SMU
SINGAPORE MANAGEMENT UNIVERSITY

Lee Kong Chian School of Business

QF 604 ECONOMETRICS OF FINANCIAL MARKETS

LECTURE 1

Professor Lim Kian Guan

SMU Master of Science in Quantitative Finance 2023-2024 T2

LECTURE OUTLINE

01 PROBABILITY DISTRIBUTIONS

02 ORDINARY LEAST SQUARES

03 STOCK INDEX FUTURES

04 HEDGING

01 Probability Distributions

- Normal probability density function (pdf) : $\frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$
- R.V. X is lognormally distributed if $\ln(X) = Y \sim N(\mu, \sigma^2)$

$$E(X) = E(e^Y) = e^{\mu+1/2\sigma^2}$$
$$Var(X) = e^{2\mu}\left[e^{2\sigma^2} - e^{\sigma^2}\right]$$

- Let $X \sim N(0,1)$ be a standardized or unit normal RV. Then RV X^2 has a chi-square distribution with one degree of freedom, i.e. χ_1^2. If $X_1, X_2, \dots, X_k$ are k-independent unit normal RVs, then $\sum_{i=1}^{k} X_i^2$ has a chi-square distribution with k degrees of freedom, i.e. χ_k^2 . The degree of freedom here refers to the number of independent observations in the sum of squares. If Y^2 and Z^2 are two independent chi-square RVs with k_1 and k_2 degrees of freedom, then their sum $Y^2 + Z^2$ is $\chi_{k_1+k_2}^2$. A chi-square RV is always positive, and its pdf has a right skew. The mean of χ_k^2 is k, and its variance is $2k$.
- **Lemma 1.1.** If $X \sim N(\mu, \sigma^2)$, X_i for $i = 1, 2, \dots, n$ are independently identically distributed RV draws of X, and $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, then

$$(n-1)\left(\frac{S^2}{\sigma^2}\right) \sim \chi_{n-1}^2$$

01 Related Distributions

- Note that $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. Another related distribution is obtained when this unit normal RV is divided by S/σ .

  Then we obtain the RV $\frac{\bar{X}-\mu}{S/\sqrt{n}}$. Since $S/\sigma \sim \sqrt{\frac{\chi^2_{n-1}}{n-1}}$, then $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)/\sqrt{\frac{\chi^2_{n-1}}{n-1}}$.

  - The normal RV in the numerator and the chi-square RV in the denominator are statistically independent.
  - $N(0,1)/\sqrt{\frac{\chi^2_{n-1}}{n-1}} \sim t_{n-1}$

- From Lemma 1.1, taking expectations, $E[(n-1)S^2/\sigma^2] = n-1$. Hence, $E[S^2] = \sigma^2$. Therefore, $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the unbiased estimator of σ^2 .

- Suppose X, Y are independent RVs, $X \sim N(\mu X, \sigma^2)$ and $Y \sim N(\mu Y, \sigma^2)$. Observations x_i and y_i are independently drawn from these two RVs. Unbiased sample variances: $S_X^2 = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}$, $S_Y^2 = \sum_{i=1}^{m} \frac{(Y_i - \bar{Y})^2}{m-1}$, their ratio is $\frac{S_X^2}{S_Y^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}{\sum_{j=1}^{m}(Y_i - \bar{Y})^2/(m-1)} \sim F_{n,m}$.

  - Interesting relationship: $t_k^2 = F_{1,k}$

01 Estimation

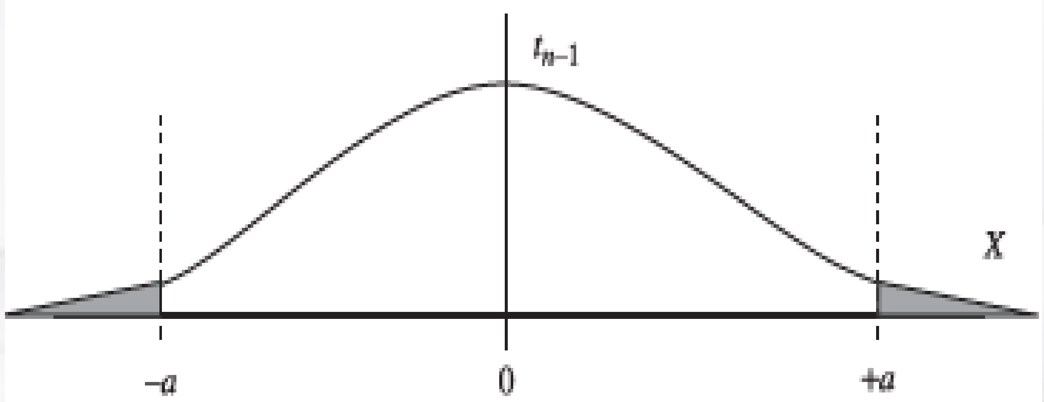
- **Point estimate of $\boldsymbol{\mu}$:** This is the sample mean of a random sample of size n =100 drawn from $X \sim N(\mu, \sigma^2)$, $\bar{x} = 0.08$. Suppose **variance of X is known, $\boldsymbol{\sigma^2 = 0.25}$**, then $z = \frac{\bar{X}-\mu}{0.5/\sqrt{100}} = \frac{\bar{X}-\mu}{0.05} \sim N(0,1)$
  - Suppose $a > 0$ such that $P(-a \le z \le +a) = 95\%$. $P(-a \le z) = 97.5\%$ and $P(z \le +a) = 97.5\% \Rightarrow a = +1.96$
  - Then, $P\left(-1.96 \le \frac{\bar{x}-\mu}{0.05} \le +1.96\right) = 0.95$
  - Rearranging: $P(\bar{x} - 1.96(0.05) \le \mu \le \bar{x} + 1.96(0.05)) = 0.95 \;\Rightarrow P(-0.018 \le \mu \le 0.178) = 0.95$
  - Interval estimate of μ at 95% confidence level is (−0.018, 0.178). Given a sample, there is 95% probability or chance that the true μ is contained in the estimated interval.
- In the above, suppose **var(X) is not known**. Suppose S^2 is the unbiased variance estimator, then $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
  - Suppose $a > 0$ such that $P(-a \le t_{n-1} \le +a) = 95\%$. $P(-a \le t_{99}) = 97.5\%$ and $P(t_{99} \le +a) = 97.5\% \Rightarrow$

    $a = +1.9842$ (found from a t-distribution table). Then $P\left(-1.9842 \le \frac{\bar{x}-\mu}{S/\sqrt{100}} \le +1.9842\right) = 0.95$
  - Rearranging: $P(\bar{x} - 1.9842(S/10) \le \mu \le \bar{x} + 1.9842(S/10)) = 0.95$
  - Suppose $\bar{x} = 0.08, S^2 = 0.36$, then $P(-0.03905 \le \mu \le 0.19905) = 0.95$
  - 95% confidence interval estimate of μ is (− 0.03905, 0.19905).

