Financial econometrics

Chapter 2,
The classical linear regression model

Overview

Section 1, Ordinary Least Squares

Finding a Line of Best Fit

- We can use the general equation for a straight line,
$$y = a + bx$$
to get the line that "best fits" the data.

- However, the equation $y = a + bx$ is completely deterministic.

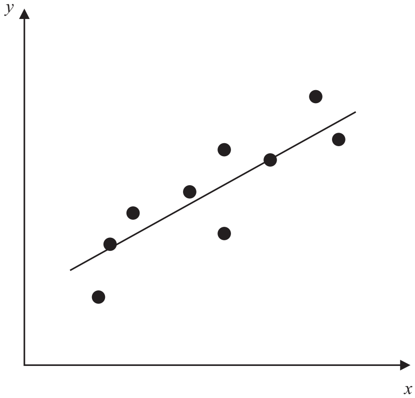
- Is this realistic? No. So what we do is to add a random disturbance term, u into the equation.
$$y_t = \alpha + \beta x_t + u_t$$
where $t = 1, 2, 3, 4, 5, \ldots$.

Why do we include a Disturbance term?

- The disturbance term can capture a number of features:
  - We always leave out some determinants of y_t.
  - There may be errors in the measurement of y_t that cannot be modelled.
  - Random outside influences on y_t which we cannot model.

Determining the Regression Coefficients

- So how do we determine what α and β are?
- Choose α and β so that the (vertical) distances from the data points to the fitted lines are minimised (so that the line fits the data as closely as possible):



Ordinary Least Squares

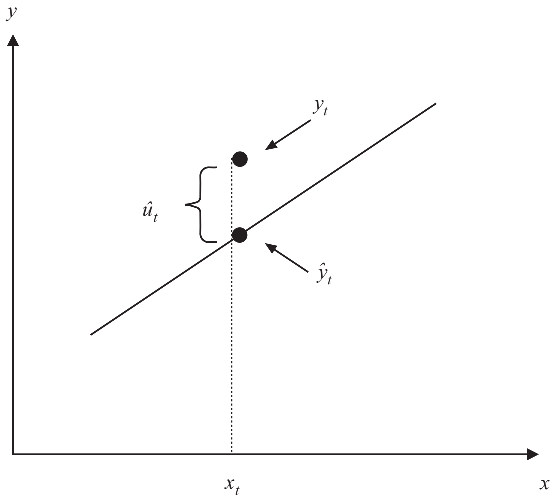
- The most common method used to fit a line to the data is known as OLS (ordinary least squares).
- What we actually do is take each distance and square it (i.e. take the area of each of the squares in the diagram) and minimise the total sum of the squares (hence least squares).
- Tightening up the notation, let

  y_t denote the actual data point t

  $\hat{y}_t$ denote the fitted value from the regression line

  $\hat{u}_t$ denote the residual, $y_t - \hat{y}_t$

Actual and Fitted Value



How OLS Works

- So min. $\hat{u}_1^2 + \hat{u}_2^2 + \hat{u}_3^2 + \hat{u}_4^2 + \hat{u}_5^2$, or minimise $\sum_{t=1}^{5} \hat{u}_t{}^2$. This is known as the residual sum of squares.
- But what was $\hat{u}_t$? It was the difference between the actual point and the line, $y_t - \hat{y}_t$.
- So minimising $\sum (y_t - \hat{y}_t)^2$ is equivalent to minimising $\sum \hat{u}_t^2$ with respect to $\hat{\alpha}$ and $\hat{\beta}$.

Deriving the OLS Estimator I

- But $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$, so let (with T the sample size)

$$L = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t)^2 .$$

- Want to minimise L w.r.t. $\hat{\alpha}$ and $\hat{\beta}$, so differentiate L w.r.t. $\hat{\alpha}$ and $\hat{\beta}$

$$\frac{\partial L}{\partial \hat{\alpha}} = -2 \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t) = 0 \qquad (1)$$

$$\frac{\partial L}{\partial \hat{\beta}} = -2 \sum_{t=1}^{T} x_t (y_t - \hat{\alpha} - \hat{\beta} x_t) = 0 \qquad (2)$$

Deriving the OLS Estimator II

- For the first derivative, from (1),

$$\sum (y_t - \hat{\alpha} - \hat{\beta} x_t) = 0 \Leftrightarrow \sum y_t - T\hat{\alpha} - \hat{\beta} \sum x_t = 0$$

- But $\sum y_t = T\bar{y}$ and $\sum x_t = T\bar{x}$. So we can write

$$T\bar{y} - T\hat{\alpha} - T\hat{\beta}\bar{x} = 0 \quad \text{or} \quad \bar{y} - \hat{\alpha} - \hat{\beta}\bar{x} = 0 \tag{3}$$

that is

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{4}$$

- For the second derivative, from (2),

$$\sum x_t (y_t - \hat{\alpha} - \hat{\beta} x_t) = 0 \tag{5}$$

Deriving the OLS Estimator III

- Substitute (4) into (5),

$$\sum x_t(y_t - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_t) = 0$$

$$\sum x_t y_t - \bar{y}\sum x_t + \hat{\beta}\bar{x}\sum x_t - \hat{\beta}\sum x_t^2 = 0$$

$$\sum x_t y_t - T\bar{x}\bar{y} + \hat{\beta}T\bar{x}^2 - \hat{\beta}\sum x_t^2 = 0$$

- Rearranging for $\hat{\beta}$,

$$\hat{\beta}\left(T\bar{x}^2 - \sum x_t^2\right) = T\bar{x}\bar{y} - \sum x_t y_t$$

- So overall we have

$$\hat{\beta} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2} \quad \textit{and} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{6}$$

- This method of finding the optimum is known as ordinary least squares.

A better look into $\hat{\beta}$

From the **nominator** of the expression for $\hat{\beta}$:

$$\begin{aligned}
\sum x_t y_t - T\bar{x}\bar{y} &= \sum x_t y_t - T\bar{x}\bar{y} - T\bar{x}\bar{y} + T\bar{x}\bar{y} \\
&= \sum x_t y_t - \sum x_t \bar{y} - \sum y_t \bar{x} + \sum \bar{x}\bar{y} \\
&= \sum \left(x_t y_t - x_t \bar{y} - y_t \bar{x} + \bar{x}\bar{y}\right) \\
&= \sum (x_t - \bar{x})(y_t - \bar{y})
\end{aligned}$$

Also (used later):

$$\begin{aligned}
\sum x_t y_t - T\bar{x}\bar{y} &= \sum_t (x_t y_t) - \bar{x} \sum y_t \\
&= \sum (x_t - \bar{x})\, y_t \qquad (7)
\end{aligned}$$

A better look into $\hat{\beta}$

From the **denominator** of the expression for $\hat{\beta}$:

$$\begin{aligned}
\sum x_t^2 - T\bar{x}^2 &= \sum x_t^2 - 2T\bar{x}^2 + T\bar{x}^2 \\
&= \sum x_t^2 - 2\bar{x}T\bar{x} + T\bar{x}^2 \\
&= \sum x_t^2 - 2\bar{x}\sum x_t + T\bar{x}^2 \\
&= \sum \left(x_t^2 - 2\bar{x}x_t + \bar{x}^2\right) \\
&= \sum (x_t - \bar{x})^2 \qquad (8)
\end{aligned}$$

A better look into $\hat{\beta}$

Therefore:

$$\hat{\beta} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2} = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2} = \frac{T}{T}\frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum (x_t - \bar{x})^2}$$

$$= \frac{\text{sample covariance}}{\text{sample variance}}$$

The Population and the Sample

- The population is the total collection of all objects or people to be studied, for example,

Interested in	Population of interest
predicting outcome of an election	the entire electorate

- A sample is a selection of just some items from the population.

- A random sample is a sample in which each individual item in the population is equally likely to be drawn.

The DGP and the PRF

- The population regression function (PRF) is a description of the model that is **thought** to be generating the actual data (DGP) and the true relationship between the variables (i.e. the true values of α and β).
- The PRF is $y_t = \alpha + \beta x_t + u_t$
- The sample regression function (SRF) is $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$
  and we also know that $\hat{u}_t = y_t - \hat{y}_t$.
- We use the SRF to infer likely values of the PRF.
- We also want to know how "good" our estimates of α and β are.

