

# Time Series: Data and Analysis

# Terminology and basic concepts

- A time series is a collection of observations made **sequentially** in time.
- Usually, these observations are taken at equally spaced intervals over time.
- An intrinsic feature of a time series is that, typically, adjacent observations are **dependent**.
- The nature of this dependence among observations of a time series is of considerable practical interest.
- Time Series Analysis is concerned with techniques for the analysis of this dependence

- since successive observations are not independent the analysis must take this into account.
- When successive observations are dependent, future values may be ‘predicted’ from past observations.
- most time series are **stochastic** in that future is only partly determined by past values, so that exact predictions are impossible
- future values have a probability distribution, which is conditioned by knowledge of past values

- A statistical phenomenon that evolves in time according to probabilistic laws is called a **stochastic process**.
- **The time series to be analysed may then be thought as a particular realisation, produced by the underlying probability mechanism**
- So when analysing a time series we regard it as a realisation of a stochastic process.
- A stochastic process is a family of random variables, defined in a probability space:  $\{X_t\}$ ,  $t = \dots -1, 0, +1, \dots$  and a TS is a sample path or realisation of a stochastic process, whose parameter (index) denotes time.

- A special class of stochastic processes, called **stationary** processes, is based on the assumption that the process is in a particular state of **statistical equilibrium**.
- A stochastic process is said to be **strictly stationary** if its properties are unaffected by a change of time origin,
- i.e if the joint probability distribution associated with  $m$  observations  $z_t, z_{t+1}, z_{t+2}, \dots$  made at any set of times  $t, t+1, t+2, \dots$ , is the same as that associated with  $m$  observations  $z_{t+k}, z_{t+1+k}, z_{t+2+k}, \dots$  made at times  $t+k, t+1+k, t+2+k, \dots$

- However, the notion of strict stationarity is very strong and rarely satisfied by TS encountered in social sciences.
- A rather weaker notion related is the so-called **weak** or **second order** stationarity or wide sense stationarity. A time series is called weakly stationary if:
- $E(X_t) = \mu$  for every  $t$  (meaning independent of time)
- $Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma_k$  for every  $t$  (meaning independent of time and only a function of the time lag. (depends only on  $k$ , the length of time separating the observations, and not  $t$ , the date of observation))

- No assumptions are made about higher order moments than those of second order. By letting  $k = 0$ , we note that the above assumption about the covariance function implies that the variance, as well as the mean, is constant.
- A strictly stationary process is weakly stationary
- If the process is Gaussian, then weak stationarity implies strict stationarity
- By symmetry  $\gamma_k = \gamma_{-k}$
- So, the graph of a stationary series will vary randomly around a constant (stable) mean value and also its variance will be constant through time.

- For a time series  $\{X_t\}$  we define as **autocovariance** (ACV) of k-order the quantity:
- $Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)] = \gamma_k, k = \pm 1, \pm 2$
- The term "auto" is prefixed because the members of the series are generated from the same stochastic process. The ACV function is an even function of  $k$ . that is because:
- $Cov(X_t, X_{t+k}) = Cov(X_{t+k}, X_t) = Cov(X_t, X_{t-k}) = \gamma_{-k}$
- Obviously,  $\gamma_0 = Cov(X_t, X_t) = Var(X_t)$



- Positive first-order autocovariance means that there is a tendency for the next observation to be towards the same side (sign) as the previous one with respect to the mean.
- For a time series  $\{X_t\}$  we define as **autocorrelation** (ACR) of k-order the quantity:  $\rho_k$
- The graph of  $\rho_k$  is called correlogram.
- Because  $\gamma_k = \gamma_{-k}$  it can be shown that  $\rho_k = \rho_{-k}$
- Obviously the ACR function has the same properties with the ACV function, and furthermore satisfies the condition  $\rho_0 = 1$

- Note that the autocorrelation function is dimensionless, that is, independent of the scale of measurement of the time series.
- **The correlogram of a stationary series will die out very fast, after the first  $k$ -lags, whereas in the case of a non-stationary series it will be very persistent, and die out very slowly.**

$$L X_t = X_{t-1}$$

$$L^2 X_t = L * L X_t = X_{t-2}$$

$$\Delta X_t = X_t - X_{t-1}$$

$$\Delta^2 X_t = \Delta \Delta X_t = \Delta (X_t - X_{t-1})$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

# Time Series Regression

$$\diamond y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t$$

## $\diamond$ 1. Basic Analysis

# Time Series vs. Cross Sectional

- Time series data has a temporal ordering, unlike cross-section data
- Will need to alter some of our assumptions to take into account that we no longer have a random sample of individuals
- Instead, we have one realization of a stochastic (i.e. random) process

# Examples of Time Series Models

- A static model relates contemporaneous variables:  $y_t = b_0 + b_1 z_t + u_t$
- A finite distributed lag (FDL) model allows one or more variables to affect  $y$  with a lag:  $y_t = a_0 + d_0 z_t + d_1 z_{t-1} + d_2 z_{t-2} + u_t$
- More generally, a finite distributed lag model of order  $q$  will include  $q$  lags of  $z$

# Finite Distributed Lag Models

- We can call  $d_0$  the impact propensity – it reflects the immediate change in  $y$
- For a temporary, 1-period change,  $y$  returns to its original level in period  $q+1$
- We can call  $d_0 + d_1 + \dots + d_q$  the long-run propensity (LRP) – it reflects the long-run change in  $y$  after a permanent change

# Trending Time Series

- Economic time series often have a trend
- Just because 2 series are trending together, we can't assume that the relation is causal
- Often, both will be trending because of other unobserved factors
- Even if those factors are unobserved, we can control for them by directly controlling for the trend



## Trends (continued)

- One possibility is a linear trend, which can be modeled as  $y_t = a_0 + a_1t + e_t$ ,  $t = 1, 2, \dots$
- Another possibility is an exponential trend, which can be modeled as  $\log(y_t) = a_0 + a_1t + e_t$ ,  $t = 1, 2, \dots$
- Another possibility is a quadratic trend, which can be modeled as  $y_t = a_0 + a_1t + a_2t^2 + e_t$ ,  $t = 1, 2, \dots$

# Detrending

- Adding a linear trend term to a regression is the same thing as using “detrended” series in a regression
- Detrending a series involves regressing each variable in the model on  $t$
- The residuals form the detrended series
- Basically, the trend has been partialled out

## Detrending (continued)

- An advantage to actually detrending the data (vs. adding a trend) involves the calculation of goodness of fit
- Time-series regressions tend to have very high  $R^2$ , as the trend is well explained
- The  $R^2$  from a regression on detrended data better reflects how well the  $x_t$ 's explain  $y_t$

```
. reg highlow trend, r
```

```
Linear regression
```

```
Number of obs = 1123  
F( 1, 1121) = 18.07  
Prob > F = 0.0000  
R-squared = 0.0240  
Root MSE = 1.4727
```

highlow	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
trend	-.0003805	.0000895	-4.25	0.000	-.0005561	-.0002049
_cons	1.372859	.0994282	13.81	0.000	1.177773	1.567946

