MFin Econometrics I
Session 4: t-distribution, Simple Linear Regression, OLS assumptions and properties of OLS estimators

Thilo Klein

University of Cambridge
Judge Business School

CAMBRIDGE Judge Business School

t-distribution

What if σ_Y is un-known? (almost always)

- Recall: $\bar{Y} \sim N(\mu_Y, \frac{{\sigma_Y}^2}{n})$ and $\frac{\bar{Y}-\mu_Y}{\frac{\sigma_Y}{\sqrt{n}}} \sim N(0,1)$
    - If $(Y_1, ..., Y_n)$ are i.i.d. (and $0 < \sigma_Y^2 < \infty$) then, when n is large, the distribution of $\bar{Y}$ is well approximated by a normal distribution: Why? *CLT*
    - If $(Y_1, ..., Y_n)$ are independently and identically drawn from a Normal distribution, then *for any value of n*, the distribution of $\bar{Y}$ is normal: Why? *Sums of normal r.v.s are normal*
- But we almost never know σ_Y. We use sample standard deviation (s_Y) to estimate the unknown σ_Y
- Consequence of using the sample s.d. in the place of the population s.d. is an increase in uncertainty. *Why?*
    - $s_Y/\sqrt{n}$ is the standard error of the mean : estimate from sample, of st. dev. of sample mean, over all possible samples of size n drawn from the population

CAMBRIDGE Judge Business School

t-distribution

Using s_Y: "estimator" of σ_Y

- ${s_Y}^2 = \frac{1}{\text{sample size-1}} \sum_{i=1}^{\text{sample size}} (Y_i - \bar{Y})^2$
  - Why sample size - 1 in this estimator?
  - Degrees of freedom (d.f.): the number of *independent observations* available for estimation. Sample size less the number of (linear) parameters estimated from the sample
- Use of an *estimator* (i.e., a random variable) for σ_Y (and thus, the use of an estimator for $\sigma_{\bar{Y}}$) motivates use of the fatter-tailed t-distribution in the place of $N(0,1)$
- The law of large numbers (LLN) applies to ${s_Y}^2$: ${s_Y}^2$ is, in fact, a "sample average" (How?)
- LLN: If sampling is such that $(Y_1, ..., Y_n)$ are i.i.d., (and if $E(Y^4) < \infty$), then ${s_Y}^2$ *converges in probability* to ${\sigma_Y}^2$:
  - ${s_Y}^2$ is an average, not of Y_i, but of its square (hence we need $E(Y^4) < \infty$)

CAMBRIDGE Judge Business School

t-distribution

Aside: t-distribution probability density

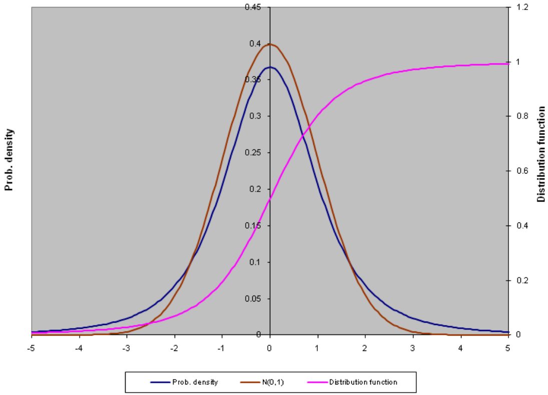
$$f(t) = \frac{\Gamma(\frac{d.f.+1}{2})}{\sqrt{d.f.\pi}\Gamma(\frac{d.f.}{2})}(1 + \frac{t^2}{d.f.})^{-\frac{d.f.+1}{2}}$$

- $d.f.$: Degrees of freedom: the only parameter of the t-distribution;
  - $\Gamma()$ (Gamma function): $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$
- Mean: $E(t) = 0$
- Variance: $V(t) = d.f./(d.f. - 2)$ for $d.f. > 2 \quad \geq 1$, converges to 1 as $d.f.$ increases (Compare: $N(0,1)$)

