Lecture Notes (1)

Introduction

1

1

Introduction to Regression Analysis

- **The regression analysis is one of the most important and widely used statistical techniques in business and economic analysis for examining the functional relationships between two or more variables. One variable is specified to be the *dependent/response variable* (*DV*), denoted by Y, and the other one or more variables are called the *independent/predictor/explanatory variables (IV)*, denoted by X_i, i=1,2, … k.**
- **There are two different situations:**
  (a) Y is a random variable and X_i are fixed, no-random variable, *e.g.* to predict the sales for a company, the Year is the fixed X_i variable.
  (b) Both X_i and Y are random variables, e.g. all survey data are of this type, in this situation, cases are selected randomly from the population, and both X_i and Y are measured.

2

Main Purposes

- **Regression analysis can be used for either of two main purposes:**

  ***Descriptive*: The kind of relationship and its strength are examined. This examination can be done graphically or by the use of descriptive equations. Tests of hypotheses and confidence intervals can serve to draw inferences regarding the relationship.**

  ***Predictive*: The equation relating Y and X_i can be used to predict the value of Y for a given value of X_i . Prediction intervals can also be used to indicate a likely range of the predicted value of Y.**

Description of Methods of Regression

- **The general form of a probabilistic model is**

  Y = *Deterministic component* + *random error*

  As you will see, the random error plays an important role in testing hypotheses and finding confidence intervals for the parameters in the model.

- **The simple regression analysis means that the value of the dependent variable Y is estimated on the basis of only one independent variable.**

$$Y = f(X) + u \ .$$

- **On the other hand, multiple regression is concerned with estimating the value of the dependent variable Y on the basis of two or more independent variables.**

$$Y = f(X_1 , X_2 \ \dots \ X_k) + u \ , \quad \text{where } k \geq 2 \ .$$

Linear Regression Model

- We begin with the simplest of probabilistic models - the linear regression model. That is, $f(X)$ is a linear function of X_1 , X_2 ... X_k,

$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u .$$

- The model can be stated as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i , \quad i = 1, 2, \ldots, n$$

where

Y_i is the value of the response variable in the ith trial

X_{ji} ($j = 1, 2, ..., k$ & $i = 1, 2, ..., n$) are independent variables.

β_j ($j = 0, 1, 2, ..., k$) are parameters (partial coefficients)

u_i is a random error with $E(u_i) = 0$, $var(u_i) = \sigma^2$, and $cov(u_i, u_j) = 0$, $i \neq j$.

- This model is ***linear in the coefficients***, it is called a ***general multiple linear regression model***.

Estimating the Model Parameters

- $E(u) = 0$ is equivalent to that $E(Y)$ equals the deterministic component of the model. That is, for simple linear regression model,

$$E(Y) = \beta_0 + \beta_1 X ,$$

where the constants β_0 and β_1 are the population parameters. It is called the population regression equation (line).

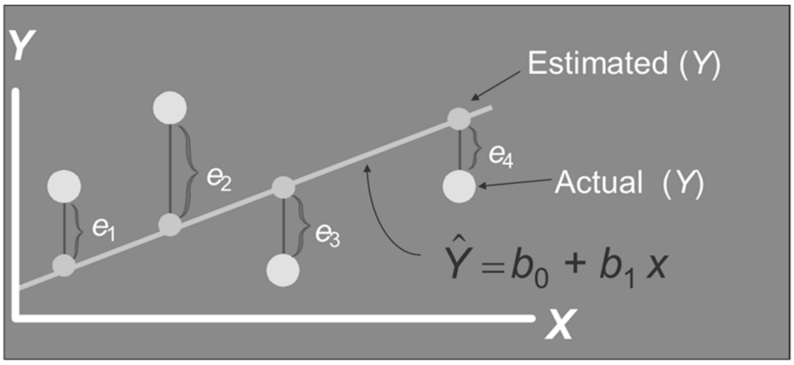
- Denoting estimates of β_0 & β_1 by b_0 and b_1 respectively, we can then estimate $E(Y)$ by $\hat{Y}$ from the sample regression equation (or the fitted regression line)

$$\hat{Y} = b_0 + b_1 X .$$

- The problem of fitting a line to a sample of points is essentially the problem of efficiently estimating the parameters β_0 and β_1 by b_0 and b_1 respectively. The best known method for doing this is called the *least squares method* (LSM).

The Least Squares Method

- The principle of least squares is illustrated in the following Figure.



The Least Squares Method (Cont.)

- **For every observed Y_i in a sample of points, there is a corresponding predicted value $\hat{Y}_i$, equal to $b_0 + b_1 X_i$. The sample deviation of the observed value Y_i from the predicted $\hat{Y}_i$ is**

  $$e_i = Y_i - \hat{Y}_i ,$$

  called a *residual*, that is,

  $$e_i = Y_i - b_0 - b_1 X_i .$$

- **We shall find b_0 and b_1 so that the sum of the squares of the errors (residuals) SSE $=\sum e_i^2 = \sum(Y_i - \hat{Y}_i)^2 = \sum(Y_i - b_0 - b_1 X_i)^2$ is a minimum.**

- ***This minimization procedure for estimating the parameters is called the method of least squares.***

The Least Squares Method (Cont.)