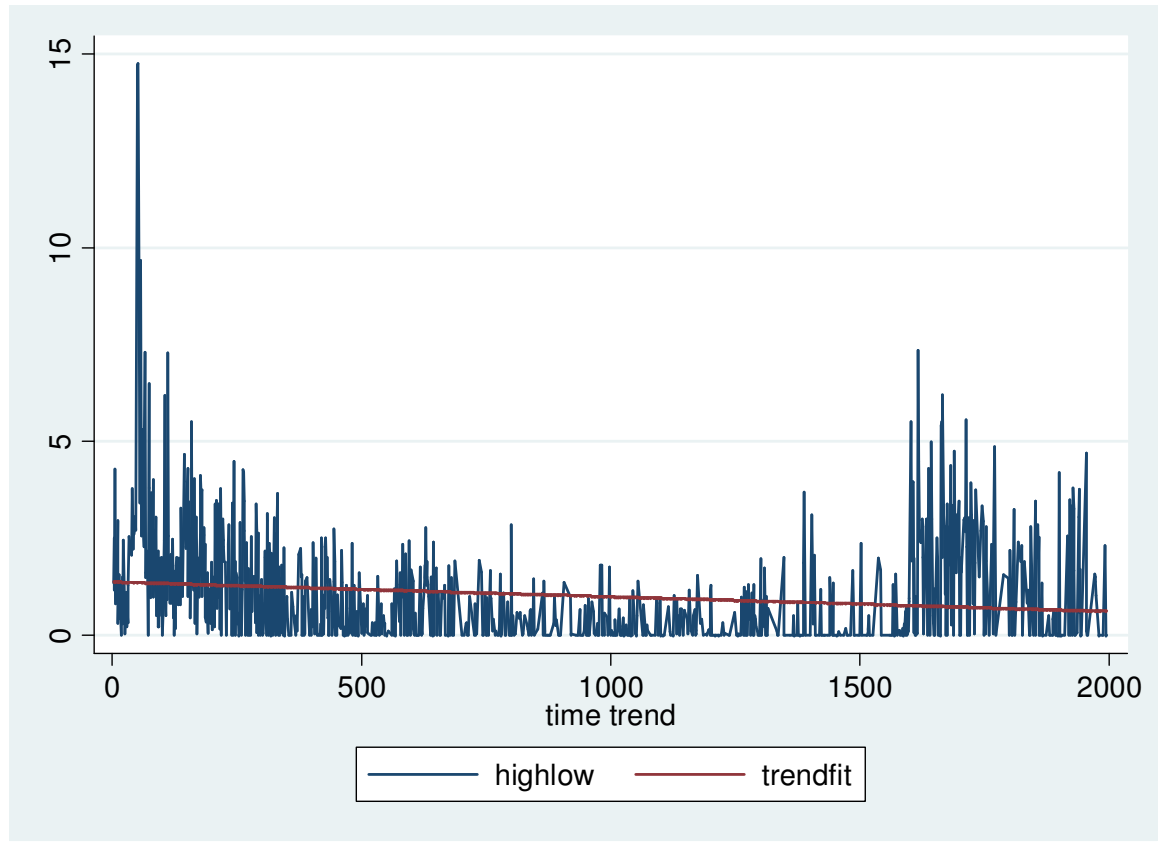
```
. gen trendsq=trend*trend
```

```
. reg highlow trend trendsq, r
```

```
Linear regression
```

```
Number of obs = 1123  
F( 2, 1120) = 70.77  
Prob > F = 0.0000  
R-squared = 0.1735  
Root MSE = 1.3558
```

highlow	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
trend	-.0040856	.0003481	-11.74	0.000	-.0047686	-.0034025
trendsq	1.94e-06	1.64e-07	11.87	0.000	1.62e-06	2.26e-06
_cons	2.417339	.1603895	15.07	0.000	2.102641	2.732036



# Seasonality

- Often time-series data exhibits some periodicity, referred to as seasonality
- Example: Quarterly data on retail sales will tend to jump up in the 4<sup>th</sup> quarter
- Seasonality can be dealt with by adding a set of seasonal dummies
- As with trends, the series can be seasonally adjusted before running the regression

```
. reg highlow weekday1- weekday5 month1- month12, r
```

```
Linear regression
```

```
Number of obs = 1123
F( 15, 1107) = 2.59
Prob > F = 0.0008
R-squared = 0.0364
Root MSE = 1.4725
```

highlow	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
weekday1	-.2671144	.1277177	-2.09	0.037	-.5177106	-.0165183
weekday2	-.0826748	.1490791	-0.55	0.579	-.3751842	.2098346
weekday3	-.2158775	.1406349	-1.54	0.125	-.4918185	.0600634
weekday4	-.1127254	.1409432	-0.80	0.424	-.3892713	.1638205
weekday5	(dropped)					
month1	-.1215709	.1640026	-0.74	0.459	-.4433618	.2002201
month2	-.1406504	.1718424	-0.82	0.413	-.477824	.1965233
month3	.8295044	.3558196	2.33	0.020	.1313476	1.527661
month4	.3554762	.2066008	1.72	0.086	-.0498971	.7608494
month5	-.0921886	.159728	-0.58	0.564	-.4055924	.2212152
month6	.3494071	.2175148	1.61	0.108	-.0773807	.7761948
month7	.1796492	.1737114	1.03	0.301	-.1611915	.52049
month8	.2445662	.1944154	1.26	0.209	-.1368981	.6260305
month9	.4577313	.1910928	2.40	0.017	.0827863	.8326763
month10	.0462309	.1808604	0.26	0.798	-.308637	.4010987
month11	.125072	.1851028	0.68	0.499	-.2381201	.488264
month12	(dropped)					
_cons	.9880832	.1590537	6.21	0.000	.6760025	1.300164

```
. test weekday1 weekday2 weekday3 weekday4 weekday5
```

```
( 1) weekday1 = 0
( 2) weekday2 = 0
( 3) weekday3 = 0
( 4) weekday4 = 0
( 5) weekday5 = 0
Constraint 5 dropped
```

```
F( 4, 1107) = 1.35
Prob > F = 0.2479
```

```
. test month1 month2 month3 month4 month5 month6 month7 month8 month9 month10 month11 month12
```

```
( 1) month1 = 0
( 2) month2 = 0
( 3) month3 = 0
( 4) month4 = 0
( 5) month5 = 0
( 6) month6 = 0
( 7) month7 = 0
( 8) month8 = 0
( 9) month9 = 0
(10) month10 = 0
(11) month11 = 0
(12) month12 = 0
Constraint 12 dropped
```

```
F( 11, 1107) = 3.07
Prob > F = 0.0005
```

# Classical linear regression model assumptions and diagnostics



## Violation of the Assumptions of the CLRM

- Recall that we assumed of the CLRM disturbance terms:
  1.  $E(u_t) = 0$
  2.  $\text{Var}(u_t) = \sigma^2 < \infty$
  3.  $\text{Cov}(u_i, u_j) = 0$
  4. The  $X$  matrix is non-stochastic or fixed in repeated samples
  5.  $u_t \sim N(0, \sigma^2)$

