Chapter 3

A brief overview of the
classical linear regression model

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 variables

- Note that there can be many x variables but we will limit ourselves to the case where there is only one x variable to start with. In our set-up, there is 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

- For simplicity, say $k=1$. This 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)
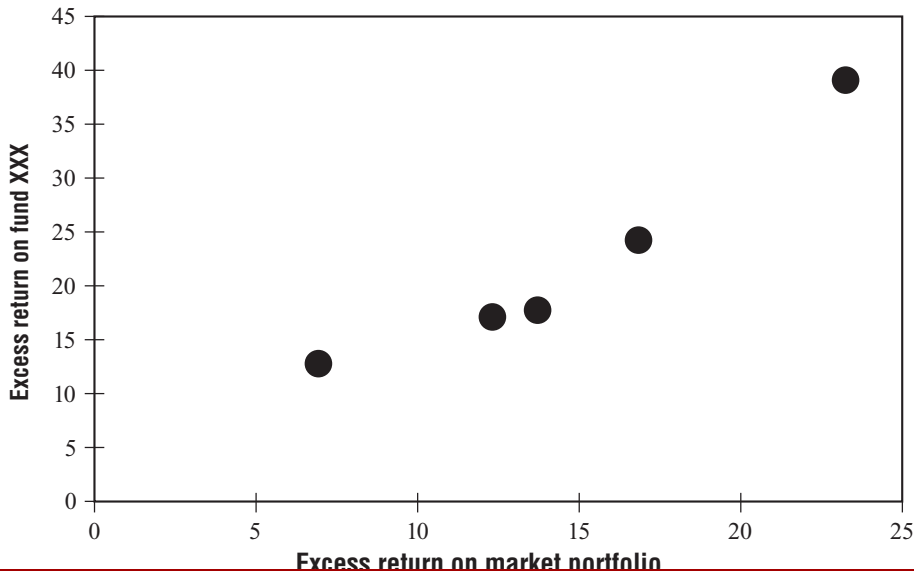
Finding a Line of Best Fit

- 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, 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 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):
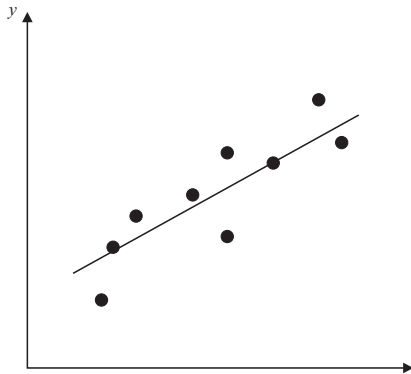
Ordinary Least Squares

- 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
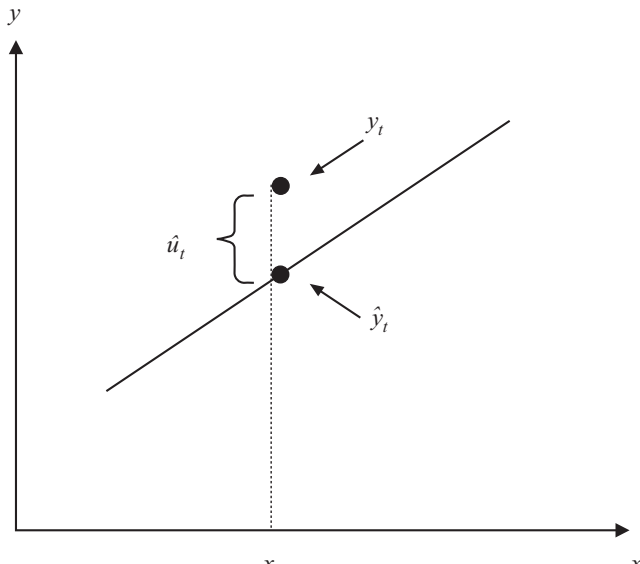
How OLS Works

- 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

- But $\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t$, so let

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

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

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

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

Deriving the OLS Estimator (Cont'd)

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

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

- From (2),

$$\sum_t x_t(y_t - \hat{\alpha} - \hat{\beta}x_t) = 0 \tag{4}$$

- From (3),

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{5}$$

Deriving the OLS Estimator (Cont'd)

