Chapter 3

A brief overview of the Classical Linear Regression Model

Outlines

- Regression: concept, simple regression, best fit line, disturbance term, PRF <> SRF
- OLS estimator
- Assumptions
- Standard errors
- Hypothesis testing: 3 approaches
- Examples.

Regression

- **Regression is probably the single most important tool at the econometrician's disposal.**

  But what is regression analysis?

- **It is concerned with describing and evaluating the relationship between a given variable (usually called the dependent variable) and one or more other variables (usually known as the independent variable(s)).**

Some Notation

- **Denote the dependent variable by y and the independent variable(s) by x_1, x_2, ... , x_k where there are k independent variables.**
- **Some alternative names for the y and x variables:**

y	x
dependent variable	**independent variables**
regressand	**regressors**
effect variable	**causal variables**
explained variable	**explanatory variable**

- **In this chapter we will limit ourselves to the case where there is only one x variable, only one y variable.**

Regression is different from Correlation

- If we say *y* and *x* are correlated, it means that we are treating *y* and *x* in a completely symmetrical way.

- In regression, we treat the dependent variable (*y*) and the independent variable(s) (*x*’s) very differently. The *y* variable is assumed to be random or “stochastic” in some way, i.e. to have a probability distribution. The *x* variables are, however, assumed to have fixed (“non-stochastic”) values in repeated samples.

Simple Regression

- Simple regression is the situation where *y* depends on only one *x* variable.

- Examples of the kind of relationship that may be of interest include:
  - How asset returns vary with their level of market risk
  - Measuring the long-term relationship between stock prices and dividends.
  - Constructing an optimal hedge ratio

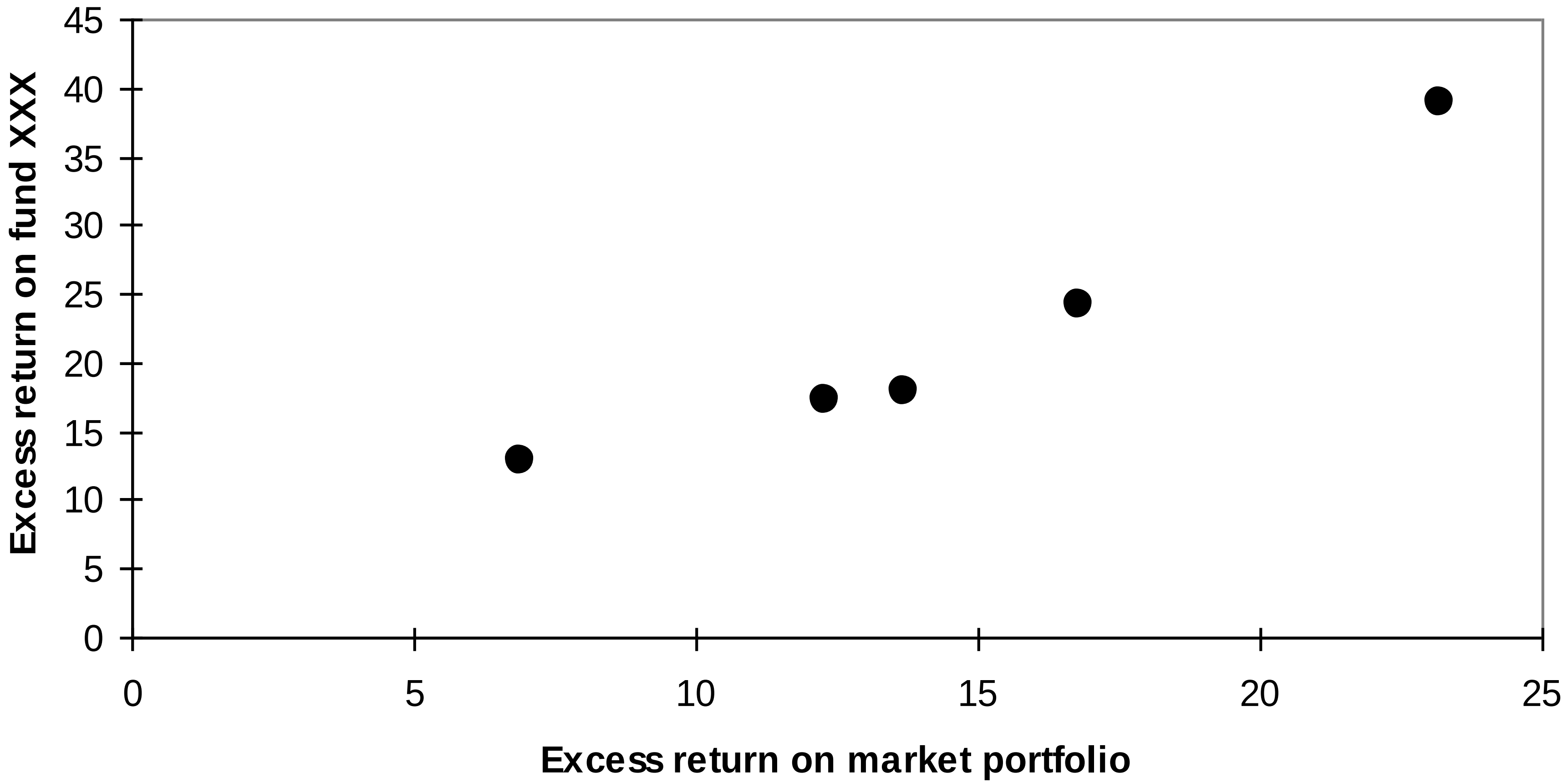
Simple Regression: An Example

- **Suppose that we have the following data on the excess returns on a fund manager's portfolio ("fund XXX") together with the excess returns on a market index:**

Year, t	Excess return $= r_{XXX,t} - rf_t$	Excess return on market index $= rm_t - rf_t$
1	17.8	13.7
2	39.0	23.2
3	12.8	6.9
4	24.2	16.8
5	17.2	12.3

- **We have some intuition that the beta on this fund is positive, and we therefore want to find whether there appears to be a relationship between *x* and *y* given the data that we have.**
- **The first stage would be to form a scatter plot of the two variables.**

Graph (Scatter Diagram)



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, this equation ($y=a+bx$) is completely deterministic.
- Is this realistic? No. So what we do is to add a random disturbance term (error), u into the equation.

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

where $t = 1,2,3,4,5$

Why do we include a Disturbance term?

- **The disturbance term (error) can capture a number of features:**

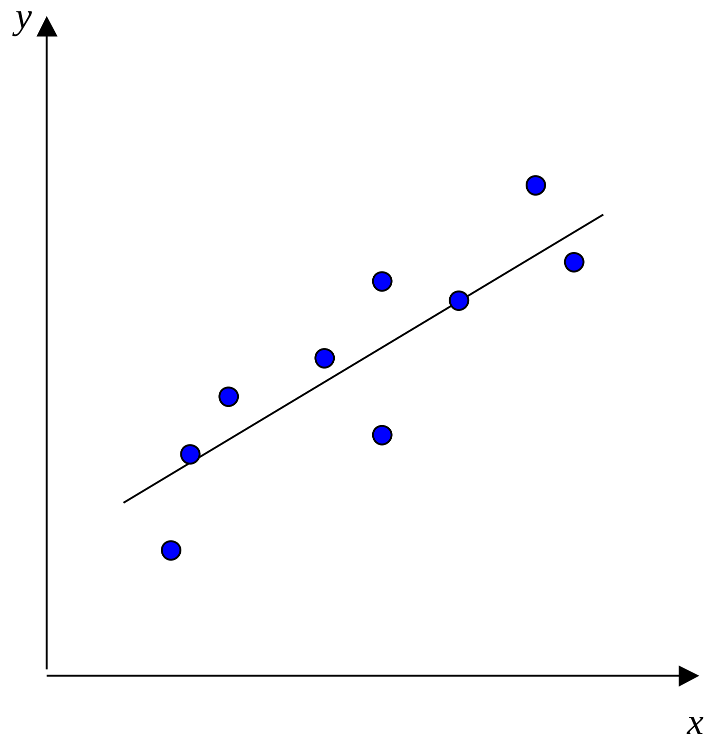
  - We always leave out some determinants of y_t