## 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

## Statistical Distributions for Diagnostic Tests

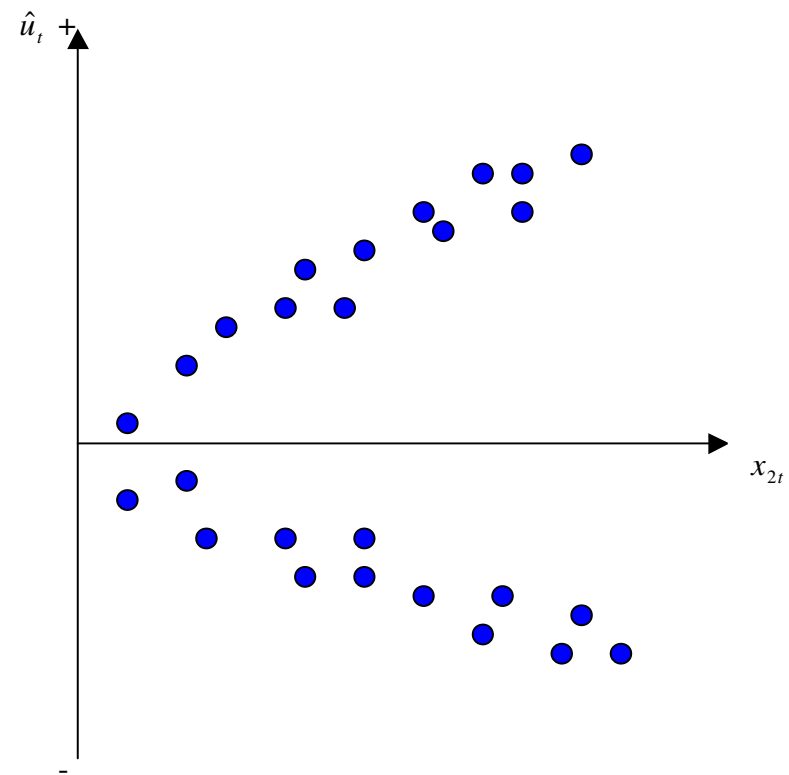
- Often, an  $F$ - and a  $\chi^2$ - version of the test are available.
- The  $F$ -test version involves estimating a restricted and an unrestricted version of a test regression and comparing the  $RSS$ .
- The  $\chi^2$ - version is sometimes called an “LM” test, and only has one degree of freedom parameter: the number of restrictions being tested,  $m$ .
- Asymptotically, the 2 tests are equivalent since the  $\chi^2$  is a special case of the  $F$ -distribution:  
$$\frac{\chi^2(m)}{m} \rightarrow F(m, T - k) \text{ as } T - k \rightarrow \infty$$
- For small samples, the  $F$ -version is preferable.

## Assumption 1: $E(u_t) = 0$

- Assumption that the mean of the disturbances is zero.
- For all diagnostic tests, we cannot observe the disturbances and so perform the tests of the residuals.
- The mean of the residuals will always be zero provided that there is a constant term in the regression.

## Assumption 2: $\text{Var}(u_t) = \sigma^2 < \infty$

- We have so far assumed that the variance of the errors is constant,  $\sigma^2$  - 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,  $\hat{u}_t$ .



## 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)

3. 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.

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

4. The test statistic is distributed as an  $F(T_1-k, T_2-k)$  under the null of homoscedasticity.
5. 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 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  $\chi^2$  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.

## 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).
- A simple illustration of GLS is as follows: Suppose that the error variance is related to another variable  $z_t$  by

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

- 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  $\text{var}(v_t) = \text{var}\left(\frac{u_t}{z_t}\right) = \frac{\text{var}(u_t)}{z_t^2} = \frac{\sigma^2 z_t^2}{z_t^2} = \sigma^2$  for known  $z_t$ .

## Other Approaches to Dealing with Heteroscedasticity

- So the disturbances from the new regression equation will be homoscedastic.
- 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.

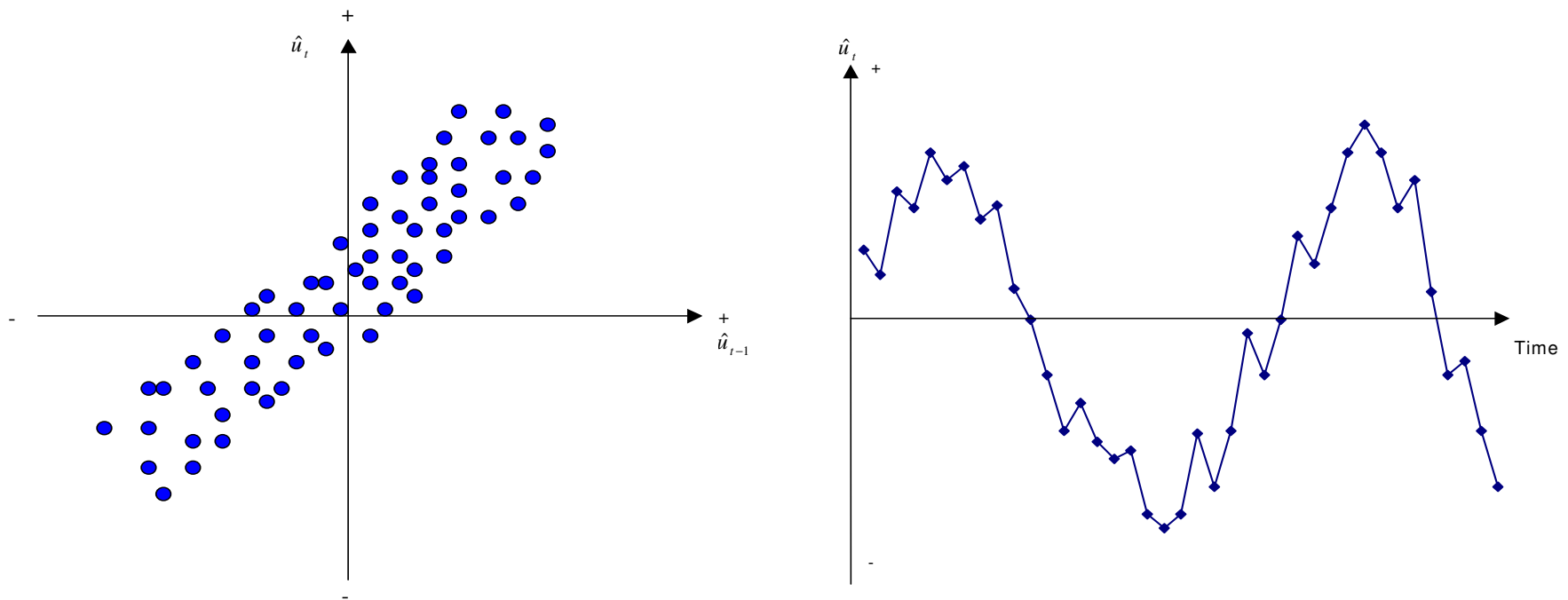
## Background – The Concept of a Lagged Value