- **Differentiating *SSE* with respect to b_0 and b_1, we have**
  $\partial(SSE)/\partial b_0 = -2\sum(y_i - b_0 - b_1 X_i)$
  $\partial(SSE)/\partial b_1 = -2\sum(y_i - b_0 - b_1 X_i)X_i$.
- **Setting the partial derivatives equal to zero and rearranging the terms, we obtain the equations (called the *normal equations*)**
  $n b_0 + b_1\sum X_i = \sum Y_i$ and $b_0 \sum X_i + b_1 \sum X_i^2 = \sum Y_i X_i$
  which may be solved simultaneously to yield computing formulas for b_0 and b_1 as follows:
  $$b_1 = SS_{xy} / SS_{xx}, \quad b_0 = \bar{Y} - b_1 \bar{X}$$
  where
  $SS_{xy} = \sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - (\sum X_i \sum Y_i)/n$
  $SS_{xx} = \sum (X_i - \bar{X})^2 = \sum X_i^2 - (\sum X_i)^2/n$

Properties of Least Squares Estimators

(1) *Gauss-Markov Theorem*

Under the conditions of the regression model, the least squares estimators b_0 and b_1 are unbiased estimators (i.e., $E(b_0) = \beta_0$ and $E(b_1) = \beta_1$) and have minimum variance among all unbiased linear estimators.

(2) The estimated value of Y (i.e. $\hat{Y} = b_0 + b_1X$) is an unbiased estimator of $E(Y) = \beta_0 + \beta_1 X$, with minimum variance in the class of unbiased linear estimators.

- **Note that the common variance σ^2 can not be estimated by LSM. We can prove that the following statistic is an unbiased point estimator of σ^2**
  $$s^2 = SSE/(n-k-1) = (SS_{yy} - b_1 SS_{xy})/(n-k-1)$$
  where k is the number of independent variables in the model.

Normal Error Regression Model

- No matter what may be the form of the distribution of the error terms ε_i (and hence of the Y_i), the LSM provides unbiased point estimators of β_0 and β_1 that have minimum variance among all unbiased linear estimators.
- To set up interval estimates and make tests, however, we need to make an assumption about the form of the distribution of the ε_i. The standard assumption is that the error terms ε_i are normally distributed, and we will adopt it here.
- Since now the functional form of the probability distribution of the error terms is specified, we can use the maximum likelihood method to obtain estimators of the parameters β_0, β_1 and σ^2. In fact, MLE and LSE for β_0 and β_1 are the same.
- A normal error term greatly simplifies the theory of regression analysis.

Inferences Concerning the Regression Coefficients

- Aside from merely estimating the linear relationship between X and Y for purposes of prediction, we may also be interested in drawing certain inferences about the population parameters, say β_0 and β_1.

- To make inferences or test hypotheses concerning these parameters, we must know the sampling distributions of b_0 and b_1. (Note that b_0 and b_1 are statistics, i.e., functions of the random sample, therefore, they are random variables)

Inferences Concerning β_1

(a) b_1 is an normal random variable for the normal error model.
(b) $E(b_1) = \beta_1$. That is, b_1 is an unbiased estimator of β_1.
(c) $Var(b_1) = \sigma^2/SS_{xx}$, which is estimated by $s^2(b_1) = s^2/SS_{xx}$, where s^2 is the unbiased estimator of σ^2.
(d) The $(1 - \alpha)$ 100% Confidence interval for β_1 (σ^2 unknown)

$$b_1 - t_{\alpha/2}\, s(b_1) < \beta_1 < b_1 + t_{\alpha/2}\, s(b_1)$$

where $t_{\alpha/2}$ is a value of the t - distribution with (n - 2) degrees of freedom, and $s(b_1)$ is the standard error of b_1, *i.e.* $s(b_1) = s/(SS_{xx})^{1/2}$.
(e) Hypothesis test of β_1
To test the null hypothesis H_0: $\beta_1 = 0$ against a suitable alternative, we can use the t distribution with n-2 degrees of freedom to establish a critical region and then base our decision on the value of

$$t = b_1/s(b_1).$$

13

Inferences Concerning β_0

(a) b_0 is an normal random variable for the normal error model.
(b) $E(b_0) = \beta_0$. That is, b_0 is an unbiased estimator of β_0.
(c) $Var(b_0) = \sigma^2 \sum X_i^2/nSS_{xx}$, which is estimated by $s^2(b_0) = s^2 \sum X_i^2/nSS_{xx}$, where s^2 is the unbiased estimator of σ^2.
(d) The $(1 - \alpha)$ 100% Confidence interval for β_0 (σ^2 unknown)

$$b_0 - t_{\alpha/2}\, s(b_0) < \beta_0 < b_0 + t_{\alpha/2}\, s(b_0)$$

where $t_{\alpha/2}$ is a value of the t - distribution with (n - 2) degrees of freedom, and $s(b_0) = s(\sum X_i^2/nSS_{xx})^{1/2}$.
(e) Hypothesis test of β_0
To test the null hypothesis H_0: $\beta_0 = 0$ against a suitable alternative, we can use the t distribution with n-2 degrees of freedom to establish a critical region and then base our decision on the value of

$$t = b_0/s(b_0).$$

14

Inferences Concerning E(Y)

(1) The sampling distribution of $\hat{Y}_i$ is normal for the normal error model.

