Session 3
$\mathbb{Q}$uantitative Analysis of Financial Markets
Simple Linear Regression

Christopher Ting

http://www.mysmu.edu/faculty/christophert/

Christopher Ting

✉: christophert@smu.edu.sg

☎: 6828 0364

: LKCSB 5036

October 22, 2018

Broad Lesson Plan

Learning Outcomes

- Gain deep insights into **simple** or **univariate OLS**:
  - classical conditions (assumptions) of simple linear regression
  - FOC (first-order condition)
  - 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

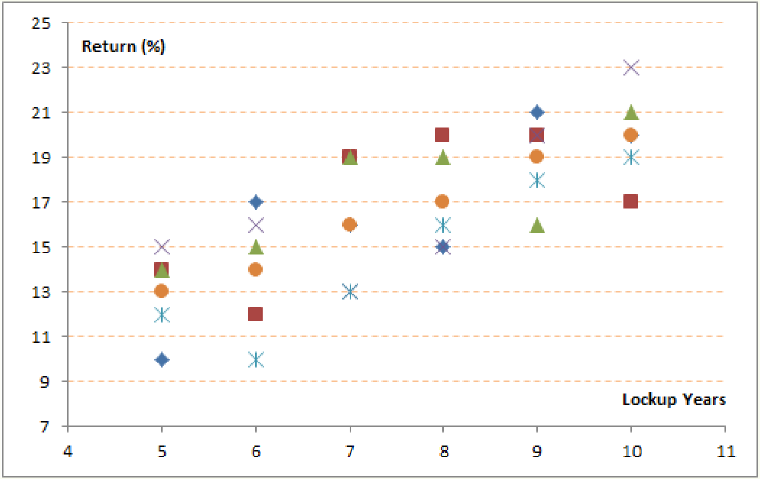
- So far we look at one time series or one set of data X.
- What about two sets of data X and Y?

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.

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**:

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

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

♣ 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

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

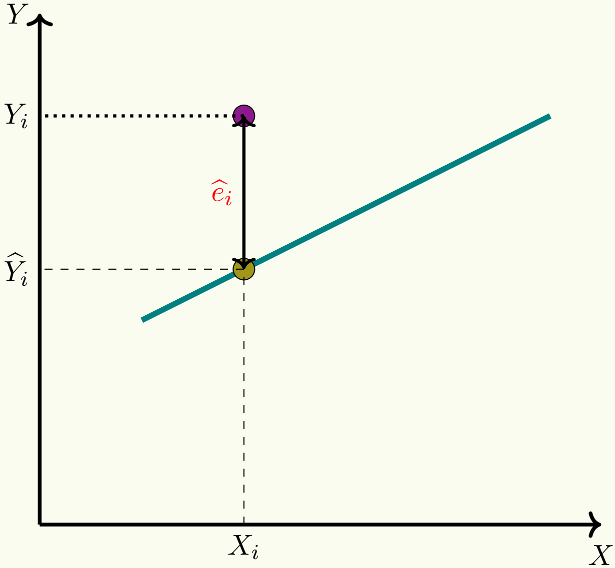
- OLS

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

- FOC

$$\begin{aligned} \sum_{i=1}^{n} \Big(Y_i - \widehat{a^*} - \widehat{b}\big(X_i - \overline{X}\big)\Big) &= 0 \\ \sum_{i=1}^{n} \big(X_i - \overline{X}\big)\Big(Y_i - \widehat{a^*} - \widehat{b}\big(X_i - \overline{X}\big)\Big) &= 0 \end{aligned}$$

Residual is the Vertical Length



Linear Estimators

- Solution of FOC's

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

- Define the weights

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

- 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:

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

Properties of Weights and OLS Estimators

- Properties of v_i

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

- Properties of w_i

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

- Finite sample properties of OLS estimators: Unbiasedness

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

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

Variance and Covariance of OLS Estimators

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

Distribution of OLS Estimators

- Slope estimator

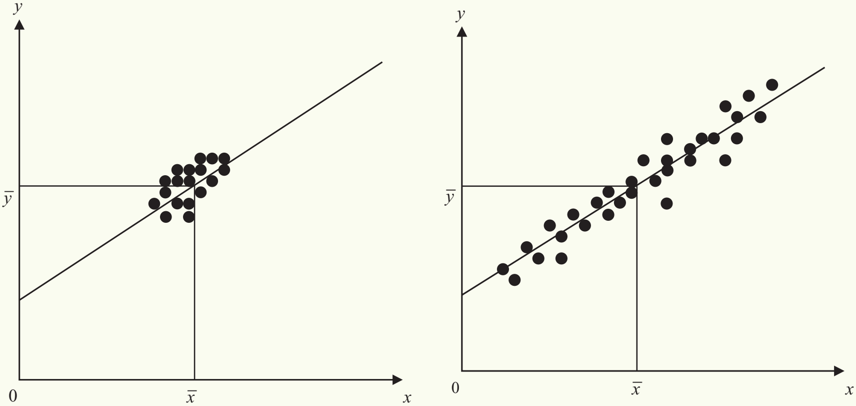
$$\widehat{b} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\,; \qquad \widehat{b} \overset{d}{\sim} N\left(b\,,\,\sigma_e^2\left(\frac{1}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\right)\right)$$