$t$	$y_t$	$y_{t-1}$	$\Delta y_t$
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

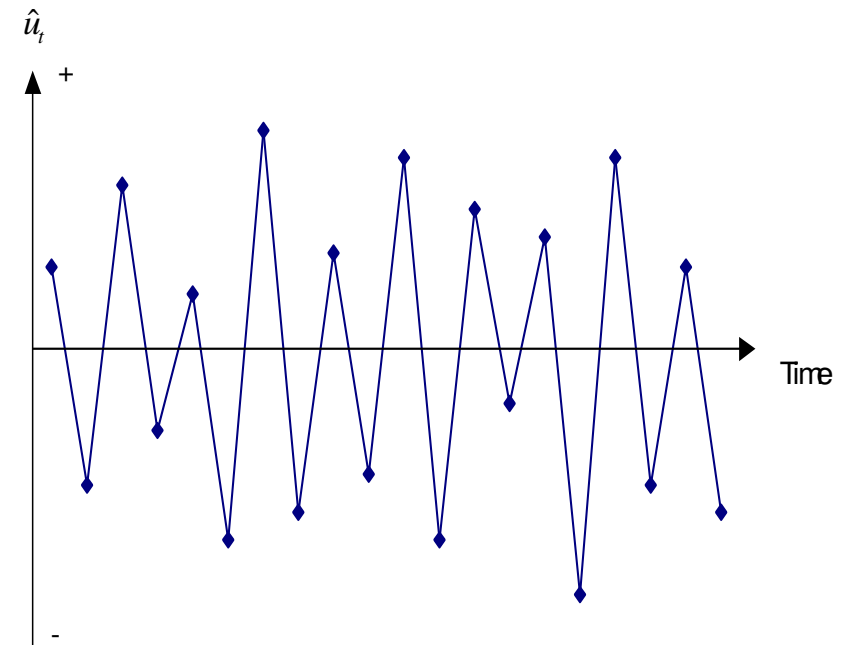
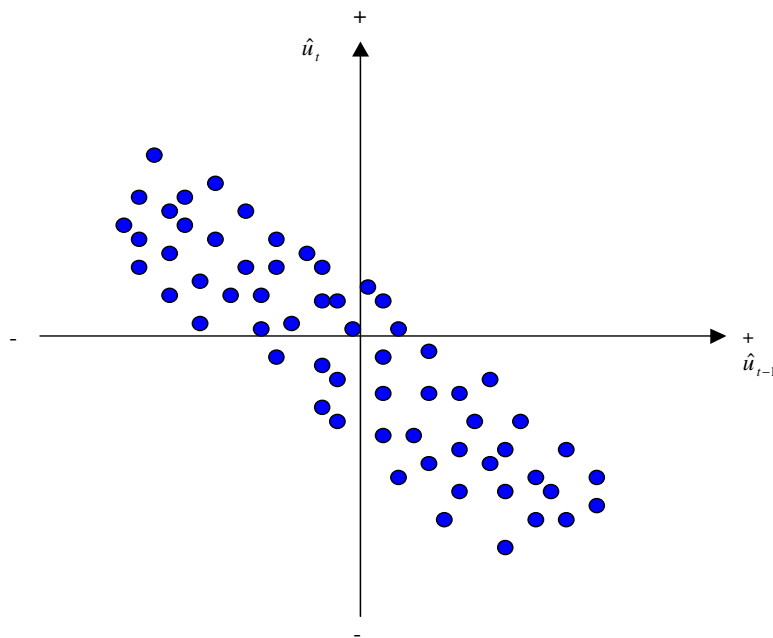
- We assumed of the CLRM's errors that  $\text{Cov}(u_i, u_j) = 0$  for  $i \neq j$ , i.e. This is essentially the same as saying there is no pattern in the errors.
- Obviously we never have the actual  $u$ 's, so we use their sample counterpart, the residuals (the  $\hat{u}$ 's).
- If there are patterns 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.

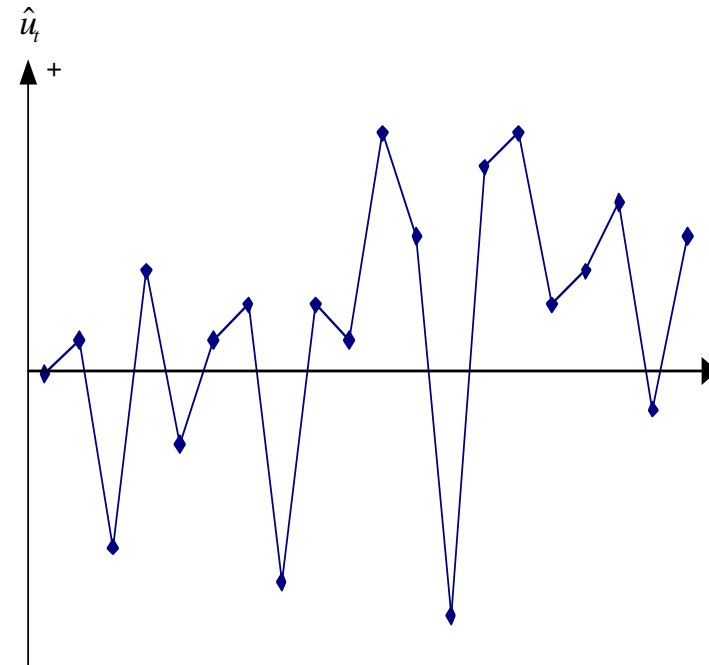
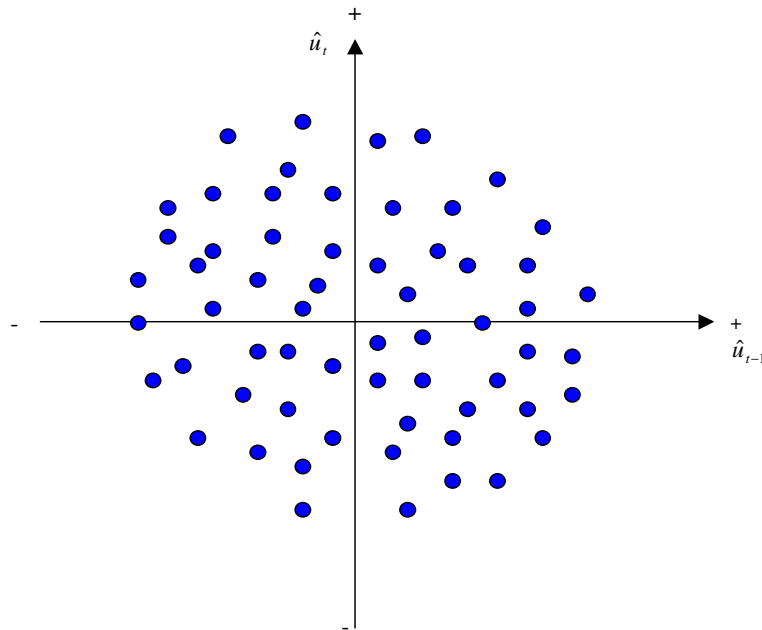
## Negative Autocorrelation