- Substitute into (4) for $\hat{\alpha}$ from (5),

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

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

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

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

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

Deriving the OLS Estimator (Cont'd)

- So overall we have

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

- 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:

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

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:
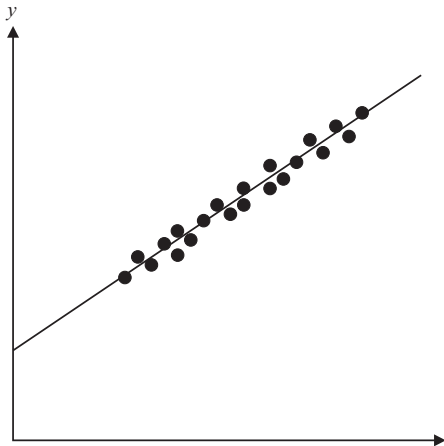
The Population and the Sample

- 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>
  predicting outcome   the entire electorate
  of an election

- 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 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 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

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

- 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

- Consistent

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

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

- Unbiased

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

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

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

Desirable properties of estimators (Cont'd)

- <u>Efficient</u>

  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 minimising the probability that it is a long way off from the true value of β.

The Assumptions Underlying the 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
(1) $\mathrm{E}(u_t) = 0 \ \forall t$	The errors have zero mean
(2) $\mathrm{var}(u_t) = \sigma^2 \ \forall t$	The variance of the errors is constant and finite over all values of x_t
(3) $\mathrm{cov}(u_t, u_s) = 0 \ \forall t \neq s$	The errors are "independent" of one another
(4) $\mathrm{cov}(u_t, x_t) = 0$	There is no relationship between the error and corresponding x variate

The Assumptions Underlying the CLRM (Cont'd)

- 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 OLS and requirements

- 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 $Var(u_t) = < \infty \ \forall t$.

- Unbiased

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

- Efficient

  The least squares estimates of $\hat{\alpha}$ and $\hat{\beta}$ are efficient if assumptions (2) and (3) hold.

The OLS Estimator as BLUE

- If assumptions (1) through (4) hold, 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; the Gauss–Markov theorem proves that the OLS estimator is best.

Precision and Standard Errors

- Any set of regression estimates of α and β 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

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

Precision and Standard Errors (Cont'd)

- 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

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

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

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

Estimating the Variance of the Disturbance Term

- The variance of the random variable u_t is given by
$$\text{Var}(u_t) = E[(u_t)\text{-}E(u_t)]^2$$

  which reduces to

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

- We could estimate this using the average of u_t^2:
$$\sigma^2 = \frac{1}{T} \sum u_t^2$$

- 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.

- 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 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.

Some Comments on the Standard Error Estimators

Consider what happens if $\sum (x_t - \bar{x})^2$ is small or large:
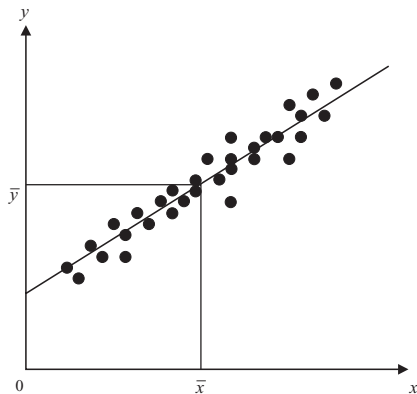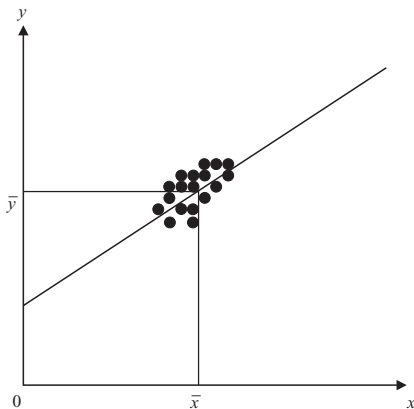
Some Comments on the Standard Error Estimators (Cont'd)

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

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

2. 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

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

Example: How to Calculate the Parameters and Standard Errors (Cont'd)

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

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

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

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

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

- We now write the results as

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

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.

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.