Linearity

- In order to use OLS, we need a model which is linear in the parameters (α and β). It does not necessarily have to be linear in the variables (y and x).
- Linear in the parameters means that the parameters are not multiplied together, divided, squared or cubed etc.
- Some models can be transformed to linear ones by a suitable substitution or manipulation, e.g. the exponential regression model

$$y_t = e^{\alpha} x_t^{\beta} e^{u_t} \Leftrightarrow \ln y_t = \alpha + \beta \ln x_t + u_t$$

- Then let $y_t' = \ln y_t$ and $x_t' = \ln x_t$

$$y_t' = \alpha + \beta x_t' + u_t$$

Linear and Non-linear Models

- This is known as the exponential regression model. Here, the coefficients can be interpreted as elasticities.
- Similarly, if theory suggests that y and x should be inversely related:
$$y_t = \alpha + \frac{\beta}{x_t} + u_t$$
then the regression can be estimated using OLS by substituting $z_t = \frac{1}{x_t}$ as
$$y_t = \alpha + \beta z_t + u_t$$
- But some models are intrinsically non-linear, e.g.
$$y_t = \alpha + x_t^{\beta} + u_t$$

Estimator or Estimate?

- Estimators are the formulae used to calculate the coefficients.
- Estimates are the actual numerical values for the coefficients.

Desirable properties of estimators I

- Consistent

  An estimator is *consistent* if the estimate will converge to its true value as the sample size increases to infinity:

$$\lim_{T\to\infty} \Pr\left[|\hat{\beta} - \beta| > \varepsilon\right] = 0, \quad \forall\, \varepsilon > 0$$

- Unbiased

  An estimator $\hat{\beta}$ is *unbiased* if

$$\mathbb{E}(\hat{\beta}) = \beta$$

  Thus on average, the estimated value will be equal to the true values.

- Efficient

Desirable properties of estimators II

An estimator $\hat{\beta}$ of parameter β is said to be efficient if it is unbiased and no other unbiased estimator has a smaller variance. If the estimator is efficient, we are minimizing the probability that it is a long way off from the true value of β.

Expected values and variance of the OLS estimators

- We now return to the population model and study the statistical properties of OLS. In other words, we now view $\hat{\beta}_0$ and $\hat{\beta}_1$ as estimators for the parameters β_0 and β_1 that appear in the population model.
- Will study properties of the distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ over different random samples from the population.

Unbiasedness of OLS

Assumption 1: linear in parameters

In the population model, the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as

$$y = \beta_0 + \beta_1 x + u, \tag{9}$$

where β_0 and β_1 are the population intercept and slope parameters, respectively.

Unbiasedness of OLS (cont'd)

- We assume that our data were obtained as a random sample

Assumption 2: random sampling

We have a random sample of size n, $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ following the population model in equation (9).

Note: Not all cross-sectional samples can be viewed as outcomes of random samples, but many can be.

- We can write (9) in terms of the random sample as

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad i = 1, 2, \ldots, n$$

u_i is the error for observation i and contains the unobservables for observation i that affect y_i

Unbiasedness of OLS (cont'd)

- The OLS slope and intercept estimates are not defined unless we have some sample variation in the explanatory variable
- We now add variation in the x_i to our list of assumptions - obvious but nevertheless necessary.

Assumption 3: sample variation in the explanatory variable

The sample outcomes on x, namely, $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$ are not all the same value.

Note: trivially, if the sample standard deviation of x_i is zero then Assumption 3 fails; otherwise it holds.

Unbiasedness of OLS (cont'd)

- In order to obtain unbiased estimators of β_0 and β_1, we need to impose the zero conditional mean assumption

Assumption 4: zero conditional mean

The error u has an expected value of zero given any value of the explanatory variable. In other words,

$$\mathbb{E}\left(u|x\right) = 0.$$

- For a random sample, this assumption implies that $\mathbb{E}\left(u_i|x_i\right) = 0$ for all $i = 1, 2, \ldots, n$.

Unbiasedness of OLS (cont'd) I

We can now show that the OLS estimator are **unbiased**:

- Use eq. (7) to rewrite the OLS estimator $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\, y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{10}$$

- because we are now interested in the behavior of $\hat{\beta}_1$ across all possible samples, $\hat{\beta}_1$ is properly viewed as a random variable
- now write $\hat{\beta}_1$ in terms of the population coefficients and errors,

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})\, y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{SST_x},$$

where SST_x stands for squared sample total variation.

Unbiasedness of OLS (cont'd) II

- for the numerator we have

$$
\begin{aligned}
&\sum_{i=1}^{n} (x_i - \bar{x}) \left(\beta_0 + \beta_1 x_i + u_i\right) \\
&= \beta_0 \sum_{i=1}^{n} (x_i - \bar{x}) + \beta_1 \sum_{i=1}^{n} (x_i - \bar{x}) \, x_i + \sum_{i=1}^{n} (x_i - \bar{x}) \, u_i
\end{aligned}
$$

- Note that for the term multiplying β,

$$
\begin{aligned}
\sum (x_i - \bar{x}) \, x_i &= \sum \left(x_i^2 - \bar{x} x_i\right) = \sum x_i^2 - \bar{x} \sum x_i \\
&= \sum x_i^2 - \bar{x} T \bar{x} \\
&= \sum x_i^2 - T \bar{x}^2 \; \ldots \; \text{identical to eq.(8)} \\
&= \sum (x_i - \bar{x})^2 = SST_x
\end{aligned}
$$

Unbiasedness of OLS (cont'd) III

- In conclusion:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\beta_0 \sum_{i=1}^{n} (x_i - \bar{x}) + \beta_1 SST_x + \sum_{i=1}^{n} (x_i - \bar{x}) u_i}{SST_x} \\
&= \frac{\beta_0}{SST_x} \sum_{i=1}^{n} (x_i - \bar{x}) + \beta_1 + \frac{\sum_{i=1}^{n} (x_i - \bar{x}) u_i}{SST_x} \\
&\overset{(11)}{=} \beta_1 + \frac{\sum_{i=1}^{n} (x_i - \bar{x}) u_i}{SST_x} \\
&= \beta_1 + \frac{1}{SST_x} \sum_{i=1}^{n} d_i u_i,
\end{aligned}$$

  with $d_i = x_i - \bar{x}$.

- this shows that conditional on the x_i the randomness of $\hat{\beta}_1$ is due entirely to the errors in the sample.

Unbiasedness of OLS (cont'd) IV

- The fact that these errors are generally different from zero is what causes $\hat{\beta}_1$ to differ from β_1.

Unbiasedness of OLS (cont'd)

Theorem

Using Assumptions 1 through 4,

$$\mathbb{E}\left(\hat{\beta}_0\right) = \beta_0 \quad \textit{and} \quad \mathbb{E}\left(\hat{\beta}_1\right) = \beta_1$$

for any values of β_0 and β_1. I.e. $\hat{\beta}_0$ is unbiased for β_0 and $\hat{\beta}_1$ is unbiased for β_1.

Unbiasedness of OLS (cont'd) I

$$
\begin{aligned}
\mathbb{E}\left(\hat{\beta}_1\right) &= \beta_1 + \mathbb{E}\left(\frac{1}{SST_x}\sum_{i=1}^{n} d_i u_i\right) \\
&= \beta_1 + \frac{1}{SST_x}\sum_{i=1}^{n} \mathbb{E}\left(d_i u_i\right) \\
&= \beta_1 + \frac{1}{SST_x}\sum_{i=1}^{n} d_i\,\mathbb{E}\left(u_i\right) \\
&= \beta_1 + \frac{1}{SST_x}\sum_{i=1}^{n} d_i \cdot 0 = \beta_1
\end{aligned}
$$

Unbiasedness of OLS (cont'd) II

- Where we used the fact that the expected value of each u_i conditional on $\{x_1, x_2, \dots x_n,\}$ is zero under assumption 2 and 4:

$$\mathbb{E}\left(u_i\right) = \mathbb{E}\left(\mathbb{E}\left(u_i | x_i\right)\right) = 0, \forall i$$

- Since unbiasedness holds for any outcome on $\{x_1, x_2, \dots, x_n\}$, unbiasedness also holds without conditioning on $\{x_1, x_2, \dots x_n\}$.

$$\begin{aligned}
\mathbb{E}\left(\hat{\beta}_0\right) &= \mathbb{E}\left(\bar{y} - \hat{\beta}_1 \bar{x}\right) = \mathbb{E}\left(\beta_0 + \beta_1 \bar{x} + \bar{u} - \hat{\beta}_1 \bar{x}\right) \\
&= \mathbb{E}\left(\beta_0 + \left(\beta_1 - \hat{\beta}_1\right) \bar{x} + \bar{u}\right) = \beta_0 + \mathbb{E}\left(\beta_1 - \hat{\beta}_1\right) \bar{x} + \mathbb{E}\left(\bar{u}\right) \\
&= \beta_0
\end{aligned}$$

as $\mathbb{E}\left(\bar{u}\right) = 0$ by assumptions 2 and 4 and we showed that $\hat{\beta}_1$ is unbiased for β_1.

