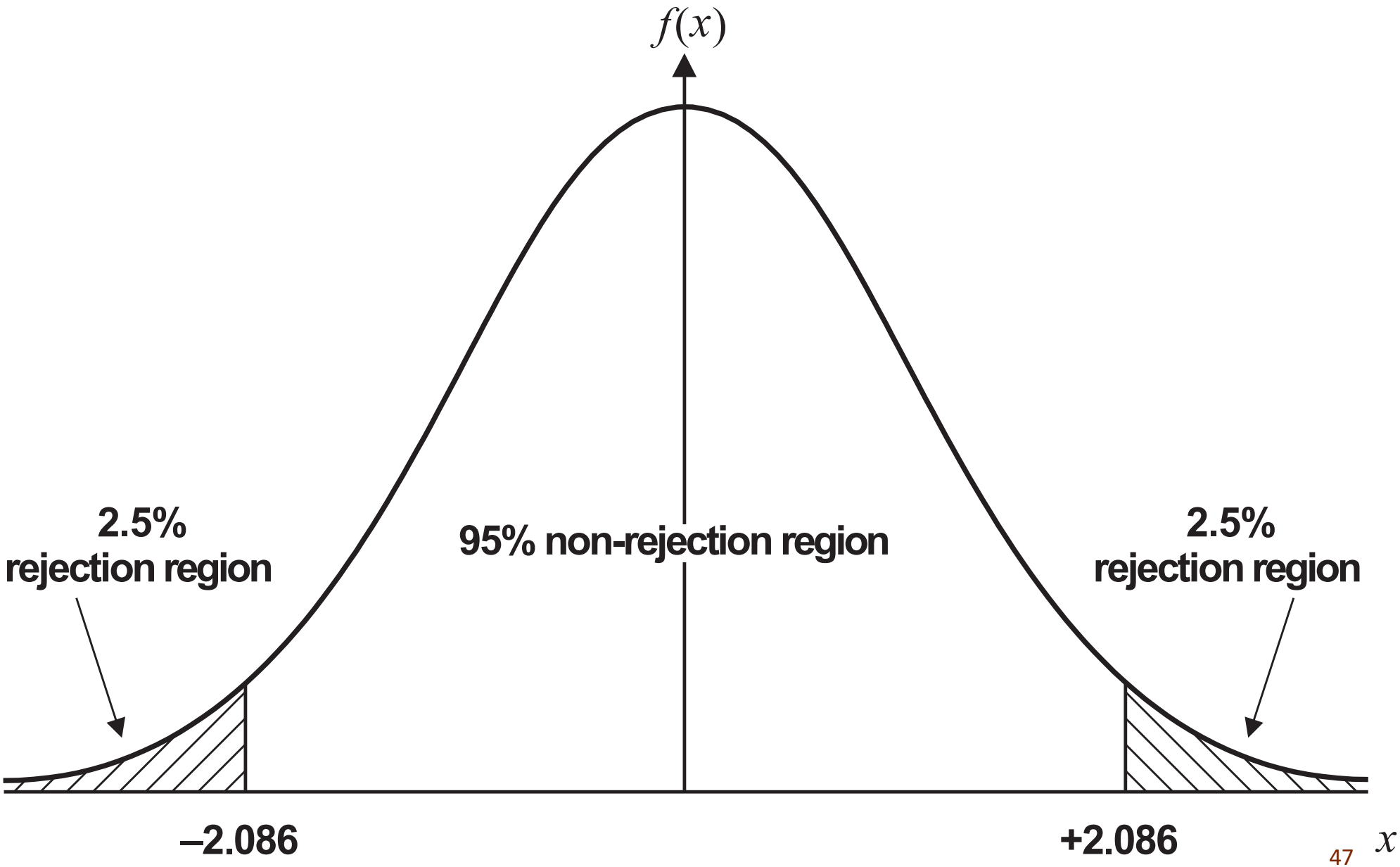
Connection between *t* and Normal Distributions

- 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

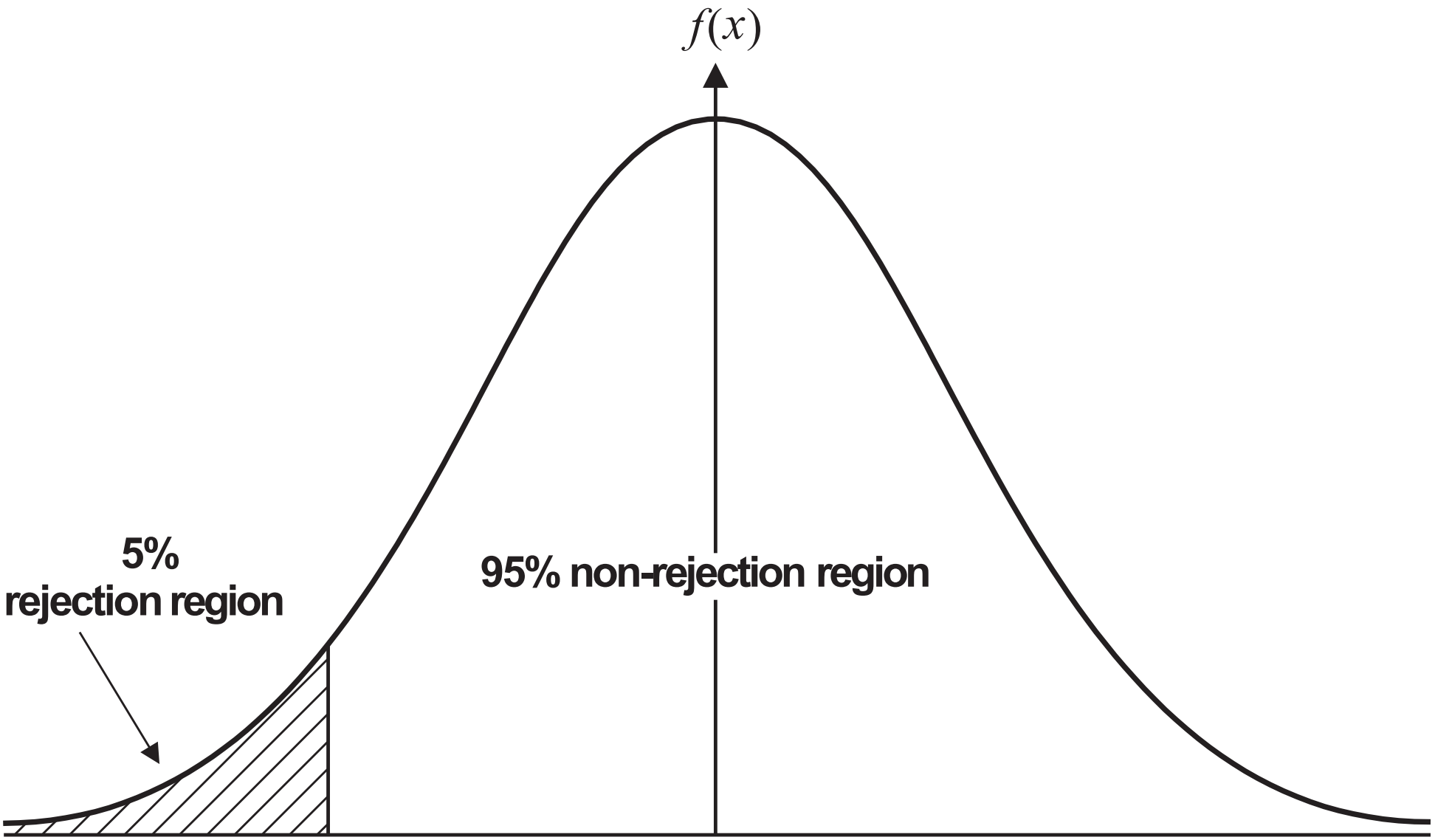
Significance level	t_∞	t_{40}	t_4
50%	0	0	0
5%	1.64	1.68	2.13
2.5%	1.96	2.02	2.78
0.5%	2.57	2.70	4.60