  - 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: idea

- 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

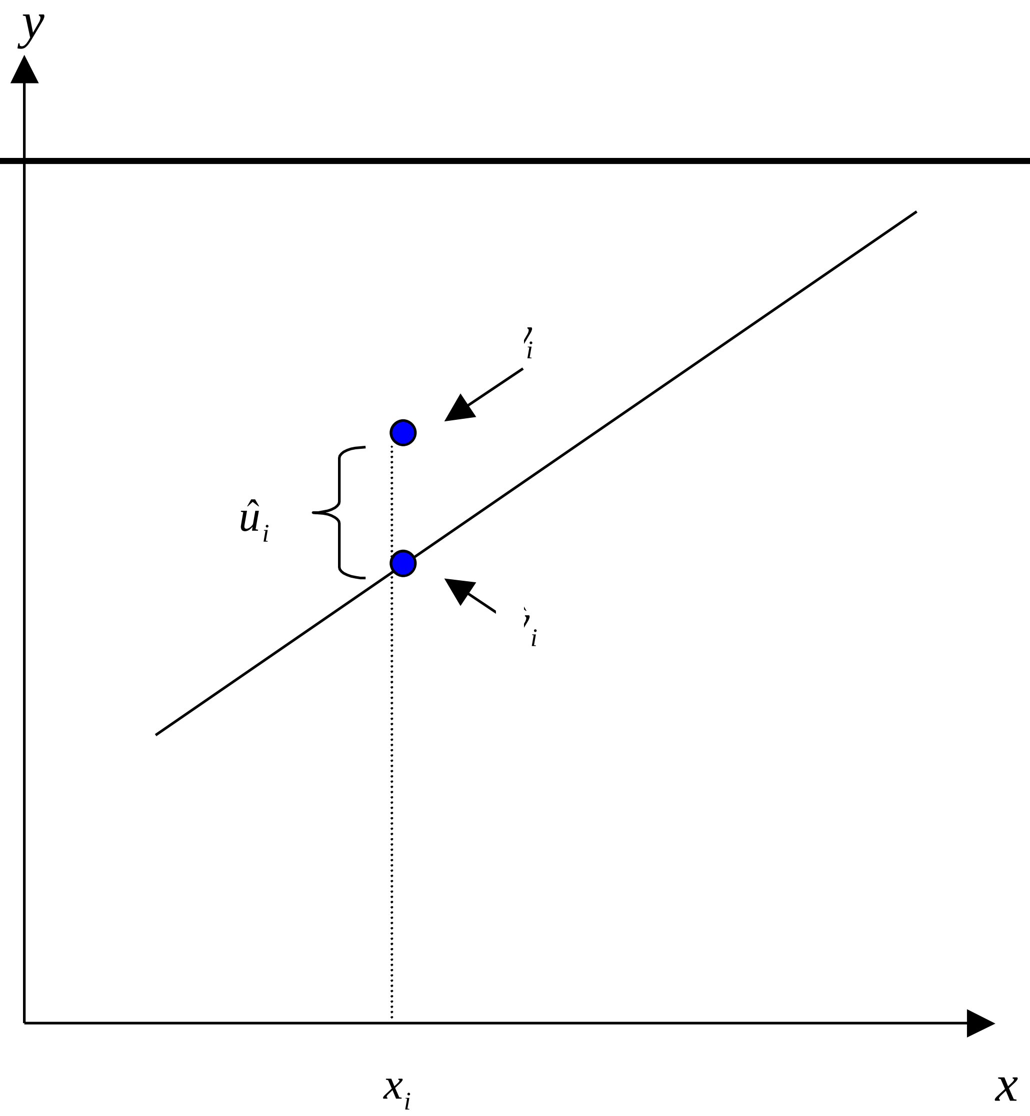
- The most common method used to fit a line to the data is known as **OLS (ordinary least squares).**
- What we actually do is take each distance and square it (i.e. take the area of each of the squares in the diagram) and minimise the total sum of the squares (hence least squares).
- Let

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 RSS**

- 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

- But $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$, so let $L = \sum_t (y_t - \hat{y}_t)^2 = \sum_i (y_t - \hat{\alpha} - \hat{\beta}x_t)^2$
- Want to minimise L with respect to (w.r.t.) $\hat{\alpha}$ and $\hat{\beta}$, so differentiate L w.r.t. $\hat{\alpha}$ and $\hat{\beta}$

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

$$\frac{\partial L}{\partial \hat{\beta}} = -2\sum_t x_t(y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \qquad (2)$$

- From (1), $\sum_t (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \Leftrightarrow \sum y_t - T\hat{\alpha} - \hat{\beta}\sum x_t = 0$
- But $\sum y_t = T\bar{y}$ and $\sum x_t = T\bar{x}$.

Deriving the OLS Estimator (cont'd)

- So we can write $T\bar{y} - T\hat{\alpha} - T\hat{\beta}\bar{x} = 0$ or $\bar{y} - \hat{\alpha} - \hat{\beta}\bar{x} = 0$ (3)
- From (2), $\sum_t x_t (y_t - \hat{\alpha} - \hat{\beta}x_t) = 0$ (4)
- From (3), $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ (5)
- Substitute into (4) for $\hat{\alpha}$ from (5),

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

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

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

Deriving the OLS Estimator (cont'd)

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

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

- So overall we have

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

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

What do We Use $\hat{\alpha}$ and $\hat{\beta}$ For?

- In the CAPM example used above, plugging the 5 observations in to make up the formulae given above would lead to the estimates
  $\hat{\alpha}$= -1.74 and $\hat{\beta}$ = 1.64. We would write the fitted line as:

$$\hat{y}_t = -1.74 + 1.64x_t$$

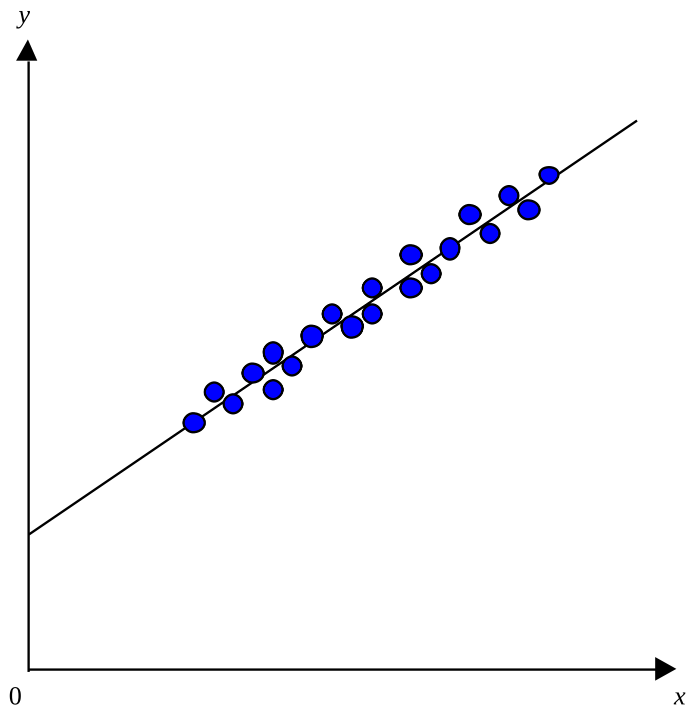
- **Question**: If an analyst tells you that she expects the market to yield a return 20% higher than the risk-free rate next year, what would you expect the return on fund XXX to be?

- **Solution**: We can say that the expected value of y = "-1.74 + 1.64 * value of x", so plug x = 20 into the equation to get the expected value for y:

$$\hat{y}_i = -1.74 + 1.64 \times 20 = 31.06$$

Accuracy of Intercept Estimate

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



The Population and the Sample

- The population is the total collection of all objects or people to be studied, for example,
- <u>Interested in</u> — <u>Population of interest</u>

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 PRF and the SRF

- The **population regression function (PRF)** is a description of the model that is thought to be generating the actual data 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 $$\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$$

  then $$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{u}_t$$

  and we also know that $\hat{u}_t = y_t - \hat{y}_t$ for the residuals.

- 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

- 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

$$z_t = \frac{1}{x_t}$$

- 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.

The Assumptions Underlying the Classical Linear Regression Model (CLRM)

- The model which we have used is known as the classical linear regression model.
- We observe data for x_t, but since y_t also depends on u_t, we must be specific about how the u_t are generated.
- We usually make the following set of assumptions about the u_t's (the unobservable error terms):
- Technical Notation — Interpretation

Technical Notation	Interpretation
1. $E(u_t) = 0$	The errors have zero mean
2. $Var(u_t) = \sigma^2$	The variance of the errors is constant and finite over all values of x_t
3. $Cov(u_i, u_j) = 0$	The errors are uncorrelated
4. $Cov(u_t, x_t) = 0$	No correlation between the error and corresponding x variate

The Assumptions Underlying the CLRM Again

- An alternative assumption to 4., which is slightly stronger, is that the x_t's are non-stochastic or fixed in repeated samples.