01 Hypothesis Testing

- Suppose the test statistic $\sim t_{n-1}$. Statistical Hypotheses are: $H_0: \mu = 1, H_A: \mu \neq 1$
  - Define the critical region (shaded): $\{t_{n-1} < -a$ or $t_{n-1} > +a\}$ for $a > 0$
  - If test statistic falls within critical region, H_0 is rejected in favour of H_A. Null hypothesis H_0 is "named" typically in situations if the statistician expects null/no effect. Not rejecting gives strong credence to the hypothesis. Otherwise H_0 is not rejected and is "accepted"
    - If H_0 is true, the probability of rejecting H_0 would be the area of the critical region, say 5% in this case
    - t-distribution is symmetrical ⇒ each of the right and left shaded tails has an area of 2.5%
    - This is a two-tailed test with a significance level of 5%

- **Type I error** (rejecting H_0 when it is true) is the significance level
- **Type II error** (not rejecting H_0 when it is false) is (1– significance level)
- ***p*-value** is 2 × the probability under H_0 of $|t_{99}|$ exceeding |sample t-value| in a two-tailed test
  - p-value < test significance level, reject H_0
  - Otherwise H_0 cannot be rejected

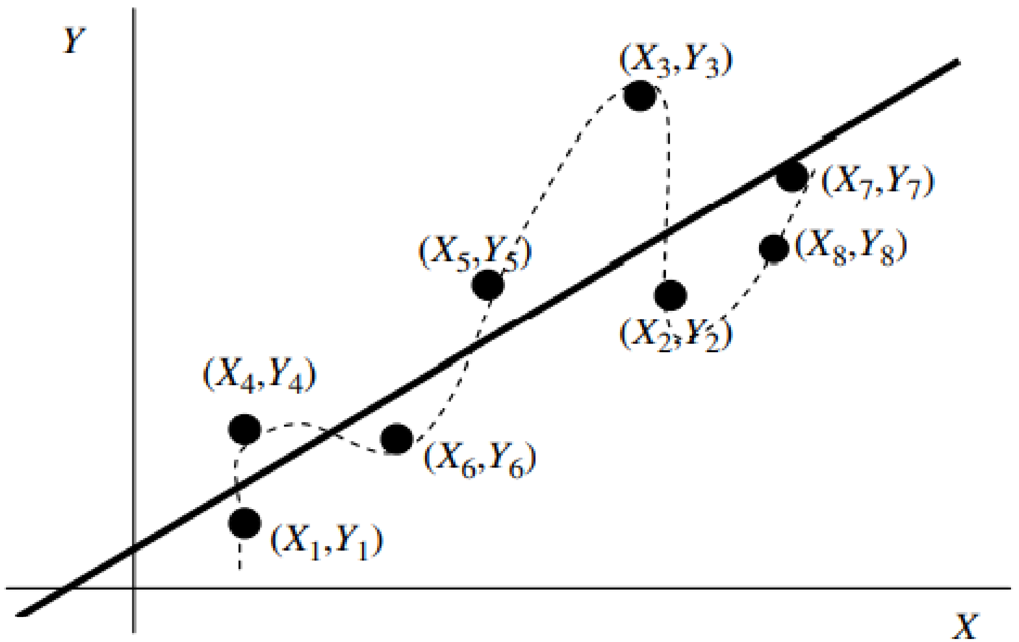


02 Simple Linear Regression

- A regression is an association between a dependent variable and other explanatory variables.

- Two major purposes:
    - provide some positive theory of how the dependent variable could be explained by other variables
    - a normative or prescriptive theory of how to use the association to predict future occurrences of the dependent variable

- The 8-parameter model is an over-fitted model with no clear economic sense.



An Association between Variables X and Y

02 Simple Linear Regression

- The idea of a regression model (need not be linear):

$$Y = f(X;\theta) + \varepsilon$$

where ε is a random error or noise, is one where parameter(s) θ are suitably estimated as $\hat{\theta}$.

  - $\hat{\theta}$ is "close to" true θ given a sample of $\{(X_i, Y_i)\}_{i=1,2,\ldots,n}$, size n, such that $\sum_{i=1}^{n} g\left[Y_i - f\left(X_i;\hat{\theta}\right)\right]$ is small in some statistical sense where $g(.)$ is a criterion function.

- A linear (bivariate) regression model is:

$$Y_i = a + bX_i + e_i$$

  - where a and b are constants.
  - Y_i : the dependent variable or regressand
  - X_i : the explanatory variable or regressor
  - e_i : a residual noise, disturbance, or innovation

02 Ordinary Least Squares(OLS)

- In a two-variable or simple linear regression model :

$$Y_i = a + bX_i + e_i\ , \quad i = 1, 2, \ldots, N. \tag{2.1}$$

  - The random variables Y_i , X_i's are observed as sample bivariate points
  - Disturbances or residual errors e_i's are not observed. $E(e_i\ ) = 0$, and $\mathrm{Var}(e_i\ )$ is assumed to be a constant σ_e^2, which is also not observed.
  - a, b are constants to be estimated.

- **Task**: Estimate parameters a , b and σ_e^2.
- The **classical assumptions** (desirable conditions) for OLS regression are:

(A1) $E(e_i) = 0$ for every i

(A2) $E\left(e_i^2\right) = \sigma_e^2$, a same constant for every i.

(A3) $E\left(e_i e_j\right) = 0$ for every $i \neq j$.