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

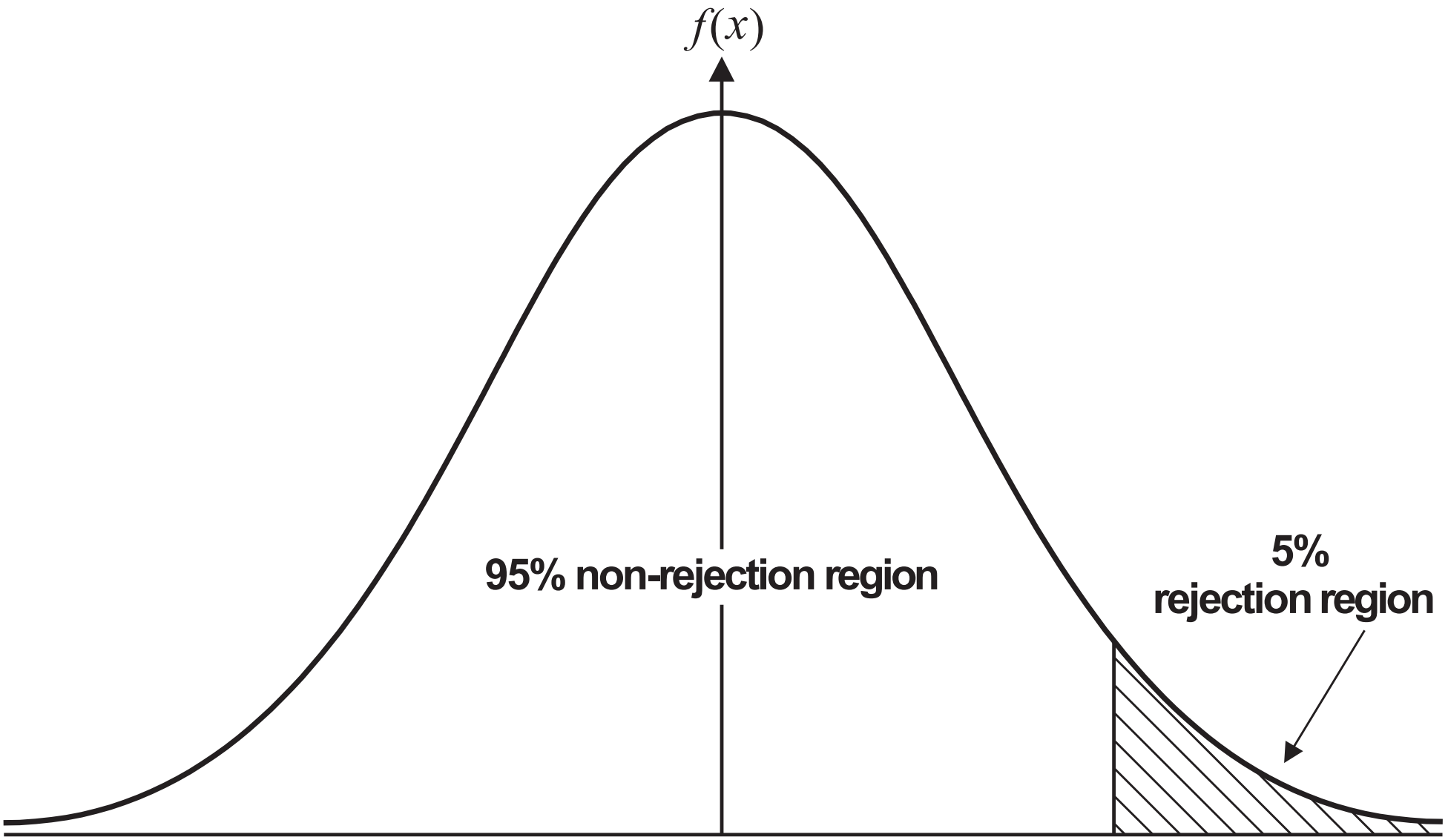
Rejection Regions for Two-Tailed Test



Rejection Region for One-Sided Lower Tail Test



Rejection Region for One-Sided Upper Tail Test



Another Example: Estimates

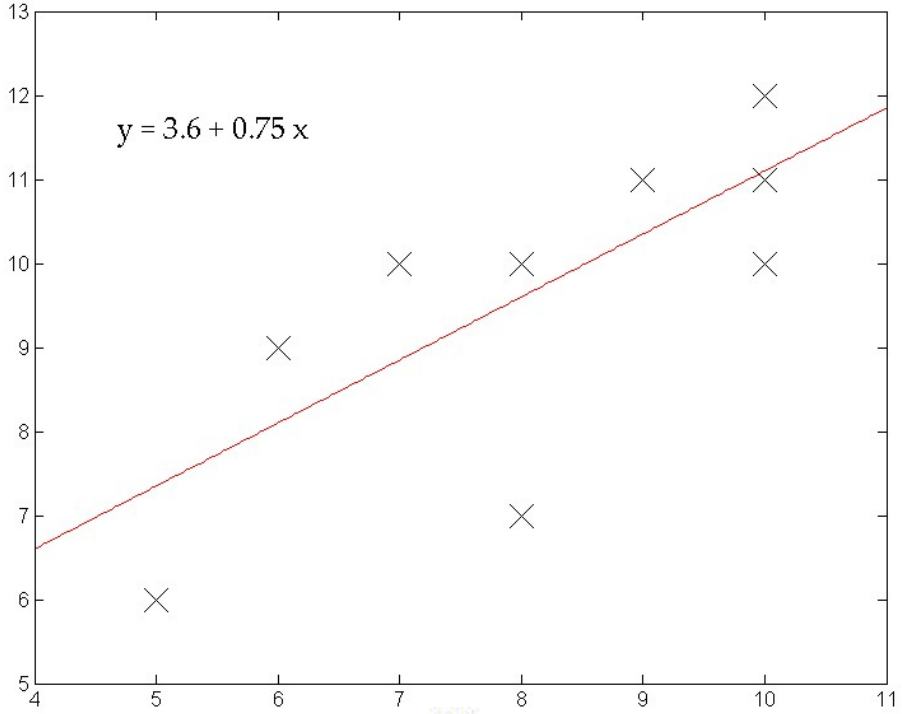
Let $X_i^* := X_i - \overline{X}$, and $Y_i^* := Y_i - \overline{Y}$.