Unbiasedness of OLS (cont'd) III

- Remember that unbiasedness says nothing about the estimate that we obtain for a given sample.
- We hope that if the sample we obtain is somehow 'typical', then our estimate should be 'near' the population value.

Unbiasedness of OLS (cont'd)

- Assumption 1 requires that y and x be linearly related, with an additive disturbance. This can certainly fail: the whole linear model setting is thus inappropriate.
- Random sampling can fail in a cross-section when samples are not representative of the underlying population. Some data sets are constructed by intentionally oversampling different parts of the population. In time-series analysis we relax Assumption 2 to deal with non-random sampling.
- Assumption 3 almost always holds in interesting regression applications. Without it, we cannot even obtain the OLS estimates.

Unbiasedness of OLS (cont'd)

- If Assumption 4 fails, then OLS estimators are biased (there are ways to study the direction of the bias).
  - The possibility that x is correlated with u is almost always a concern in simple regression analysis with non-experimental data
  - Using simple regression when u contains factors affecting y that are also correlated with x can result in **spurious correlation**
  - That is, we find a relationship between y and x that is really due to other unobserved factors that affect y and also happen to be correlated with x.
  - In addition to omitted variables, there are other reasons for x to be correlated with u in the simple regression model

Variances of the OLS estimators

- it is important to know how far we can expect $\hat{\beta}_1$ to be away from β_1, on average.
- The variance of the OLS estimators can be computed under Assumptions 1 to 4 but this turns to be (quite a lot) complicated: we add an assumption (typical in cross-sectional analysis):

Assumption 5: homoskedasticity

The error u has the same variance given any value of the explanatory variable. In other words,

$$\mathbb{V}\left[u|x\right] = \sigma^2$$

Variances of the OLS estimators (cont'd)

Note:

- If we were to assume that u and x are independent, then the distribution of u given x does not depend on x, and so $\mathbb{E}[u|x] = 0 = \mathbb{E}[u]$ and $\mathbb{V}[u|x] = \sigma^2$.
- In general, without independence:

$$\mathbb{V}[u|x] = \mathbb{E}\left(u^2|x\right) - [\mathbb{E}(u|x)]^2 = \mathbb{E}\left(u^2|x\right)$$

  that is $\sigma^2 = \mathbb{E}\left(u^2|x\right) = \mathbb{E}\left(u^2\right)$ means that σ^2 is also the **conditional** and **unconditional** expectation of u^2.
- Recall the law of total variance

$$\mathbb{V}[u] = \mathbb{E}(\mathbb{V}[u|x]) + \mathbb{V}[\mathbb{E}(u|x)],$$

  and note that $\sigma^2 = \mathbb{E}\left(u^2\right) = \mathbb{V}[u]$. This means that σ^2 is the unconditional variance of u and can be referred to as **error variance**.

Variances of the OLS estimators (cont'd)

- A common way to write Assumption 4 and 5 is

$$\mathbb{E}\left[y|x\right] = \beta_0 + \beta_1 x, \qquad \mathbb{V}\left[y|x\right] = \sigma^2.$$

- When $\mathbb{V}\left[u|x\right]$ depends on x, the error term is said to exhibit **heteroskedasticity**.

Variances of the OLS estimators (cont'd) I

Theorem (Sampling variances of the OLS estimators)

Under the assumptions 1 through 5,

$$\mathbb{V}\left[\hat{\beta}_1\right] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sigma^2 / SST_x$$

and

$$\mathbb{V}\left[\hat{\beta}_0\right] = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{SST_x}$$

where these are conditional on the sample values $\{x_1, \ldots x_n\}$*, and* n *is the sample size.*

Variances of the OLS estimators, proof I

$$
\begin{aligned}
\mathbb{V}\left[\hat{\beta}_1\right] &= \mathbb{V}\left[\beta_1 + 1/SST_x \sum_{i=1}^{n} d_i u_i\right] = (1/SST_x)^2 \sum_{i=1}^{n} d_i^2 \mathbb{V}\left[u_i\right] \\
&= (1/SST_x)^2 \left(\sum_{i=1}^{n} d_i^2 \sigma^2\right) \\
&= \sigma^2 \left(1/SST_x^2\right) \left(\sum_{i=1}^{n} d_i^2\right) = \sigma^2 \left(1/SST_x\right)^2 SST_x \\
&= \sigma^2 / SST_x
\end{aligned}
$$

Variances of the OLS estimators, proof II

$$
\begin{aligned}
\mathbb{V}\left[\hat{\beta}_0\right] &= \mathbb{V}\left[\bar{y} - \hat{\beta}_1 \bar{x}\right] = \mathbb{V}\left[\bar{y}\right] + \bar{x}^2 \mathbb{V}\left[\hat{\beta}_1\right] - 2\bar{x}\,\mathbb{C}\mathrm{ov}\left(\bar{y}, \hat{\beta}_1\right) \\
&= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{SST_x} = \frac{\sigma^2\left(SST + n\bar{x}^2\right)}{nSST_x} = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{nSST_x} \\
&= \frac{\sigma^2 n^{-1} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n \left(x_i - \bar{x}\right)^2},
\end{aligned}
$$

since

$$
\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2,
$$

which means that

$$
SST_x + n\bar{x}^2 = \sum_{i=1}^n x_i^2,
$$

Variances of the OLS estimators, proof III

and

$$
\begin{aligned}
\mathrm{Cov}(\bar{y}, \hat{\beta}_1) &= \mathbb{C}\mathrm{ov}\left\{\frac{1}{n}\sum_{i=1}^{n} y_i, \frac{\sum_{j=1}^{n}(x_j - \bar{x})y_j}{\sum_{j=1}^{n}(x_j - \bar{x})^2}\right\} \\
&= \frac{1}{n}\frac{1}{\sum_{j=1}^{n}(x_j - \bar{x})^2}\mathbb{C}\mathrm{ov}\left\{\sum_{i=1}^{n} y_i, \sum_{j=1}^{n}(x_j - \bar{x})y_j\right\} \\
&= \frac{1}{n\sum_{j=1}^{n}(x_j - \bar{x})^2}\sum_{j=1}^{n}(x_j - \bar{x})\sum_{i=1}^{n}\mathbb{C}\mathrm{ov}(y_i, y_j) \\
&= \frac{1}{n\sum_{j=1}^{n}(x_i - \bar{x})^2}\sum_{i=1}^{n}(x_i - \bar{x})\sigma^2 \\
&= \frac{\sigma^2}{n\sum_{j=1}^{n}(x_i - \bar{x})^2}\cdot 0 = 0,
\end{aligned}
$$

Variances of the OLS estimators, proof IV

because the sum of the deviations from the mean is zero:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0. \tag{11}$$

Variances of the OLS estimators (cont'd) I

Remarks:

- The formulas for the variance of the OLS estimators turn invalid in presence of heteroskedasticity.
- The larger the error variance, the larger $\mathbb{V}\left[\hat{\beta}_1\right]$: does is make sense?
- As the variability in the x_i increases, the variance of $\mathbb{V}\left[\hat{\beta}_1\right]$ decreases: does is make sense?
- What happens to $\mathbb{V}\left[\hat{\beta}_1\right]$ if $\bar{x} = 0$?
- So far, all of this is still of little practical utility as σ^2 is unknown!

Estimating the error variance I

- In the following will see how to use the data to estimate σ^2.
- And use and plug-in approach to compute estimates of $\mathbb{V}\left[\hat{\beta}_1\right]$ and $\mathbb{V}\left[\hat{\beta}_0\right]$.
- Why? to compute standard errors and thus confidence intervals (with a further hypothesis).