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

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

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

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

where  $\hat{\rho}$  is the estimated correlation coefficient. Since  $\hat{\rho}$  is a correlation, it implies that

$$-1 \leq \hat{\rho} \leq 1$$

- 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

Reject  $H_0$  :

positive  
autocorrelation

Inconclusive

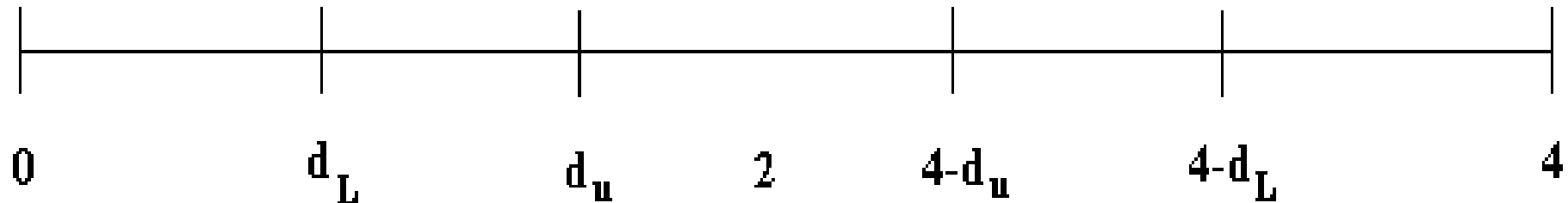
Do not reject

$H_0$ : No evidence  
of autocorrelation

Inconclusive

Reject  $H_0$  :

negative  
autocorrelation



### Conditions which Must be Fulfilled for DW to be a Valid Test

1. Constant term in regression
2. Regressors are non-stochastic
3. No lags of dependent variable

## Another Test for Autocorrelation: The Breusch-Godfrey Test

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

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

- The null and alternative hypotheses are:

$$H_0 : \rho_1 = 0 \text{ and } \rho_2 = 0 \text{ and } \dots \text{ and } \rho_r = 0$$

$$H_1 : \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or } \dots \text{ or } \rho_r \neq 0$$

- 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  $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-r}$

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.

## 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.
- 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.

## “Remedies” for 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.
- But such procedures that “correct” for autocorrelation require assumptions about the form of the autocorrelation.
- If these assumptions are invalid, the cure would be more dangerous than the disease! - see Hendry and Mizon (1978).
- However, it is unlikely to be the case that the form of the autocorrelation is known, and a more “modern” view is that residual autocorrelation presents an opportunity to modify the regression.

## Dynamic Models

- All of the models we have considered so far have been static, e.g.

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

- But we can easily extend this analysis to the case where the current value of  $y_t$  depends on previous values of  $y$  or one of the  $x$ 's, e.g.

$$y_t = \beta_1 + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \gamma_1 y_{t-1} + \gamma_2 x_{2t-1} + \dots + \gamma_k x_{kt-1} + u_t$$

- We could extend the model even further by adding extra lags, e.g.

$$x_{2t-2}, y_{t-3}.$$



## Why Might we Want/Need To Include Lags in a Regression?

- Inertia of the dependent variable
- Over-reactions
- Measuring time series as overlapping moving averages
- However, other problems with the regression could cause the null hypothesis of no autocorrelation to be rejected:
  - Omission of relevant variables, which are themselves autocorrelated.
  - If we have committed a “misspecification” error by using an inappropriate functional form.
  - Autocorrelation resulting from unparameterised seasonality.

## 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  $\Delta y_t$ ; similarly for the  $x$ -variables,  $\Delta x_{2t} = x_{2t} - x_{2t-1}$  etc.

- The model would now be

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

- Sometimes the change in  $y$  is purported to depend on previous values of  $y$  or  $x_t$  as well as changes in  $x$ :

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \beta_3 x_{2t-1} + \beta_4 y_{t-1} + u_t$$

## The Long Run Static Equilibrium Solution

- One interesting property of a dynamic model is its long run or static equilibrium solution.
- “Equilibrium” implies that the variables have reached some steady state and are no longer changing, i.e. if  $y$  and  $x$  are in equilibrium, we can say

$$y_t = y_{t+1} = \dots = y \text{ and } x_t = x_{t+1} = \dots = x$$

Consequently,  $\Delta y_t = y_t - y_{t-1} = y - y = 0$  etc.

- So the way to obtain a long run static solution is:
  1. Remove all time subscripts from variables
  2. Set error terms equal to their expected values,  $E(u_t)=0$
  3. Remove first difference terms altogether
  4. Gather terms in  $x$  together and gather terms in  $y$  together.
- These steps can be undertaken in any order

## The Long Run Static Equilibrium Solution: An Example

If our model is

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{2t} + \beta_3 x_{2t-1} + \beta_4 y_{t-1} + u_t$$

then the static solution would be given by

$$0 = \beta_1 + \beta_3 x_{2t-1} + \beta_4 y_{t-1}$$

$$\beta_4 y_{t-1} = -\beta_1 - \beta_3 x_{2t-1}$$

$$y = \frac{-\beta_1}{\beta_4} - \frac{\beta_3}{\beta_4} x_2$$

## Problems with Adding Lagged Regressors to “Cure” Autocorrelation

- Inclusion of lagged values of the dependent variable violates the assumption that the RHS variables are non-stochastic.
- What does an equation with a large number of lags actually mean?
- Note that if there is still autocorrelation in the residuals of a model including lags, then the OLS estimators will not even be consistent.

## Multicollinearity

- This problem occurs when the explanatory variables are very highly correlated with each other.
- Perfect multicollinearity  
Cannot estimate all the coefficients
  - e.g. suppose  $x_3 = 2x_2$   
and the model is  $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + u_t$
- Problems if Near Multicollinearity is Present but Ignored
  - $R^2$  will be high but the individual coefficients will have high standard errors.
  - The regression becomes very sensitive to small changes in the specification.
  - Thus confidence intervals for the parameters will be very wide, and significance tests might therefore give inappropriate conclusions.

## Measuring Multicollinearity

- The easiest way to measure the extent of multicollinearity is simply to look at the matrix of correlations between the individual variables. e.g.

Corr	$x_2$	$x_3$	$x_4$
$x_2$	-	0.2	<u>0.8</u>
$x_3$	0.2	-	0.3
$x_4$	<u>0.8</u>	0.3	-

- But another problem: if 3 or more variables are linear  
- e.g.  $x_{2t} + x_{3t} = x_{4t}$
- Note that high correlation between  $y$  and one of the  $x$ 's is not multicollinearity.

## Solutions to the Problem of Multicollinearity

- “Traditional” approaches, such as ridge regression or principal components. But these usually bring more problems than they solve.
- Some econometricians argue that if the model is otherwise OK, just ignore it
- The easiest ways to “cure” the problems are
  - drop one of the collinear variables
  - transform the highly correlated variables into a ratio
  - go out and collect more data e.g.
    - a longer run of data
    - switch to a higher frequency



## Adopting the Wrong Functional Form

- We have previously assumed that the appropriate functional form is linear.
- This may not always be true.
- We can formally test this using Ramsey's RESET test, which is a general test for mis-specification of functional form.