- A fifth assumption is required if we want to make inferences about the population parameters (the actual α and β) from the sample parameters ($\hat{\alpha}$ and $\hat{\beta}$)

- Additional Assumption

5. u_t is normally distributed

Properties of the OLS Estimator

- If assumptions 1. through 4. hold, then the estimators $\hat{\alpha}$ and $\hat{\beta}$ determined by OLS are known as **Best Linear Unbiased Estimators** (BLUE).

  What does the acronym stand for?

- "Estimator" - $\hat{\beta}$ is an estimator of the true value of β.
- "Linear" - $\hat{\beta}$ is a linear estimator
- "Unbiased" - On average, the actual value of the $\hat{\alpha}$ 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.

Precision and Standard Errors

- Any set of regression estimates of $\hat{\alpha}$ and $\hat{\beta}$ are specific to the sample used in their estimation.
- Recall that the estimators of α and β from the sample parameters ($\hat{\alpha}$ and $\hat{\beta}$) are given by

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

- What we need is some measure of the reliability or precision of the estimators ($\hat{\alpha}$ and $\hat{\beta}$). The precision of the estimate is given by its **standard error**. Given assumptions 1 - 4 above, then the standard errors can be shown to be given by

$$SE(\hat{\alpha}) = s\sqrt{\frac{\sum x_t^2}{T\sum (x_t - \bar{x})^2}} = s\sqrt{\frac{\sum x_t^2}{T\sum x_t^2 - T^2\bar{x}^2}},$$

$$SE(\hat{\beta}) = s\sqrt{\frac{1}{\sum (x_t - \bar{x})^2}} = s\sqrt{\frac{1}{\sum x_t^2 - T\bar{x}^2}}$$

where ***s* is the estimated standard deviation of the residuals.**

Estimating the Variance of the Disturbance Term

- The variance of the random variable u_t is given by

  $$\text{Var}(u_t) = \text{E}[(u_t)\text{-E}(u_t)]^2$$

  which reduces to

  $$\text{Var}(u_t) = \text{E}(u_t^2)$$

- We could estimate this using the average of u_t^2:

  $$s^2 = \frac{1}{T}\sum u_t^2$$

- Unfortunately this is not workable since u_t is not observable. We can use the sample counterpart to u_t, which is $\hat{u}_t$:

  $$s^2 = \frac{1}{T}\sum \hat{u}_t^2$$

  But this estimator is a biased estimator of σ^2.

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, s is the estimated standard deviation of the residuals

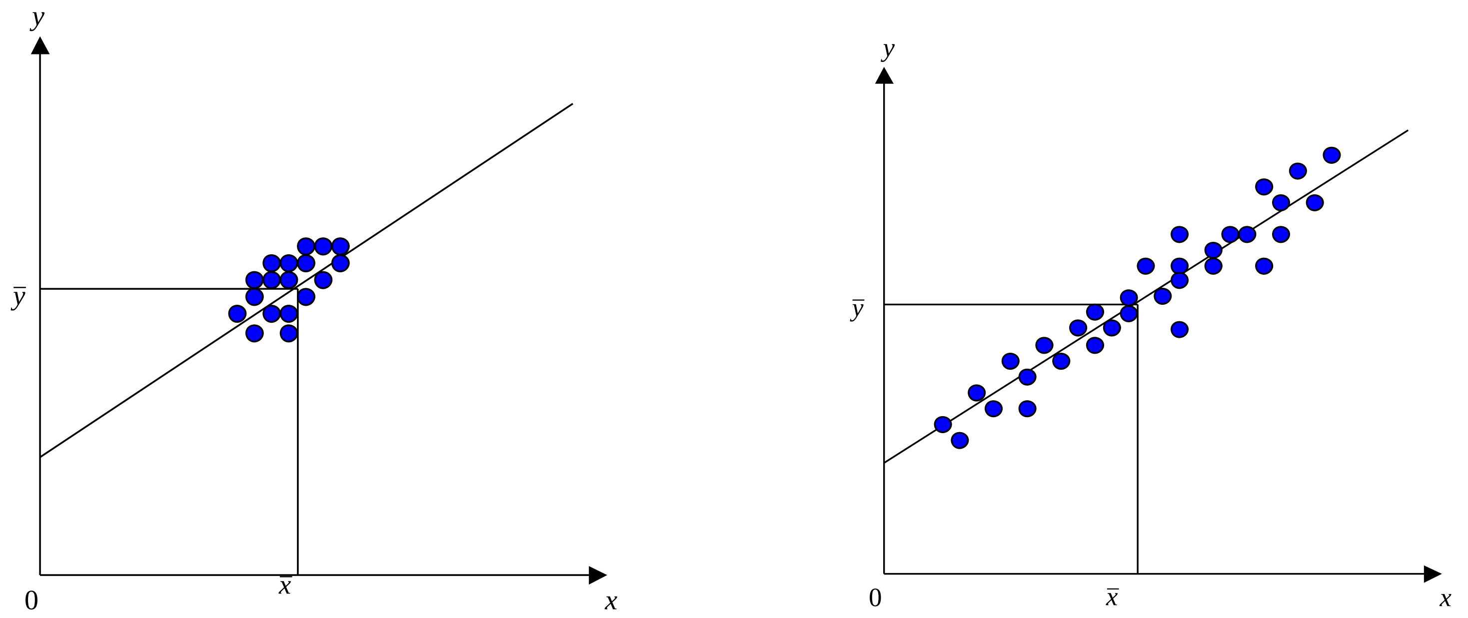
Some Comments on the **Standard Error Estimators**

1. Both SE($\hat{\alpha}$) and SE($\hat{\beta}$) depend on s^2 (or s). The greater the variance s^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.

Some Comments on the Standard Error Estimators (cont'd)

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



Some Comments on the Standard Error Estimators (cont'd)

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

T appears implicitly since the sum $\sum (x_t - \bar{x})^2$ is from t = 1 to T.

4. The term $\sum x_t^2$ appears in the SE($\hat{\alpha}$).

The reason is that $\sum x_t^2$ measures how far the points are away from the y-axis.

Example: How to Calculate the Parameters and Standard Errors

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

$$\sum x_t y_t = 830102,\ T = 22,\ \bar{x} = 416.5,\ \bar{y} = 86.65,$$

$$\sum x_t^2 = 3919654,\ RSS = 130.6$$

- Calculations: $\hat{\beta} = \dfrac{830102 - (22 * 416.5 * 86.65)}{3919654 - 22 * (416.5)^2} = 0.35$

$$\hat{\alpha} = 86.65 - 0.35 * 416.5 = -59.12$$

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

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

Example (cont'd)

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

$$SE(\hat{\alpha}) = 2.55 * \sqrt{\frac{3919654}{(22 \times 3919654) - (22^2 \times 416.5^2)}} = 3.35$$

$$SE(\hat{\beta}) = 2.55 * \sqrt{\frac{1}{3919654 - (22 \times 416.5^2)}} = 0.0079$$

- We now write the results as

$$\hat{y}_t = -59.12 + 0.35 x_t$$
$$(3.35) \quad (0.0079)$$

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 **3 ways** to conduct a hypothesis test: via the **test of significance** approach, or **confidence interval** approach, or **p-value**

The OLS Estimators Probability Distribution

- **We assume that** $u_t \sim \mathrm{N}(0,\sigma^2)$

- Since the least squares estimators are linear combinations of the random variables
  i.e. $\hat{\beta} = \sum w_t y_t$

- The weighted sum of normal random variables is also normally distributed, so
  $$\hat{\alpha} \sim \mathrm{N}(\alpha, \mathrm{Var}(\hat{\alpha}))$$
  $$\hat{\beta} \sim \mathrm{N}(\beta, \mathrm{Var}(\hat{\beta}))$$

- **Note:** What if the errors are not normally distributed? Will the parameter estimates still be normally distributed? **Yes, if the other assumptions of the CLRM hold, and the sample size is sufficiently large.**

The Probability Distribution of the Least Squares Estimators (cont'd)

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

$$\frac{\hat{\alpha}-\alpha}{\sqrt{\operatorname{var}\left(\hat{\alpha}\right)}} \sim N\left(0,1\right) \text{ and } \frac{\hat{\beta}-\beta}{\sqrt{\operatorname{var}\left(\hat{\beta}\right)}} \sim N\left(0,1\right)$$

- But var($\hat{\alpha}$) and var($\hat{\beta}$) are unknown, so

$$\boxed{\frac{\hat{\alpha}-\alpha}{SE(\hat{\alpha})} \sim t_{T-2} \quad \text{and} \quad \frac{\hat{\beta}-\beta}{SE(\hat{\beta})} \sim t_{T-2}}$$

Testing Hypotheses:
The Test of Significance Approach

- Assume the regression equation is given by ,

$$y_t = \alpha + \beta x_t + u_t \qquad \text{for } t=1,2,\ldots,T$$

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

1. Estimate $\hat{\alpha}$, $\hat{\beta}$ and $SE(\hat{\alpha})$, $SE(\hat{\beta})$ in the usual way

2. Calculate the test statistic. This is given by the formula

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

where β^* is the value of β under the null hypothesis.