Estimating the error variance (cont'd) I

- Since $\sigma^2 = \mathbb{E}\left(u^2\right)$, an unbiased estimator for σ^2 is $\frac{1}{n}\sum_{i=1}^{n} u_i^2$
- ...this is not feasible as we do not observe the **errors** u_i.
- however we do have estimates of u_i, namely the OLS **residuals** $\hat{u}_i$
- replace the errors with the residuals and estimate σ^2 with

$$\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{SSR}{n}$$

this is a **true** estimator (a computable rule for any sample). SSR stands for 'sum of squared residuals'.

Estimating the error variance (cont'd)

The estimator SSR/n is biased:

- This because it ignores two restrictions that we implicitly made in obtaining the OLS estimators (the two normal equations)

$$\sum_i \hat{u}_i = 0 \qquad \sum_i x_i \hat{u}_i = 0$$

- To see that there are actually restrictions on the $\hat{u}_i, i = 1, \dots, n$ think that if we know the first $n-2$ residuals the remaining two are not 'free to be whatever' but may turn the above two restrictions true.
- In particular, given the restrictions above and $n-2$ residuals, the other two residuals are implied by the first order OLS conditions.
- We say that there are $n-2$ **degrees of freedom** *in the OLS residuals*.

Estimating the error variance (cont'd) I

- We thus make the degrees of freedom adjustment:

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-2} \sum_i \hat{u}_i^2 = SSR \frac{1}{n-2}$$

- The estimator for the residuals' variance is often called 's-squared': we often write s^2 (s) in place of $\hat{\sigma}^2$ ($\sqrt{\hat{\sigma}}$).
- Under assumptions 1 through 5: $\mathbb{E}\left(\hat{\sigma}^2\right) = \sigma^2$ (we do not prove it).
- Accordingly, the natural estimator for σ is

$$s = \hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

  and is called the **standard error of the regression**, or root mean squared error (RMSE).
- Note that $\hat{\sigma}$ is *not* unbiased estimator of σ, but *is* consistent for σ.
- Interpretation: (i) $\hat{\sigma}$ is an estimate of the standard deviation in the unobservables affecting y, (ii) is an estimate of the standard deviation in y after the effect of x has been taken out.

Estimating the variance of the OLS estimators

The natural estimators of the standard deviation of the the OLS estimators, $\hat{\beta}_0$ and $\hat{\beta}_0$ are

$$\mathrm{se}\left(\hat{\beta}_1\right) = \hat{\sigma}/\sqrt{SST_x}, \tag{12}$$

$$\mathrm{se}\left(\hat{\beta}_0\right) = \hat{\sigma}\sqrt{\frac{\sum_i x_i^2}{nSST_x}}. \tag{13}$$

These are called **standard errors** of $\hat{\beta}_1$ and $\hat{\beta}_2$, respectively.

Estimating the variance of the OLS estimators (cont'd)

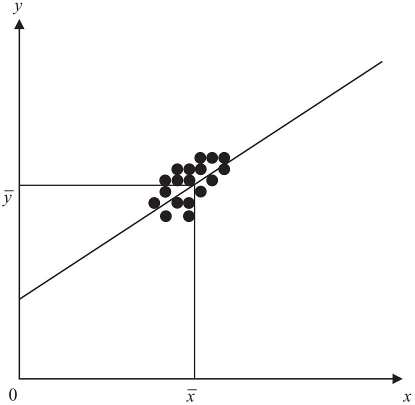
- For computing the standard errors, the following equalities hold:

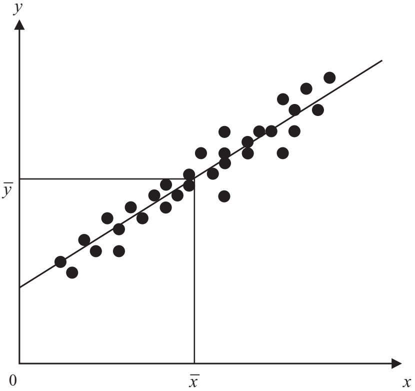
$$
\begin{aligned}
\mathrm{se}\left(\hat{\beta}_0\right) &= \hat{\sigma}\sqrt{\frac{\sum x_t^2}{n\sum\left(x_t-\bar{x}\right)^2}} = \hat{\sigma}\sqrt{\frac{\sum x_t^2}{n\left[\sum x_t^2 - n\bar{x}^2\right]}} \\
\mathrm{se}\left(\hat{\beta}_1\right) &= \hat{\sigma}\sqrt{\frac{1}{\sum\left(x_t-\bar{x}\right)^2}} = \hat{\sigma}\sqrt{\frac{1}{\sum x_t^2 - n\bar{x}^2}}
\end{aligned}
$$

... these are basically alternative ways to write eq.s. (12), (13) that you can encounter.

Some Comments on the Standard Error Estimators I

Consider what happens if $\sum (x_t - \bar{x})^2$ is small or large:





1. The larger the sample size, T, the smaller will be the coefficient variances. T appears explicitly in SE($\hat{\alpha}$) and implicitly in SE($\hat{\beta}$).

Some Comments on the Standard Error Estimators II

2. The sample size appears implicitly on $\text{se}\left(\hat{\beta}_0\right)$ and $\text{se}\left(\hat{\beta}_1\right)$ since the sum $\sum (x_t - \bar{x})^2$ is over the sample.
3. The term $\sum x_t^2$ appears in the $\text{se}\left(\hat{\beta}_0\right)$ only.
4. The reason is that $\sum x_t^2$ measures how far the points are away from the y-axis.

Estimating the Variance of the Disturbance Term (cont'd)

- An unbiased estimator of σ is given by
$$s = \sqrt{\frac{\sum \hat{u}_t^2}{T-2}}$$
where $\sum \hat{u}_t^2$ is the residual sum of squares and T is the sample size.

- Some Comments on the Standard Error Estimators
  1. Both $\mathrm{SE}(\hat{\alpha})$ and $\mathrm{SE}(\hat{\beta})$ depend on s^2 (or s). The greater the variances^2, then the more dispersed the errors are about their mean value and therefore the more dispersed y will be about its mean value.
  2. The sum of the squares of x about their mean appears in both formulae. The larger the sum of squares, the smaller the coefficient variances.

(A rather trivial) Example: How to Calculate the Parameters

Example

- Assume we have the following data calculated from a regression of y on a single variable x and a constant over 22 observations.
- Data:

$$\begin{aligned} \sum x_t y_t &= 830102, \;\; T = 22, \;\; \bar{x} = 416.5, \;\; \bar{y} = 86.65, \\ \sum x_t^2 &= 3919654, \;\; RSS = 130.6 \end{aligned}$$

- Calculations

$$\begin{aligned} \hat{\beta} &= \frac{830102 - (22 \times 416.5 \times 86.65)}{3919654 - 22 \times (416.5)^2} = 0.35 \\ \hat{\alpha} &= 86.65 - 0.35 \times 416.5 = -59.12 \end{aligned}$$

(A rather trivial) Example: How to Calculate the SE

Example

- We write $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$

$$\hat{y}_t = -59.12 + 0.35 x_t$$

- se (*regression*), $s = \sqrt{\frac{\sum \hat{u}_t^2}{T-2}} = \sqrt{\frac{130.6}{20}} = 2.55$

$$\begin{aligned} \text{se}\left(\hat{\alpha}\right) &= 2.55 \times \sqrt{\frac{3919654}{22 \times (3919654 - 22 \times 416.5^2)}} = 3.35 \\ \text{se}\left(\hat{\beta}\right) &= 2.55 \times \sqrt{\frac{1}{3919654 - 22 \times 416.5^2}} = 0.0079 \end{aligned}$$

- We now write the results as

$$\begin{array}{c} \hat{y}_t = -59.12 + 0.35 x_t \\ \qquad (3.35) \quad (0.0079) \end{array}$$

OLS assumptions

- The following generates **great** confusion and needs to be discussed carefully.
- There are two it formulations of the OLS assumptions.
- We used the assumption set as in (Wooldridge, 2015).
- Otherwise OLS assumptions appear as in (Brooks, 2014).