- Essentially the method works by adding higher order terms of the fitted values (e.g.  $\hat{y}_t^2, \hat{y}_t^3$  etc.) into an auxiliary regression:

Regress  $\hat{u}_t$  on powers of the fitted values:

Obtain  $R^2$  from this regression. The test statistic is given by  $TR^2$  and is distributed as a

$$\chi^2(p-1)$$

- So if the value of the test statistic is greater than a  $\chi^2(p-1)$  then reject the null hypothesis that the functional form was correct.

## But what do we do if this is the case?

- The RESET test gives us no guide as to what a better specification might be.
- One possible cause of rejection of the test is if the true model is

In this case the remedy is obvious.  $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{2t}^2 + \beta_4 x_{2t}^3 + u_t$

- Another possibility is to transform the data into logarithms. This will linearise many previously multiplicative models into additive ones:

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

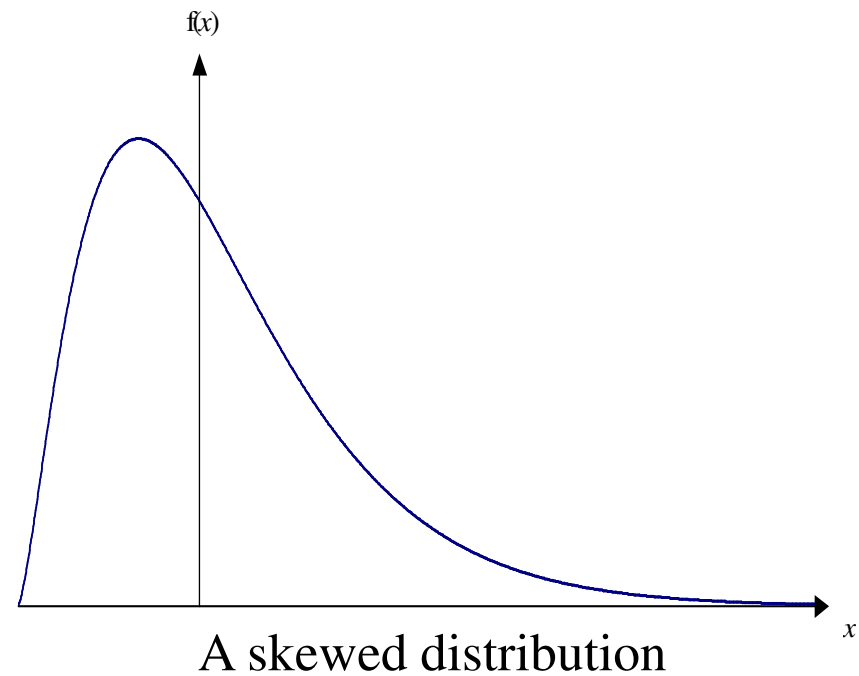
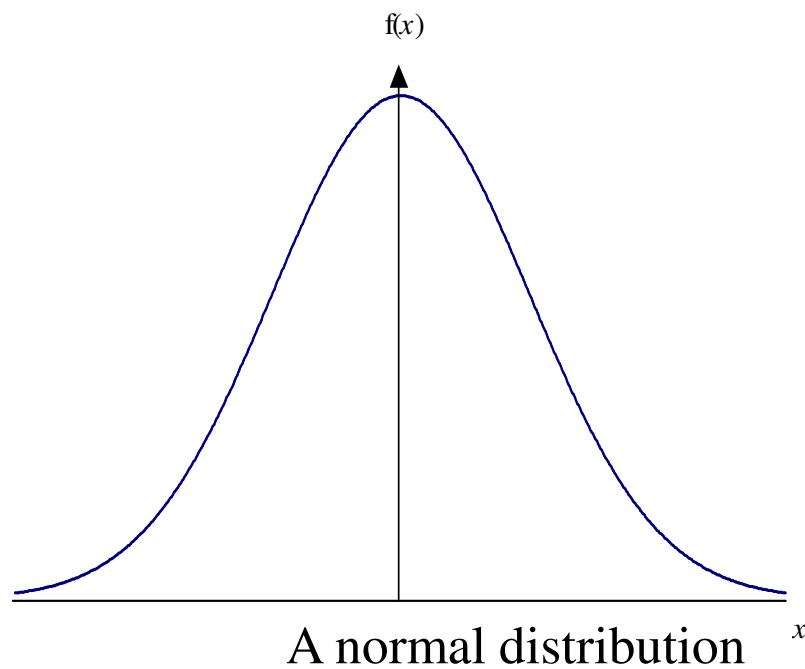
## Testing the Normality Assumption

- Why did we need to assume normality for hypothesis testing?

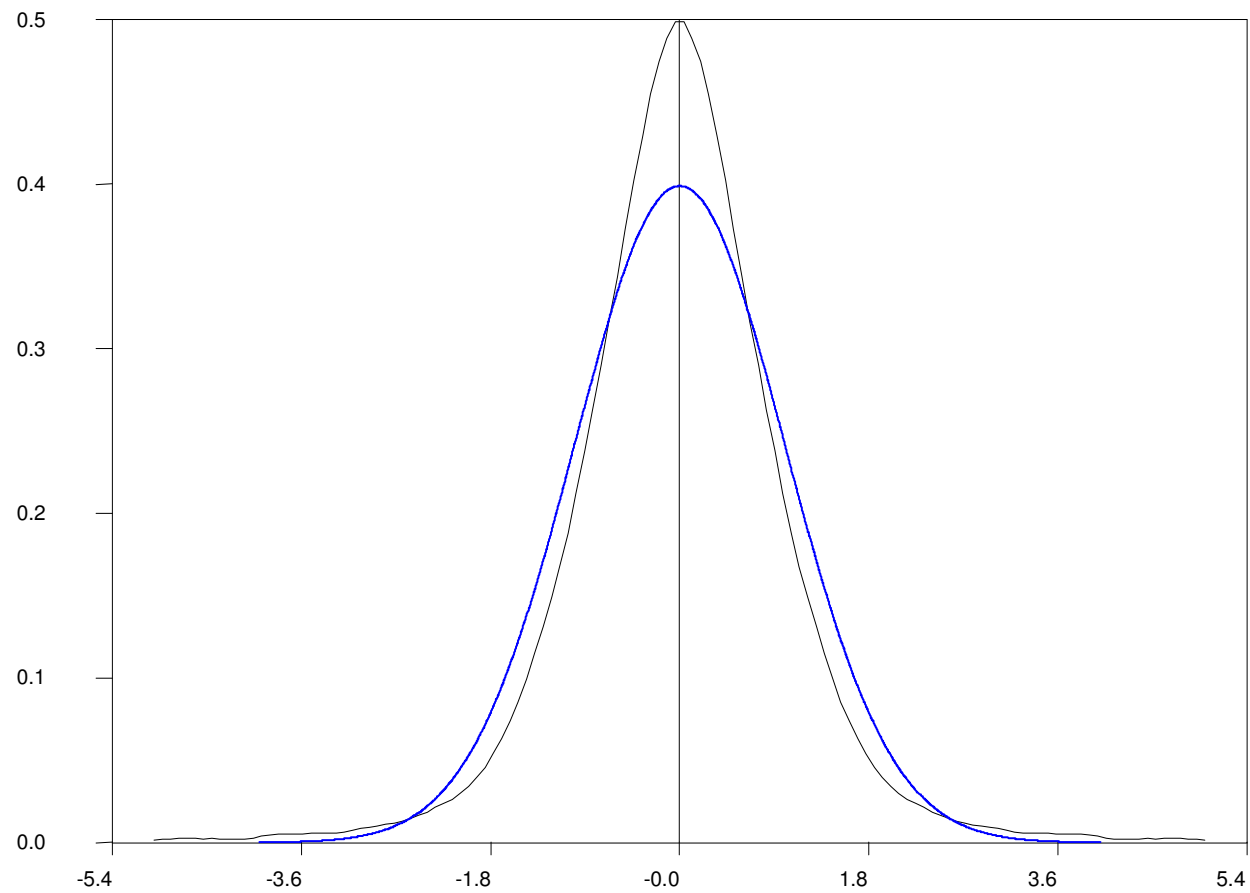
### Testing for Departures from Normality