CAMBRIDGE Judge Business School

t-distribution

t-distribution pdf and CDF: d.f.=3



t-distribution

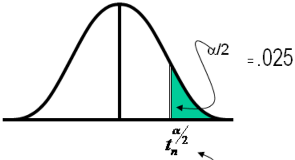
t - statistic

$$t_{d.f.} = \frac{\bar{Y} - \mu}{s_Y / \sqrt{n}} \qquad d.f. = n - 1$$

- test statistic for the sample average same form as z-statistic: Mean 0, Variance $\rightarrow 1$
- If the r.v. Y is is normally distributed in population, the test-statistic above for $\bar{Y}$ is t-distributed with $d.f. = n - 1$ degrees of freedom

t-distribution

t - distribution, table and use



Degrees of Freedom	$t_{.100}$	$t_{.05}$	$t_{.025}$	$t_{.01}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.92	4.303	6.965	9.925
.	.	.	.	.	.
.	.	.	.	.	.
20	1.325	1.725	2.086	2.528	2.845
.	.	.	.	.	.
.	.	.	.	.	.
200	1.286	1.653	1.972	2.345	2.601
∞	1.282	1.645	1.96	2.326	2.576

t-distribution

$t \rightarrow$ SND

- If d.f. is moderate or large (> 30, say) differences between the t-distribution and $N(0,1)$ *critical values* are negligible
- Some 5% critical values for 2-sided tests:

degree of freedom	5%t-distribution critical value
10	2.23
20	2.09
30	2.04
60	2.00
∞	**1.96**

CAMBRIDGE Judge Business School

t-distribution

Comments on t distribution, contd.

- The t distribution is only relevant when the sample size is small. But then, for the t distribution to be correct, the populn. distrib. of Y must be Normal Why?
  - The popln r.v. ($\bar{Y}$) must be Normally distributed for the test-statistic (with σ_Y estimated by s_Y) to be t-distributed
  - If sample size small, CLT does not apply; the popln distrib. of Y must be Normal for the popln. distrib. of $\bar{Y}$ to be Normal (sum of Normal r.v.s is Normal)
- So if sample size small and σ_Y estimated by s_Y, then test-statistic to be t-distributed if popln. is Normally distributed
- If sample size large (e.g., > 30), the popln. distrib. of $\bar{Y}$ is Normal irrespective of distrib. of Y - CLT (we saw that as d.f. increases, distrib. of $t_{d.f.}$ converges to $N(0, 1)$)
- Finance / Management data: Normality assumption dubious. e.g., earnings, firm sizes etc.

CAMBRIDGE Judge Business School

t-distribution

Comments on t distribution, contd.

- So, with large samples,
- or, with small samples, but with σ known, and a normal population distribution

$$
\begin{aligned}
Pr(-z_{\alpha/2} \leq \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) &= 1-\alpha \\
Pr(\bar{Y} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) &= 1-\alpha \\
Pr(\bar{Y} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y} + 1.96\frac{\sigma}{\sqrt{n}}) &= 0.95
\end{aligned}
$$

CAMBRIDGE Judge Business School

t-distribution

Comments on t distribution, contd.

- With small samples (<30 d.f.),
- drawn from an approximately normal population distribution,
- with the unknown σ estimated with s_Y,
- for test statistic $\frac{\bar{Y}-\mu}{s_Y/\sqrt{n}}$:

$$Pr(\bar{Y} - t_{d.f.,\alpha/2}\frac{s_Y}{\sqrt{n}} \leq \mu \leq \bar{Y} + t_{d.f.,\alpha/2}\frac{s_Y}{\sqrt{n}}) = 1 - \alpha$$

Q: What is $t_{d.f.,\alpha/2}$?

CAMBRIDGE Judge Business School

Simple linear regression model

Population Linear Regression Model: Bivariate

$$Y = \beta_0 + \beta_1 X + u$$

- Interested in conditional probability distribution of Y given X
- Theory / Model : the conditional mean of Y given the value of X is a linear function of X
  - X : the independent variable or regressor
  - Y : the dependent variable
  - β_0 : intercept
  - β_1 : slope
  - u : regression disturbance, which consists of effects of factors other than X that influence Y, as also *measurement errors* in Y
  - n : sample size, $Y_i = \beta_0 + \beta_1 X_i + u_i \quad i = 1, \cdots, n$

CAMBRIDGE Judge Business School

Simple linear regression model

Linear Regression Model - Issues

- Issues in estimation and inference for linear regression estimates are, at a general level, the same as the issues for the sample mean
- Regression coefficient is just a glorified mean
- Estimation questions:
  - How can we estimate our model? i.e., "draw" a line/plane/surface through the data ?
  - Advantages / disadvantages of different methods ?
- Hypothesis testing questions:
  - How do we test the hypothesis that the population parameters are *zero*?
    - Why test if they are zero?
  - How can we construct confidence intervals for the parameters?