OLS assumptions (cont'd)

The assumptions in (Wooldridge, 2015) read:

Assumptions A

For the disturbance term we assume:

- A1 The model is linear in the parameters$y = \beta_0 + \beta_1 x + u$
- A2 $\{(x_i, y_i), i = 1, 2, \ldots, n\}$ is a random sample
- A3 $\{x_t, t = 1, 2, \ldots, n\}$ are not the same
- A4 $\mathbb{E}(u_t | x_t) = 0 \ \forall t$
- A5 $\mathbb{V}[u_t | x_t] = \sigma^2 \ \forall t$

OLS assumptions (cont'd)

The assumptions in (Brooks, 2014) read:

Assumptions B

For the disturbance term we assume:

- B1 $\mathbb{E}(u_t) = 0$
- B2 $\mathbb{V}[u_t] = \sigma^2 < \infty$
- B3 $\mathbb{C}\text{ov}(u_i, u_j) = 0 \; \forall i \neq j$
- B4 $\mathbb{C}\text{ov}(u_t, x_t) = 0 \; \forall t$
- B5 ($u_t \sim \mathcal{N}(0, \sigma^2)$ - not discussed so far)

Note: the assumptions in (Wooldridge, 2015) and (Brooks, 2014) are similar but not the same.

OLS assumptions (cont'd)

- By the law of iterated expectations:

$$\mathbb{E}(u) = \mathbb{E}(\mathbb{E}[u|x])$$

therefore $\mathbb{E}[u|x] =$ implies $\mathbb{E}(u) = 0$, that is A4 $\to$ B1.

- The converse does not hold, unless u and x are independent (which however does not figure neither in A1-5 nor B1-5). B1 does not imply A4.

OLS assumptions (cont'd)

- From A4 we get $\mathbb{E}[u|x] = \mathbb{E}[u] = 0$
- Use A4 and compute $\mathbb{C}\text{ov}(u, x)$

$$\mathbb{C}\text{ov}(u, x) = \mathbb{E}(ux) - \mathbb{E}(u)\,\mathbb{E}(x) = \mathbb{E}(ux) \tag{14}$$

$$\mathbb{E}(ux) = \mathbb{E}(\mathbb{E}(xu|x)) = \mathbb{E}(x\,\mathbb{E}(u|x)) = 0 \tag{15}$$

- It turns out that A4 implies $\mathbb{C}\text{ov}(u, x) = 0$, which is made explicit in B4.
- Thus A4 implies both B1 and B4: A4 is preferred.

OLS assumptions (cont'd) I

Does B4 imply A4?

- $\mathbb{C}\text{ov}\,(u, x) \neq 0$ implies that $\mathbb{E}\,(u|x) \neq 0$.
  Reason: If x and u are correlated then $\mathbb{E}\,(u|x)$ must depend on x and so cannot be zero.

- $\mathbb{C}\text{ov}\,(u, x) = 0$ does *not* imply that $\mathbb{E}\,(u|x) = 0$.
  Reason: $\mathbb{C}\text{ov}\,(u, x)$ measures only *linear* dependence between u an x. But any nonlinear dependence between u an x will also cause $\mathbb{E}\,(u|x)$ to depend on x, and hence differ from zero. So B4 is not enough to ensure A4.

OLS assumptions (cont'd) II

- $[\mathbb{C}\text{ov}\,(u,x) = 0$ and $\mathbb{E}\,(u) = 0]$ imply $\mathbb{E}\,(u|x) = 0$.
  Reason: look at eq. (14) and (15),

$$0 \overset{B4}{=} \mathbb{C}\text{ov}\,(u,x) \overset{By\,def.}{=} \mathbb{E}\,(ux) - \mathbb{E}\,(u)\,\mathbb{E}\,(x) \overset{B1}{=} \mathbb{E}\,(ux) = \mathbb{E}\,(x\,\mathbb{E}\,(u|x))$$

  to be true it needs has to be $\mathbb{E}\,(x\,\mathbb{E}\,(u|x)) = 0$.

$\rightarrow$ Thus B1 and B4 imply A4.

OLS assumptions (cont'd) I

A key is assumption A2

- The assumption of random sampling implies that the sample observations are statistically independent.
- It thus means that the error terms u_i and u_j are statistically independent, and hence have zero covariance, for any two observations x_i and x_j.

$$\text{Random sampling} \rightarrow \mathbb{C}\text{ov}\left(u_i, u_j | x_i, x_j\right) = \mathbb{C}\text{ov}\left(u_i, u_j\right) = 0 \qquad (16)$$

- It also means that the dependent variable values y_i and y_j are statistically independent, and hence have zero covariance with any two observations i and j.

$$\text{Random sampling} \rightarrow \mathbb{C}\text{ov}\left(y_i, y_j | x_i, x_j\right) = \mathbb{C}\text{ov}\left(y_i, y_j\right) = 0 \qquad (17)$$

- Assumption A2 thus implies B3.

OLS assumptions (cont'd) II

- However A2, is stronger than B3 and although stronger than necessary for simple regression, is usually appropriate for cross-sectional regression models.
- A common, almost universal characteristic of time-series data sets is that the sample observations exhibit a high degree of time dependence, and therefore the data cannot be assumed to be generated by random sampling. A2 is restrictive w.r.t. B3.
- I.e. depending on the context, one uses A2 or B3.

Summary I

$$\mathbb{V}[u|x] \overset{\text{Ass. 5}}{=} \sigma^2 \qquad \Rightarrow \mathbb{V}[u|x] \overset{\substack{\text{Ass. 5}\\ \text{LTV}\\ \text{Ass.4}}}{=} \mathbb{V}[u] = \sigma^2$$

$$\Rightarrow \mathbb{V}[u|x] \overset{\substack{\text{Ass.5}\\ \text{Ass.4}}}{=} \mathbb{E}\left(u^2|x\right) \overset{\text{LTE}}{=} \mathbb{E}\left(u^2\right)$$

$$\Rightarrow \mathbb{V}[y|x] \overset{\substack{\text{Ass.4}\\ \text{Ass,1}}}{=} \sigma^2$$

$$\mathbb{E}(u|x) \overset{\text{Ass.4}}{=} 0 \qquad \Rightarrow \mathbb{E}(u|x) \overset{\text{LTE}}{=} \mathbb{E}(u) = 0 \tag{18}$$

$$\Rightarrow \mathbb{C}\text{ov}(u,x) \overset{\substack{\text{Ass.4}\\ (18)}}{=} \mathbb{E}(ux) = 0$$

$$\Rightarrow \mathbb{E}(y|x) \overset{\substack{\text{Ass.4}\\ \text{Ass.1}}}{=} \beta_0 + \beta_1 x$$

Summary II

$$\text{Random sampling} \overset{\text{By. def}}{=} \{(y_i, x_i)\}_{i=1,\dots,n} \text{ are i.i.d.}$$

$$
\begin{aligned}
&\overset{\text{i.e.}}{=} (x_i, y_i) \perp (x_j, y_j) \\
&\overset{\substack{\text{Ass.2} \\ \text{Ass.1}}}{\Rightarrow} (\varepsilon_i, x_i) \perp (\varepsilon_j, x_j) && (19) \\
&\Rightarrow \mathbb{E}\left(\varepsilon_i | x_1, \dots, x_n\right) \overset{(19)}{=} \mathbb{E}\left(\varepsilon_i | x_i\right) \overset{\text{Ass.4}}{=} 0 \\
&\Rightarrow \mathbb{E}\left(\varepsilon_i^2 | x_1, \dots, x_n\right) \overset{(19)}{=} \mathbb{E}\left(\varepsilon_i^2 | x_i\right) \overset{\text{Ass.5}}{=} \sigma^2 \\
&\Rightarrow \mathbb{E}\left(u_i, u_j | x_1, \dots, x_n\right) \overset{(19)}{=} \mathbb{E}\left(u_i | x_i\right) \mathbb{E}\left(u_j | x_j\right) \overset{\text{Ass.4}}{=} 0 && (20)
\end{aligned}
$$

$+$ eq. (16), (17).

The OLS Estimator as BLUE I

- If assumptions A1-A5 (or B1-B4), then the estimators and determined by OLS are known as Best Linear Unbiased Estimators (BLUE).

  What does the acronym stand for?