Observation	X_i	Y_i	X_i^*	Y_i^*	X_i^{*2}	Y_i^{*2}	$X_iX_i^*$	$X_iY_i^*$	$X_i^*Y_i^*$	
1	10	11	2	1.4	4	1.96	20	14	2.8	
2	7	10	-1	0.4	1	0.16	-7	2.8	-0.4	
3	10	12	2	2.4	4	5.76	20	24	4.8	
4	5	6	-3	-3.6	9	12.96	-15	-18	10.8	
5	8	10	0	0.4	0	0.16	0	3.2	0	
6	8	7	0	-2.6	0	6.76	0	-20.8	0	
7	6	9	-2	-0.6	4	0.36	-12	-3.6	1.2	
8	7	10	-1	0.4	1	0.16	-7	2.8	-0.4	
9	9	11	1	1.4	1	1.96	9	12.6	1.4	
10	10	10	2	0.4	4	0.16	20	4	0.8	
Average	8	9.6	0	0	28	30.4	28	21	21	Total

$$\widehat{b} = \frac{21}{28} = 0.75\,, \qquad \widehat{a} = 9.6 - 0.75 \times 8 = 3.6$$

$$Y_i = 3.6 + 0.75X_i$$

Regression Result



$$\widehat{\sigma}_e^2 = \frac{1}{10-2}\sum_{i=1}^{10}\widehat{e}_i^2 = 1.83125$$

- For a estimate, the standard error is _____
- For b estimate, the standard error is _____

Estimation with Asymptotically Large Sample

- When X_i and Y_i are stationary, by the **Law of Large Numbers**,

$$\lim_{n \to \infty} X_n = \mu_X , \qquad \lim_{n \to \infty} Y_n = \mu_Y$$

- When n is asymptotically large, the biased second-order estimators approach the population variances σ^2_X, σ^2_Y, and covariance σ_{XY} .

$$S_X^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad S_Y^2 := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

$$S_{XY} := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

- When n is asymptotically large, OLS slope estimator is expressed as

$$\widehat{b} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{S_{XY}}{S_X^2}. \qquad (2)$$

Consistent Properties of OLS

- Covariance between X_i and Y_i when $Y_i = a + bX_i + e_i$ is

$$\begin{aligned} \mathbf{C}(X_i, Y_i) &= \mathbf{C}(X_i, a + bX_i + e_i) \\ &= b\mathbf{V}(X_i) + \mathbf{C}(e_i, X_i) \\ &= b\mathbf{V}(X_i) \end{aligned}$$

$$\Rightarrow \qquad b = \frac{\mathbf{C}(X_i, Y_i)}{\mathbf{V}(X_i)}$$

- Hence from (2), $\lim_{n \to \infty} b = b$.
- Implications
  - : OLS $\hat{b}$ estimator is consistent: $\lim_{n \to \infty} \hat{b} = b$
  - OLS $\hat{a}$ estimator is consistent: since $\hat{a} = \overline{Y} - \hat{b}\overline{X}$,

Decomposition

- 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

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

Coefficient of Determination

- The population correlation coefficient is $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$. The sample estimate r_{XY} is

$$r_{XY} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}} = \frac{S_{XY}}{S_X S_Y}.$$

- The OLS slope estimator is then

$$\widehat{b} = \frac{S_{XY}}{S_X^2} = \frac{r_{XY} S_X S_Y}{S_X^2} = r_{XY} \frac{S_Y}{S_X}.$$

- Consequently, ESS : $= r_{XY}^2 \frac{S_Y^2}{S_X^2} \times n S_X^2 = r_{XY}^2 n S_Y^2.$

- Coefficient of determination R^2 is defined as

$$R^2 := \frac{\text{ESS}}{\text{TSS}} = \frac{r_{XY}^2 n S_Y^2}{n S_Y^2} = r_{XY}^2.$$

Forecasting: Point Estimate

- The OLS forecast of Y_{n+1} given X_{n+1} is

$$\widehat{Y}_{n+1} = \widehat{a} + \widehat{b}\,X_{n+1} = (\overline{Y} - \widehat{b}\,\overline{X}) + \widehat{b}\,X_{n+1} = \overline{Y} + \widehat{b}(X_{n+1} - \overline{X}).$$

  - Now, by summing up and then dividing by n, we obtain

$$\overline{Y} = a + b\,\overline{X} + \frac{1}{n}\sum_{i=1}^{n} e_i.$$

  - The **point forecast** is thus given by

$$\widehat{Y}_{n+1} = a + b\,\overline{X} + \frac{1}{n}\sum_{i=1}^{n} e_i + \widehat{b}(X_{n+1} - \overline{X}).$$

Forecasting Error

- The true Y_{n+1} is $a + bX_{n+1} + e_{n+1}$, so the forecast error is

$$\begin{aligned} Y_{n+1} - \widehat{Y}_{n+1} &= b(X_{n+1} - \overline{X}) - \widehat{b}(X_{n+1} - \overline{X}) + e_{n+1} - \frac{1}{n}\sum_{i=1}^{n} e_i \\ &= -(\widehat{b} - b)(X_{n+1} - \overline{X}) + e_{n+1} - \frac{1}{n}\sum_{i=1}^{n} e_i. \end{aligned}$$

- The forecast error conditional on X_{n+1} is normally distributed.
- **The OLS forecast is unbiased**:

$$\mathbb{E}(Y_{n+1} - \widehat{Y}_{n+1} | X_{n+1}) = 0.$$

Properties of the OLS Forecast

- Variance of the OLS Forecast

$$\begin{aligned}\mathbb{V}(Y_{n+1} - \widehat{Y}_{n+1} | X_{n+1}) &= (X_{n+1} - \overline{X})^2 \mathbb{V}(\widehat{b}) + \sigma_e^2 + \frac{1}{n}\sigma_e^2 \\ &= \sigma_e^2 \left(1 + \frac{1}{n} + \frac{(X_{n+1} - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}\right)\end{aligned}$$

- The *t*-statistic of the forecast

$$t_{n-2} = \frac{Y_{n+1} - \widehat{Y}_{n+1}}{\widehat{\sigma}_e \sqrt{1 + \frac{1}{n} + \frac{(X_{n+1} - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}}}$$

Point Forecast and Confidence Interval

- The point forecast is

$$\widehat{Y}_{n+1} = \widehat{a} + \widehat{b}\,X_{n+1} \qquad (3)$$

- Since the forecast is a random variable, it has a confidence Interval associated with it.
- With 95% probability, the forecast value falls within the **confidence interval** bounded by

$$\widehat{Y}_{n+1} \pm t_{n-2,\,97.5\%} \times \widehat{\sigma}_e \sqrt{1 + \frac{1}{n} + \frac{\left(X_{n+1} - \overline{X}\right)^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}}.$$

Application: Hedging with Futures

- An institutional investor holds a portfolio of Japanese stocks that has returns following closely those of the Nikkei 225 stock index returns $\Delta S_t / S_{t-1}$.
- Contract size of Nikkei 225 futures traded on SGX is ¥500.
- To hedge against a potential bear market going forward, the investor forms a **hedged portfolio** such that the change of hedged portfolio's value $\Delta P_t := P_t - P_{t-1}$ is

$$\Delta P_t = f \times \Delta S_t - h \times 500 \times \Delta F_t.$$

- f is a constant **proportional factor** that equates the unhedged value of the portfolio to S_t.
- h is the number of contracts, and F_t is the futures price.