The Test of Significance Approach (cont'd)

3. 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 *T*-2 degrees of freedom.

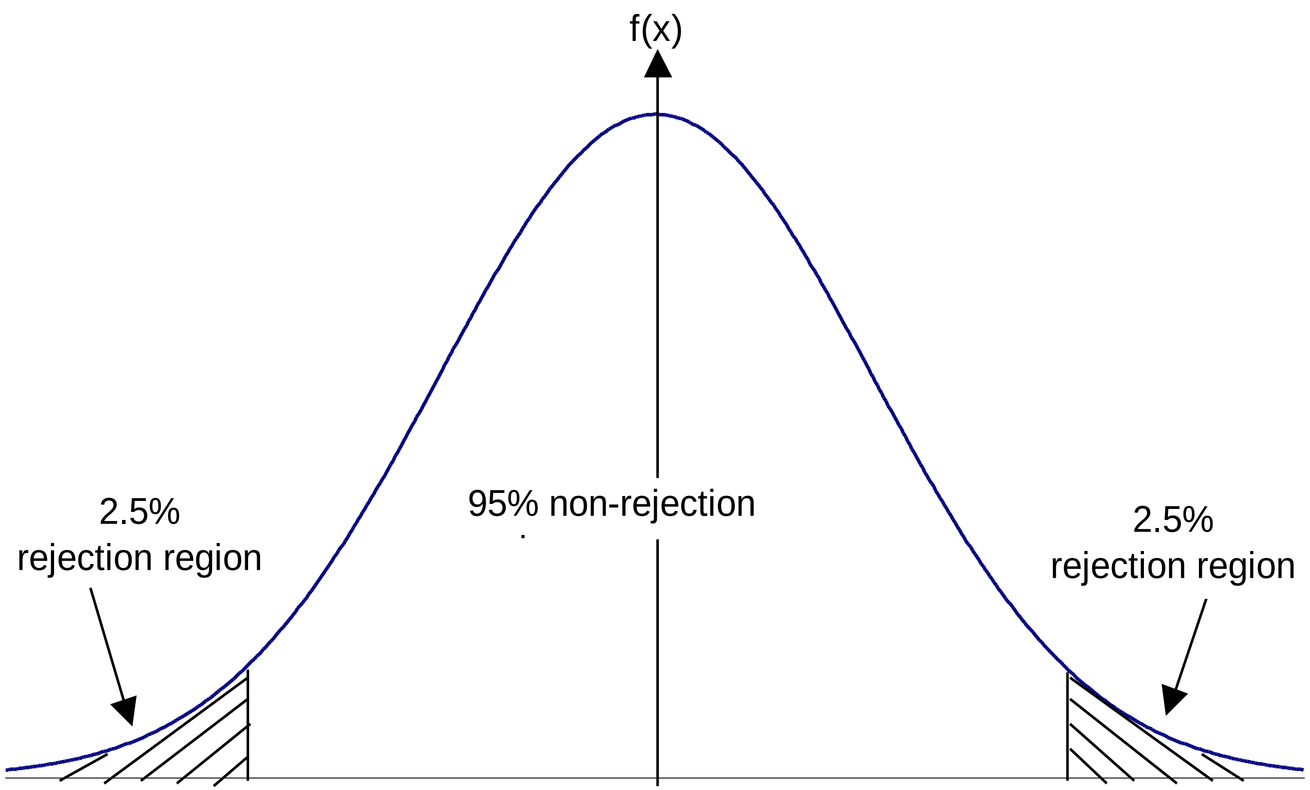
4. We need to choose a "significance level", often denoted α. This is also sometimes called the size of the test and it determines the region where we will reject or not reject the null hypothesis that we are testing. It is conventional to use a significance level of 5%.

5. Given a significance level, we can determine a rejection region and non-rejection region.

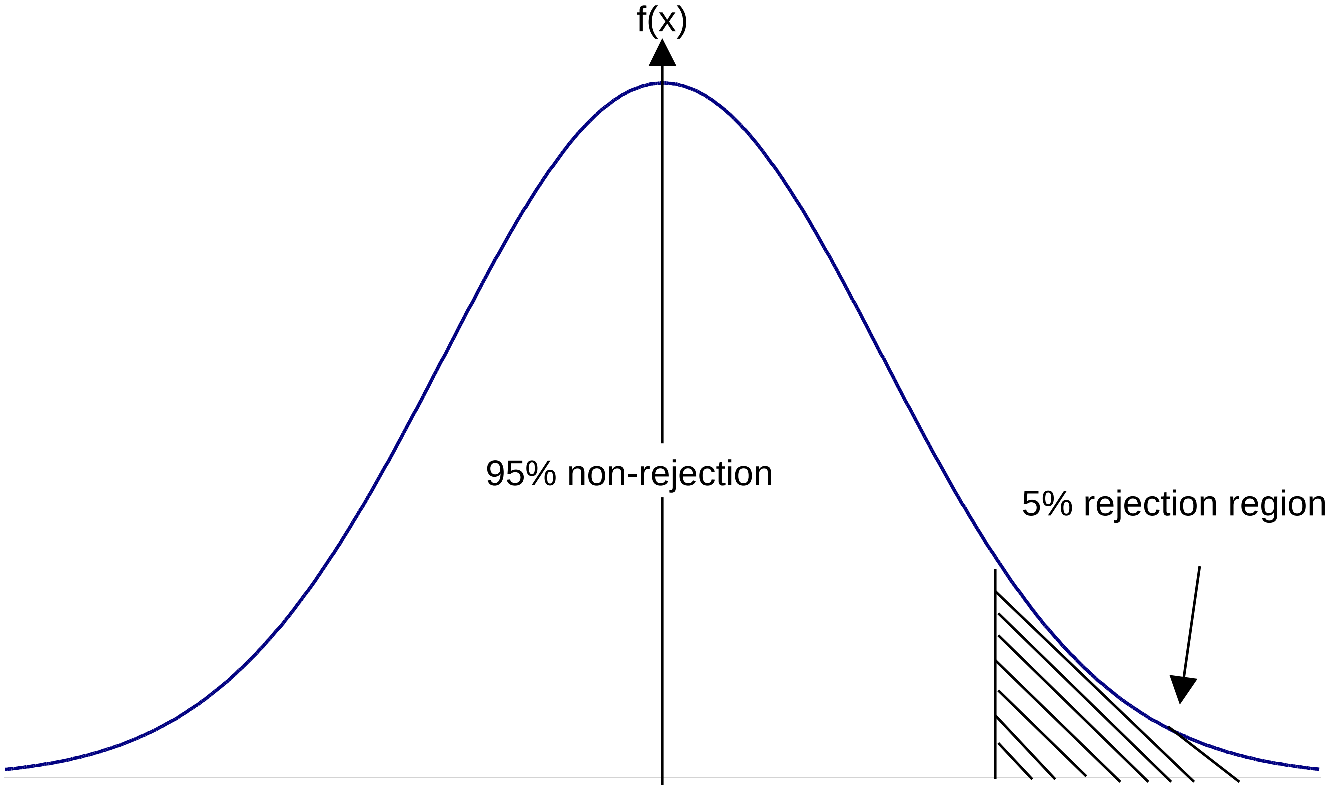
6. Use the *t*-tables to obtain a critical value or values with which to compare the test statistic.