- 'Estimator' – $\hat{\alpha}$ and $\hat{\beta}$ are estimators of the true value of α and β
- 'Linear' – $\hat{\alpha}$ and $\hat{\beta}$ are linear estimators
- 'Unbiased' – on average, the actual values of $\hat{\alpha}$ and $\hat{\beta}$ will be equal to their true values
- 'Best' – means that the OLS estimator $\hat{\beta}$ has minimum variance among the class of linear unbiased estimators.
- "Estimator" -$\hat{\beta}$ is an estimator of the true value of β.
- "Linear" -$\hat{\beta}$ is a linear estimator

The OLS Estimator as BLUE II

- "Unbiased" - On average, the actual value of $\hat{\alpha}$ and $\hat{\beta}$'s will be equal to the true values.
- "Best" - means that the OLS estimator $\hat{\beta}$ has minimum variance among the class of linear unbiased estimators. The **Gauss-Markov theorem** proves that the OLS estimator is best.

Properties of OLS and requirements I

- Consistent

  The least squares estimators $\hat{\alpha}_{OLS}$ and $\hat{\beta}_{OLS}$ are consistent – that is, the estimates will converge to their true values as the sample size increases to infinity – under the assumptions $E(u) = 0$, $E(x_t u_t) = 0$ and $Var(u_t) = \sigma^2 < \infty \ \forall t$.

- Unbiased

  The least squares estimates of $\hat{\alpha}$ and $\hat{\beta}$ are unbiased – that is $\mathbb{E}(\hat{\alpha}) = \alpha$ and $\mathbb{E}(\hat{\beta}) = \beta$ so that on average the estimated value will be equal to the true values – under the assumption that $\mathbb{E}(u_t) = 0$.

- Efficient

  The least squares estimates of $\hat{\alpha}$ and $\hat{\beta}$ are efficient if assumptions A2 and A5 (or B2 and B3) hold.

An Introduction to Statistical Inference

- We want to make inferences about the likely population values from the regression parameters.

- Example: Suppose we have the following regression results:

$$\hat{y}_t = 20.3 + 0.5091 x_t$$
$$(14.38) \quad (0.2561)$$

- $\hat{\beta} = 0.5091$ is a single (point) estimate of the unknown population parameter, β. How "reliable" is this estimate?

- The reliability of the point estimate is measured by the coefficient's standard error.

- Can we test specific hypotheses for β, e.g. $H0 : \beta = 0$?

Section 2,
Inference of the linear regression model

Hypothesis Testing: Some Concepts

- We can use the information in the sample to make inferences about the population.
- We will always have two hypotheses that go together, the null hypothesis (denoted H_0) and the alternative hypothesis (denoted H_1).
- The null hypothesis is the statement or the statistical hypothesis that is actually being tested. The alternative hypothesis represents the remaining outcomes of interest.
- For example, suppose given the regression results above, we are interested in the hypothesis that the true value of β is in fact 0.5. We would use the notation

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

  This would be known as a two sided test.

One-Sided Hypothesis Tests

- Sometimes we may have some prior information that, for example, we would expect $\beta > 0.5$ rather than $\beta < 0.5$. In this case, we would do a one-sided test:

  H_0: $\beta = 0.5$
  H_1: $\beta < 0.5$

  or we could have had

  H_0: $\beta = 0.5$
  H_1: $\beta > 0.5$

- There are two ways to conduct a hypothesis test: via the test of significance approach or via the confidence interval approach.

Normality of the errors

Normality of the errors

On top of the earlier hypotheses, furthermore assume that

$$u_t \sim \mathcal{N}\left(0, \sigma^2\right), \quad \forall t.$$

- This hypothesis is not necessary deriving the OLS estimators, nor for proving their properties.
- Normality is only introduced for the purpose of statistical inference.

The Probability Distribution of the Least Squares Estimators I

- We assume that $u_t \sim \mathcal{N}\left(0, \sigma^2\right)$.
- The least squares estimators are linear combinations of the random variables (i.e. $\hat{\beta} = \sum w_t y_t$): what is their distribution?
- The weighted sum of normal random variables is also normally distributed, so

$$\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \mathbb{V}\left[\hat{\beta}_0\right]\right)$$
$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \mathbb{V}\left[\hat{\beta}_1\right]\right)$$

- What if the errors are not normally distributed? Will the parameter estimators still be normally distributed?
- Yes, if the other assumptions of the Linear regression model still hold, and the sample size is sufficiently large (Central limit theorem).

The Probability Distribution of the Least Squares Estimators II

- Standard normal variates can be constructed from $\hat{\alpha}$ and $\hat{\beta}$:

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\mathbb{V}[\hat{\beta}_0]}} \sim \mathcal{N}(0,1) \text{ and } \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\mathbb{V}[\hat{\beta}_1]}} \sim \mathcal{N}(0,1).$$

- But $\text{var}(\beta_0)$ and $\text{var}(\beta_1)$ are unknown, so:

$$\frac{\hat{\beta}_0 - \beta_0}{\text{se}(\hat{\beta}_0)} \sim t_{n-2} \text{ and } \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \sim t_{n-2}.$$

Testing Hypotheses: The Test of Significance Approach I

- Assume the regression equation is given by,

$$y_t = \alpha + \beta x_t + u_t \text{ for } t = 1, 2, ..., n.$$

- The steps involved in doing a test of significance are:

1. Estimate $\hat{\alpha}$, $\hat{\beta}$
2. Obtain the residuals and estimate s^2
3. Estimate $\text{se}(\hat{\alpha})$, $\text{se}\left(\hat{\beta}\right)$ in the usual way.
4. Calculate the test statistic: This is given by the formula

$$\textit{test statistic} = \frac{\hat{\beta} - \beta^*}{\text{se}\left(\hat{\beta}\right)}$$

where β^* is the value of β under the null hypothesis (analogoust for testing an hypothesis on α).

Testing Hypotheses: The Test of Significance Approach II

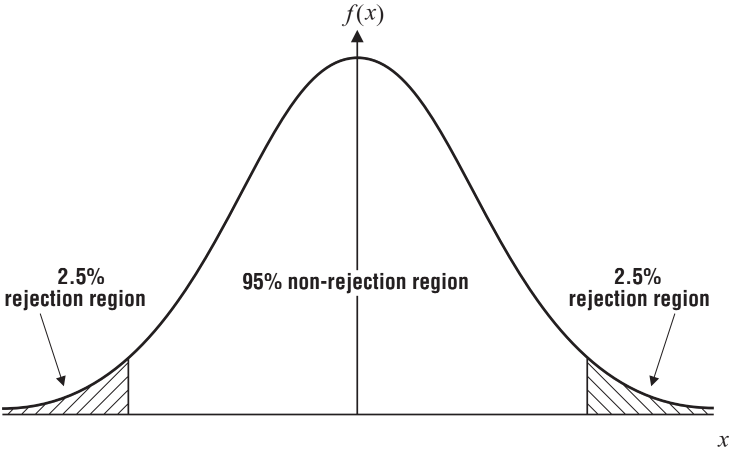
5. We need some tabulated distribution with which to compare the estimated test statistics. Test statistics derived in this way can be shown to follow a *t*-distribution with *n*-2 degrees of freedom.
   As the number of degrees of freedom increases, we need to be less cautious in our approach since we can be more sure that our results are robust.

$$\textit{test statistic} = \frac{\hat{\beta} - \beta^*}{\mathrm{se}\left(\hat{\beta}\right)} \sim t_{n-2}$$

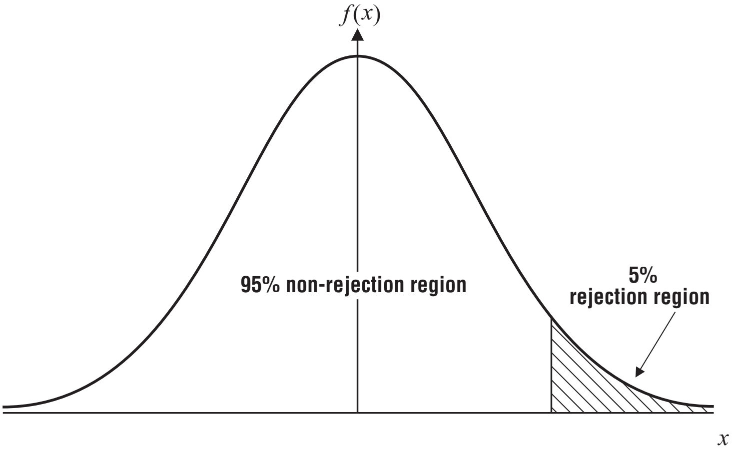
6. We need to choose a "significance level", often denoted α. This is also sometimes called the size of the test and it determines the region where we will reject or not reject the null hypothesis that we are testing. It is conventional to use a significance level of 5%.
   Intuitive explanation is that we would only expect a result as extreme as this or more extreme 5% of the time as a consequence of chance alone.
   Conventional is to use a 5% size of test, but 10% and 1% are also commonly used.