CAMBRIDGE Judge Business School

Simple linear regression model

Regression Coefficients of the Simple Linear Regression Model

- How can we estimate β_0 and β_1 from data?
  - $\bar{Y}$ is the ordinary least squares (OLS) estimator of μ_Y:
  - Can show that $\bar{Y}$ solves:

$$min_X \sum_{i=1}^{n} (Y_i - X)^2$$

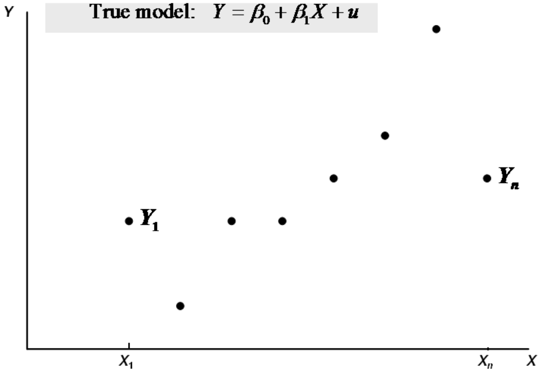
- By analogy, OLS estimators of the unknown parameters β_0 and β_1, solve:

$$min_{\hat{\beta}_0,\, \hat{\beta}_1} \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$$

CAMBRIDGE Judge Business School

Simple linear regression model

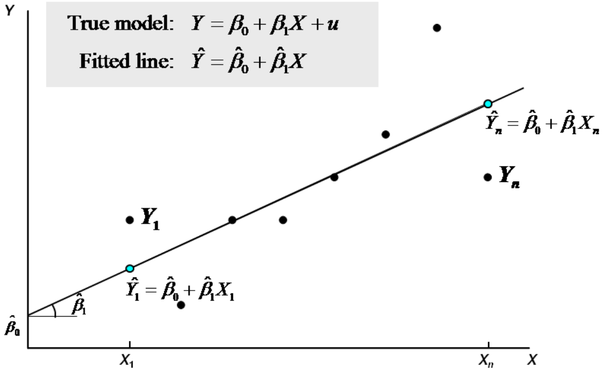
OLS in pictures (1)



CAMBRIDGE Judge Business School

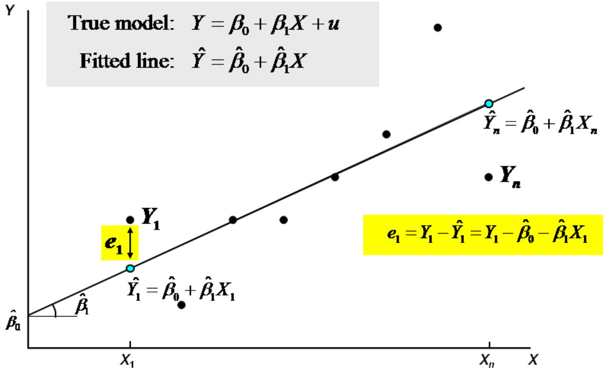
Simple linear regression model

OLS in pictures (2)



Simple linear regression model

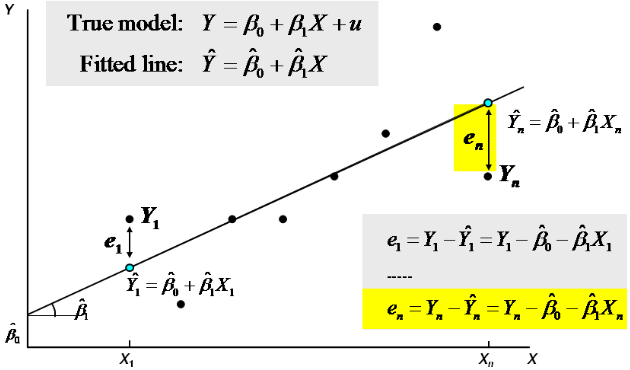
OLS in pictures (3)



CAMBRIDGE Judge Business School

Simple linear regression model

OLS in pictures (4)



Simple linear regression model

OLS method of obtaining regression coefficients

- The sum : $e_1^2 + .. + e_n^2$, is the Residual Sum of Squares (RSS), a measure of total *error*
  - RSS is a function of both $\hat{\beta}_0$ and $\hat{\beta}_1$ (How?)

$$\begin{aligned} \text{RSS} &= (Y_1 - \hat{\beta}_0 - \hat{\beta}_1 X_1)^2 + .. + (Y_n - \hat{\beta}_0 - \hat{\beta}_1 X_n)^2 \\ &= \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \end{aligned}$$

- Idea: Find values of $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimise RSS
- $\frac{\partial RSS(\cdot)}{\partial \hat{\beta}_0} = -2 \sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$
- $\frac{\partial RSS(\cdot)}{\partial \hat{\beta}_1} = -2 \sum X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$

CAMBRIDGE Judge Business School

Simple linear regression model

Ordinary Least Squares, contd.