Application: Hedging with Futures (cont'd)

- How many contracts h should the investor short?
- In effect, the investor wants to minimize the risk or variance of ΔP_t:

$$\mathbb{V}(\Delta P_t) = f^2 \times \mathbb{V}(\Delta S_t) + h^2 \times (500)^2 \times \mathbb{V}(\Delta F_t) \\ -2 \times f \times h \times 500 \times \mathbb{C}(\Delta S_t, \Delta F_t)$$

Solution to Hedging

- The FOC for minimizing $\mathbf{V}(\Delta P_t)$ with respect to h yields

$$2h \times (500)^2 \times \mathbb{V}(\Delta F_t) - 2 \times (500f) \times \mathbb{C}(\Delta S_t\,,\, \Delta F_t) = 0.$$

- The risk-minimizing "optimal" hedge is to short

$$h^* = \frac{f \times \mathbb{C}(\Delta S_t\,,\, \Delta F_t)}{500 \times \mathbb{V}(\Delta F_t)}.$$

- Estimation: Run the following simple linear regression

$$\Delta S_t = a + b\Delta F_t + e_t.$$

- Since $b = \dfrac{\mathbb{C}(\Delta S_t\,,\, \Delta F_t)}{\mathbb{V}(\Delta F_t)}$, the number of contracts to short is

$$h^* = \widehat{b} \times \frac{f}{500}.$$

Tutorial

On January 19, 2018, the value of the portfolio is ¥78 billion, the Nikkei 225 index is 23,808.06, and the OLS estimate for b is 0.71575. How many contracts should the fund manager short?

Takeaways

- Scatter plot gives an intuitive view of whether X could explain Y.
- Parameter estimates are obtained by minimizing the sum of squared errors.
- Each residual is the vertical distance from the data point to the OLS fitted line.
- OLS estimators are BLUE.
- Covariance divided by variance of explanatory variable = slope of OLS line.
- Variance decomposition: TSS = ESS + RSS
- R^2 of simple OLS regression = square of correlation coefficient.
- t statistic's degrees of freedom = $n - 2$.
- Many many applications!

3. Discrete dependent variables

OLS is not the only widely used estimation algorithm

Probit / logit models

- Used for binary dependent variables

Multinomial probit / logit

- Categorical dependent variables

Ordered probit / logit

- Discrete dependent variables

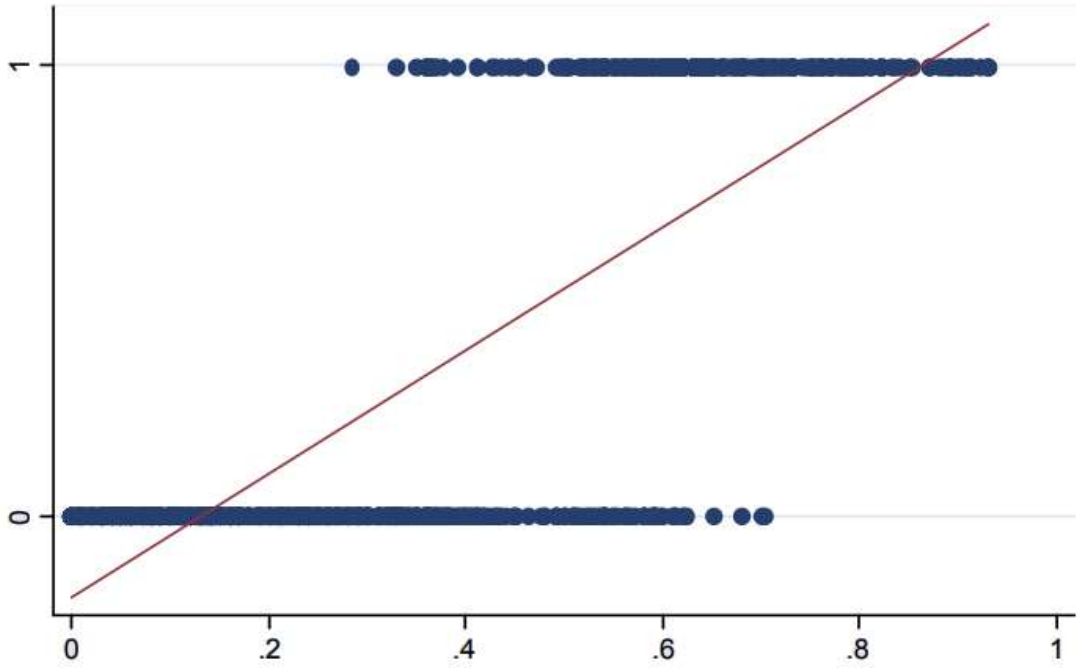
Probit / logit models

$Y_i = \alpha + \beta X_i$ where Y is either 1 or 0

Estimation options:

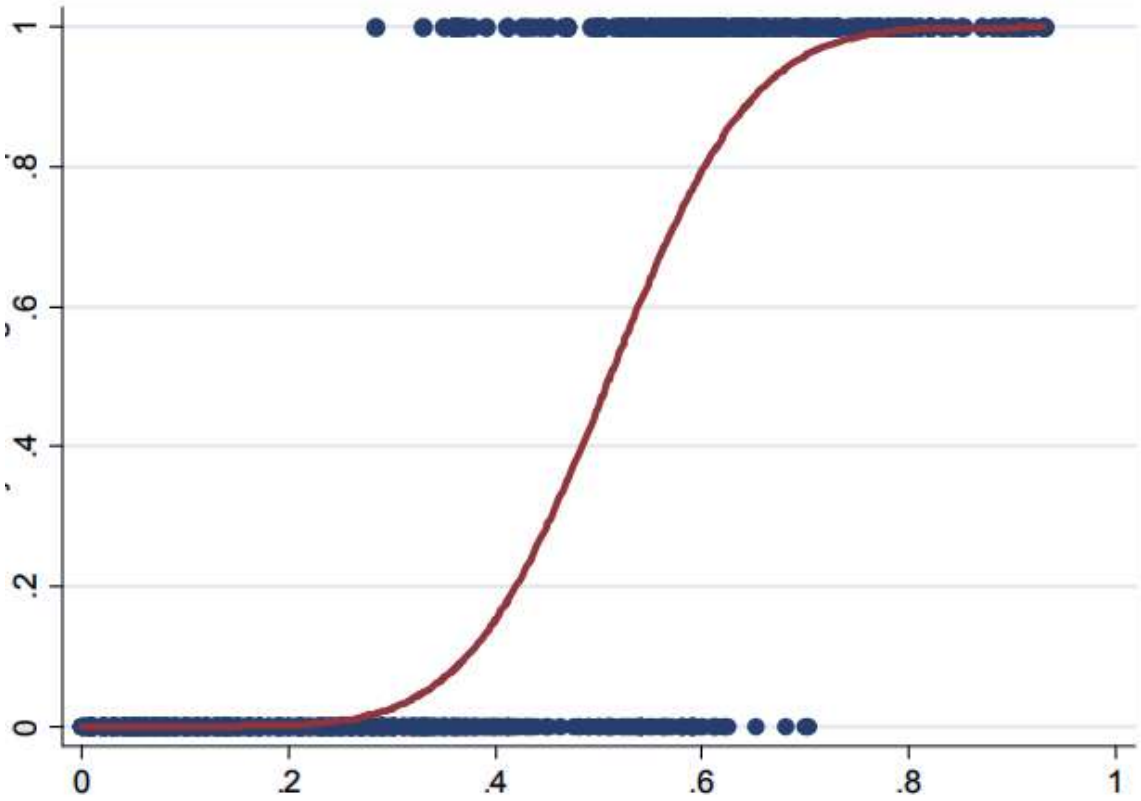
1. Just run OLS; estimate will still be unbiased and BLUE
2. Use a probit or logit model, which exploits fact that Y_i is binary.