The Probability Distribution of the Least Squares Estimators

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

- 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

  $$\hat{\alpha} \sim N(\alpha, Var(\alpha))$$
  $$\hat{\beta} \sim N(\beta, Var(\beta))$$

- What if the errors are not normally distributed? Will the parameter estimators still be normally distributed?

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

- Yes, if the other assumptions of the CLRM hold, and the sample size is sufficiently large.

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

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

- But var(α) and var(β) are unknown, so

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

Testing Hypotheses:
The Test of Significance Approach

- Assume the regression equation is given by,

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

- 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.

  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.
  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.

Testing Hypotheses:
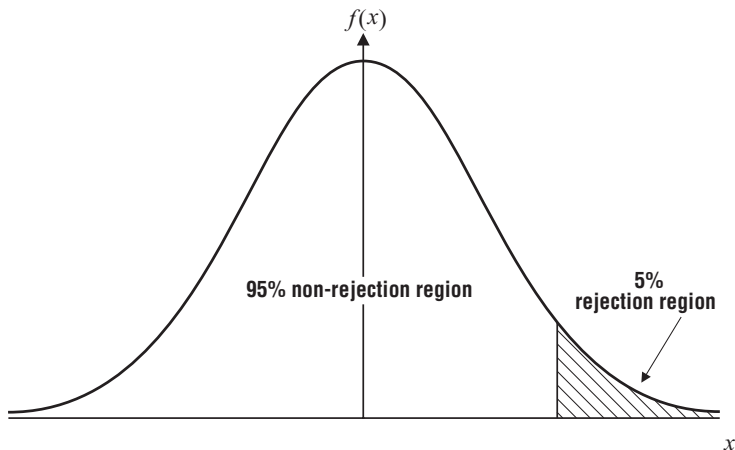# The Test of Significance Approach (Cont'd)

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%.
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 to use a 5% size of test, but 10% and 1% are also commonly used.

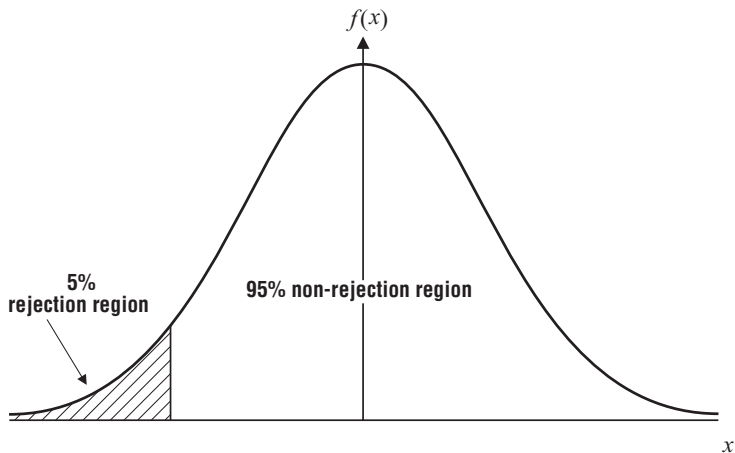# Determining the Rejection Region for a Test of Significance

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

$f(x)$

2.5% rejection region

95% non-rejection region

2.5% rejection region
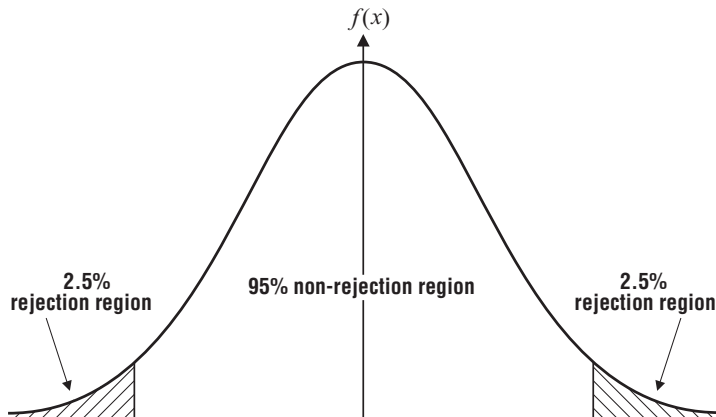
The Rejection Region for a 1-Sided Test (Upper Tail)

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.