- $\frac{\partial RSS}{\partial \hat{\beta}_0} = 0$ and $\frac{\partial RSS}{\partial \hat{\beta}_1} = 0$
- Solving these two equations together:
- $\hat{\beta}_1 = \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum(X_i - \bar{X})^2} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{Cov(X,Y)}{Var(X)}$
- $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

CAMBRIDGE Judge Business School

Simple linear regression model

Linear Regression Model: Interpretation

- Exercise

Linear regression model: Evaluation

Measures of Fit (i): Standard Error of the Regression (SER) and Root Mean Squared Error (RMSE)

- Standard error of the regression (SER) is an estimate of the dispersion (st.dev.) of the distribution of the disturbance term, u;
- Equivalently, of Y, conditional on X
- How close are Y values to $\hat{Y}$ values? can develop confidence intervals around any prediction
  - $SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(e_i - \bar{e})^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} {e_i}^2}$
  - SER converges to root mean squared error (RMSE)
  - $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$
  - RMSE denominator has n SER has $(n-2)$
  - Why?

CAMBRIDGE Judge Business School

Linear regression model: Evaluation

Measures of Fit (ii): R^2

- How much of the variance in Y can we explain with our model?
- Without the model, the best estimate of Y_i is the sample mean $\bar{Y}$
- With the model, the best estimate of Y_i is conditional on X_i and is the fitted value $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$
- How much does the error in estimate of Y reduce with the model?

CAMBRIDGE Judge Business School

Linear regression model: Evaluation

Goodness of fit

- The model: $Y_i = \hat{Y}_i + e_i$

$$\begin{aligned}
Var(Y) &= Var(\hat{Y} + e) \\
&= Var(\hat{Y}) + Var(e) + 2Cov(\hat{Y}, e) \\
&\overset{\text{Why?}}{=} Var(\hat{Y}) + Var(e) \\
\frac{1}{n}\sum(Y - \bar{Y})^2 &= \frac{1}{n}\sum(\hat{Y} - \bar{\hat{Y}})^2 + \frac{1}{n}\sum(e - \bar{e})^2 \\
\sum(Y - \bar{Y})^2 &= \sum(\hat{Y} - \bar{\hat{Y}})^2 + \sum(e - \bar{e})^2
\end{aligned}$$

- Total Sum of Squares (TSS) = Explained Sum of Squares (ESS) + Residual Sum of Squares (RSS)

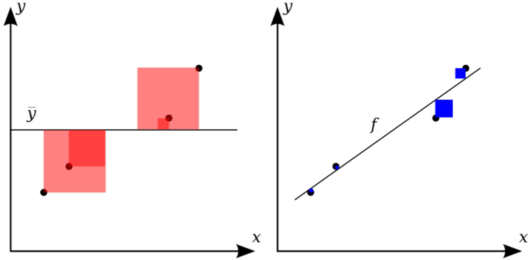
$$R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

CAMBRIDGE Judge Business School

Linear regression model: Evaluation

Goodness of fit

- $TSS = ESS + RSS$
- $R^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum e_i^2}{\sum(Y_i - \bar{Y})^2}$
- $\sqrt{R^2} = \frac{Cov(Y, \hat{Y})}{st.dev.(Y) st.dev.(\hat{Y})} = r_{Y,\hat{Y}}$



CAMBRIDGE Judge Business School

Properties of OLS Estimators

After regression estimates

- In practical terms, we wish to:
  - quantify sampling uncertainty associated with $\hat{\beta}_1$
  - use $\hat{\beta}_1$ to test hypotheses such as $\beta_1 = 0$
  - construct confidence intervals for β_1
- all these require knowledge of the sampling distribution of the OLS estimators (based on the probability framework of regression)