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

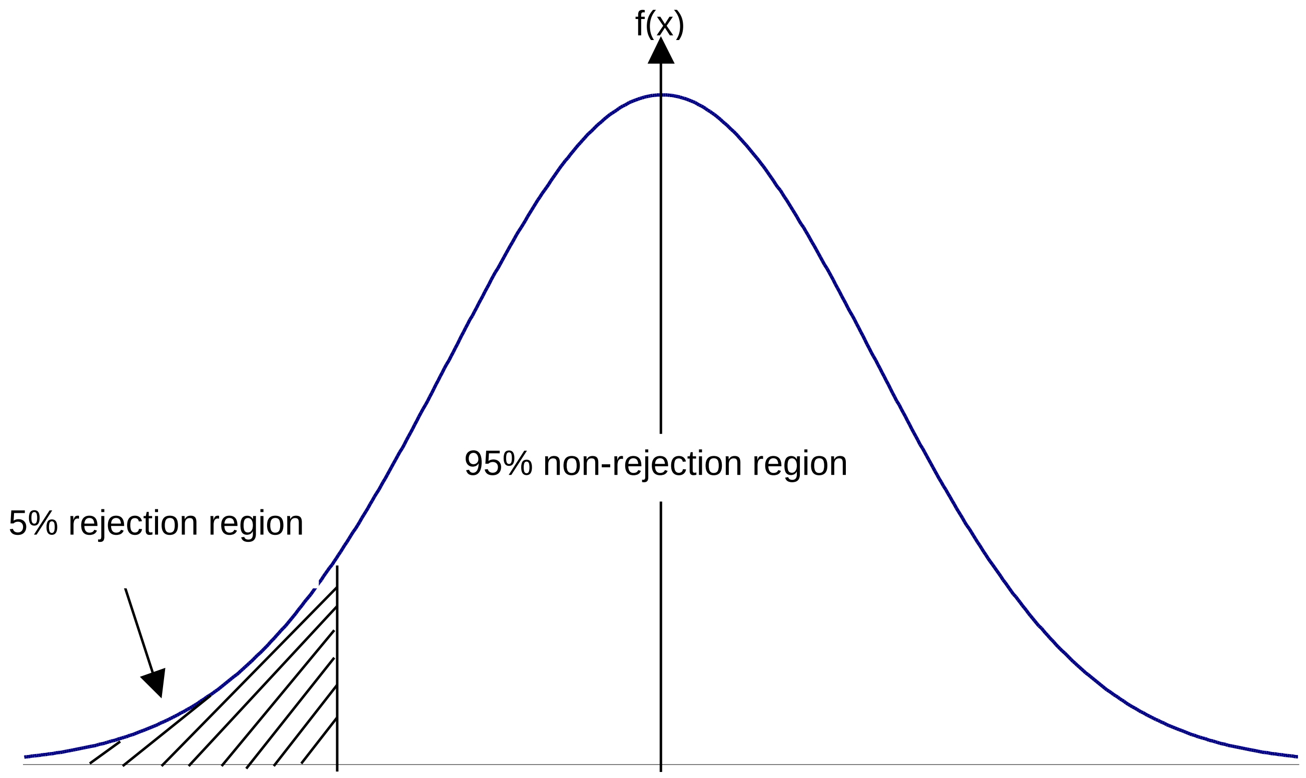
Rejection Region for a 2-tailed test



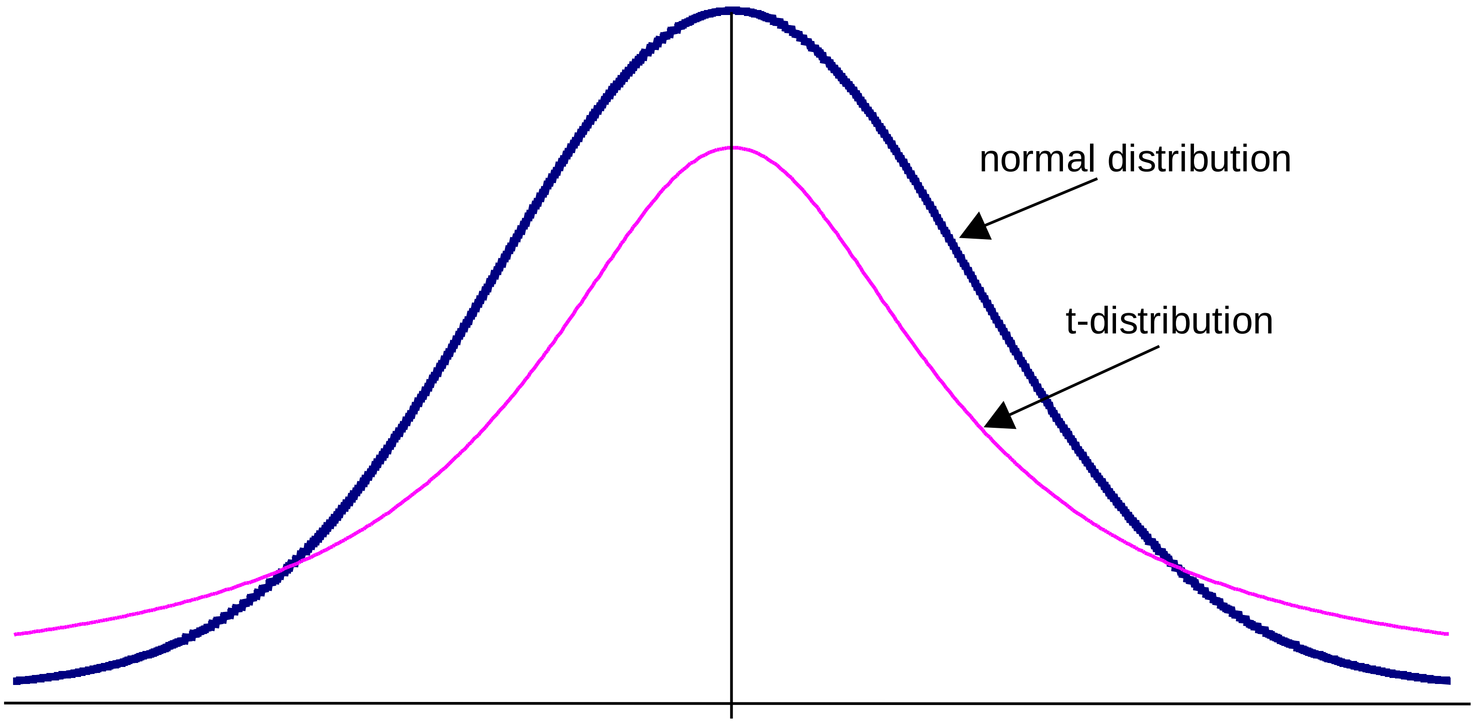
The Rejection Region for a 1-sided Test (Upper Tail)



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



What Does the *t*-Distribution Look Like?



How to Carry out a Hypothesis Test Using Confidence Intervals (2-tailed test)

1. Calculate $\hat{\alpha}$, $\hat{\beta}$ and $SE(\hat{\alpha})$, $SE(\hat{\beta})$ as before.

2. Choose a significance level, α, (again the convention is 5%). This is equivalent to choosing a (1-α)×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 *T*-2 degrees of freedom.

4. The confidence interval is given by $(\hat{\beta} - t_{crit} \times SE(\hat{\beta}), \hat{\beta} + t_{crit} \times SE(\hat{\beta}))$

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 (1st method)

- 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

$$-t_{crit} \le \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})} \le +t_{crit}$$

- Rearranging, we would not reject if

$$-t_{crit} \times SE(\hat{\beta}) \le \hat{\beta} - \beta^* \le +t_{crit} \times SE(\hat{\beta})$$

$$\hat{\beta} - t_{crit} \times SE(\hat{\beta}) \le \beta^* \le \hat{\beta} + t_{crit} \times SE(\hat{\beta})$$