Fitting a linear line to a dichotomous variable



Estimated line can continue infinitely above or below [0,1]

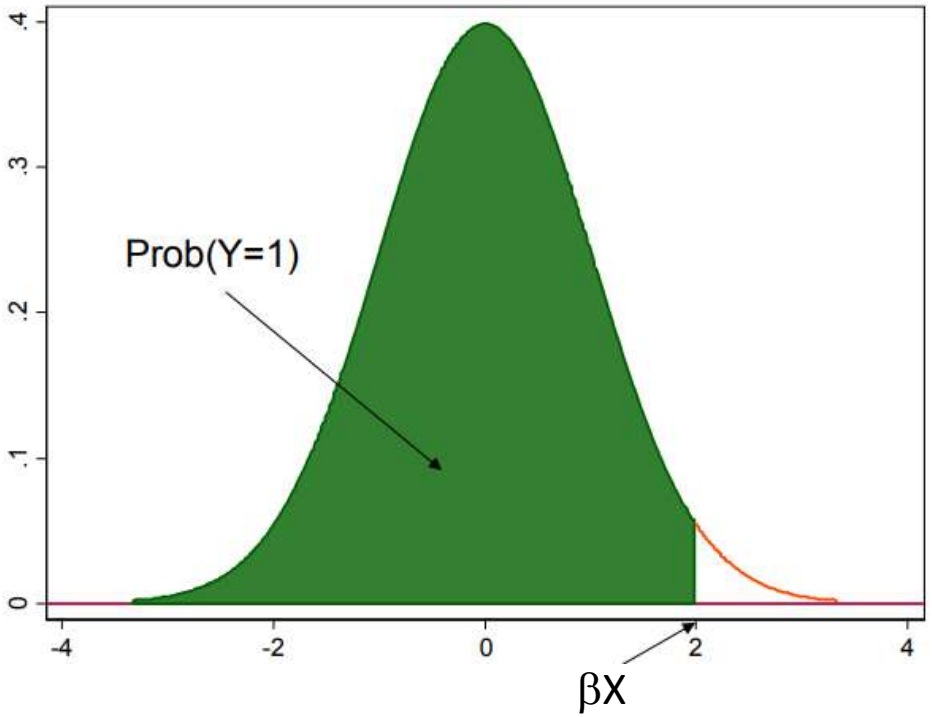
In this case, a nonlinear curve may suit the data better



What curve have we seen in this class which is:

1. Naturally between [0,1]
2. Takes above shape

CDF of normal distribution is used to determine probability of observing a 1 in the probit model



Estimating probit/logit models

Probit models are estimated using maximum likelihood (MLE), which is an iterative computational process

- Tries different values of coefficients repeatedly to estimate the parameters of a probability distribution
- Objective function is to maximize the likelihood of observing the dataset
- For a probit model, a normal distribution is used for the CDF, while in a logit function is used in the case of the logit model
- Density function of the logit model is very similar to standard normal (it may be computationally more efficient), but with thinner tails

Interpreting output of a logit model

01

Estimated coefficients for a logit model are log(odds-ratio)

02

Odds-ratio of an event is p(1-p) where p is the probability of the event

03

For the logit model, p is the probability of observing a 1

Extending concept of discrete dependent variables

- Returning to example of bond ratings, recall we can have 21 bond ratings e.g. with Moodys

- We can code each bond rating as “1”, “2”, “3’, etc in order of their probability of default. E.g. “1” corresponds to “AAA”, “21” to “CCC” etc

- Note there will be a natural ordering, where “1” is “better” than “2” in terms of default probability, etc

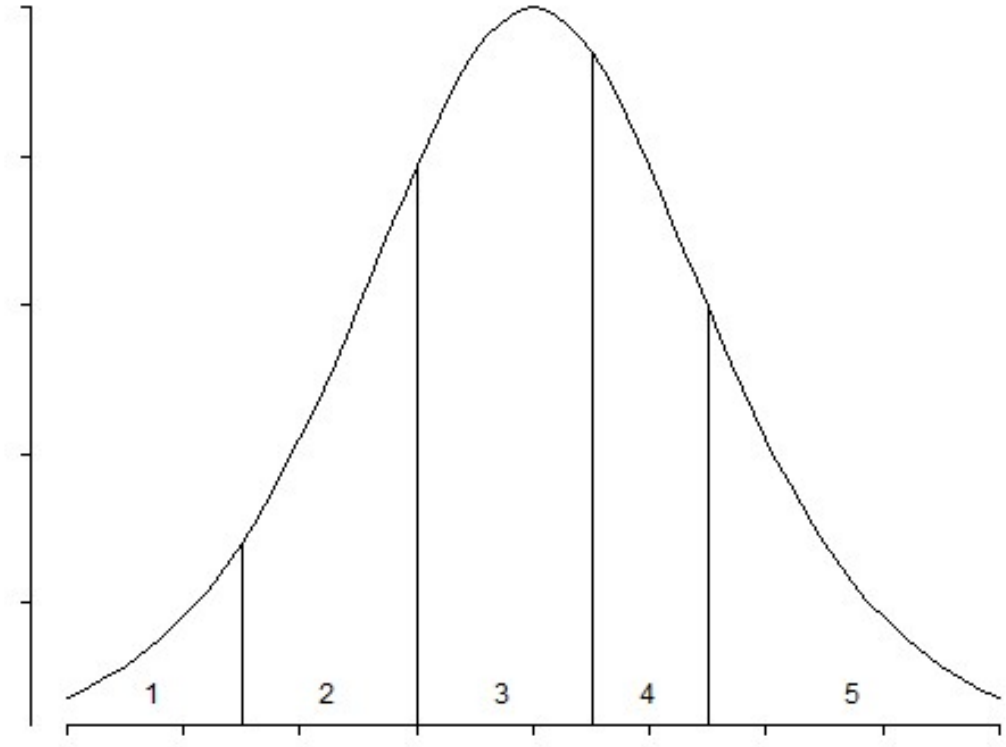
Ordered logit/probit models are a more general case of logit/probit

- Both ordered logit / probit models are a form of classification algorithm where the dependent value exists on an arbitrary scale, where only the relative ordering between different values is significant

- E.g. to model “customer satisfaction” from “5 – very good, 4 – good, 3 – neutral, etc”

- Estimation similar to logit/probit models, except multiple breakpoints on x-axis are estimated instead of just 1

Estimation of ordered logit/probit model by MLE



What if dependent variable is purely categorical?