(2) $\hat{Y}_i$ is an unbiased estimator of $E(Y_i)$.

Because $E(Y_i) = \beta_0 + \beta_1 X_i$ and

$E(\hat{Y}_i) = E(b_0 + b_1 X_i) = \beta_0 + \beta_1 X_i = E(Y_i)$.

(3) The variance of $\hat{Y}_i$: $var(\hat{Y}_i) = \sigma^2 [(1/n) + (X_i - \bar{X})^2/SS_{xx}]$

& the estimated variance of $\hat{Y}_i$: $s^2(\hat{Y}_i) = s^2 [(1/n) + (X_i - \bar{X})^2/SS_{xx}]$

(4) The $(1 - \alpha)$ 100% confidence interval for the mean response $E(Y_i)$ is as follows $\hat{Y}_i - t_{\alpha/2, (n-2)} s(\hat{Y}_i) < E(Y_i) < \hat{Y}_i + t_{\alpha/2, (n-2)} s(\hat{Y}_i)$

- **Note that the confidence limits for $E(Y_i)$ are not sensitive to moderate departures from the assumption that the error terms are normally distributed. Indeed, the limits are not sensitive to substantial departures from normality if the sample size is large.**

Some Considerations

- ***<u>Effects of Departures From Normality</u>***

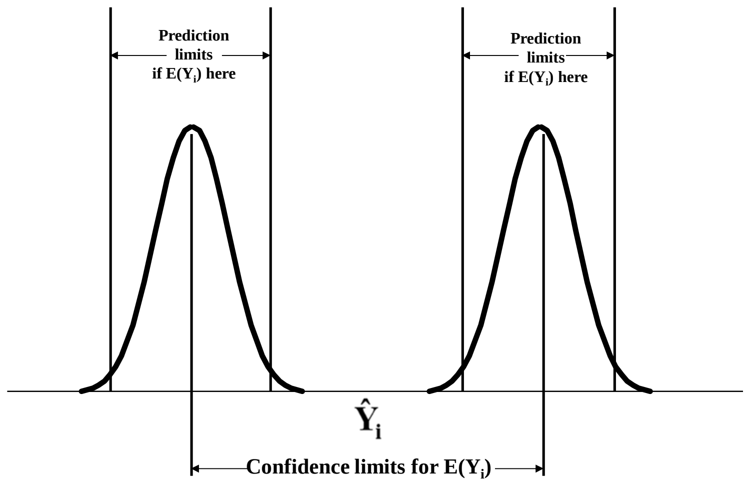
  If the probability distributions of Y are not normal but do not depart seriously, the sampling distributions of b_0 and b_1 will be approximately normal (CLT). Even if the distributions of Y are far from normal, the estimators b_0 and b_1 generally have the property of asymptotic normality as the sample size increases. Thus, with sufficiently large samples, the confidence interval and decision rules given earlier still apply even if the probability distributions of Y depart far from normality.

Prediction of New Observation

- The distinction between estimation of the mean response $E(Y_i)$, discussed in the preceding section, and prediction of a new response $Y_{i(new)}$, discussed now, is basic. In the former case, we estimate the mean of the distribution of Y. In the present case, we predict an individual outcome draw from the distribution of Y.
- ***<u>Prediction Interval for $Y_{i(new)}$</u>***

  When the regression parameters are unknown, they must be estimated. The mean of the distribution of Y is estimated by , as usual, and the variance of the distribution of Y is estimated by MSE (i.e. s^2). From the Figure in next page, we can see that there are two probability distributions of Y, corresponding to the upper and lower limits of a confidence interval for E(Y).

Prediction Interval



Prediction Interval (cont.)

- **Since we cannot be certain of the location of the distribution of Y, prediction limits for $Y_{i(new)}$ clearly must take account of two elements: (a) variation in possible location of the distribution of Y; and (b) variation within the probability distribution of Y. That is, $\text{var}(\text{pred}_i)=\text{var}(Y_{i(new)}-\hat{Y}_i)=\text{var}(Y_{i(new)})+\text{var}(\hat{Y}_i)=\sigma^2+\text{var}(\hat{Y}_i)$.**
- **An unbiased estimator of var(pred) is as follows $s^2(\text{pred}_i)=s^2+s^2(\hat{Y}_i)=s^2[1+(1/n)+(X_i-\bar{X})^2/SS_{xx}]$**
- **The (1 - α) 100% prediction interval for $Y_{i(new)}$ is as follows**

$$\hat{Y}_i - t_{\alpha/2,(n-2)}\, s(\text{pred}_i) < Y_{i(new)} < \hat{Y}_i + t_{\alpha/2,(n-2)}\, s(\text{pred}_i)$$

Comments on Prediction Interval

- **The prediction limits, unlike the confidence limits for a mean response $E(Y_i)$, are sensitive to departures from normality of the error terms distribution.**
- **Prediction intervals resemble confidence intervals. However, they differ conceptually. A confidence interval represents an inference on a parameter and is an interval that is intended to cover the value of the parameter. A prediction interval, on other hand, is a statement about the value to be taken by a random variable, the new observation $Y_{i(new)}$.**

Types of Data

- There are broadly 3 types of data that can be employed in quantitative analysis of financial problems: (a) Time series data; (b) Cross-sectional data; (c) Panel data (combination of (a) & (b)).