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

Constructing Tests of Significance and Confidence Intervals: An Example

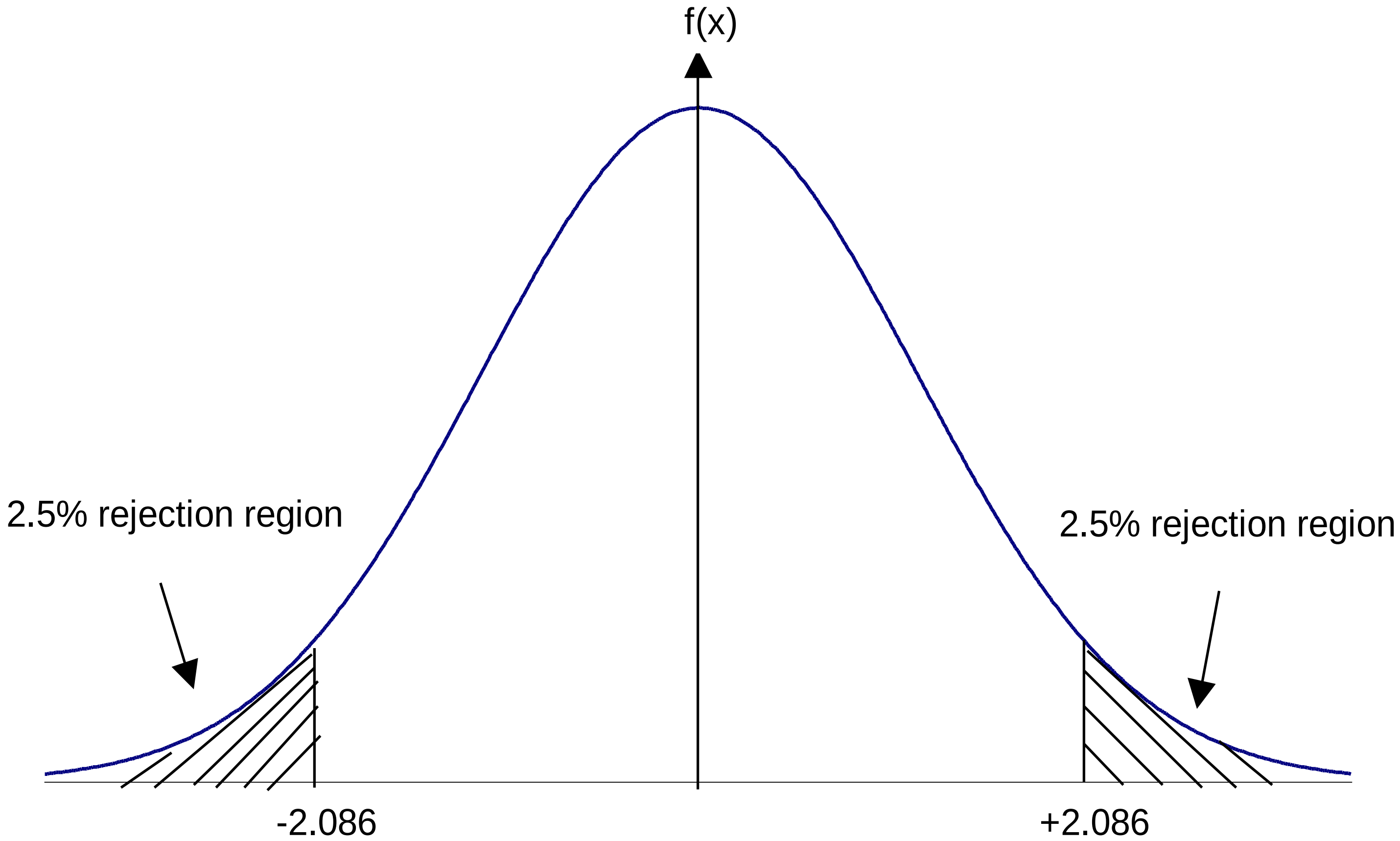
- Using the regression results above,

$$\hat{y}_t = \underset{(14.38)}{20.3} + \underset{(0.2561)}{0.5091} x_t \quad , \quad T=22$$

- 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 $t_{crit} = t_{2.5\%}$

Determining the Rejection Region



Performing the Test

- The hypotheses are:

$$H_0 : \beta = 1$$
$$H_1 : \beta \neq 1$$

Test of significance approach

$$test\ stat = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$$
$$= \frac{0.5091 - 1}{0.2561} = -1.917$$

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

Confidence interval approach

$$\hat{\beta} \pm t_{crit} \times SE(\hat{\beta})$$
$$= 0.5091 \pm 2.086 \times 0.2561$$
$$= (-0.0251, 1.0433)$$

Since 1 lies within the confidence interval, do not reject H_0

Testing other Hypotheses

- What if we wanted to test $H_0 : \beta = 0$ or $H_0 : \beta = 2$?

- Note that we can test these at 5% significance level with the confidence interval approach. For example, test:

$$
\begin{array}{ll}
 & H_0 : \beta = 0 \\
\text{vs.} & H_1 : \beta \neq 0
\end{array}
$$

$$
\begin{array}{ll}
 & H_0 : \beta = 2 \\
\text{vs.} & H_1 : \beta \neq 2
\end{array}
$$

Changing the significance level

- If we use a different significance level, the test of significance approach is better than a confidence interval.

- For example, we wanted to use a 10% significance level. Using the test of significance approach,

$$test\ stat = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$$

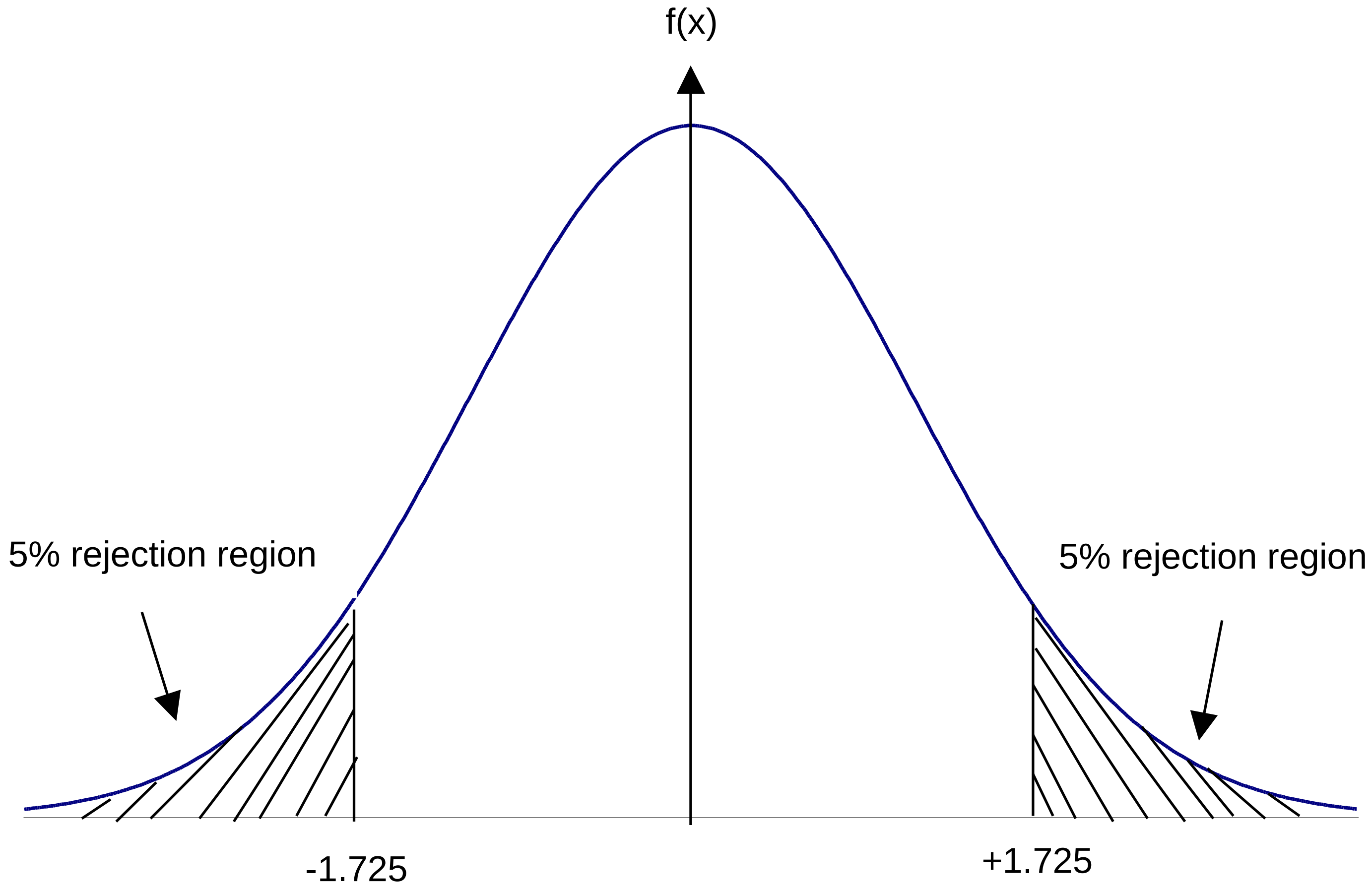
$$= \frac{0.5091 - 1}{0.2561} = -1.917$$

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

- $t_{20;10\%}$ = 1.725. So now, as the test statistic lies in the rejection region, we would reject H_0.