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

02 Ordinary Least Squares(OLS)

- In addition to assumptions (A1) through (A4), we could also add a distributional assumption to the random variables, e.g. (A5) $e_i \sim N(0, \sigma_e^2)$
- The scalar value $\hat{e}_i$ indicates measure of the vertical distance between the point (X_i, Y_i) and the fitted regression line. In Figure 2.5, the solid line provides a better fit than the dotted line.
- The OLS method of estimating a and b is to find $\hat{a}$ and $\hat{b}$ so as to minimize the residual sum of squares (RSS), $\sum_{i=1}^{N} \hat{e_i}^2$, which is different from minimizing the sum of squares of random variables e_i (We do not observe the e_i's.)

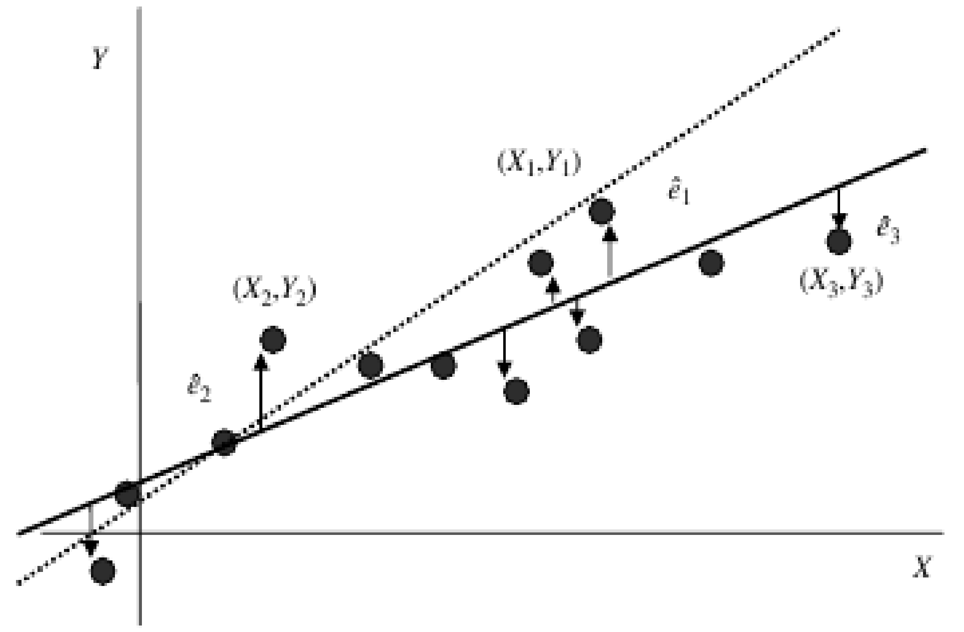


Figure 2.5: OLS Regression of Observations (X_i, Y_i)

02 Ordinary Least Squares(OLS)

- The key criterion in OLS: $\min\limits_{\hat{a},\hat{b}} \sum_{i=1}^{N} \hat{e_i}^2 \equiv \sum_{i=1}^{N} \left(Y_i - \hat{a} - \hat{b}X_i\right)^2$.
- The First Order Conditions(FOC) yield the following two equations:

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

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

- From the above two normal equations, we can obtain:
  - $$\hat{a} = \bar{Y} - \hat{b}\bar{X} \qquad (2.2)$$

  It also shows that the fitted OLS line $\bar{Y} = \hat{a} + \hat{b}\bar{X}$ passes through $(\bar{X}, \bar{Y})$, the “centroid".

  - $$\hat{b} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})Y_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} \qquad (2.3)$$

  where $x_i = X_i - \bar{X}$ and $y_i = Y_i - \bar{Y}$.

02 Properties of OLS estimators

- From Eq.(2.1), with stochastic Y_i and X_i, $cov(e_i, X_i) = 0$, then

$$cov(Y_i, X_i) = cov(a + bX_i + e_i, X_i) = b\,\mathrm{var}(X_i) + cov(e_i, X_i) = b\,\mathrm{var}(X_i)$$

In terms of population moments, $b = cov(Y_i, X_i)/\mathrm{var}(X_i)$.

- But in Eq.(2.3), $\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} = S_{XY}/S_X^2$

Thus, when $N \to \infty$ for large sample, $\lim_{N\to\infty} \hat{b} = b$ under conditions of stationarity.

If $\hat{b}$ is an estimator of b, and $\lim_{N\to\infty} \hat{b} = b$, $\hat{b}$ is said to be a consistent estimator.

We can show likewise that $\lim_{N\to\infty} \hat{a} = a$, and hence $\hat{a}$ is also consistent.

02 Properties of OLS estimators

Important question is: what are the desirable properties of the OLS estimators, assuming (A1) to (A5), in finite sample?

- $\hat{b} = \sum_{i=1}^{N}\left(x_i / \sum_{j=1}^{N} x_j^2\right)(a + bX_i + e_i) = b + \sum_{i=1}^{N}\left(x_i / \sum_{j=1}^{N} x_j^2\right) e_i \qquad (2.4)$

Taking unconditional expectation and invoking (A4) and (A1), then

$$E(\hat{b}) = b + \sum_{i=1}^{N} E\left(x_i / \sum_{j=1}^{N} x_j^2\right) E(e_i) = b \text{ since } E(e_i) = 0.$$

- $\hat{a} = \bar{Y} - \hat{b}\bar{X}$

$$= \bar{Y} - \left(b + \sum_{i=1}^{N}\left(x_i / \sum_{j=1}^{N} x_j^2\right) e_i\right)\bar{X}$$

$$= \bar{Y} - b\bar{X} - \sum_{i=1}^{N}\left(x_i\bar{X} / \sum_{j=1}^{N} x_j^2\right) e_i$$

Taking unconditional expectation and invoking (A4) and (A1), then $E(\hat{a}) = E(\bar{Y}) - bE(\bar{X})$. Since $E(Y_i) = a + bE(X_i)(\forall_i)$, then $E(\bar{Y}) = a + bE(\bar{X})$. Hence $E(\hat{a}) = a$.