- It is also possible for dependent variable to be purely discrete and categorical

- In this case, each value of dependent variable is merely a label, and larger or smaller values of dependent variable have no meaning

- For example, we may encode a discrete dependent variable to take values of 1, 2, 3, 4, 5, where 1 = blue, 2 = red, 3 = green, 4 = yellow and 5 = black

- In this case, 1 <2 < 3 < … does not mean anything. The numbers are just **class labels**

We may use multinomial logit and probit models for classification problems

We wish to forecast which class / category a dependent variable will take, given independent variable values

Econometric methodology is to construct a score function that constructs a score from a set of weights that are linearly combined with the explanatory variables (features)

$$\text{score}(\mathbf{X}_i, k) = \boldsymbol{\beta}_k \cdot \mathbf{X}_i$$

Multinomial estimation pseudo-algorithm

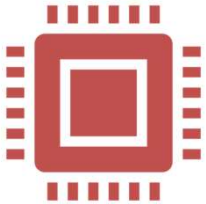
- Estimate coefficients for each score function independently (number of score functions is # categories - 1)
- Each score function gives (probability of an outcome) / (probability of base case)
- All probabilities must sum to 1, which gives us probability of base case
- Pick alternative with highest probability
- Estimation framework is also MLE

Methodology roadmap

Nature of Dependent Variable	Methodology
Continuous	OLS [note: weighted OLS, GMM, etc are all also possible, for future reference]
Discrete 1/0	Probit/Logit
Multivalued Discrete, with a value order [i.e. lower numbers are 'better' or worse']	Ordered Probit/Logit
Categorical labels	Multinomial Probit/Logit

4. Estimation Practicum

- Fixed Effects [class 3]
- Clustering of standard errors [class 3]
- Interaction terms [class 4]
- Squared terms [class 4]
- Dealing with outliers [class 3]
  - Winsorization
  - Using log(..) function
- Orthogonalization of independent variables [class 4]
- Robustness Testing [class 4]
- Trying to determine causality with lags [class 4]



In this section, we implement some regressions in python using both OLS as well as logit models, as well as explore some practical issues with outliers, fixed effects, and standard error estimation

Python packages used are fully open source (pandas, numpy, statsmodels), and are also widely used in industry or research academia.

There are many python versions. One convenient version that will allow you to follow along is here (https://www.anaconda.com/products/individual). Notwithstanding, we make no judgement that any version of python is better or worse; if you already have it installed, just use pre-existing

Datasets and sample code

- On E-Learn:
  - **corpfund.csv** contains panel data on corporate performance (e.g. EBITDA, net income, revenue, etc) for US listed firms
  - **stockmetadata.csv** contains company descriptions such as tickers, industry, geographical location, etc
  - **class3.py** contains python code that ingests and merges the data, as well as runs statistical analysis
  - Do not share data with anyone who is not taking this class

Data ingestation and merging steps

```
1  import pandas as pd
2  import numpy as np
3  import statsmodels.api as sm
4  import statsmodels.discrete.discrete_model as smdiscrete
5  pd.set_option('use_inf_as_na', True)
6
7  meta_df = pd.read_csv("stockmetadata.csv")
8  fdata_df = pd.read_csv("corpfund.csv")
9  fdata_df = fdata_df[fdata_df['dimension']=='ARQ']
10 fdata_df['datekey'] = pd.to_datetime(fdata_df['datekey'])
11 df_left = pd.merge(fdata_df, meta_df, on='ticker', how='left')
12 df_left = df_left.set_index('datekey')
```

1. Lines 7 & 8: read in both csv files as python dataframes
2. A python dataframe is a table with (usually) time on the x-axis and various variables in columns
3. In line 9, we only keep “as reported data” to avoid forward bias due to corporate restatements
4. In line 11, we merge both dataframes based on ticker
5. In line 12, we set index of the dataframe to be data-date

```
In [2]: df_left
Out[2]:
        ticker dimension calendardate reportperiod lastupdated_x      accoci  ... firstpricedate lastpricedate firstquarter lastquarter                                secfilings                companysite
datekey                                                                       ...
2010-06-07      A         ARQ  2010-03-31   2010-04-30   2020-09-01 -239000000.0  ...      1999-11-18     2020-10-09   1997-06-30   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...          http://www.agilent.com
2010-09-07      A         ARQ  2010-06-30   2010-07-31   2020-09-01 -225000000.0  ...      1999-11-18     2020-10-09   1997-06-30   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...          http://www.agilent.com
2010-12-20      A         ARQ  2010-09-30   2010-10-31   2020-09-01  -88000000.0  ...      1999-11-18     2020-10-09   1997-06-30   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...          http://www.agilent.com
2011-03-09      A         ARQ  2010-12-31   2011-01-31   2020-09-01  -63000000.0  ...      1999-11-18     2020-10-09   1997-06-30   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...          http://www.agilent.com
2011-06-07      A         ARQ  2011-03-31   2011-04-30   2020-09-01  278000000.0  ...      1999-11-18     2020-10-09   1997-06-30   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...          http://www.agilent.com
...           ...         ...         ...          ...          ...          ...  ...             ...            ...          ...          ...                                         ...                             ...
2020-10-05    STTK        ARQ  2019-12-31   2019-12-31   2020-10-09      54000.0  ...      2020-10-09     2020-10-09   2018-12-31   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...       http://www.shattucklabs.com
2020-10-08    STTK        ARQ  2020-06-30   2020-06-30   2020-10-09      18000.0  ...      2020-10-09     2020-10-09   2018-12-31   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...       http://www.shattucklabs.com
2020-10-08    KRON        ARQ  2020-06-30   2020-06-30   2020-10-09     164000.0  ...      2020-10-09     2020-10-09   2018-12-31   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...             http://kronosbio.com
2020-09-18    SPRB        ARQ  2019-12-31   2019-12-31   2020-10-09          0.0  ...      2020-10-09     2020-10-09   2018-12-31   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...  http://www.sprucebiosciences.com
2020-10-05    SPRB        ARQ  2020-06-30   2020-06-30   2020-10-09          0.0  ...      2020-10-09     2020-10-09   2018-12-31   2020-06-30  https://www.sec.gov/cgi-bin/browse-edgar?actio...  http://www.sprucebiosciences.com

[207494 rows x 138 columns]
```

This is what dataset looks like after running code from previous slide

Fixed effects

- We want to estimate relationship between E/P ratio and operating profit margin for each of these data point
- However, we note different industries may have completely different operating margins on average, as well as different relationships between operating margin and E/P ratio
- One solution is to introduce an indicator variable (“dummy variable”) for each industry
- This allows each industry to effectively have its own y-intercept

Fixed effects specification

OLS without fixed effects:

$$Y_i = \alpha + \beta X_i + \varepsilon$$

OLS with fixed effects:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 * I_1 + \beta_3 * I_2 + \beta_4 * I_3 + \ldots \beta_{12} * I_{12} + \varepsilon$$

Where

I_1 = 1 if firm is in sector 1, 0 otherwise

I_2 = 1 if firm is in sector 2, 0 otherwise

…

I_{12} = 1 if firm is in sector 12, 0 otherwise

Consequently:

Effective y-intercept for datapoints in sector 1 = $\alpha + \beta_2$

Effective y-intercept for datapoints in sector 2 = $\alpha + \beta_3$, etc

```
15  industrydummies = pd.get_dummies(df_left['sicsector'])
16  industrydummies.sum()       #purely for exploring the data, has no other purpose
17  industrydummies.describe()      #purely for exploring the data, has no other purpose
18
19  data_w_dummies = pd.concat([df_left, industrydummies], axis=1)
20
21  data_w_dummies.drop(['Wholesale Trade'], inplace=True, axis=1)  #drop 1 dummy variable
22  data_w_dummies['epratio'] = data_w_dummies['eps']/data_w_dummies['price']   #generate dependent variable
23  data_w_dummies['operatingmargin'] = data_w_dummies['opinc'] / data_w_dummies['revenue']     #generate independent var
```