Determining the Rejection Region for a Test of Significance

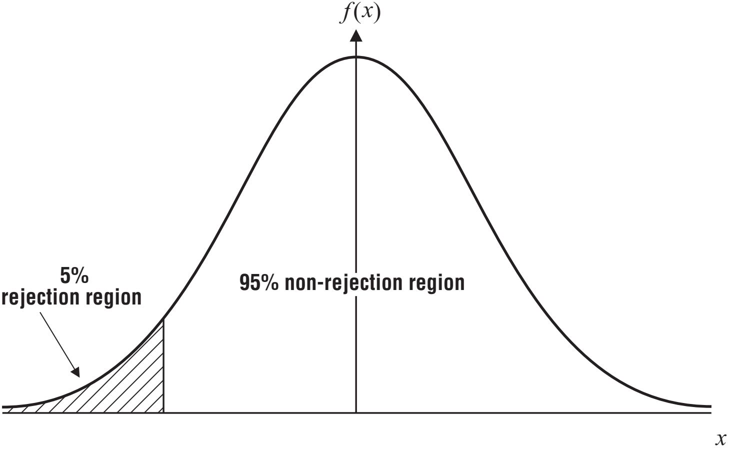
5. Given a significance level, we can determine a rejection region and non-rejection region. For a 2-sided test:



The Rejection Region for a 1-Sided Test (Upper Tail)



The Rejection Region for a 1-Sided Test (Lower Tail)



The Test of Significance Approach: Drawing Conclusions

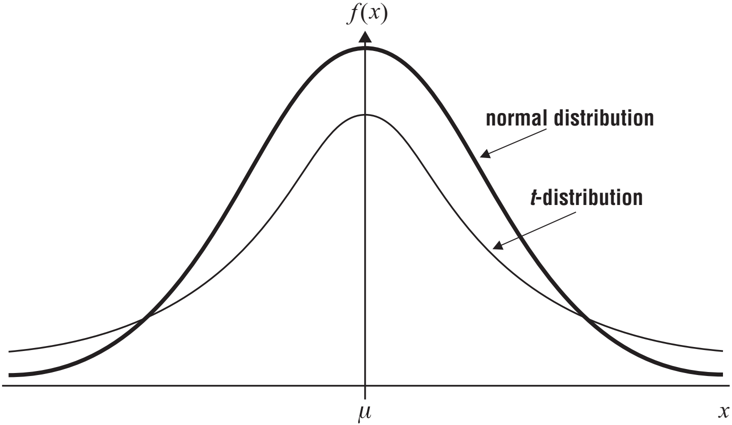
6. Use the t-tables to obtain a critical value or values with which to compare the test statistic. (Or compute the quantiles with R!)

7. Finally perform the test:
   If the test statistic lies in the rejection region then reject the null hypothesis (H_0), otherwise do not reject H_0.

A Note on the t and the Normal Distribution

- You should all be familiar with the normal distribution and its characteristic "bell" shape.
- We can scale a normal variate to have zero mean and unit variance by subtracting its mean and dividing by its standard deviation.
- There is, however, a specific relationship between the t- and the standard normal distribution. Both are symmetrical and centred on zero. The t-distribution has another parameter, its degrees of freedom. We will always know this (for the time being from the number of observations -2).

What Does the t-Distribution Look Like?



Comparing the t and the Normal Distribution

- In the limit, a t-distribution with an infinite number of degrees of freedom is a standard normal, i.e. $t_\infty = \mathcal{N}(0, 1)$
- Examples from statistical tables:

Significance level	$\mathcal{N}(0,1)$	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 had to estimate σ^2, the variance of the disturbances.

The Confidence Interval Approach to Hypothesis Testing

- An example of its usage: We estimate a parameter, say to be 0.93, and a "95% confidence interval" to be (0.77, 1.09). This means that we are 95% confident that the interval containing the true (but unknown) value of β.
- Confidence intervals are almost invariably two-sided, although in theory a one-sided interval can be constructed.

How to Carry out a Hypothesis Test Using Confidence Intervals

1. Calculate $\hat{\alpha}$, $\hat{\beta}$ and $\mathrm{se}(\hat{\alpha})$, $\mathrm{se}\left(\hat{\beta}\right)$ as before.
2. Choose a significance level α, the convention is 5%. This is equivalent to choosing a $(1\text{-}\alpha)\times 100\%$ confidence interval, i.e. 5% significance level = 95% confidence interval
3. Use the *t*-tables to find the appropriate critical value, which will again have *n*-2 degrees of freedom.
4. The confidence interval is given by $(\hat{\beta} - t_{crit} \times \mathrm{se}\left(\hat{\beta}\right), \hat{\beta} + t_{crit} \times \mathrm{se}\left(\hat{\beta}\right))$
5. Perform the test: If the hypothesised value of β (β^*) lies outside the confidence interval, then reject the null hypothesis that $\beta = \beta^*$, otherwise do not reject the null.

Confidence Intervals Versus Tests of Significance

- Note that the Test of Significance and Confidence Interval approaches always give the same answer.
- Under the test of significance approach, we would not reject H_0 that $\beta = \beta^*$ if the test statistic lies within the non-rejection region, i.e. if

$$-t_{crit} \leq \frac{\hat{\beta}-\beta^*}{\mathrm{se}\left(\hat{\beta}\right)} \leq +t_{crit}$$

- Rearranging, we would not reject if

$$-t_{crit} \times \mathrm{se}\left(\hat{\beta}\right) \leq \hat{\beta} - \beta^* \leq +t_{crit} \times \mathrm{se}\left(\hat{\beta}\right))$$
$$\hat{\beta} - t_{crit} \times \mathrm{se}\left(\hat{\beta}\right) \leq \beta^* \leq \hat{\beta} + t_{crit} \times \mathrm{se}\left(\hat{\beta}\right))$$

- But this is just the rule under the confidence interval approach.

Constructing Tests of Significance and Confidence Intervals: An Example

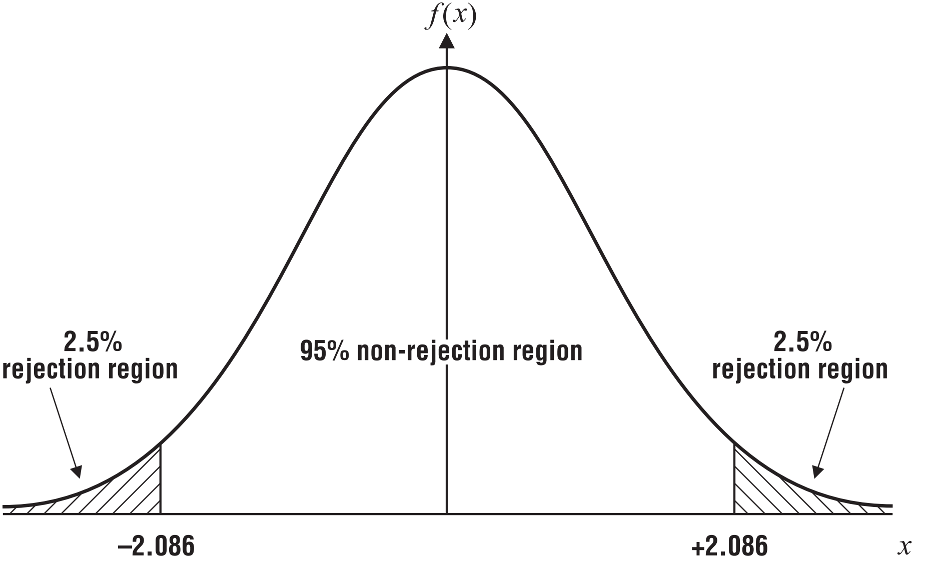
- With the following regression results

$$\begin{array}{c} \hat{y}_t = 20.3 + 0.5091 x_t \\ \quad\quad (14.38)\,(0.2561) \end{array} \quad , \qquad n = 22$$

- Using both the test of significance and confidence interval approaches, test the hypothesis that $\beta = 1$ against a two-sided alternative.