- *The Bera Jarque normality test*
- A normal distribution is not skewed and is defined to have a coefficient of kurtosis of 3.
- The kurtosis of the normal distribution is 3 so its excess kurtosis ( $b_2-3$ ) is zero.
- Skewness and kurtosis are the (standardised) third and fourth moments of a distribution.

## Normal versus Skewed Distributions



## 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.
- It can be proved that the coefficients of skewness and kurtosis can be expressed respectively as:

$$b_1 = \frac{E[u^3]}{(\sigma^2)^{3/2}} \quad \text{and} \quad b_2 = \frac{E[u^4]}{(\sigma^2)^2}$$

- The Bera Jarque test statistic is given by

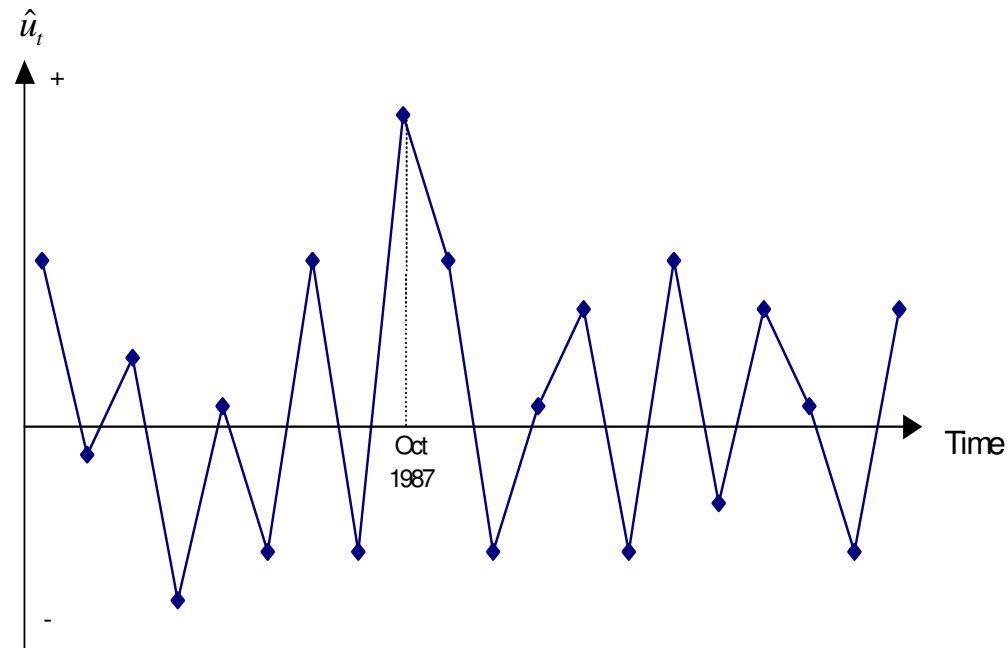
$$W = T \left[ \frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24} \right] \sim \chi^2(2)$$

- We estimate  $b_1$  and  $b_2$  using the residuals from the OLS regression,  $\hat{u}$ .

## What do we do if we find evidence of Non-Normality?

- It is not obvious what we should do!
- Could use a method which does not assume normality, but difficult and what are its properties?
- Often the case that one or two very extreme residuals causes us to reject the normality assumption.
- An alternative is to use dummy variables.  
e.g. say we estimate a monthly model of asset returns from 1980-1990, and we plot the residuals, and find a particularly large outlier for October 1987:

## What do we do if we find evidence of Non-Normality? (cont'd)



- Create a new variable:  
 $D87M10_t = 1$  during October 1987 and zero otherwise.  
This effectively knocks out that observation. But we need a theoretical reason for adding dummy variables.



## **Omission of an Important Variable or Inclusion of an Irrelevant Variable**

### **Omission of an Important Variable**

- Consequence: The estimated coefficients on all the other variables will be biased and inconsistent unless the excluded variable is uncorrelated with all the included variables.
- Even if this condition is satisfied, the estimate of the coefficient on the constant term will be biased.
- The standard errors will also be biased.

### **Inclusion of an Irrelevant Variable**

- Coefficient estimates will still be consistent and unbiased, but the estimators will be inefficient.

## Parameter Stability Tests

- So far, we have estimated regressions such as  $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$
- We have implicitly assumed that the parameters ( $\beta_1$ ,  $\beta_2$  and  $\beta_3$ ) are constant for the entire sample period.
- We can test this implicit assumption using parameter stability tests. The idea is essentially to split the data into sub-periods and then to estimate up to three models, for each of the sub-parts and for all the data and then to “compare” the *RSS* of the models.
- There are two types of test we can look at:
  - Chow test (analysis of variance test)
  - Predictive failure tests

## The Chow Test

- The steps involved are:
  1. Split the data into two sub-periods. Estimate the regression over the whole period and then for the two sub-periods separately (3 regressions). Obtain the RSS for each regression.
  2. The restricted regression is now the regression for the whole period while the “unrestricted regression” comes in two parts: for each of the sub-samples.

We can thus form an F-test which is the difference between the *RSS*'s.

The statistic is 
$$\frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2k}{k}$$

## The Chow Test (cont'd)

where:

$RSS$  = RSS for whole sample

$RSS_1$  = RSS for sub-sample 1

$RSS_2$  = RSS for sub-sample 2

$T$  = number of observations

$2k$  = number of regressors in the “unrestricted” regression (since it comes in two parts)

$k$  = number of regressors in (each part of the) “unrestricted” regression

3. Perform the test. If the value of the test statistic is greater than the critical value from the F-distribution, which is an  $F(k, T-2k)$ , then reject the null hypothesis that the parameters are stable over time.