02 Properties of OLS estimators

- Using assumptions (A1) to (A4), the variance of $\hat{b}$ is

$$E\left(\hat{b}-b\right)^2 = E\left[\sum_{i=1}^{N}\left(x_i/\sum_{j=1}^{N}x_j^2\right)e_i\right]^2$$

$$=\sum_{i=1}^{N}\left[E\left(x_i/\sum_{j=1}^{N}x_j^2\right)^2 E\left(e_i^2\right)\right] = \sigma_e^2\sum_{i=1}^{N}E\left[x_i^2/\left(\sum_{j=1}^{N}x_j^2\right)^2\right] = \sigma_e^2\,E\left[1/\left(\sum_{j=1}^{N}x_j^2\right)\right]$$

Note: the unconditional variance involves expectations of functions of RVs X_1, X_2, and so on. These expectations are complicated moments of X, even as the distribution of X itself is not specified. It is a similar situation in the case of the variance of $\hat{a}$.

02 Properties of OLS estimators

For inferences and computations of estimator variances in OLS, the conditional distributions of the estimators given X (realized X_1, X_2, ….. , X_N) are used.

From Eq.(2.4), conditional mean of $\hat{b}$ is

- $E[\hat{b}|X] = b + \sum_{i=1}^{N}\left(x_i / \sum_{j=1}^{N} x_j^2\right) E[e_i|X] = b$

Conditional Variance of $\hat{b}$ is

- $$E\left[\left(\hat{b}-b\right)^2|X\right] = \sum_{i=1}^{N}\sum_{j=1}^{N}\left(x_i/\sum_{k=1}^{N} x_k^2\right)\left(x_j/\sum_{k=1}^{N} x_k^2\right) E\left[e_i e_j|X\right]$$
  $$= \sum_{i=1}^{N}\left(x_i/\sum_{k=1}^{N} x_k^2\right) E\left[e_i^2|X\right]$$
  $$= \sigma_e^2/\sum_{i=1}^{N} x_i^2 \qquad (2.5)$$

02 Properties of OLS estimators

From Eq.(2.2), conditional mean of $\hat{a}$ is

- $E[\hat{a}|X] = E[\bar{Y}|X] - \bar{X}E[\hat{b}|X] = (a + b\bar{X}) - \bar{X}b = a$

Conditional Variance of $\hat{a}$ is

- $Var[\hat{a}|X] = Var(\bar{Y}|X) - 2\bar{X}cov(\bar{Y}, \hat{b}|X) + \bar{X}^2 var(\hat{b}|X)$

$$= \frac{\sigma_e^2}{N} - 2\frac{\bar{X}}{N}cov + \left(\sum_{i=1}^{N} a + bX_i + e_i\, , b + \sum_{i=1}^{N} \frac{x_i e_i}{\sum_{j=1}^{N} x_j^2} |X\right) \bar{X}^2\, \sigma_e^2 / \sum_{i=1}^{N} x_i^2$$

$$= \sigma_e^2 \left(\frac{1}{N} + \bar{X}^2 / \sum_{i=1}^{N} x_i^2\right) \qquad (2.6)$$

since the middle term gives $\sum_{i=1}^{N} cov(e_i\, , \frac{x_i}{\sum_{j=1}^{N} x_j^2}\, e_i \mid X) = 0$ as $\sum_{i=1}^{N} x_i = 0$.

Variance of $\hat{a}$ and $\hat{b}$ conditional on X reduce toward 0 as $N \to \infty$.

02 Properties of OLS estimators

In Summary:

- Classical assumptions (A1) to (A4) apply under the general context of stochastic X and e, in finite sample, OLS method → unbiased estimators $\hat{a}$ and $\hat{b}$.
- The unbiasedness property holds under both unconditional expectations and expectations conditional on X.
- As sample size N ↑ : estimates is accurate and forecasting can be done
- However, testing if the estimates are significantly different from some hypothesized values requires use of conditional on X distributions of the estimators in order to obtain conveniently computed conditional variances. Together with (A5), statistical inferences can be made.

02 Forecasting

In forecast or prediction of Y_i based on an observed X_i , a useful result is that the best forecast of Y_i , in the sense of minimum mean square error, is the conditional expectation $E(Y_i|X_i)$.

Lemma: For a bivariate distribution of RVs Y_i and X_i , for any function $g(X_i)$

$$E\left(Y_i - g(X_i)\right)^2 \geq E\left(Y_i - E(Y_i|X_i)\right)^2$$

Note: $E(Y_i|X_i)$ is not necessarily a linear forecast.

Proof is shown at the end of p.38 of the book.

02 Forecasting

- The best forecast of Y_{N+1} with linear regression model is $E(Y_{N+1}) = a + bX_{N+1}$
- OLS forecast of Y_{N+1} is $\hat{Y}_{N+1} = \hat{a} + \hat{b}X_{N+1}$
  - The forecast or prediction error:

$$Y_{N+1} - \hat{Y}_{N+1} = bx_{N+1} - \hat{b}x_{N+1} + e_{N+1} - \frac{1}{N}\sum_{i=1}^{N} e_i = -(\hat{b} - b)x_{N+1} + e_{N+1} - \frac{1}{N}\sum_{i=1}^{N} e_i$$

  - $Var(Y_{N+1} - \hat{Y}_{N+1}|x_{N+1}) = x_{N+1}^2 Var(\hat{b}) + \sigma_e^2 + \frac{1}{N}\sigma_e^2 = \sigma_e^2\left(1 + \frac{1}{N} + \frac{x_{N+1}^2}{\sum_{i=1}^{N} x_i^2}\right)$

    Note: $Var(\hat{b})$ is a conditional variance.

  - The forecast error (conditional on x_{N+1}) under (A5) is normally distributed:

$$\frac{Y_{N+1} - \hat{Y}_{N+1}}{\hat{\sigma}_e\sqrt{\left(1 + \frac{1}{N} + \frac{x_{N+1}^2}{\sum_{i=1}^{N} x_i^2}\right)}} \sim t_{N-2}$$

02 Gauss-Markov Theorem

- Important result that justifies use of OLS in $Y_i = a + b X_i + e_i$
- Let linear estimators of a and b take the form:

$$\hat{A} = \sum_{i=1}^{N} \theta_i(X) Y_i \quad \text{and} \quad \hat{B} = \sum_{i=1}^{N} \gamma_i(X) Y_i$$

where θ_i and γ_i are fixed or deterministic functions of X only (and not in Y's or a or b).