- The first step is to obtain the critical value. We want $t_{crit} = t_{20;5\%}$

Determining the Rejection Region



Performing the Test I

- The hypotheses are:

$$H_0 : \beta = 1$$

$$H_1 : \beta \neq 1$$

Test of significance approach

$$\textit{test stat} = \frac{\hat{\beta} - \beta^*}{\text{se}\left(\hat{\beta}\right)} = \frac{0.5091 - 1}{0.2561} = -1.917$$

Do not reject H_0 since test statistic lies within non-rejection region

Confidence interval approach

Find $t_{crit} = t_{20;5\%} = \pm 2.086$

$$\hat{\beta} \pm t_{crit} \cdot \text{se}\left(\hat{\beta}\right) = 0.5091 \pm 2.086 \cdot 0.2561 = (-0.0251, 1.0433)$$

Do not reject H_0 since 1 lies within the confidence interval

Testing other Hypotheses

- What if we wanted to test H_0: $\beta = 0$ or H_0: $\beta = 2$?
- Note that we can test these with the confidence interval approach.
- For interest (!), test

$$\begin{aligned} & H_0\colon\ \beta = 0 \\ \text{vs.}\quad & H_1\colon\ \beta \neq 0 \end{aligned}$$

$$\begin{aligned} & H_0\colon\ \beta = 2 \\ \text{vs.}\quad & H_1\colon\ \beta \neq 2 \end{aligned}$$

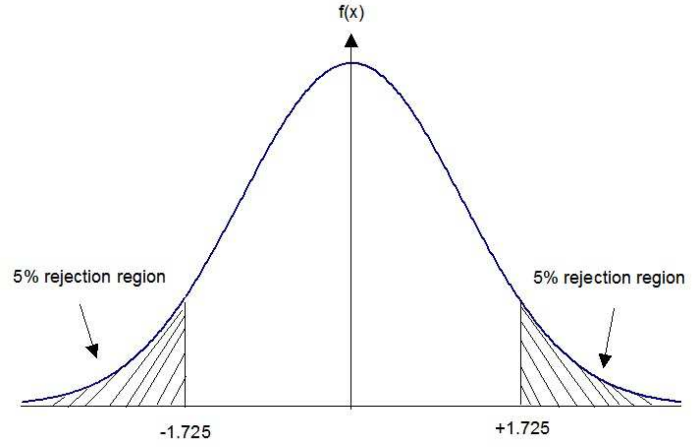
Changing the Size of the Test

- But note that we looked at only a 5% size of test. In marginal cases (e.g. H_0: $\beta = 1$), we may get a completely different answer if we use a different size of test. This is where the test of significance approach is better than a confidence interval.
- For example, say we wanted to use a 10% size of test. Using the test of significance approach,

$$\begin{aligned} \textit{test stat} &= \frac{\hat{\beta} - \beta^*}{\text{se}\left(\hat{\beta}\right)} \\ &= \frac{0.5091 - 1}{0.2561} = -1.917 \end{aligned}$$

as above. The only thing that changes is the critical t-value.

Changing the Size of the Test: The New Rejection Regions



Changing the Size of the Test: The Conclusion

- $t_{20;10\%} = 1.725$. So now, as the test statistic lies in the rejection region, we would reject H_0.
- Caution should therefore be used when placing emphasis on or making decisions in marginal cases (i.e. in cases where we only just reject or not reject).

Some More Terminology

- If we reject the null hypothesis at the 5% level, we say that the result of the test is statistically significant.
- Note that a statistically significant result may be of no practical significance. E.g. if a shipment of cans of beans is expected to weigh 450g per tin, but the actual mean weight of some tins is 449g, the result may be highly statistically significant but presumably nobody would care about 1g of beans.

The Errors That We Can Make Using Hypothesis Tests

- We usually reject H_0 if the test statistic is statistically significant at a chosen significance level.
- There are two possible errors we could make:
  1. Rejecting H_0 when it was really true. This is called a type I error.
  2. Not rejecting H_0 when it was in fact false. This is called a type II error.

		Reality	
		H_0 is true	H_0 is false
Result of test	Significant (reject H_0)	Type I error $= \alpha$	$\surd$
	Insignificant (do not reject H_0)	$\surd$	Type II error $= \beta$

The Trade-off Between Type I and Type II Errors I

- The probability of a type I error is just α, the significance level or size of test we chose. To see this, recall what we said significance at the 5% level meant: it is only 5% likely that a result as or more extreme as this could have occurred purely by chance.

- Note that there is no chance for a free lunch here! What happens if we reduce the size of the test (e.g. from a 5% test to a 1% test)? We reduce the chances of making a type I error ... but we also reduce the probability that we will reject the null hypothesis at all, so we increase the probability of a type II error:

Reduce size of test (e.g. 5% to 1%) $\rightarrow$ More strict criterion for rejection $\rightarrow$ Reject null hypothesis less often

$\nearrow$ Less likely to falsely reject $\rightarrow$ Lower chance of type I error

$\searrow$ More likely to incorrectly not reject $\rightarrow$ Higher chance of type II error

The Trade-off Between Type I and Type II Errors II

- So there is always a trade off between type I and type II errors when choosing a significance level. The only way we can reduce the chances of both is to increase the sample size.

A Special Type of Hypothesis Test: The t-ratio

- Recall that the formula for a test of significance approach to hypothesis testing using a t-test was

$$\textit{test statistic} = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

- If the test is $\begin{aligned} H_0 &: \beta_i = 0 \\ H_1 &: \beta_i \neq 0 \end{aligned}$

  i.e. a test that the population coefficient is zero against a two-sided alternative, this is known as a t-ratio test.

  Since $\beta_i^* = 0$, $\textit{test stat} = \frac{\hat{\beta}_i}{\mathrm{se}\left(\hat{\beta}_i\right)}$.

- The ratio of the coefficient to its SE is known as the t-ratio or t-statistic.

The t-ratio: An Example

- Suppose that we have the following parameter estimates, standard errors and t-ratios for an intercept and slope respectively.

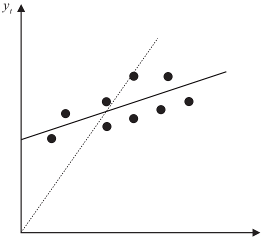
Coefficient	1.10	-4.40
se	1.35	0.96
t-ratio	0.81	-4.63

Compare this with a t_{crit} with 15-3	$=$	12 d.f.	
($2\frac{1}{2}$% in each tail for a 5% test)	$=$	2.179	5%
	$=$	3.055	1%

- Do we reject H_0: $\beta_1 = 0$? (No)
  H_0: $\beta_2 = 0$? (Yes)

What Does the t-ratio tell us?

- If we reject H_0, we say that the result is significant. If the coefficient is not "significant" (e.g. the intercept coefficient in the last regression above), then it means that the variable is not helping to explain variations in y. Variables that are not significant are usually removed from the regression model.
- In practice there are good statistical reasons for always having a constant even if it is not significant. Look at what happens if no intercept is included:



CAPM Example

- Now we are engaged with the tools for testing the significance of the CAPM
- Five years of monthly data: Jan. 2002 to April 2013.
- Asset under interest, FORD (ticker: F).

	Estimate	se	tStat	pValue
α	-0.3199	1.0864	-0.2944	0.7689
β	2.0262	0.2377	8.5227	0.0000

Table: Regression results.

- **Question**: how to interpret the results?

Section 3, References

Slides

Disclaimer:

- Some slides from Chris Brooks' book (Brooks, 2014) (copyrighted)
- Some slides from Christophe Hurlin's (University of Orleans), financial econometrics course (2019), available online.
- Some slides original (made ad-hoc for this course)

Bibliography I

[Bro14] Chris Brooks. *Introductory Econometrics for Finance*. 3rd ed. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139540872.

[Woo15] Jeffrey M Wooldridge. *Introductory econometrics: A modern approach*. Cengage learning, 2015.