- Intercept estimator

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

Effect of the Variance of X

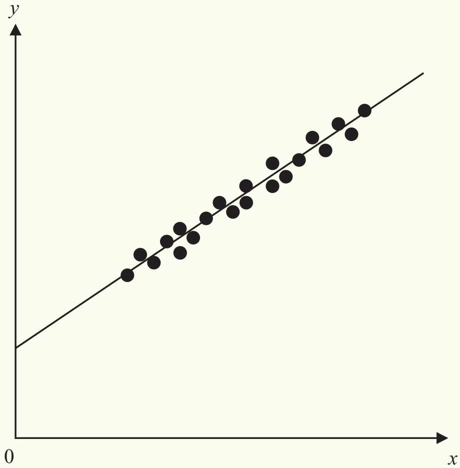
- What happens if $\sum_{i=1}^{n}(X_i - \overline{X})^2 = n\widehat{\sigma}_X^2$ is big or small?



- 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 $\boldsymbol{X}'$ to both sides of $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}$ to obtain
$$\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{X}\widehat{\beta} + \boldsymbol{X}'\boldsymbol{e}.$$
- By the classical assumption (A4), $\boldsymbol{X}'\boldsymbol{e} = \mathbf{0}$.
- Note that $\boldsymbol{X}'\boldsymbol{X}$ is a ________ $\times$ ________ matrix.
- Suppose $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exists.
- Multiply $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ to both sides of $\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{X}'\boldsymbol{X}\widehat{\beta}$ to obtain
$$(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$
which is
$$\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$

Tutorial

Proposition 1

Given the data matrix $\boldsymbol{X}$, the estimator $\widehat{\boldsymbol{\beta}}$ is unbiased.

- Proof:

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

- It follows that

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

□

Conditional Variance of y

Proposition 2

Given the data matrix $\boldsymbol{X}$, the variance of $\boldsymbol{y}$ is the variance of the error σ_e^2.

Proof:

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

$\square$

Variance of $\widehat{\beta}$

Proposition 3

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 $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$. Then

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

$\square$

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 $\tilde{\beta} = \boldsymbol{C}\boldsymbol{y}$ be another linear estimator of β with

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

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

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

Proof of Gauss-Markov Theorem (cont'd)

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

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

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

Properties of Residuals

† Once the estimates $\widehat{a}$ and $\widehat{b}$ are obtained, we can compute the **residuals**:

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

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

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

† Mean and variance of $\widehat{e}_i$ conditional on X_i

$$\begin{aligned} \mathbb{E}_{X_i}\big(\widehat{e}_i\big) &= \mathbb{E}_{X_i}(Y_i) - \widehat{a} - \widehat{b}\, X_i; \\ \mathbb{V}_{X_i}\big(\widehat{e}_i\big) &= \sigma_e^2 \left(1 - \frac{1}{n} - \frac{(X_i - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right). \end{aligned}$$

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

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

Lockup and Hedge Fund Return

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

	Lockup X	Return (%) 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 $\widehat{b}$:

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

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

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

t Statistics

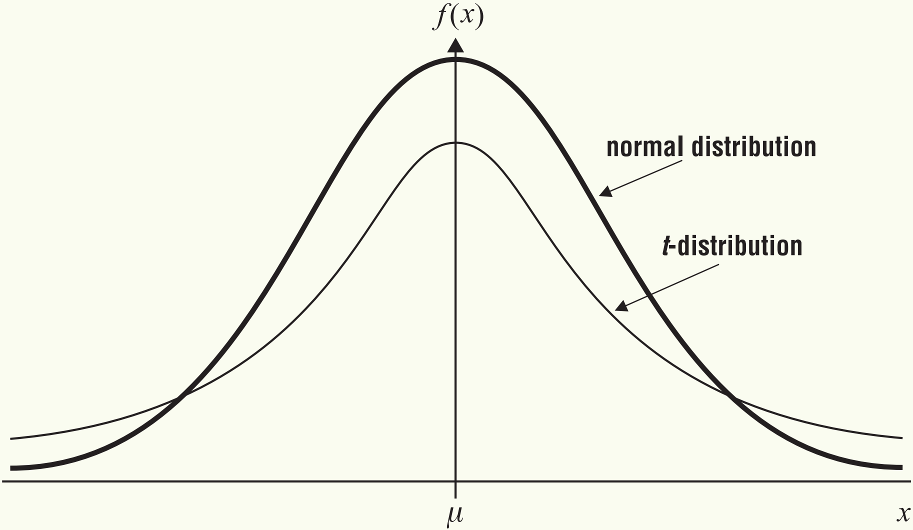
- To test the null hypothesis $H_0 : b = 0$,

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

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

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

What Does the *t*-Distribution Look Like?