7. Finally perform the test. If the test statistic lies in the rejection region then reject the null hypothesis (H_0), else 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?
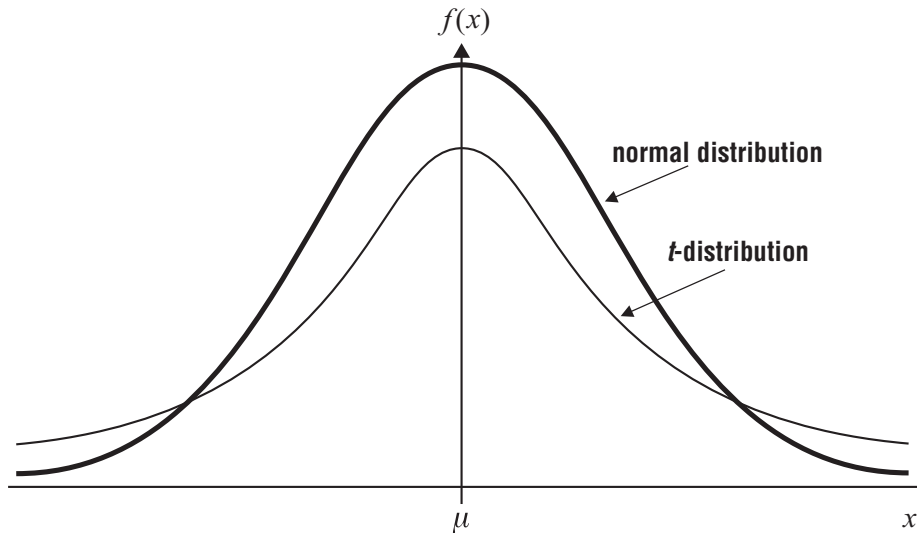
Comparing the t and the Normal Distribution

- In the limit, a t-distribution with an infinite number of degrees of freedom is a standard normal, i.e. $t(\infty) = N(0,1)$

- Examples from statistical tables:

Significance level	$N(0,1)$	$t(40)$	$t(4)$
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 $SE(\hat{\alpha})$, $SE(\hat{\beta})$ as before.

2. Choose a significance level, α, (again the convention is 5%). This is equivalent to choosing a $(1-\alpha) \times 100\%$ 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 *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

- 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

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

- Rearranging, we would not reject if

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

- 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,

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

- 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 $t_{crit} = t_{20;5\%}$

Determining the Rejection Region
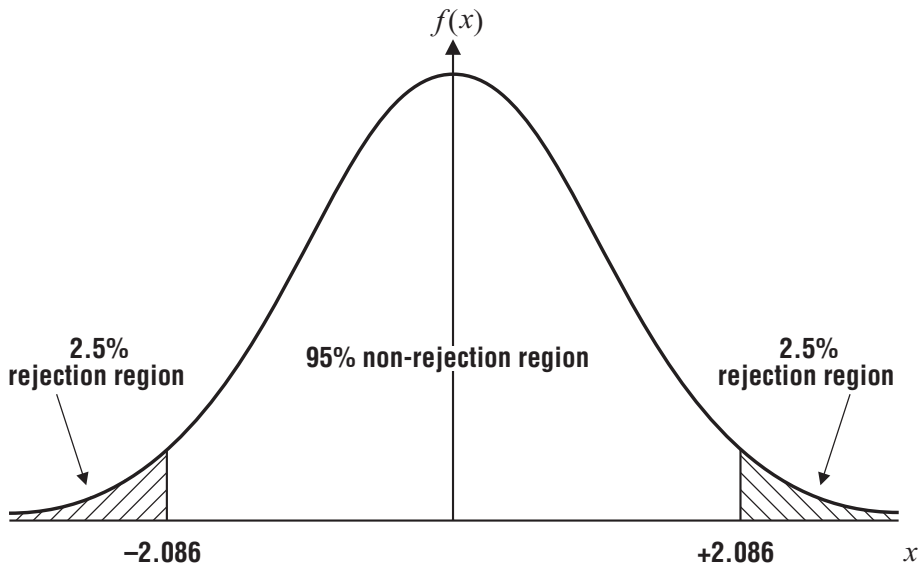
Performing the Test