Note that these estimators are expressed as linear functions (conditional on X) of Y

- Gauss-Markov theorem states that amongst all linear and unbiased (conditional on X) estimators of the form in $\hat{A}$ and $\hat{B}$ of $Y_i = a + b X_i + e_i$, the OLS estimators $\hat{a}, \hat{b}$ have the minimum variances, i.e.

$$Var(\hat{a}) \leq Var(\hat{A})$$

$$Var(\hat{b}) \leq Var(\hat{B})$$

given the assumptions (A1) to (A4).

Proof is shown on pp. 40 – 43 of the book.

02 Characterization of OLS estimators

- Conditional on X, $\hat{b}$ is a linear estimator with w_i's as fixed weights (function of X) on the Y_i's.

$$\hat{b} = \sum_{i=1}^{N} w_i Y_i$$

where $w_i = x_i / \sum_{i=j}^{N} x_j^2$.

- Similarly, conditional on X, $\hat{a} = \bar{Y} - \hat{b}\bar{X}$ is a linear function of Y.
  The weights for $\hat{a}$, v_i's, are as follows.

$$\hat{a} = \frac{1}{N}\sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} w_i \bar{X} Y_i = \sum_{i=1}^{N} v_i Y_i$$

where $v_i = \left(\frac{1}{N} - \frac{x_i \bar{X}}{\sum_{i=1}^{N} x_i^2}\right)$.

02 Characterization of OLS estimators

- Probability distribution of $\hat{b}$:

$$\hat{b} \sim N\left(b, \sigma_e^2\left(\frac{1}{\sum x_i^2}\right)\right)$$

- Probability distribution of $\hat{a}$:

$$\hat{a} \sim N\left(a, \sigma_e^2\left(\frac{1}{N} + \frac{\bar{X}^2}{\sum x_i^2}\right)\right)$$

- The Covariance between the estimators $\hat{a}$ and $\hat{b}$:

$$Cov(\hat{a}, \hat{b}) = E(\hat{a} - a)(\hat{b} - b) = E\left(\sum v_i e_i\right)\left(\sum w_i e_i\right) = \sigma_e^2 \sum v_i w_i = \sigma_e^2\left(-\frac{\bar{X}^2}{\sum x_i^2}\right)$$

02 Characterization of OLS estimators

- The estimated residual:

$$\hat{e}_i = Y_i - \hat{a} - \hat{b}X_i = Y_i - \left(\bar{Y} - \hat{b}\bar{X}\right) - \hat{b}X_i = (Y_i - \bar{Y}) - \hat{b}(X_i - \bar{X}) = y_i - \hat{b}x_i$$

- Important : distinguish estimated residual $Y_i - \hat{Y}_i$ from the actual unobserved e_i

$$\sum_{i=1}^{N} \hat{e}_i = 0 \quad \text{and} \quad \sum_{i=1}^{N} X_i \hat{e}_i = 0$$

- With the Gauss- Markov theorem, OLS estimators under the classical conditions are called best linear unbiased estimators (BLUE) for the linear regression model.
- Some other estimators like the Stein estimators can be more efficient but are biased.

02 Decomposition

- Residual Sum of Squares(RSS): RSS = $\sum \hat{e}_i^2 = \sum\left(Y_i - \hat{Y}_i\right)^2$
- Explained Sum of Squares (ESS): ESS = $\sum\left(\hat{Y}_i - \bar{Y}\right)^2 = r_{XY}^2 N S_Y^2$
- Total Sum of Squares (TSS): TSS = ESS+RSS = $\sum(Y_i - \bar{Y})^2 = N S_Y^2$
- The coefficient of determination: ESS/TSS = $R^2 = r_{XY}^2$
- $R^2 = r_{XY}^2 = 1 - \mathrm{RSS/TSS}$, $0 \le R^2 \le 1$

R^2 determines the degree of fit of the linear regression line to the data points in the sample.

02 Statistical Inference of Parameters

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \sim N\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} \sigma_e^2 \left(\frac{1}{N} + \frac{\bar{X}^2}{\sum x^2} \right) & -\sigma_e^2 \left(\frac{\bar{X}^2}{\sum x^2} \right) \\ -\sigma_e^2 \left(\frac{\bar{X}^2}{\sum x^2} \right) & \sigma_e^2 \left(\frac{1}{\sum x^2} \right) \end{pmatrix} \right)$$

So, $\frac{\hat{b}-b}{s.e.(\hat{b})} \equiv Z \sim N(0,1)$.

- For testing null hypothesis $H_0: b = 1$, employ sample estimate of σ_e^2 using $\hat{\sigma}_e^2$. Use

$$t_{N-2} = \frac{\hat{b} - 1}{\hat{\sigma}_e \sqrt{\frac{1}{\sum x^2}}}$$

- For testing null hypothesis $H_0: a = 0$, use

$$t_{N-2} = \frac{\hat{a} - 0}{\hat{\sigma}_e \sqrt{\frac{1}{N} + \frac{\bar{X}^2}{\sum x^2}}}$$

03 Stock Index Futures

Let's take the Nikkei 225 as an example.

- A December Nikkei 225 Index futures contract
- The traded "price" in September → an index price or a notional price
  - September: a notional price of 12,000 → buy/long *N* of the Nikkei 225 Index December futures contracts
  - December: a notional price of 14,000
  - A big profit: 2000 points × Yen value per point per contract × number of contracts *N*
- At Maturity: the index futures notional price converges to the underlying Nikkei 225 stock index number

03 Cost of Carry Model

Assumption: a large diversified portfolio of major Japanese stocks has the same return as N225 index portfolio return