- If we reject the null hypothesis at the 5% level, we say that the result of the test is statistically significant.

Changing the significance level :
The New Rejection Regions



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 $= \alpha$	√
	Insignificant (do not reject H_0)	√	Type II error $= \beta$

The Trade-off Between Type I and Type II Errors

- The probability of a type I error is just α, the significance level (or size of test) we chose.
- 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 → more strict criterion for rejection → reject null hypothesis less often ↗ less likely to falsely reject

↘ more likely to incorrectly not reject

- 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

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

- If the test is $H_0 : \beta_i = 0$
  $H_1 : \beta_i \neq 0$

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

  Since $\beta_i^* = 0$, $test\ stat = \dfrac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$

- 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

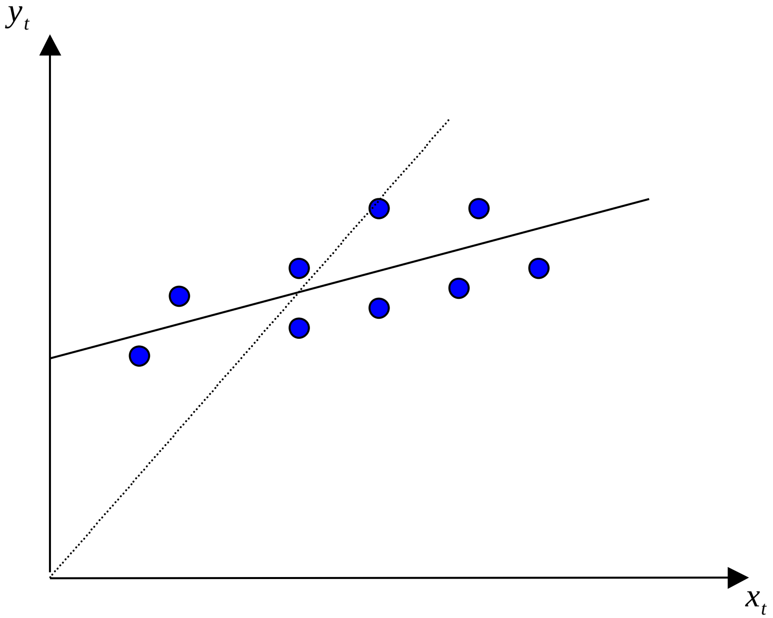
Compare this with a t_{crit} with 15-2	=	13 d.f.	
(2½% in each tail for a 5% test)	=	2.160	5%
	=	3.012	1%

- 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:



The p-value

- This is equivalent to choosing an infinite number of critical t-values from tables. It gives us the marginal significance level where we would be indifferent between rejecting and not rejecting the null hypothesis.

- If the test statistic is large in absolute value, the p-value will be small, and vice versa. The p-value gives the plausibility of the null hypothesis.

  e.g. a test statistic is distributed as a $t_{62} = 1.47$.

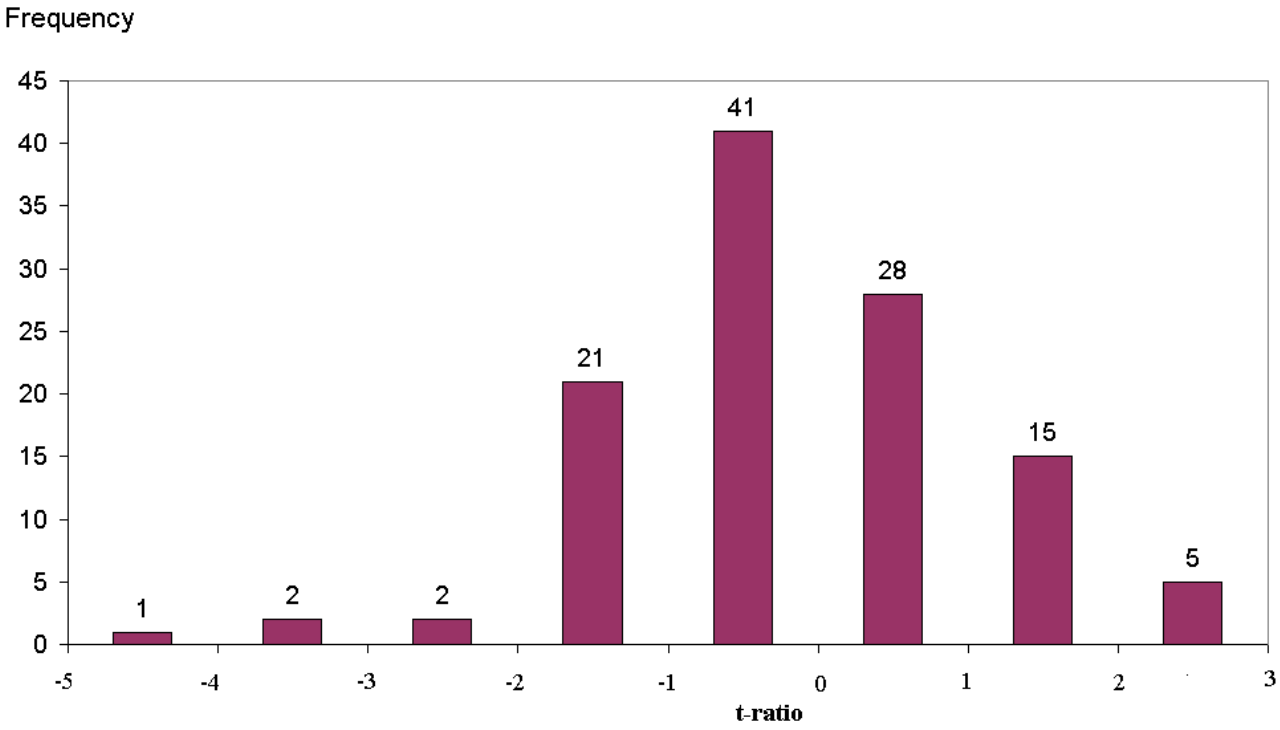
  The p-value = 0.12 (from the computer software)

- Do we reject at the 5% level?...........................No
- Do we reject at the 10% level?.........................No
- Do we reject at the 20% level?.........................Yes

An Example of the Use of a Simple *t*-test to Test a Theory in Finance
Can US mutual funds beat the market?

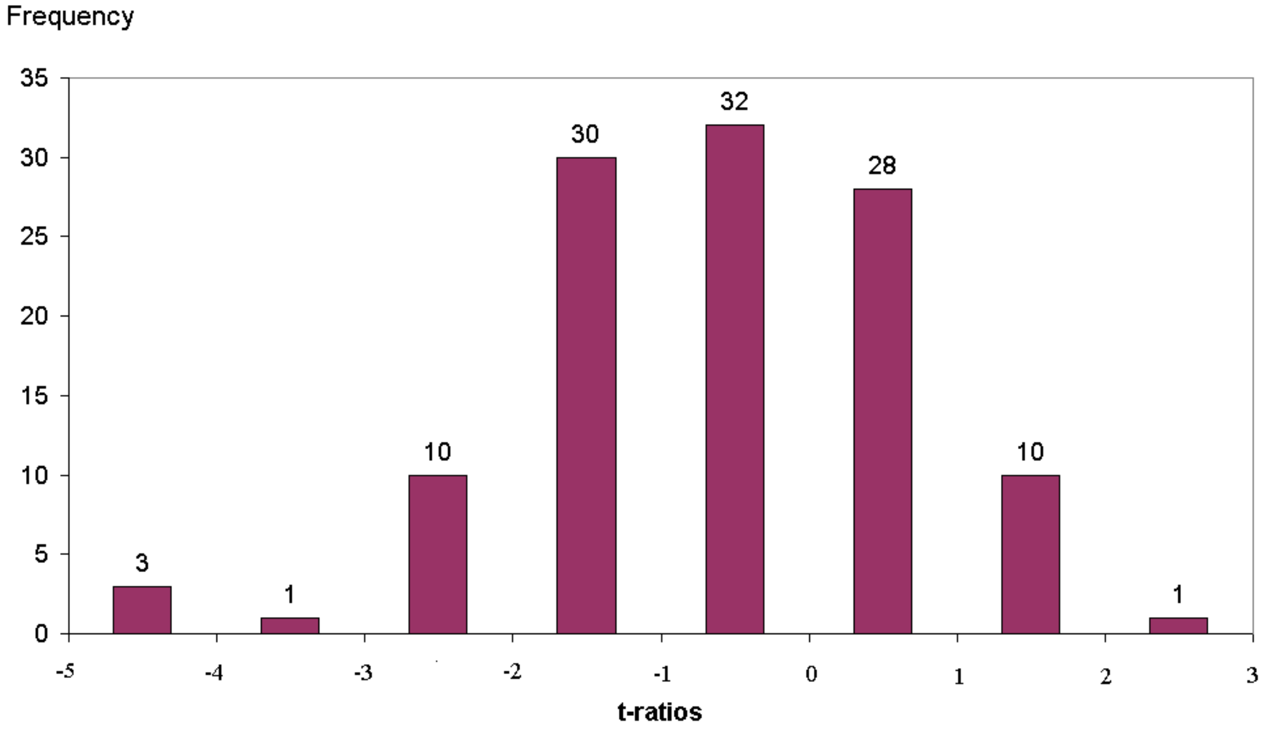
- Testing for the presence and significance of **abnormal returns** ("**Jensen's alpha**" - Jensen, 1968).
- The Data: Annual Returns on the portfolios of 115 mutual funds from 1945-1964.
- The model: $R_{jt} - R_{ft} = \alpha_j + \beta_j (R_{mt} - R_{ft}) + u_{jt}$ for $j = 1, \ldots, 115$
- We are interested in the significance of α_j.
- The null hypothesis is H_0: $\alpha_j = 0$.