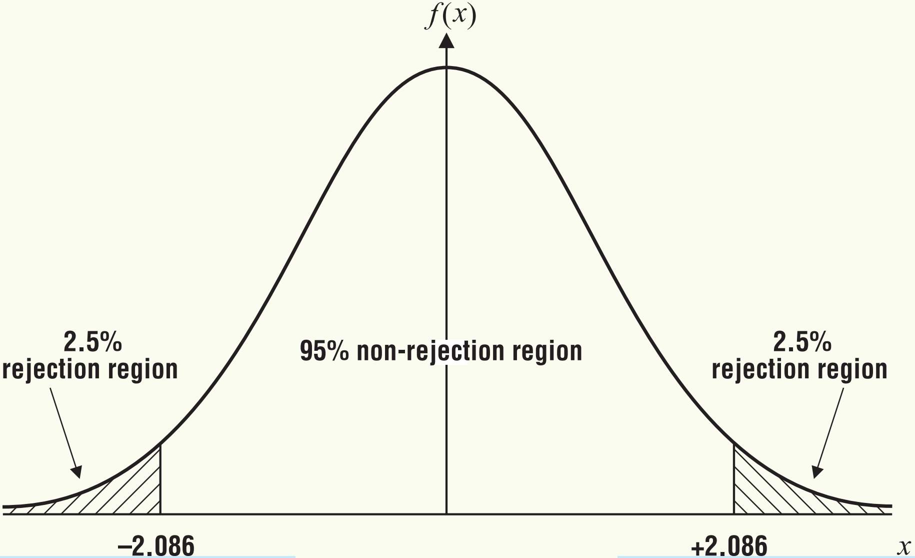
Connection between *t* and Normal Distributions

- 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

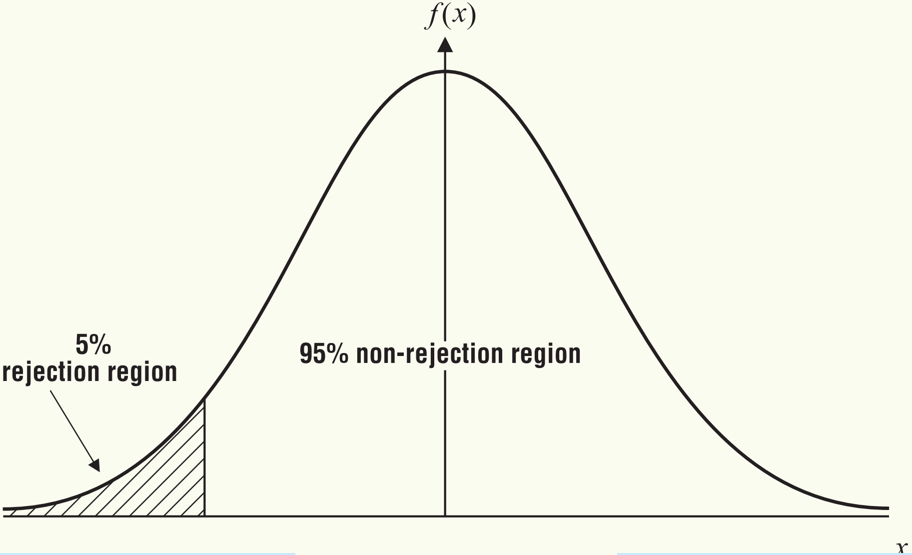
Significance level	t_∞	t_{40}	t_4
50%	0	0	0
5%	1.64	1.68	2.13
2.5%	1.96	2.02	2.78
0.5%	2.57	2.70	4.60