## A Chow Test Example

- Consider the following regression for the CAPM  $\beta$  (again) for the returns on Glaxo.
- Say that we are interested in estimating Beta for monthly data from 1981-1992. The model for each sub-period is

- 1981M1 - 1987M10

$$0.24 + 1.2R_{Mt} \quad T = 82 \quad RSS_1 = 0.03555$$

- 1987M11 - 1992M12

$$0.68 + 1.53R_{Mt} \quad T = 62 \quad RSS_2 = 0.00336$$

- 1981M1 - 1992M12

$$0.39 + 1.37R_{Mt} \quad T = 144 \quad RSS = 0.0434$$

## A Chow Test Example - Results

- The null hypothesis is

$$H_0: \alpha_1 = \alpha_2 \text{ and } \beta_1 = \beta_2$$

- The unrestricted model is the model where this restriction is not imposed

$$\begin{aligned} \text{Test statistic} &= \frac{0.0434 - (0.0355 + 0.00336)}{0.0355 + 0.00336} \times \frac{144 - 4}{2} \\ &= 7.698 \end{aligned}$$

Compare with 5%  $F(2, 140) = 3.06$

- We reject  $H_0$  at the 5% level and say that we reject the restriction that the coefficients are the same in the two periods.

## The Predictive Failure Test

- Problem with the Chow test is that we need to have enough data to do the regression on both sub-samples, i.e.  $T_1 \gg k$ ,  $T_2 \gg k$ .
- An alternative formulation is the predictive failure test.
- What we do with the predictive failure test is estimate the regression over a “long” sub-period (i.e. most of the data) and then we predict values for the other period and compare the two.

To calculate the test:

- Run the regression for the whole period (the restricted regression) and obtain the RSS
- Run the regression for the “large” sub-period and obtain the RSS (called  $RSS_1$ ). Note we call the number of observations  $T_1$  (even though it may come second).

$$\text{Test Statistic} = \frac{RSS - RSS_1}{RSS} \times \frac{T_1 - k}{T_2}$$

where  $T_2$  = number of observations we are attempting to “predict”. The test statistic will follow an  $F(T_2, T_1 - k)$ .

## Backwards versus Forwards Predictive Failure Tests

- There are 2 types of predictive failure tests:
  - Forward predictive failure tests, where we keep the last few observations back for forecast testing, e.g. we have observations for 1970Q1-1994Q4. So estimate the model over 1970Q1-1993Q4 and forecast 1994Q1-1994Q4.
  - Backward predictive failure tests, where we attempt to “back-cast” the first few observations, e.g. if we have data for 1970Q1-1994Q4, and we estimate the model over 1971Q1-1994Q4 and backcast 1970Q1-1970Q4.



## Predictive Failure Tests – An Example

- We have the following models estimated:

For the CAPM  $\beta$  on Glaxo.

- 1980M1-1991M12

$$0.39 + 1.37R_{Mt} \quad T = 144 \quad RSS = 0.0434$$

- 1980M1-1989M12

$$0.32 + 1.31R_{Mt} \quad T_1 = 120 \quad RSS_1 = 0.0420$$

Can this regression adequately “forecast” the values for the last two years?

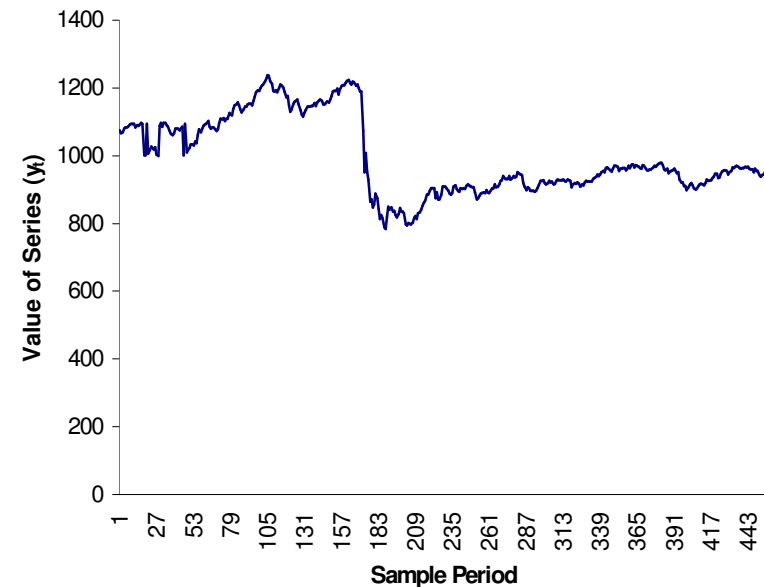
$$\text{TestStatistic} = \frac{0.0434 - 0.0420}{0.0420} \times \frac{120 - 2}{24} = 0.164$$

- Compare with  $F(24, 118) = 1.66$ .

So we do not reject the null hypothesis that the model can adequately predict the last few observations.

## How do we decide the sub-parts to use?

- As a rule of thumb, we could use all or some of the following:
  - Plot the dependent variable over time and split the data accordingly to any obvious structural changes in the series, e.g.
- Split the data according to any known important historical events (e.g. stock market crash, new go
- Use all but the last few observations and do a predictive failure test on those.



## A Strategy for Building Econometric Models

Our Objective:

- To build a statistically adequate empirical model which
  - satisfies the assumptions of the CLRM
  - is parsimonious
  - has the appropriate theoretical interpretation
  - has the right “shape” - i.e.
    - all signs on coefficients are “correct”
    - all sizes of coefficients are “correct”
  - is capable of explaining the results of all competing models

## 2 Approaches to Building Econometric Models

- There are 2 popular philosophies of building econometric models: the “specific-to-general” and “general-to-specific” approaches.
- “Specific-to-general” was used almost universally until the mid 1980’s, and involved starting with the simplest model and gradually adding to it.
- Little, if any, diagnostic testing was undertaken. But this meant that all inferences were potentially invalid.
- An alternative and more modern approach to model building is the “LSE” or Hendry “general-to-specific” methodology.
- The advantages of this approach are that it is statistically sensible and also the theory on which the models are based usually has nothing to say about the lag structure of a model.

## The General-to-Specific Approach

- First step is to form a “large” model with lots of variables on the right hand side
- This is known as a GUM (generalised unrestricted model)
- At this stage, we want to make sure that the model satisfies all of the assumptions of the CLRM
- If the assumptions are violated, we need to take appropriate actions to remedy this, e.g.
  - taking logs
  - adding lags
  - dummy variables
- We need to do this before testing hypotheses
- Once we have a model which satisfies the assumptions, it could be very big with lots of lags & independent variables

## The General-to-Specific Approach: Reparameterising the Model

- The next stage is to reparameterise the model by
  - knocking out very insignificant regressors
  - some coefficients may be insignificantly different from each other, so we can combine them.
- At each stage, we need to check the assumptions are still OK.
- Hopefully at this stage, we have a statistically adequate empirical model which we can use for
  - testing underlying financial theories
  - forecasting future values of the dependent variable
  - formulating policies, etc.