CAMBRIDGE Judge Business School

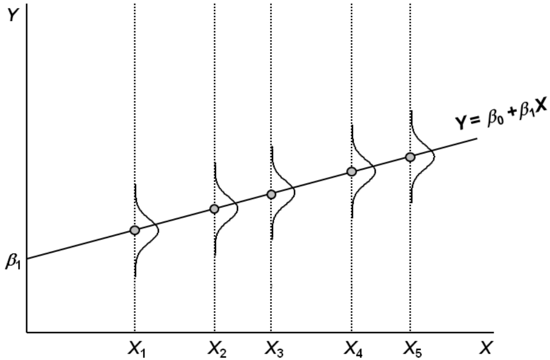
Properties of OLS Estimators

Properties of Estimators and the Least Squares Assumptions

- What kinds of estimators would we like?
- unbiased, efficient, consistent
- Under what conditions can these properties be guaranteed?
- We focus on the sampling distribution of $\hat{\beta}_1$ (Why not $\hat{\beta}_0$?)
  - The results below do hold for the sampling distribution of $\hat{\beta}_0$ too.

Properties of OLS Estimators

Assumption 1: $E(u|X = x) = 0$



CAMBRIDGE Judge Business School

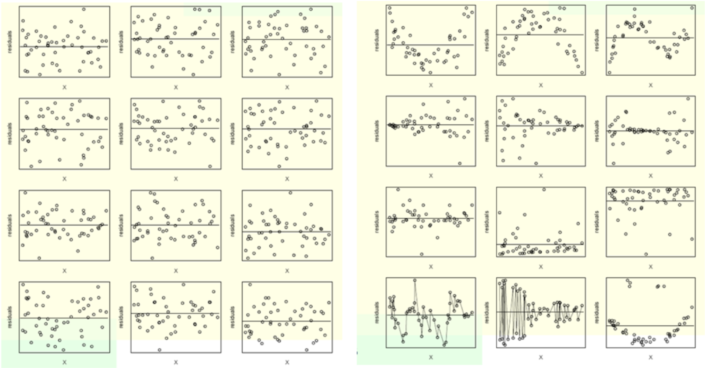
Properties of OLS Estimators

Assumption 1: $E(u|X = x) = 0$

- Conditional on X, u does not tend to influence Y either positively or negatively.
- Implication : Either X is not random, or,
- If X is random, it is distributed independently of the disturbance term, u: $Cov(X, u) = 0$
- This will be true if there are no relevant omitted variables in the regression model (i.e., those that are correlated to X)

Properties of OLS Estimators

Assumption 1: Residual plots that pass and that fail



Properties of OLS Estimators

Aside: include a constant in the regression

- Suppose $E(u_i) = \mu_u \neq 0$
- Suppose $u_i = \mu_u + v_i$ where $v_i \sim N(0, {\sigma_V}^2)$
- Then $Y_i = \beta_0 + \beta_1 X_i + v_i + \mu_u = (\beta_0 + \mu_u) + \beta_1 X_i + v_i$

CAMBRIDGE Judge Business School

Properties of OLS Estimators

Assumption 2: (X_i, Y_i), $i = 1, ..., n$ are i.i.d.

- This arises naturally with simple random sampling procedure
- Because most estimators are linear functions of observations,
- Independence between observations helps in obtaining the sampling distributions of the estimators

CAMBRIDGE Judge Business School

Properties of OLS Estimators

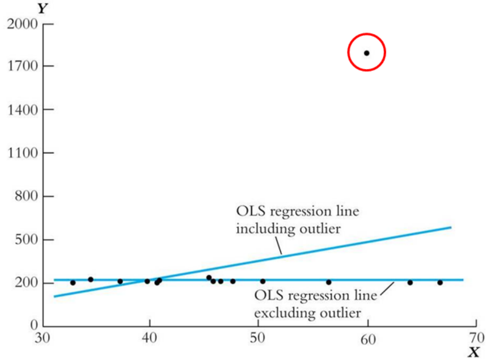
Assumption 3: Large outliers are rare

- A large outlier is an *extreme* value of X or Y
- Technically, $E(X^4) < \infty$ and $E(Y^4) < \infty$
  - Note: If X and Y are bounded, then they have finite fourth moments (income, etc.)
- Rationale : a large outlier can influence the results significantly