- **Time series data**
  Time series data, as the name suggests, are data that have been collected over a period of time on one or more variables. Time series data have associated with them a particular frequency of observation or collection of data points. It is also generally a requirement that all data used in a model be of the same frequency of observation.

- The data may be quantitative (e.g. exchange rates, stock prices, number of shares outstanding), or qualitative (e.g. day of the week).

Time series data

- **Examples of *time series data***

Series	***Frequency***
– GNP or unemployment	monthly, or quarterly
– government budget deficit	annually
– money supply	weekly
– value of a stock market index	as transactions occur

- **Problems that could be tackled using time series data**

  - How the value of a country's stock index has varied with that country's macroeconomic fundamentals.
  - How the value of a company's stock price has varied when it announced the value of its dividend payment.
  - The effect on a country's currency of an increase in its interest rate

- In all of the above cases, it is clearly the time dimension which is the most important, and the analysis will be conducted using the values of the variables over time.

Cross-sectional Data

- **Cross-sectional data**

  Cross-sectional data are data on one or more variables collected at a single point in time, e.g.

  - A poll of usage of internet stock broking services
  - Cross-section of stock returns on the New York Stock Exchange
  - A sample of bond credit ratings for UK banks

- **Problems that could be tackled using cross-sectional data**

  - The relationship between company size and the return to investing in its shares
  - The relationship between a country's GDP level and the probability that the government will default on its sovereign debt.

Panel Data

- **Panel data**

  Panel Data has the dimensions of both time series and cross-sections, e.g. the daily prices of a number of blue chip stocks over two years. The estimation of panel regressions is an interesting and developing area, and will not be discussed in this course.

- **Notation**

  It is common to denote each observation by the letter t and the total number of observations by T for time series data, and to denote each observation by the letter i and the total number of observations by N for cross-sectional data.

- **Quality of Data**

  The researchers should always keep in mind that the results of research are only as good as the quality of the data.

Steps involved in formulating an econometric model

- **Step 1**a and 1b: general statement of the problem.
- **Step 2**: collection of data relevant to the model
- **Step 3**: choice of estimation method relevant to the model proposed in step 1
- **Step 4**: statistical evaluation of the model
- **Step 5**: evaluation of the model from a theoretical perspective
- **Step 6**: use of model

- It is important to note that the process of building a robust empirical model is an iterative one, and it is certainly not an exact science. Often, the final preferred model could be very different from the one originally proposed, and need not be unique in the sense that another researcher with the same data and the same initial theory could arrive at a different final specification.
- Textbook pages 11~12

Statistical Package - EViews

- EViews is an excellent interactive program, which provide an excellent tool to do time series data analysis. It is simple to use, menu-driven, and will be sufficient to estimate most of the models required for this course.
- One of the most important features of EViews that makes it useful for model-building is the wealth of diagnostic (misspecification) tests, that are automatically computed, making it possible to test whether the model is econometrically/statistically valid or not.

- Textbook pages 14~23

- Data – UK Average House Price (Ukhp13.xls)

Violation of the Assumptions of the CLRM

- Recall that we assumed of the CLRM disturbance/error terms:

1. $\text{Var}(u_t) = \sigma^2 < \infty$
2. $\text{Cov}(u_i, u_j) = 0$ for $i \neq j$
3. $u_t \sim N(0,\sigma^2)$

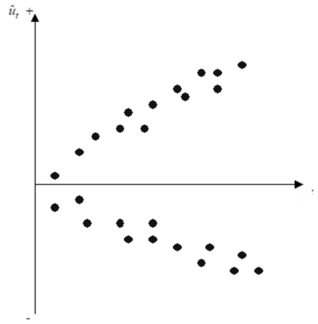
Investigating Violations of the Assumptions of the CLRM

- We will now study these assumptions further, and in particular look at:
  - How we test for violations
  - Causes
  - Consequences
    - in general we could encounter any combination of 3 problems:
      - the coefficient estimates are wrong
      - the associated standard errors are wrong
      - the distribution that we assumed for the test statistics will be inappropriate
  - Solutions
    - the assumptions are no longer violated
    - we work around the problem so that we use alternative techniques which are still valid

Assumption: Var(u_t) = $\sigma^2 < \infty$

- We have so far assumed that the variance of the errors is constant, σ^2 - this is known as homoscedasticity. If the errors do not have a constant variance, we say that they are heteroscedastic e.g. say we estimate a regression and calculate the residuals, $\hat{u}_t$.

- Var(u_t) = E(u_t –E(u_t))^2 = E(u_t^2)



Detection of Heteroscedasticity: The GQ Test

- Graphical methods

- Formal tests: There are many of them: we will discuss Goldfeld-Quandt test and White’s test

The Goldfeld-Quandt (GQ) test is carried out as follows.

1. Split the total sample of length T into two sub-samples of length T_1 and T_2. The regression model is estimated on each sub-sample and the two residual variances are calculated.