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

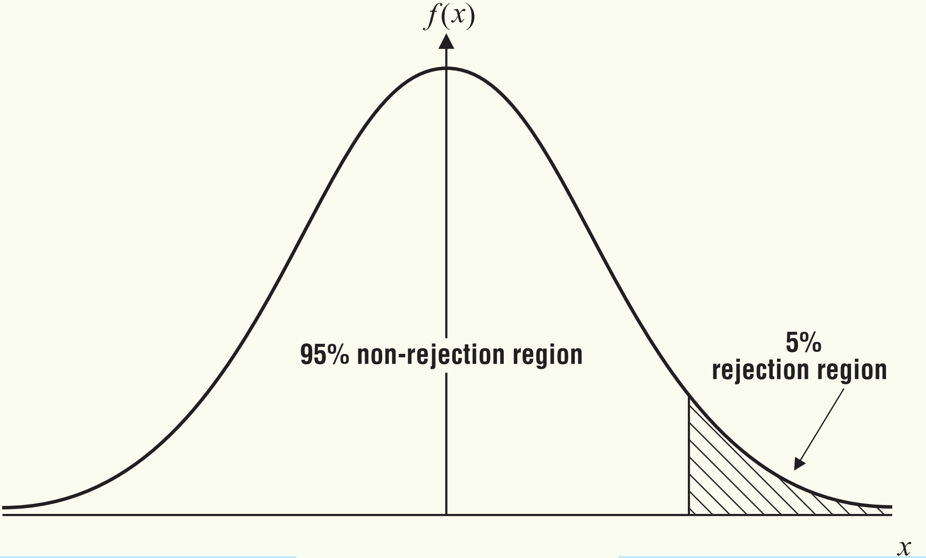
Rejection Regions for Two-Tailed Test



Rejection Region for One-Sided Lower Tail Test



Rejection Region for One-Sided Upper Tail Test



Another Example: Estimates

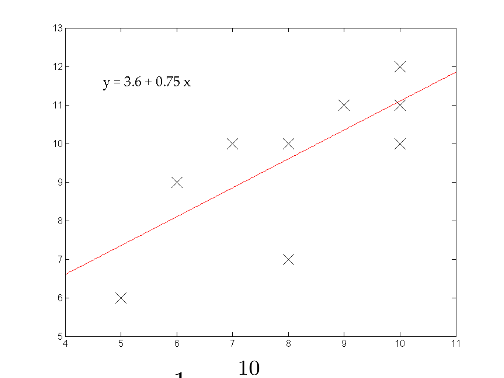
Let $X_i^* := X_i - \overline{X}$, and $Y_i^* := Y_i - \overline{Y}$.

Observation	X_i	Y_i	X_i^*	Y_i^*	X_i^{*2}	Y_i^{*2}	$X_i X_i^*$	$X_i Y_i^*$	$X_i^* Y_i^*$	
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

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

$$Y_i = 3.6 + 0.75 X_i$$

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**,

$$\lim_{n\to\infty} \overline{X}_n = \mu_X \,, \qquad \lim_{n\to\infty} \overline{Y}_n = \mu_Y$$

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

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

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

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

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

Consistent Properties of OLS

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

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

$$\Longrightarrow \qquad b = \frac{\mathbb{C}(X_i, Y_i)}{\mathbb{V}(X_i)}$$

⋈ Hence from (2), $\lim\limits_{n\to\infty} \widehat{b} = b$.

⋈ Implications:

- OLS $\widehat{b}$ estimator is consistent: $\lim\limits_{n\to\infty} \widehat{b} = b$
- OLS $\widehat{a}$ estimator is consistent: Since $\widehat{a} = \overline{Y} - \widehat{b}\,\overline{X}$,

$$\lim_{n\to\infty} \widehat{a} = \mu_Y - b\,\mu_X = a$$

Decomposition

- Consider

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

- **TSS = ESS + RSS**

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

- ESS can be expressed as

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

Coefficient of Determination

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

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

- 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 $:= r_{XY}^2 \dfrac{S_Y^2}{S_X^2} \times n S_X^2 = r_{XY}^2 n S_Y^2$.
- Coefficient of determination R^2 is defined as

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

Illustrative Example: Goodness of Fit

- From Slides 35 and 46, we can compute the following quantities
- Sample correlation coefficient: ______________
- TSS: ______________
- ESS: ______________
- RSS: ______________
- R^2: ______________

Forecasting: Point Estimate

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

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

- 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

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

Forecasting Error

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

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

§ The forecast error conditional on X_{n+1} is normally distributed.

§ **The OLS forecast is unbiased**:

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

Properties of the OLS Forecast

▷ Variance of the OLS Forecast

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

▷ The t-statistic of the forecast

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

Point Forecast and Confidence Interval

- The point forecast is

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

- 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

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

Illustrative Example: Forecast

Continuing from Slides 35 and 46, suppose $X_{11} = 2$.

- The forecast $\widehat{Y}_{11}$: ________________

- The forecast standard error: $\widehat{Y}_{11}$: ________________

- At the 95% confidence level, the forecast lower bound: ________________

- At the 95% confidence level, the forecast upper bound: ________________

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:

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

Solution to Hedging

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

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

- The risk-minimizing “optimal” hedge is to short

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

- Estimation: Run the following simple linear regression

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

- Since $b = \dfrac{\mathbb{C}\big(\Delta S_t\,,\,\Delta F_t\big)}{\mathbb{V}\big(\Delta F_t\big)}$, 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!