CAMBRIDGE Judge Business School

Properties of OLS Estimators

Assumption 3: OLS sensitive to outliers



Properties of OLS Estimators

Sampling distribution of $\hat{\beta}_1$

- If the three Least Squares Assumptions (mean zero disturbances, i.i.d. sampling, no large outliers) hold,
- then the exact (finite sample) sampling distribution of $\hat{\beta}_1$ is such that:
  - $\hat{\beta}_1$ is unbiased, that is, $E(\hat{\beta}_1) = \beta_1$
  - $Var(\hat{\beta}_1)$ can be determined
    - Other than its mean and variance, the exact distribution of $\hat{\beta}_1$ is complicated and depends on the distribution of (X, u)
  - $\hat{\beta}_1$ is consistent: $\hat{\beta}_1 \rightarrow_p \beta_1$ $\quad plim(\hat{\beta}_1) = \beta_1$
  - So when n is large, $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{Var(\hat{\beta}_1)}} \sim N(0, 1)$ (by CLT)
- This parallels the sampling distribution of $\bar{Y}$

CAMBRIDGE Judge Business School

Properties of OLS Estimators

Mean of the sampling distribution of $\hat{\beta}_1$

- $Y = \beta_0 + \beta_1 X + u$
- unbiasedness

$$\begin{aligned}
\hat{\beta}_1 &= \frac{Cov(X,Y)}{Var(X)} = \frac{Cov(X,[\beta_0 + \beta_1 X + u])}{Var(X)} \\
&= \frac{Cov(X,\beta_0) + Cov(X,\beta_1 X) + Cov(X,u)}{Var(X)} \\
&= \frac{0 + \beta_1 Cov(X,X) + Cov(X,u)}{Var(X)} \\
&= \beta_1 + \frac{Cov(X,u)}{Var(X)}
\end{aligned}$$

CAMBRIDGE Judge Business School

Properties of OLS Estimators

Unbiasedness of $\hat{\beta}_1$

- $\hat{\beta}_1 = \frac{Cov(X,Y)}{Var(X)} = \beta_1 + \frac{Cov(X,u)}{Var(X)}$
- To investigate unbiasedness, take Expectation
- $E(\hat{\beta}_1) = \beta_1 + \frac{1}{Var(X)} E(Cov(X,u)) = \beta_1$
  - Expected value of $Cov(X,u)$ is zero (Why?)
- $\hat{\beta}_1$ is an unbiased estimator of β_1

CAMBRIDGE Judge Business School

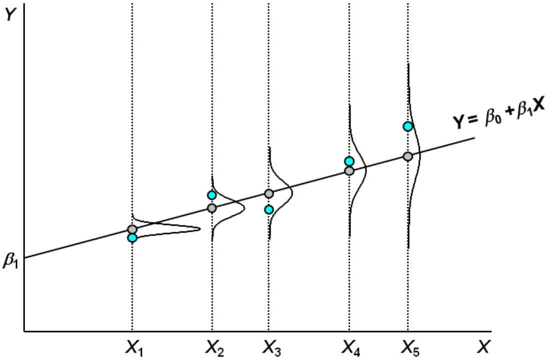
Linear regression model: further assumptions

Homoscedasticity and Normality of residuals

- u is homoscedastic
- $u \sim N(0, {\sigma_u}^2)$
- These assumptions are more restrictive
- However, if these assumptions are not violated, then other desirable properties obtain

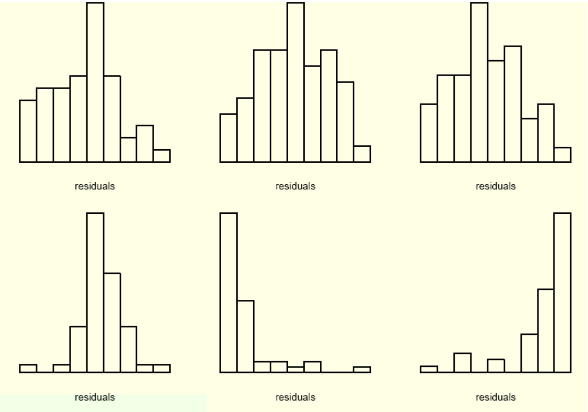
Linear regression model: further assumptions

Heteroscedasticity: one example



Linear regression model: further assumptions

Normality of disturbances: histograms of residuals that 'pass' and 'fail'



Linear regression model: further assumptions

Normality of disturbances 2: q-q plots of residuals that 'pass' and 'fail'