2. The null hypothesis is that the variances of the disturbances are equal,

$$H_0: \sigma_1^2 = \sigma_2^2$$

The GQ Test (Cont'd)

4. The test statistic, denoted GQ, is simply the ratio of the two residual variances where the larger of the two variances must be placed in the numerator.

$$GQ = \frac{s_1^2}{s_2^2}$$

5. The test statistic is distributed as an $F(T_1\text{-}k,\ T_2\text{-}k)$ under the null of homoscedasticity.

A problem with the test is that the choice of where to split the sample is that usually arbitrary and may crucially affect the outcome of the test.

Detection of Heteroscedasticity using White's Test

- White's general test for heteroscedasticity is one of the best approaches because it makes few assumptions about the form of the heteroscedasticity.
- The White's test is carried out as follows:

1. Assume that the regression we carried out is as follows

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

And we want to test $\text{Var}(u_t) = \sigma^2$. We estimate the model, obtaining the residuals, $\hat{u}_t$

2. Then run the auxiliary regression

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 x_{2t} + \alpha_3 x_{3t} + \alpha_4 x_{2t}^2 + \alpha_5 x_{3t}^2 + \alpha_6 x_{2t} x_{3t} + v_t$$

Performing White's Test for Heteroscedasticity

3. Obtain R^2 from the auxiliary regression and multiply it by the number of observations, T. It can be shown that

$$T R^2 \sim \chi^2 (m)$$

where m is the number of regressors in the auxiliary regression excluding the constant term.

4. If the χ^2 test statistic from step 3 is greater than the corresponding value from the statistical table then reject the null hypothesis that the disturbances are homoscedastic.

(Refer to Box 5.1 and example 5.1 on pages183~184.
Testing for Heteroscedasticity using EViews – Bloodpre.xls)

Consequences of Using OLS in the Presence of Heteroscedasticity

- OLS estimation still gives unbiased coefficient estimates, but they are no longer BLUE.
- This implies that if we still use OLS in the presence of heteroscedasticity, our standard errors could be inappropriate and hence any inferences we make could be misleading.
- Whether the standard errors calculated using the usual formulae are too big or too small will depend upon the form of the heteroscedasticity.

How Do we Deal with Heteroscedasticity?

- If the form (i.e. the cause) of the heteroscedasticity is known, then we can use an estimation method which takes this into account (called generalised least squares, GLS). GLS is also known as weighted least squares (WLS).
- A simple illustration of GLS is as follows: Suppose that the error variance is related to another variable z_t by

$$\mathrm{var}(u_t) = \sigma^2 z_t^2$$

- To remove the heteroscedasticity, divide the regression equation by z_t

$$\frac{y_t}{z_t} = \beta_1 \frac{1}{z_t} + \beta_2 \frac{x_{2t}}{z_t} + \beta_3 \frac{x_{3t}}{z_t} + v_t$$

where $v_t = \frac{u_t}{z_t}$ is an error term.

- Now $\mathrm{var}(v_t) = \mathrm{var}\left(\frac{u_t}{z_t}\right) = \frac{\mathrm{var}(u_t)}{z_t^2} = \frac{\sigma^2 z_t^2}{z_t^2} = \sigma^2$ for known z_t.
- So the disturbances from the new regression equation will be homoscedastic.

Other Approaches to Dealing with Heteroscedasticity

- Other solutions include:

1. Transforming the variables into logs or reducing by some other measure of "size".

2. Use White's heteroscedasticity consistent standard error estimates.

The effect of using White's correction is that in general the standard errors for the slope coefficients are increased relative to the usual OLS standard errors.

This makes us more "conservative" in hypothesis testing, so that we would need more evidence against the null hypothesis before we would reject it.

Assumption: Cov(u_i, u_j) = 0 for $i \neq j$

- Assumption 3 that is made of the CLRM's disturbance terms is assumed that the errors are uncorrelated with one another.
- If the errors are not uncorrelated with one another, it would be stated that they are "autocorrelated" or that they are "serially correlated". A test of this assumption is therefore required.
- Before we proceed to see how formal tests for autocorrelation are formulated, the concept of the lagged value of a variable needs to be defined.
- The lagged value of a variable (which may be y_t, x_t or u_t) is simply the value that the variable took during a previous period, e.g. the value of y_t lagged one period, written y_{t-1}, can be constructed by shifting all of the observations forward one period in a spreadsheet, as illustrated in the table below:

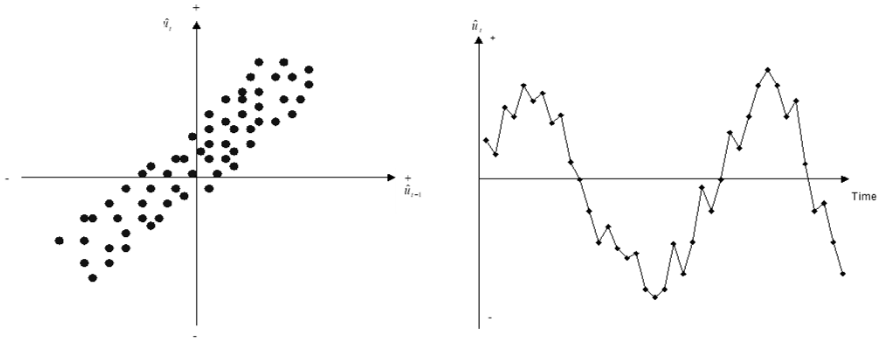
The Concept of a Lagged Value

t	y_t	y_{t-1}	Δy_t
1989M09	0.8	-	-
1989M10	1.3	0.8	1.3-0.8=0.5
1989M11	-0.9	1.3	-0.9-1.3=-2.2
1989M12	0.2	-0.9	0.2--0.9=1.1
1990M01	-1.7	0.2	-1.7-0.2=-1.9
1990M02	2.3	-1.7	2.3--1.7=4.0
1990M03	0.1	2.3	0.1-2.3=-2.2
1990M04	0.0	0.1	0.0-0.1=-0.1
.	.	.	.
.	.	.	.
.	.	.	.