- The hypotheses are:

  $H_0 : \beta = 1$

  $H_1 : \beta \neq 1$

Test of significance approach

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

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

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

Confidence interval approach

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

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

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

$$H_0: \beta = 0$$
$$\text{vs.} \quad H_1: \beta \neq 0$$

$$H_0: \beta = 2$$
$$\text{vs.} \quad H_1: \beta \neq 2$$

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,

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

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

Changing the Size of the Test: The New Rejection Regions
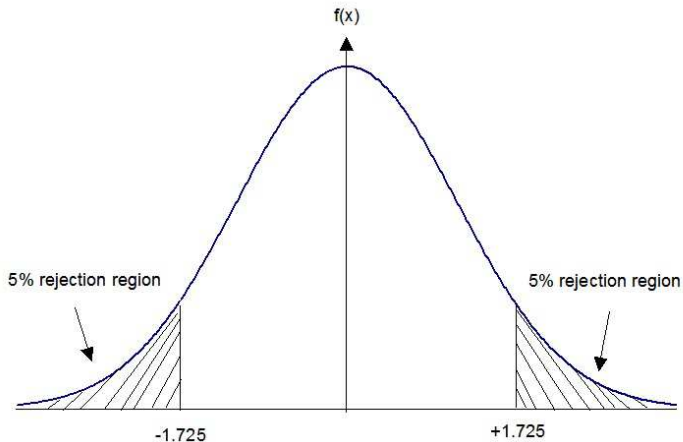
Changing the Size of the Test: The Conclusion

- $t_{20;10\%} = 1.725$. 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 $= \alpha$	$\sqrt{}$
	Insignificant (do not reject H_0)	$\sqrt{}$	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. 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:

				Less likely to falsely	Lower $\rightarrow$chance of
Reduce size$\rightarrow$	More strict $\rightarrow$	Reject null$\nearrow$	reject	type I error
of test (e.g.	criterion for	hypothesis$\searrow$			
5% to 1%)	rejection	less often	More likely to incorrectly not reject	Higher $\rightarrow$chance of type II error
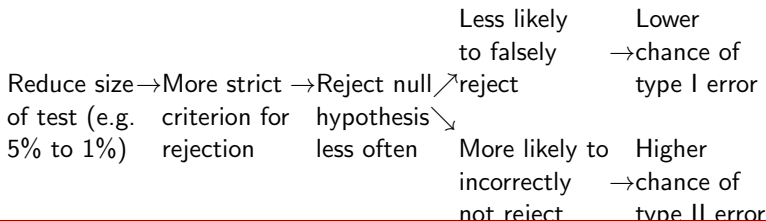
The Trade-off Between Type I and Type II Errors (Cont'd)

- 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

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

- 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 $\beta_i^* = 0$, $\text{test stat} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$

- 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 15-3	=	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:
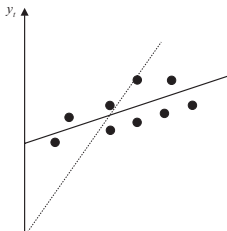
An Example of the Use of a Simple t-test to Test a Theory in Finance

- 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, ..., 115

- We are interested in the significance of α_j.

- The null hypothesis is H_0: α_j=0 .

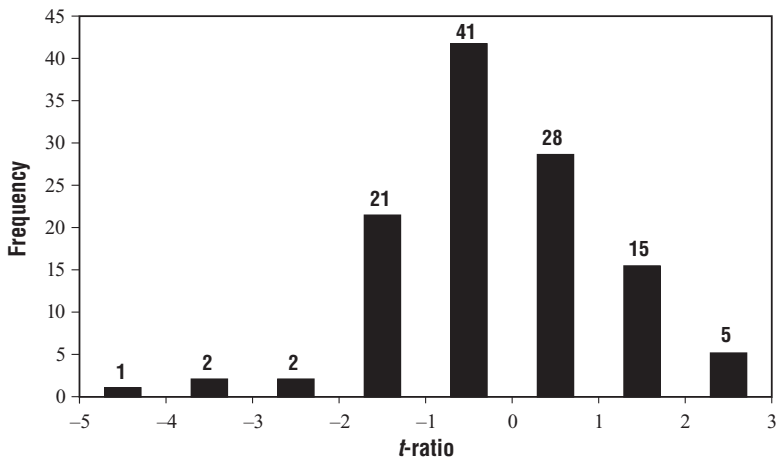# Frequency Distribution of t -ratios of Mutual Fund Alphas (gross of transactions costs)