- $\boldsymbol{S_t}$: the stock index value at current time *t*
- $\boldsymbol{r_{t,T}}$: the effective risk-free interest rate or the cost of carry over [*t*, *T*]
- $\boldsymbol{\alpha}$: a constant multiplier reflecting proportionate relationship of the portfolio value to the index notional value $\boldsymbol{S_t}$
- $\boldsymbol{\alpha S_t}$: cost of this portfolio
- $\boldsymbol{\alpha S_t(1 + r_{t,T})}$: the portfolio cost if held till maturity T
- $\boldsymbol{D}$: an aggregate amount of dividends over the period [*t*, *T*]
- $\boldsymbol{d \times \alpha S_t}$: the dividends issued to the arbitrageur's portfolio amount, assume this is perfectly anticipated. $\boldsymbol{d^*} \times \boldsymbol{\alpha S_t}$: the present value of these dividends
- $\boldsymbol{D^*} = \boldsymbol{d^* S_t}$
- Net cost of carry of the stocks as at time T, $\boldsymbol{\alpha[S_t - D^*](1 + r_{t,T})}$

03 Cost of Carry Model

- N225 index futures notional price for contract with maturity at T, at t is $F_{t,T}$
- At t, **if** $F_{t,T} > [S_t - D^*](1 + r_{t,T})$:
  (1) The arbitrageur would enter a buy or long position in the stocks.
  (2) The arbitrageur sells an index futures contract at notional price $F_{t,T}$.
- Assume: the currency value per point per contract is 1. Also assume $\alpha = 1$.
- At T, $S_T = F_{T,T}$, the arbitrageur
  (1) Sell the portfolio at ¥ S_T in Yen, gaining $\boldsymbol{S_T - [S_t - D^*](1 + r_{t,T})}$
  (2) Cash-settle the index futures trade, gaining $\boldsymbol{F_{t,T} - F_{T,T}}$ **or** $\boldsymbol{F_{t,T} - S_T}$
  (3) The net gain (arbitrage profit) is the sum of the two terms:

$$\text{¥}\,\boldsymbol{F_{t,T} - [S_t - D^*](1 + r_{t,T}) > 0}$$

- Check the converse case if $F_{t,T} < [S_t - D^*](1 + r_{t,T})$

SMU Singapore Management University | Lee Kong Chian School of Business

03 Cost of Carry Model

- Suppose we add transaction costs of 0.5%. Let $\tau = T - t$ be the term-to-maturity in terms of fraction of a year. Assume

  $r_{t,T} \approx 0.005\tau$, *and* $D^* \approx 0.01\tau\, S_t/(1+0.005\tau)$

- Cash-and-carry arbitrage:

  $¥0.995S_T - [1.005S_t - D^*](1+r_{t,T}) + F_{t,T} - S_T$

  or approximately $F_{t,T} - 1.01\,[S_t - D^*](1+r_{t,T})$

  Fair price: $F_{t,T} = 1.01\,[S_t - D^*](1+r_{t,T})$

- Reverse Cash-and-carry arbitrage:

  $¥[0.995S_t - D^*](1+r_{t,T}) - 1.005S_T + S_T - F_{t,T}$

  or approximately $0.99[S_t - D^*](1+r_{t,T}) - F_{t,T}$

  Fair price: $F_{t,T} = 0.99[S_t - D^*](1+r_{t,T})$

- Implication: No arbitrage money could be made if $F_{t,T}$ falls within

  $\left(0.99[S_t - D^*](1+r_{t,T}),\, 1.01\,[S_t - D^*](1+r_{t,T})\right)$

For a given futures price $F_{t,T}$ (without transaction costs), fair price $F_t^* = [S_t - D^*](1+r_{t,T})$

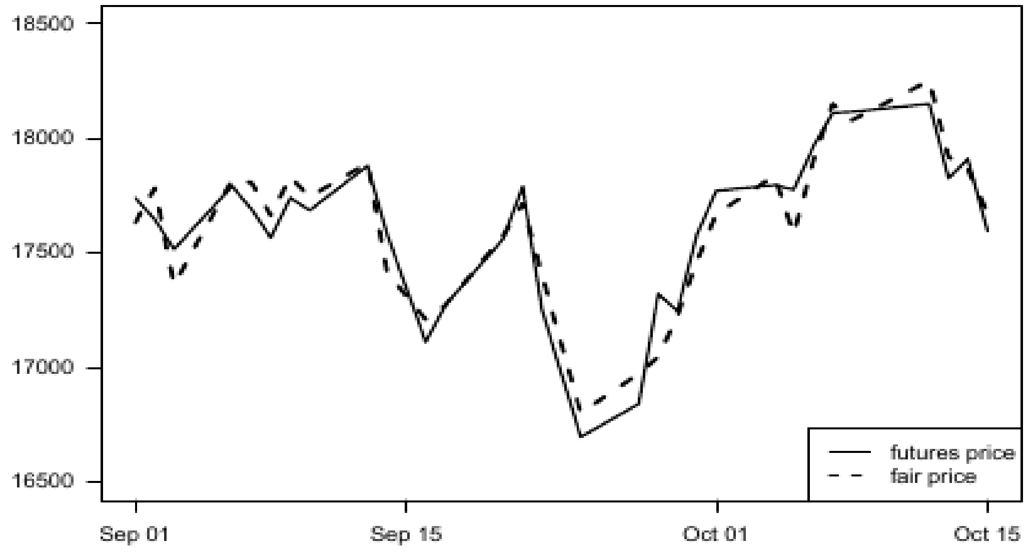


Figure 2.6: Prices of Nikkei 225 December Futures Contract from 9/1/1999 to 10/15/1999

03 Cost of Carry Model

Percentage difference or deviation from from fair price F_t^* is $p_t = (F_{t,T} - F_t^*)/F_t^*$

Figure 2.7 : possible negative serial correlat in the change of p_t within the no-arbitrage bounds.