Autocorrelation

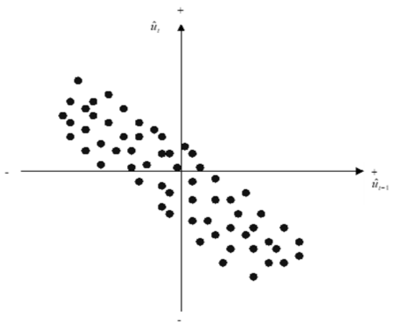
- We assumed of the CLRM's errors that Cov (u_i , u_j) = 0 for $i \neq j$.
  This is essentially the same as saying there is no pattern in the errors.
- Obviously we never have the actual u's, so we use their sample counterpart, the residuals $\hat{u}_t$.
- If there are patterns in the residuals from a model, we say that they are autocorrelated.
- Some stereotypical patterns we may find in the residuals are given on the next 3 slides.

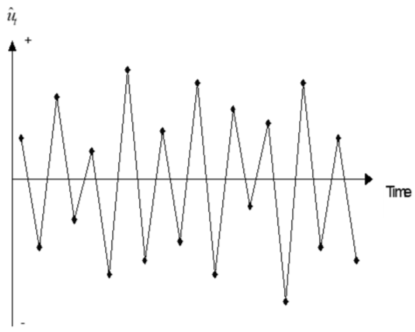
Positive Autocorrelation



Positive Autocorrelation is indicated by a cyclical residual plot over time.

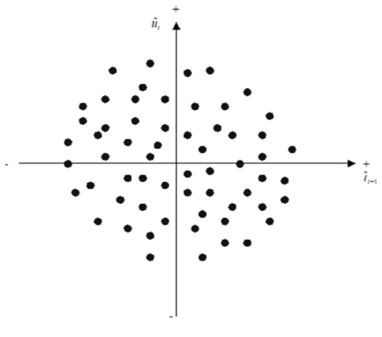
Negative Autocorrelation

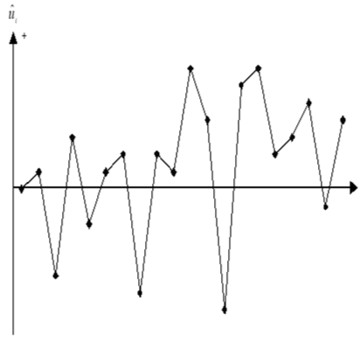




Negative autocorrelation is indicated by an alternating pattern where the residuals cross the time axis more frequently than if they were distributed randomly

No pattern in residuals – No autocorrelation





No pattern in residuals at all: this is what we would like to see

Detecting Autocorrelation: The Durbin-Watson Test

The Durbin-Watson (DW) is a test for first order autocorrelation - i.e. it assumes that the relationship is between an error and the previous one

$$u_t = \rho\, u_{t-1} + v_t \qquad (1)$$

where $v_t \sim N(0, \sigma_v^2)$.

- The DW test statistic actually tests

$$H_0 : \rho=0 \text{ and } H_1 : \rho\neq 0$$

- The test statistic is calculated by

$$DW = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^{T} \hat{u}_t^2}$$

The Durbin-Watson Test: Critical Values

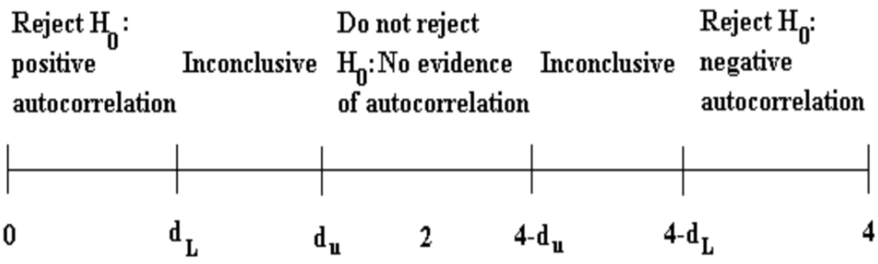
- We can also write

$$DW \approx 2(1-\hat{\rho}) \qquad (2)$$

where $\hat{\rho}$ is the estimated correlation coefficient. Since $\hat{\rho}$ is a correlation, it implies that $-1 \leq \hat{\rho} \leq 1$. (refer to pages 194~196).

- Rearranging for DW from (2) would give $0 \leq DW \leq 4$.
- If $\hat{\rho} = 0$, $DW = 2$. So roughly speaking, do not reject the null hypothesis if DW is near 2 $\rightarrow$ i.e. there is little evidence of autocorrelation
- Unfortunately, DW has 2 critical values, an upper critical value (d_u) and a lower critical value (d_L), and there is also an intermediate region where we can neither reject nor not reject H_0.