Fixed effects example code

- Line 15, we generate 1 dummy variable for each sector
- Line 19, we merge the dummy variable dataframe into the main dataframe
- Line 21, we drop 1 dummy variable, which will be the control group
- Line 22, we generate our dependent variable (E/P ratio)
- Line 23, we generate our independent variable, which is operating margin

```
In [3]: industrydummies
Out[3]:
          Agriculture Forestry And Fishing  Construction  Finance Insurance And Real Estate  Manufacturing  Mining  Retail Trade  Services  Transportation Communications Electric Gas And Sanitary Service  Wholesale Trade
datekey
2010-06-07                                0             0                                  0              1       0             0         0                                                  0                        0
2010-09-07                                0             0                                  0              1       0             0         0                                                  0                        0
2010-12-20                                0             0                                  0              1       0             0         0                                                  0                        0
2011-03-09                                0             0                                  0              1       0             0         0                                                  0                        0
2011-06-07                                0             0                                  0              1       0             0         0                                                  0                        0
...                                     ...           ...                                ...            ...     ...           ...       ...                                                ...                      ...
2020-10-05                                0             0                                  0              1       0             0         0                                                  0                        0
2020-10-08                                0             0                                  0              1       0             0         0                                                  0                        0
2020-10-08                                0             0                                  0              1       0             0         0                                                  0                        0
2020-09-18                                0             0                                  0              1       0             0         0                                                  0                        0
2020-10-05                                0             0                                  0              1       0             0         0                                                  0                        0

[207494 rows x 9 columns]
```

This is what ‘industrydummies’ dataframe looks like after line 15 in the previous slide

```
26  #initial analysis
27  result = sm.OLS(data_w_dummies['epratio'], sm.add_constant(data_w_dummies[['operatingmargin']]), missing='drop').fit()
28  result.summary()
29  result = sm.OLS(data_w_dummies['epratio'], sm.add_constant(data_w_dummies[['operatingmargin', 'Agriculture Forestry And Fishing', 'Construction','Finance Insurance And Real Estate',
30  result.summary()
```

OLS regression with and without dummy variables

- Line 27 and 28, we run the OLS regression without dummy variables
- Line 29 and 30, we include the dummy variables

```
In [4]: result = sm.OLS(data_w_dummies['epratio'], sm.add_constant(data_w_dummies[['operatingmargin']]),
   ...: result.summary()
Out[4]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                epratio   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                 4.803e-06
Date:                Sat, 10 Oct 2020   Prob (F-statistic):              0.998
Time:                        16:58:52   Log-Likelihood:            -6.4512e+05
No. Observations:              115861   AIC:                         1.290e+06
Df Residuals:                  115859   BIC:                         1.290e+06
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const              0.2554      0.186      1.372      0.170      -0.110       0.620
operatingmargin  4.402e-06      0.002      0.002      0.998      -0.004       0.004
==============================================================================
Omnibus:                   651923.203   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):   60232700557916.930
Skew:                         331.783   Prob(JB):                         0.00
Kurtosis:                  111701.024   Cond. No.                         92.7
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

No Fixed Effects

```
   ...: result.summary()
Out[5]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                epratio   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.3172
Date:                Sat, 10 Oct 2020   Prob (F-statistic):              0.970
Time:                        16:59:55   Log-Likelihood:            -6.4512e+05
No. Observations:              115861   AIC:                         1.290e+06
Df Residuals:                  115851   BIC:                         1.290e+06
Df Model:                           9
Covariance Type:            nonrobust
================================================================================================================================
                                                                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------------------------
const                                                            0.7960      0.372      2.141      0.032       0.067       1.525
operatingmargin                                              -1.244e-05      0.002     -0.006      0.995      -0.004       0.004
Agriculture Forestry And Fishing                                -0.7124      3.933     -0.181      0.856      -8.422       6.997
Construction                                                    -0.7724      1.732     -0.446      0.656      -4.168       2.623
Finance Insurance And Real Estate                               -0.6747      0.531     -1.270      0.204      -1.716       0.366
Manufacturing                                                   -0.7171      0.525     -1.367      0.172      -1.745       0.311
Mining                                                          -0.7509      1.226     -0.612      0.540      -3.154       1.653
Retail Trade                                                    -0.7757      0.927     -0.837      0.402      -2.592       1.040
Services                                                        -0.7737      0.689     -1.122      0.262      -2.125       0.577
Transportation Communications Electric Gas And Sanitary Service -0.7695      0.822     -0.936      0.349      -2.380       0.841
==============================================================================
Omnibus:                   651916.510   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):   60226461523137.062
Skew:                         331.771   Prob(JB):                         0.00
Kurtosis:                  111695.238   Cond. No.                     1.96e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.96e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

With Fixed Effects

We note t-statistic for estimated coefficient on ‘operatingmargin’ is very low

This might be due to the effect of outliners overly influencing estimation

Two ways to explore the effect of outliers are:

- Winsorize data [albeit at expense of deleting observations arbitrarily (“datamining”)]
- Apply logarithm to both dependent and independent variables. This will naturally reduce impact of outliers

```
31  data_w_dummies['lnoperatingmargin'] = np.log(data_w_dummies['operatingmargin'])
32  result = sm.OLS(data_w_dummies['epratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin']]), missing='drop').fit()
33  result.summary()
34  data_w_dummies['lnepratio'] = np.log(data_w_dummies['epratio'])
35  result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin']]), missing='drop').fit()
36  result.summary()
37
38  #with dummy variables
39  result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And Fishing', 'Construction','Finance Insurance And Real Estate'
40  result.summary()
```

- Lines 31 and 34: Generate log versions of both dependent and independent variables
- Line 35: Run regression with log variables
- Line 39: Same regression, but with dummy variables included

```
In [6]: result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin']]), missing='drop').fit()^M
   ...: result.summary()^M
   ...:
Out[6]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:              lnepratio   R-squared:                       0.071
Model:                            OLS   Adj. R-squared:                  0.071
Method:                 Least Squares   F-statistic:                     8846.
Date:                Sat, 10 Oct 2020   Prob (F-statistic):               0.00
Time:                        17:08:40   Log-Likelihood:            -1.5681e+05
No. Observations:              115861   AIC:                         3.136e+05
Df Residuals:                  115859   BIC:                         3.136e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const             -3.7721      0.006   -602.657      0.000      -3.784      -3.760
lnoperatingmargin  0.2638      0.003     94.054      0.000       0.258       0.269
==============================================================================
Omnibus:                    18365.631   Durbin-Watson:                   1.036
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           158342.851
Skew:                           0.510   Prob(JB):                         0.00
Kurtosis:                       8.635   Cond. No.                         5.93
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
"""
```