- Most differences are within 1%.
- Practically little arbitrage, then
  (1) deviation of p_t from zero is due to random disturbances
  (2) change in p_t or Δp_t would display negative daily correlation

p_t

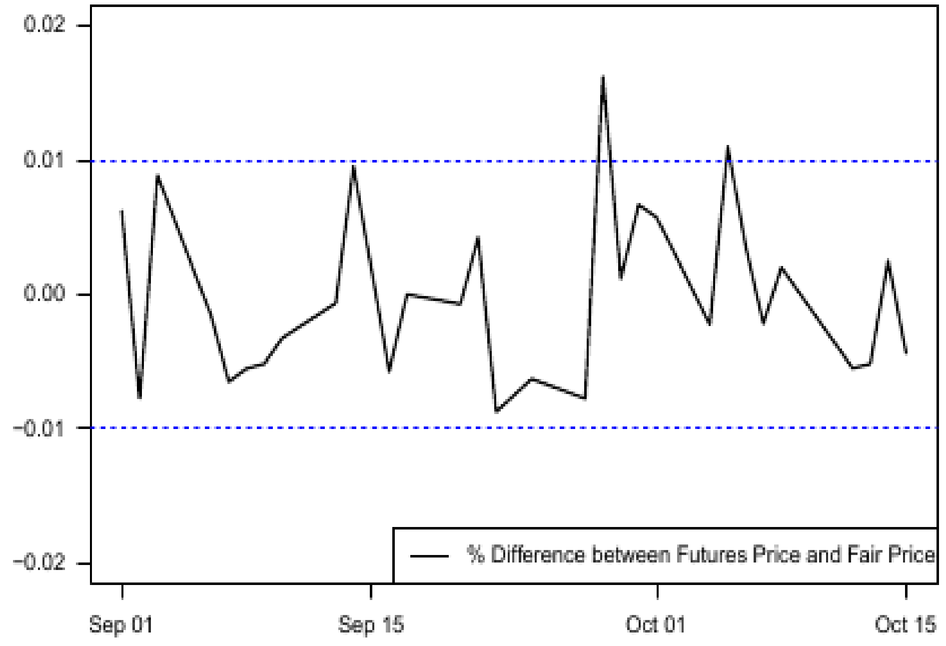


Figure 2.7: Percentage Difference between Futures Price and Fair Price

03 Cost of Carry Model

- Instead of p_t, we also can use the variable $\ln(F_t/F_t^*)$, which can be assumed to follow a normal distribution.

Table 2.1: Regression of $\Delta \ln(F_t/F_t^*)$ on Its Lag: $\Delta q_{t+1} = a + b\Delta q_t + e_{t+1}$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	0.00005	0.0014	0.032	0.9751
Lagged Δq_t	-0.57340	0.1534	-3.739	0.0009***
R-squared	0.3497	$F\,(d.f.\,1, 26)$-statistic		13.98
Adjusted R-squared	0.3247	Prob(F-statistic)		0.00092
S.E. of regression	0.0076	Sum squared resid.		0.00151

Note: *** indicates significance at the 2-tail 0.1% level.

- Estimated $\hat{a} = 0.00005$ with a two-tailed p-value of 0.9751 ⇒ cannot reject H_0: $a = 0$ at any reasonable significance level
- Estimated coefficient of lagged Δq_t, regression slope b, $\hat{b} = -0.57340$ a two-tailed p-value of 0.0009
- The statistical evidence: $b \neq 0$ (basically $b < 0$ in this case). The presence of reversals in Δq_t across days
- Coefficient of Determination: $R^2 = 0.3497$

  A common problem using R^2 is that its value increases when more regressors are added, regardless of whether the added variables help in fact to explain the regressand or not.

  To mitigate the issue, use

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(N-1)}{N-k-1} = 0.3247$$

03 Quantiles

- Quantiles: RV X values that divide the cumulative distribution function cdf of the RV into equal intervals.

  Deciles: if $\text{cdf}(X = x_1) = 0.10$, $\text{cdf}(X = x_2) = 0.20$, $\text{cdf}(X = x_3) = 0.30, \ldots, x_1, x_2, x_3 \ldots$ are called the deciles.

- Suppose there is a sample of 28 estimated residual errors $\hat{e}_j$ for j = 1,2,…,28.

  Order $\hat{e}_1 < \hat{e}_2 < \ldots < \hat{e}_{28}$. We want to check if $\hat{e}_j$ follows a normal distribution $N(\mu,\sigma^2)$.

- Assume $\hat{e}_j \sim N(\mu,\sigma^2)$. Then empirical cdf of $\hat{e}_j = j/28 = M(\hat{e}_j) = \Phi(z_j)$ where $z_j = (\hat{e}_j - \mu) / \sigma$.

  Hence $\hat{e}_j = M^{-1}(j/28)$, and $\hat{e}_j = \mu + \sigma\, \Phi^{-1}(j/28)$. Or, $M^{-1}(j/28) = \mu + \sigma\, \Phi^{-1}(j/28)$.

- Note cdf of z_j is $\Phi(z_j) = \int_{-\infty}^{z_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_j^2}$.

- The quantile-quantile (Q-Q) plot: a graphical technique for visually comparing two probability distributions by plotting their quantiles side by side. In above, $(\Phi^{-1}(j/28), M^{-1}(j/28))$ is a Q-Q plot.

  If $\hat{e}_j \sim N(\mu,\sigma^2)$ is true, then plot will lie on a straight line with intercept μ and slope σ.

03 Quantiles

Compute the Normal Q-Q plot of the estimated residuals from Table 2.1, the plot is shown in Figure 2.8.

- Two outliers with values 0.01162 and 0.02294

- Most of the other values of $\widehat{e}_t$ appear to fall on the straight-line → closeness to the normal distribution.

- Several points with $\Phi^{-1}(n/28)$ below -1 have $M^{-1}(n/28)$ values below the straight line → a fatter left tail than that of the normal distribution. For same low cdf, sample distribution has more negative quantiles.

- Similarly, Several points with $\Phi^{-1}(n/28)$ above +1 have $M^{-1}(n/28)$ values above the straight line → a fatter right tail. For same high cdf, sample distribution has more positive quantiles.