The Durbin-Watson Test: Interpreting the Results



- Discuss Example 5.2 on page 197.

Another Test for Autocorrelation: The Breusch-Godfrey Test

- It is a more general test for r^{th} order autocorrelation:

  $$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \ldots + \rho_r u_{t-r} + v_t \quad , \quad v_t \sim \mathrm{N}(0, \sigma_v^2)$$

- The null and alternative hypotheses are:

  $H_0 : \rho_1 = 0$ and $\rho_2 = 0$ and ... and $\rho_r = 0$

  $H_1 : \rho_1 \neq 0$ or $\rho_2 \neq 0$ or ... or $\rho_r \neq 0$

- The test is carried out as follows:
  1. Estimate the linear regression using OLS and obtain the residuals, $\hat{u}_t$
  2. Regress $\hat{u}_t$ on all of the regressors from stage 1 (the x's) plus $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-r}$ Obtain R^2 from this regression.
  3. It can be shown that $(T-r)R^2 \sim \chi^2(r)$
- If the test statistic exceeds the critical value from the statistical tables, reject the null hypothesis of no autocorrelation.

(Refer to Box 5.4 on page 198)

Consequences of Ignoring Autocorrelation if it is Present

- The coefficient estimates derived using OLS are still unbiased, but they are inefficient, i.e. they are not BLUE, even in large sample sizes.
- MSE may seriously underestimate the variance of the error terms.
- Confidence intervals and tests using the t and F distributions, discussed earlier, are no longer strictly applicable.
- Thus, if the standard error estimates are inappropriate, there exists the possibility that we could make the wrong inferences.
- R^2 is likely to be inflated relative to its "correct" value for positively correlated residuals.

Dealing with Autocorrelation

- If the form of the autocorrelation is known, we could use a GLS procedure – i.e. an approach that allows for autocorrelated residuals e.g., Cochrane-Orcutt approach (Refer to Box 5.5 on page 201).
- Discuss equations (5.18 ~ 5.23) on pages 199~200.
- Autocorrelation in EViews –Capm.xls and Sale.xls
- Example – CAPM
  - Y = excess return on shares in Company A (percentage)
  - X1 = excess return on a stock index (percentage)
  - X2 = the sales of Company A (thousands of dollars)
  - X3 = the debt of Company A (thousands of dollars)

Example - CAPM

Dependent Variable: Y
Method: Least Squares
Date: 15/08/10 Time: 18:03
Sample: 1 120
Included observations: 120

	Coefficient	Std. Error	t-Statistic	Prob.
C	2.529897	1.335296	1.894633	0.0606
X1	1.747189	0.202429	8.631119	0.0000
X2	-0.000303	0.000939	-0.322231	0.7479
X3	-0.022015	0.011795	-1.866527	0.0645

R-squared	0.410147	Mean dependent var	2.143089
Adjusted R-squared	0.394892	S.D. dependent var	1.295861
S.E. of regression	1.008033	Akaike info criterion	2.886644
Sum squared resid	117.8712	Schwarz criterion	2.979561
Log likelihood	-169.1987	Hannan-Quinn criter.	2.924378
F-statistic	26.88639	Durbin-Watson stat	2.197432
Prob(F-statistic)	0.000000		

Example – Sale

- Consider the time series data in the following table which gives sales data for the 35 year history of a company.

A Firm's Annual Sales Revenue

Year (T)	Sales (Y)	Year (T)	Sales (Y)	Year (T)	Sales (Y)
1	4.8	13	48.4	25	100.3
2	4.0	14	61.6	26	111.7
3	5.5	15	65.5	27	108.2
4	15.6	16	71.4	28	115.5
5	23.1	17	83.4	29	119.2
6	23.3	18	93.6	30	125.2
7	31.4	19	94.2	31	136.3
8	46.0	20	85.4	32	146.8
9	46.1	21	86.2	33	146.1
10	41.9	22	89.9	34	151.4
11	45.5	23	89.2	35	150.9
12	53.5	24	99.1		

Example – Sale (cont.)

Dependent Variable: SALE
Method: Least Squares
Date: 15/08/10 Time: 17:52
Sample: 1 35
Included observations: 35

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.397143	2.205591	0.180062	0.8582
YEAR	4.295714	0.106861	40.19900	0.0000

R-squared	0.979987	Mean dependent var	77.72000
Adjusted R-squared	0.979381	S.D. dependent var	44.46515
S.E. of regression	6.384903	Akaike info criterion	6.601195
Sum squared resid	1345.310	Schwarz criterion	6.690072
Log likelihood	-113.5209	Hannan-Quinn criter.	6.631875
F-statistic	1615.959	Durbin-Watson stat	0.821097
Prob(F-statistic)	0.000000		

Models in First Difference Form

- Another way to sometimes deal with the problem of autocorrelation is to switch to a model in first differences.

- Denote the first difference of y_t, i.e. y_t - y_{t-1} as Δy_t; similarly for the x-variables, $\Delta x_{2t} = x_{2t} - x_{2t-1}$ etc.