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

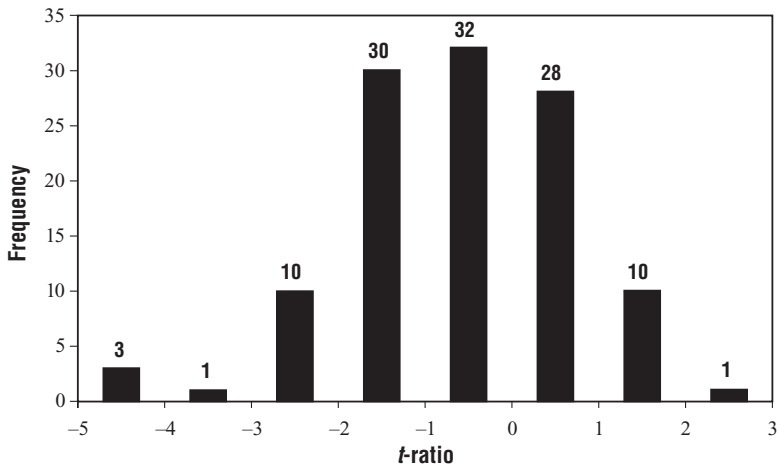# Frequency Distribution of t -ratios of Mutual Fund Alphas (net of transactions costs)

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

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	Min	Max	Median
Average monthly return	1.0%	0.6%	1.4%	1.0%
Std dev of returns	5.1%	4.3%	6.9%	5.0%

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

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

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.

The Overreaction Hypothesis and the UK Stock Market

- Motivation

  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.

- How Can This be Explained?

  2 suggestions
  1. A manifestation of the size effect

     DeBondt & Thaler did not believe this a sufficient explanation, but Zarowin (1990) found that allowing for firm size did reduce the subsequent return on the losers.
  2. Reversals reflect changes in equilibrium required returns

The Overreaction Hypothesis and the UK Stock Market   (Cont'd)

> Ball & Kothari (1989) find the CAPM beta of losers to be considerably higher than that of winners.

- Another interesting anomaly: the January effect.
    - Another possible reason for the superior subsequent performance of losers.
    - Zarowin (1990) finds that 80% of the extra return available from holding the losers accrues to investors in January.
- Example study: Clare and Thomas (1995)

  Data:

  Monthly UK stock returns from January 1955 to 1990 on all firms traded on the London Stock exchange.

Methodology

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

$$U_{it} = R_{it} - R_{mt} \ n = 12, \ 24 \text{ or } 36 \text{ months}$$

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

$$\bar{R}_i = \frac{1}{n} \sum_{t=1}^{n} U_{it}$$

Portfolio Formation

- Then rank the stocks from highest average return to lowest and from 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 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 $\bar{R}_{D_t} = \bar{R}_p^L - \bar{R}_p^W$.

- Then perform the regression
$$\bar{R}_{D_t} = \alpha_1 + \eta_t \qquad \text{(Test 1)}$$

- Look at the significance of α_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:

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

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 (3.37): $\hat{\alpha}_1$	-0.00031	0.0014^{**}	0.0013
	(0.29)	(2.01)	(1.55)
Coefficients for (3.38): $\hat{\alpha}_2$	-0.00034	0.00147^{**}	0.0013^{*}
	(-0.30)	(2.01)	(1.41)
Coefficients for (3.38): $\hat{\beta}$	-0.022	0.010	-0.0025
	(-0.25)	(0.21)	(-0.06)
Panel B: all months except January			
Coefficient for (3.37): $\hat{\alpha}_1$	-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, $\bar{R}_{D_t}$, and regress this on 12 month-of-the-year dummies:

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

- 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

The Exact Significance Level or 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 $= 0.12$.

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

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

- Do we reject at the 20% level?.........................Yes