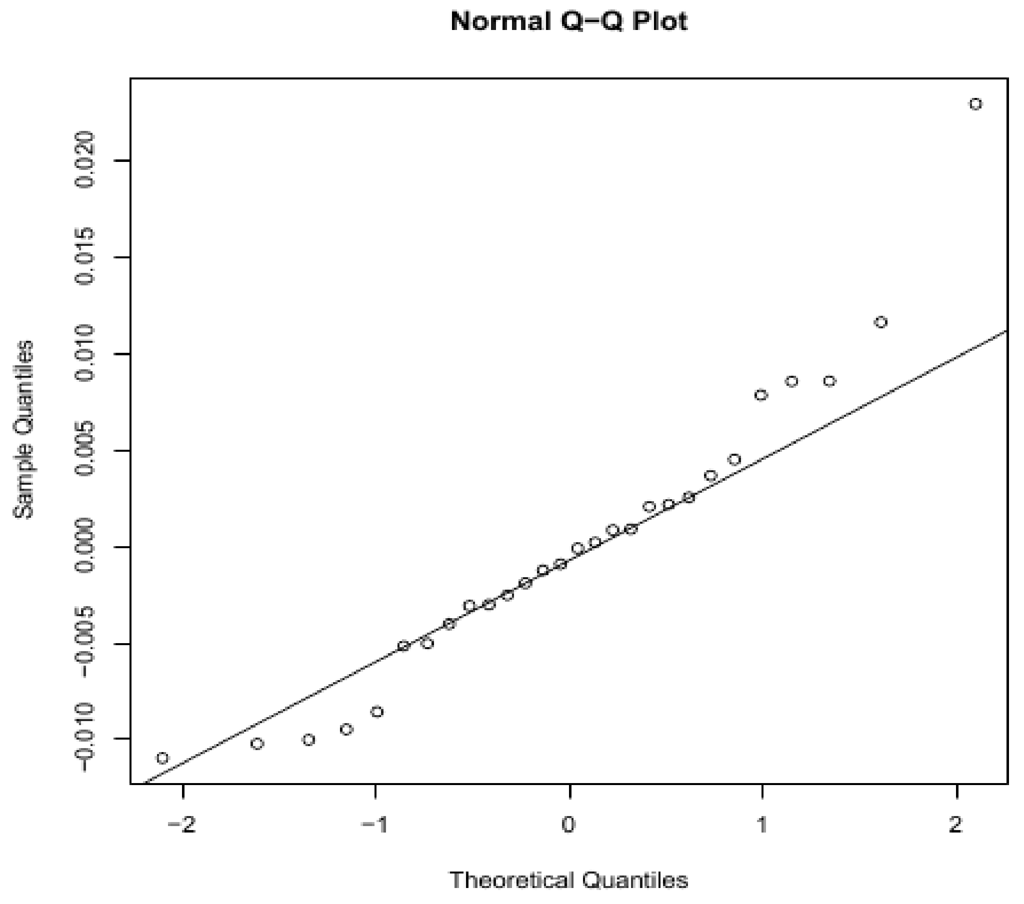


Figure 2.8: Normal Q-Q Plot of Residuals

04 Hedging

- $\boldsymbol{V_t}$: a total current value of the investor's original stock position
- $\boldsymbol{S_t}$: the N225 Index value $\boldsymbol{f}$: a constant factor $\rightarrow \boldsymbol{\Delta V_{t+1} = f \Delta S_{t+1}}$
- $\boldsymbol{\Delta V_{t+1}/V_t = \Delta S_{t+1}/S_t}$: portfolio return rate

- Form a hedged portfolio comprising ¥$\boldsymbol{P_t}$ = ¥ $\boldsymbol{f} \times \boldsymbol{S_t}$, and $\boldsymbol{h}$ number of short positions in N225 Index futures contracts, with maturity **T**, notional traded price $\boldsymbol{F_{t,T}}$, actual price value of 500 $\times$ $\boldsymbol{F_{t,T}}$ Yen
- At the end of the risky period, **T**, hedged portfolio in Yen value change would be:

$$\Delta \boldsymbol{P_{t+1}} = \boldsymbol{f} \times \Delta \boldsymbol{S_{t+1}} - \boldsymbol{h} \times \mathbf{500} \times \Delta \boldsymbol{F_{t+1,T}}$$

where $\Delta \boldsymbol{P_{t+1}} = \boldsymbol{P_{t+1}} - \boldsymbol{P_t}$, $\Delta \boldsymbol{S_{t+1}} = \boldsymbol{S_{t+1}} - \boldsymbol{S_t}$, $\Delta \boldsymbol{F_{t+1,T}} = \boldsymbol{F_{t+1,T}} - \boldsymbol{F_{t,T}}$

- $\boldsymbol{Var(\Delta P_t) = f^2 \times Var(\Delta S_t) + h^2 \times 500^2 \times Var(\Delta F_{t,T}) - 2h \times 500f \times Cov(\Delta S_t, \Delta F_{t,T})}$

04 Hedging

- The FOC for minimizing Var(ΔP_t) with respect to decision variable h yields:

$$2 \times h(500^2)Var(\Delta F_{t,T}) - 2 \times (500f)Cov(\Delta S_t, \Delta F_{t,T}) = 0$$

or a risk-minimizing "optimal" hedge of

$$h^* = \frac{f \times Cov(\Delta S_t, \Delta F_{t,T})}{500\, Var(\Delta F_{t,T})}$$

- h^*: can be estimated by substituting in the sample estimates of the covariance in the numerator and of the variance in the denominator.
- The optimal hedge can also be estimated through the following linear regression employing OLS method:

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

where e_t is residual error that is uncorrelated with ΔF_t. Short form: $F_t = F_{t,T}$

Lee Kong Chian School of Business

04 Hedging

- $b = \frac{Cov(\Delta S_t, \Delta F_t)}{Var(\Delta F_t)} = 500/f \times h^*$
- h^* estimate is found as $\hat{b} \times f/500$ number of the futures contracts to short
- $\hat{b} = 0.71575$ with a 10 billion Yen portfolio value and spot N225 Index on September 1, 1999, at 17479, $f = 10 \text{ billion}/17479 = 572{,}115$.
- The number of futures contract to short in this case is estimated as:

$$h^* = \hat{b} \times \frac{f}{500} = 0.71575 \times \frac{572{,}115}{500} \approx 819$$

number of N225 futures contracts.

Table 2.2: OLS Regression of Change in Nikkei Index on Change in Nikkei Futures Price: $\Delta S_t = a + b\Delta F_t + e_t$. Sample size = 29

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	4.666338	24.01950	0.194	0.8474
ΔF_t	0.715750	0.092666	7.724	0.0000***

R-squared	0.6884	F $(d.f.\ 1, 27)$-statistic	59.6597
Adjusted R-squared	0.6769	Prob(F-statistic)	0.0000
S.E. of regression	129.325	Sum squared resid.	451573.2

Practice Exercise (not graded)

N225_Hedging.ipynb

N225.csv