- The model would now be

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + ... + \beta_k \Delta x_{kt} + u_t$$

Testing the Normality Assumption

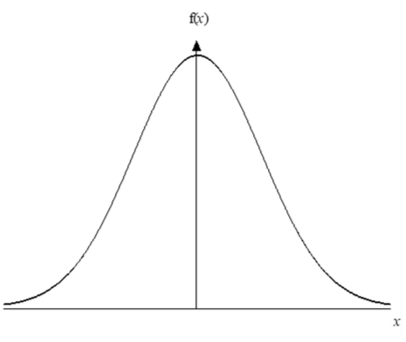
- Testing for Departures from Normality- ***Bera Jarque normality test***
- Skewness and kurtosis are the (standardised) third and fourth moments of a distribution. The skewness and kurtosis can be expressed respectively as:

$$S = \frac{E[u^3]}{\left(\sigma^2\right)^{3/2}} \qquad K = \frac{E[u^4]}{\left(\sigma^2\right)^2}$$

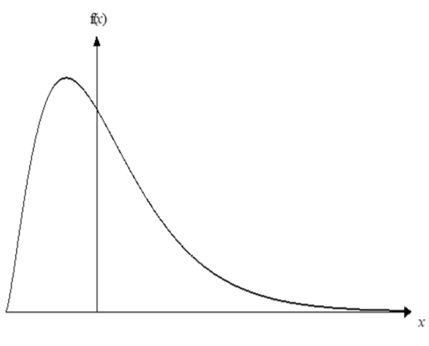
- The Skewness of the normal distribution is 0. The kurtosis of the normal distribution is 3 so its excess kurtosis (K -3) is zero.
- We estimate S and K using the residuals from the OLS regression, $\hat{u}$,

  b_1 (estimator of S) and (b_2 -3) (estimator of (K-3)) may be distributed approximately normally, N(0, 6/T) and N(0, 24/T) respectively for large sample sizes.

Normal versus Skewed Distributions

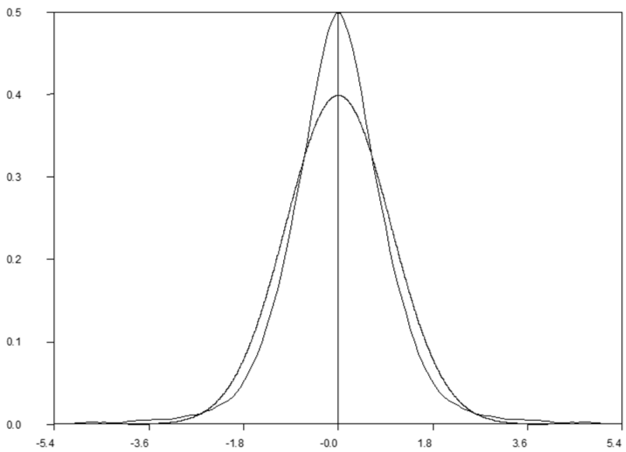


A normal distribution



A skewed distribution

Leptokurtic versus Normal Distribution



Testing for Normality

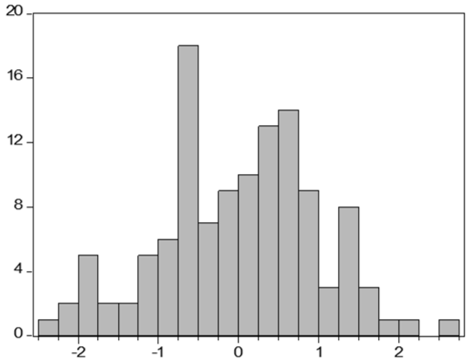
- Bera and Jarque formalise this by testing the residuals for normality by testing whether the coefficient of skewness and the coefficient of excess kurtosis are jointly zero.

- The Bera Jarque test statistic is given by

$$W = T\left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24}\right] \sim \chi^2(2)$$

- Testing for non-normality using EViews - Capm.xls and Sales.xls

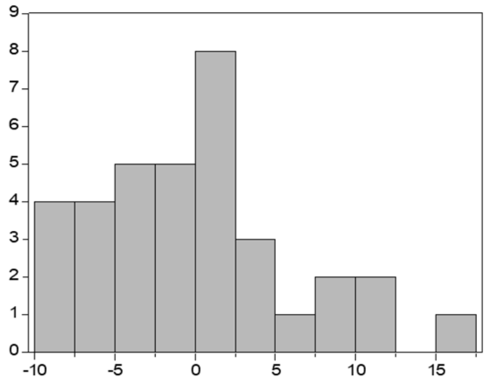
Testing for Normality - Camp.xls



Series: Residuals	
Sample 1 120	
Observations 120	
Mean	-9.25e-17
Median	0.037869
Maximum	2.665635
Minimum	-2.404670
Std. Dev.	0.995246
Skewness	-0.102244
Kurtosis	2.793525
Jarque-Bera	0.422234
Probability	0.809679

57

Testing for Normality - Sales.xls



Series: Residuals	
Sample 1 35	
Observations 35	
Mean	-1.09e-14
Median	-0.385714
Maximum	15.88000
Minimum	-9.998571
Std. Dev.	6.290306
Skewness	0.686581
Kurtosis	2.911288
Jarque-Bera	2.761274
Probability	0.251418

58