Frequency Distribution of *t*-ratios of Mutual Fund Alphas (gross of transactions costs)



Source Jensen (1968). Reprinted with the permission of Blackwell publishers.

Frequency Distribution of *t*-ratios of Mutual Fund Alphas (net of transactions costs)



Source Jensen (1968). Reprinted with the permission of Blackwell publishers.

Example 2. Can UK Unit Trust Managers "Beat the Market"?

- We now perform a variant on Jensen's test in the context of the UK market, considering monthly returns on 76 equity unit trusts. The data cover the period January 1979 – May 2000 (257 observations for each fund). Some summary statistics for the funds are:

	Mean	Minimum	Maximum	Median
Average monthly return, 1979-2000	1.0%	0.6%	1.4%	1.0%
Standard deviation of returns over time	5.1%	4.3%	6.9%	5.0%

- Jensen Regression Results for UK Unit Trust Returns, January 1979-May 2000

$$R_{jt} - R_{ft} = \alpha_j + \beta_j (R_{mt} - R_{ft}) + \varepsilon_{jt}$$

Can UK Unit Trust Managers "Beat the Market"? Results

Estimates of	Mean	Minimum	Maximum	Median
α	-0.02%	-0.54%	0.33%	-0.03%
β	0.91	0.56	1.09	0.91
t-ratio on α	-0.07	-2.44	3.11	-0.25

- In fact, gross of transactions costs, 9 funds of the sample of 76 were able to significantly out-perform the market by providing a significant positive alpha, while 7 funds yielded significant negative alphas.

Example 3. The Overreaction Hypothesis and the UK Stock Market

- <u>Motivation</u>
  Two studies by DeBondt and Thaler (1985, 1987) showed that **stocks which experience a poor performance over a 3 to 5 year period tend to outperform stocks which had previously performed relatively well.**

- <u>How Can This be Explained?</u> 2 suggestions

  - A manifestation of the **size effect: tendency of small firms to generate higher returns to larger firms, on average**

  - Reversals reflect changes in equilibrium required returns: **losers are likely more risky with higher beta**.

- <u>Another interesting anomaly</u>: the **January effect**.

  – Zarowin (1990) finds that 80% of the **extra return available from holding the losers accrues to investors in January**.

Methodology

- **Example study: Clare and Thomas (1995)**

  Data:

  Monthly UK stock returns from January 1955 to 1990 (**36 years**) on all firms traded on the London Stock exchange.

- Calculate the **monthly excess return** of the stock over the market over a 12, 24 or 36 month period for each stock *i*:

$$U_{it} = R_{it} - R_{mt}$$

- Calculate the **average monthly excess return** for the stock *i* over the first 12, 24, or 36 month period:

$$\bar{R}_i = \frac{1}{n}\sum_{t=1}^{n} U_{it} \qquad n = 12, 24 \text{ or } 36 \text{ months}$$

Portfolio Formation

- Then rank the stocks from highest average return to lowest and **form 5 portfolios:**

Portfolio 1:	Best performing 20% of firms
Portfolio 2:	Next 20%
Portfolio 3:	Next 20%
Portfolio 4:	Next 20%
Portfolio 5:	Worst performing 20% of firms.

- Use the same sample length *n* to monitor the performance of each portfolio.

Portfolio Formation and Portfolio Tracking Periods

- How many samples of length *n* have we got?

  n = 1, 2, or 3 years.

- If *n* = 1 year:

  Estimate for year 1

  Monitor portfolios for year 2

  Estimate for year 3

  Monitor for year 3

  ….

  Monitor portfolios for year 36

- **So if *n* = 1, we have 18 INDEPENDENT (non-overlapping) observation / tracking periods.**

Constructing Winner and Loser Returns

- Similarly, ***n* = 2 gives 9 independent periods and *n* = 3 gives 6 independent periods.**
- Calculate monthly portfolio returns assuming an equal weighting of stocks in each portfolio.
- Denote the **mean return for each month over the 18, 9 or 6 periods for the winner and loser portfolios** respectively as $\bar{R}_p^W$ and $\bar{R}_p^L$ respectively.
- Define the difference between these as $\bar{R}_{Dt} = \bar{R}_p^L - \bar{R}_p^W$.
- Then perform the first regression and look at the significance of α_1:

$$\bar{R}_{Dt} = \alpha_1 + \eta_t \qquad \text{(Test 1)}$$

Allowing for Differences in the Riskiness of the Winner and Loser Portfolios

- **Problem**: Significant and positive α_1 could be due to higher return being required on loser stocks due to loser stocks being more risky.

- **Solution:** Allow for risk differences by regressing against the market risk premium:

$$\bar{R}_{Dt} = \alpha_2 + \beta(R_{mt} - R_{ft}) + \eta_t \qquad \text{(Test 2)}$$

where

R_{mt} is the return on the FTA All-share

R_{ft} is the return on a UK government 3 month t-bill.

Is there an Overreaction Effect in the UK Stock Market? Results

Panel A: All Months	n = 12	n = 24	n =36
Return on Loser	0.0033	0.0011	0.0129
Return on Winner	0.0036	-0.0003	0.0115
Implied annualised return difference	-0.37%	1.68%	1.56%
Coefficient for (3.47): $\hat{\alpha}_1$	-0.00031	0.0014**	0.0013
	(0.29)	(2.01)	(1.55)
Coefficients for (3.48): $\hat{\alpha}_2$	-0.00034	0.00147**	0.0013*
	(-0.30)	(2.01)	(1.41)
$\hat{\beta}$	-0.022	0.010	-0.0025
	(-0.25)	(0.21)	(-0.06)
Panel B: All Months Except January			
Coefficient for (3.47): $\hat{\alpha}_1$	-0.0007	0.0012*	0.0009
	(-0.72)	(1.63)	(1.05)

Notes: *t*-ratios in parentheses; * and ** denote significance at the 10% and 5% levels respectively. Source: Clare and Thomas (1995). Reprinted with the permission of Blackwell Publishers.

Testing for Seasonal Effects in Overreactions

- **Is there evidence that losers out-perform winners more at one time of the year than another?**
- To test this, calculate the difference between the winner & loser portfolios as previously, $\bar{R}_{Dt}$, and **regress this on 12 month-of-the-year dummies**:

$$\bar{R}_{Dt} = \sum_{i=1}^{12} \delta_i M_i + v_t$$

- Significant out-performance of losers over winners in,
  - June (for the 24-month horizon), and
  - January, April and October (for the 36-month horizon)
  - winners appear to stay significantly as winners in
    - March (for the 12-month horizon).

Conclusions

- **Evidence of overreactions in stock returns**.
- Losers tend to be small so we can attribute most of the overreaction in the UK to the size effect.

Comments

- Small samples
- No diagnostic checks of model adequacy