Log variables, no FE

```
In [7]: result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And Fishi
   ...: 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit()^M
   ...: result.summary()
Out[7]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:              lnepratio   R-squared:                       0.085
Model:                            OLS   Adj. R-squared:                  0.084
Method:                 Least Squares   F-statistic:                     1188.
Date:                Sat, 10 Oct 2020   Prob (F-statistic):               0.00
Time:                        17:09:31   Log-Likelihood:            -1.5595e+05
No. Observations:              115861   AIC:                         3.119e+05
Df Residuals:                  115851   BIC:                         3.120e+05
Df Model:                           9
Covariance Type:            nonrobust
================================================================================================================================
                                                                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------------------------------------------
const                                                           -3.6332      0.009   -419.187      0.000      -3.650      -3.616
lnoperatingmargin                                                0.2969      0.003     96.466      0.000       0.291       0.303
Agriculture Forestry And Fishing                                 0.6587      0.058     11.417      0.000       0.546       0.772
Construction                                                     0.3404      0.025     13.352      0.000       0.290       0.390
Finance Insurance And Real Estate                               -0.1784      0.008    -21.688      0.000      -0.195      -0.162
Manufacturing                                                   -0.0182      0.008     -2.360      0.018      -0.033      -0.003
Mining                                                          -0.0450      0.018     -2.489      0.013      -0.080      -0.010
Retail Trade                                                     0.1256      0.014      9.187      0.000       0.099       0.152
Services                                                        -0.2630      0.010    -26.005      0.000      -0.283      -0.243
Transportation Communications Electric Gas And Sanitary Service -0.1285      0.012    -10.640      0.000      -0.152      -0.105
==============================================================================
Omnibus:                    18757.262   Durbin-Watson:                   1.031
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           159948.347
Skew:                           0.533   Prob(JB):                         0.00
Kurtosis:                       8.657   Cond. No.                         51.5
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
"""
```

Log variables, with FE

We note t-statistic is significantly improved, but could this be a result of intra-cluster correlations?

- Observations within same industry are likely to be highly correlated versus observations in different industries
- This violates OLS assumption of iid datapoints
- Estimates may therefore greatly overstate efficiency (inflated t-statistics, artificially low p-values, downward biases standard errors)
- We can adjust the standard error estimates by clustering errors within each industry
- Background: See for e.g. https://economics.mit.edu/files/13927 for more information

For next couple of slides, we focus on empirical application rather than theoretical proof of forgoing

```
42  #clustering standard variables
43  data_w_dummies.dropna(subset = ['lnepratio', 'lnoperatingmargin'], inplace=True)    #because of a bug in python where fillna is not working perfectly
44  #note that we can cannot cluster by str variables in python, hence using siccode instead of sicsector
45  result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And Fishing', 'Construction','Finance Insurance And Real Es
    ,'Trade', 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit(cov_type='cluster', cov_kwds={'groups': data_w_dummies['siccode']})
46  result.summary()
```

Line 43: We drop all observations with missing data to avoid a bug in python

Line 45: Run OLS estimation on log variables with FE, and standard errors clustered by sic_code

```
In [8]: result = sm.OLS(data_w_dummies['lnepratio'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And Fishir
   ...: 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit(cov_type='cluster', cov_kwds=
   ...: result.summary()
   ...:
Out[8]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                lnepratio   R-squared:                       0.085
Model:                            OLS   Adj. R-squared:                  0.084
Method:                 Least Squares   F-statistic:                     44.65
Date:                Sat, 10 Oct 2020   Prob (F-statistic):           2.30e-80
Time:                        17:22:14   Log-Likelihood:            -1.5595e+05
No. Observations:              115861   AIC:                         3.119e+05
Df Residuals:                  115851   BIC:                         3.120e+05
Df Model:                           9
Covariance Type:              cluster
==============================================================================================================================
                                                                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------------------
const                                                         -3.6332      0.034   -107.420      0.000      -3.700      -3.567
lnoperatingmargin                                              0.2969      0.016     18.678      0.000       0.266       0.328
Agriculture Forestry And Fishing                               0.6587      0.357      1.847      0.065      -0.040       1.358
Construction                                                   0.3404      0.064      5.292      0.000       0.214       0.467
Finance Insurance And Real Estate                             -0.1784      0.127     -1.405      0.160      -0.427       0.070
Manufacturing                                                 -0.0182      0.052     -0.351      0.726      -0.120       0.083
Mining                                                        -0.0450      0.048     -0.947      0.344      -0.138       0.048
Retail Trade                                                   0.1256      0.074      1.699      0.089      -0.019       0.270
Services                                                      -0.2630      0.085     -3.105      0.002      -0.429      -0.097
Transportation Communications Electric Gas And Sanitary Service    -0.1285      0.058     -2.215      0.027      -0.242      -0.015
==============================================================================
Omnibus:                    18757.262   Durbin-Watson:                   1.031
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           159948.347
Skew:                           0.533   Prob(JB):                         0.00
Kurtosis:                       8.657   Cond. No.                         51.5
==============================================================================

Warnings:
[1] Standard Errors are robust to cluster correlation (cluster)
```

Estimated t-statistic is several times lower after clustering!

```
48  #generate categorical variable for probit/logit analysis
49  data_w_dummies['paydividend'] = data_w_dummies['dps']>0
50  data_w_dummies['paydividend']   #purely for describing data
51  data_w_dummies['paydividend'].mean()    #purely for describing data
52  data_w_dummies['paydividend'] = data_w_dummies['paydividend'].astype(int)   #formatting the data for estimation models
53
54  #try using OLS anyway
55  result = sm.OLS(data_w_dummies['paydividend'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Forestry And
     Trade', 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit(cov_type='cluste
56  result.summary()
57
58  #use logit instead
59  result = smdiscrete.Logit(data_w_dummies['paydividend'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture For
     'Mining','Retail Trade', 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit
60  result.summary()
```

Using a binary dependent variable:

- Line 49 to 52: Generate a 1/0 variable on whether the company pays a dividend
- Line 55 runs an estimation with 1/0 dependent variable using OLS anyway
- Line 59 estimates a logit model

```
In [9]: result = smdiscrete.Logit(data_w_dummies['paydividend'], sm.add_constant(data_w_dummies[['lnoperatingmargin', 'Agriculture Fores
   ...: ail Trade', 'Services','Transportation Communications Electric Gas And Sanitary Service']]), missing='drop').fit(cov_type='clust
   ...: result.summary()
Optimization terminated successfully.
         Current function value: 0.638094
         Iterations 5
Out[9]:
<class 'statsmodels.iolib.summary.Summary'>
"""
                           Logit Regression Results
==============================================================================
Dep. Variable:            paydividend   No. Observations:               115861
Model:                          Logit   Df Residuals:                   115851
Method:                           MLE   Df Model:                            9
Date:                Sat, 10 Oct 2020   Pseudo R-squ.:                0.07838
Time:                        17:25:58   Log-Likelihood:                -73930.
converged:                       True   LL-Null:                       -80218.
Covariance Type:             cluster   LLR p-value:                     0.000
=============================================================================================================
                                                coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------------------------------------
const                                          0.2637      0.088      2.989      0.003       0.091       0.437
lnoperatingmargin                              0.3209      0.044      7.249      0.000       0.234       0.408
Agriculture Forestry And Fishing               0.0979      0.498      0.196      0.844      -0.879       1.075
Construction                                  -0.0196      0.258     -0.076      0.940      -0.525       0.486
Finance Insurance And Real Estate              1.2749      0.170      7.494      0.000       0.941       1.608
Manufacturing                                  0.3926      0.116      3.373      0.001       0.164       0.621
Mining                                         0.2640      0.293      0.900      0.368      -0.311       0.839
Retail Trade                                   0.4485      0.156      2.868      0.004       0.142       0.755
Services                                      -0.2165      0.148     -1.460      0.144      -0.507       0.074
Transportation Communications Electric Gas And Sanitary Service     1.0202      0.208      4.903      0.000       0.612       1.428
=============================================================================================================
"""
```

Estimated coefficients are log(odds_ratios)