Time Series Econometrics

Nicky Grant

ECON30401 Lecture Notes 3

Semester 1 2016

1 Inference in Time Series

In the first two lectures we defined the notion of a process $Y_t(t = ..-2,-1,0,1,2,..)$, some unobserved random variable (i.e data generating mechanism) which generates the data we observe $\{y_t\}_{t=1}^{T}$. So far we have studied the properties of different theoretical processes Y_t (specifically ARMA models) by studying the form of dynamics they can capture. Namely we derived the theoretical Auto-Correlation Function (ACF) of some key ARMA processes- for example the AR(1) and MA(1) and observed how they permit very different dependence properties. Understanding the theoretical properties of alternative processes is a crucial part to understanding the notion of inference from sample data and is the topic we now move on to.

In reality we observe some realisation of time series data, for example UK GDP Growth which came from some unknown process, the exact form of which is unknown to us. **This lecture is concerned with methods to infer the form of the true process $Y_t(t = ..-2,-1,0,1,2,..)$ based on a realisation from this process, namely the sample data of size T $\{y_t\}_{t=1}^{T}$.** From our perspective the true process could take any form and the aim of inference is to determine the form of this process as closely as possible. To achieve this we discuss some simple methods to select and test between different processes using our sample data to try and unearth the form of the underlying process.

Often the first step in an analysis of time series data is to undertake a visual examination of the properties of the observed series by graphing the data over time. Below is the graph of UK GDP between 1984Q1 and 2012Q4. [1] By eye-balling the graph of the series we may have some notion of some general properties of the process of UK GDP growth. For example the mean looks to be around 1 and there is clear evidence of positive persistence in the series. However this still gives no indication more specifically of the form of the process, which may be useful for example to predict future GDP growth. Suppose for sake of example that the true process of UK GDP is $Y_t = 0.4 + 0.5Y_{t-1} - 0.1\varepsilon_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim WN(1)$ (i.e an ARMA(1,1), **the question is then how to use a realisation (namely the sample data) to unearth this process which is unknown to us?**

[1] 'Q' refers to Quarter.

At the broadest level the task is quite daunting, we have a set of T observations, say for example $T = 240$ observations of monthly UK GDP data from September 1995 to September 2015 with which to unearth this process which

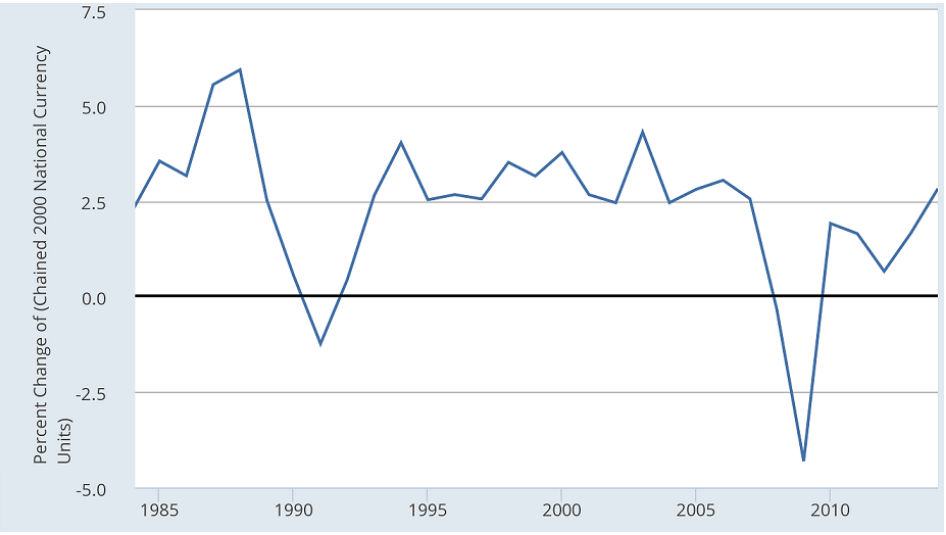


Figure 1.1: **Annual Growth (%) in Total UK GDP (1984Q1-2012Q4)**

Source: St Louis Fed

is completely unknown. From our perspective the true model could take any form- i.e it could be a particular ARMA(1,1) or an AR(5) or an ARMA(5,16) or an even more complex model.

Intuitively at the most fundamental level it should be apparent that in order to make any inference on the true process we need a sufficiently large data set. For example if we only have data on 3 quarterly observations of UK GDP Growth it is going to be extremely difficult for any form of estimation or model selection to unearth the true process. Simply because with 3 data points it is difficult to distinguish between different processes.

In order to be able to infer the form of the dynamics (i.e the correlation patterns, the ACF) we need sufficiently long data sets with which to be able to reliably estimate correlations of Y_t with past observations. Secondly in order to reliably estimate the true process we require the process to be stable over time, i.e that the process does not change dramatically over time. Namely it would be desirable if the true process we are trying to determine is stationary. In this lecture we make this assumption which is often the starting point in many theoretical and empirical studies of time series. [2]

[2] We will refer to these more general issues in the PC Labs and lectures.

In this lecture we will study

1. Methods to estimate correlations in the data and perform tests on the dependence properties of the underlying process. To do this we introduce and discuss the **Sample Autocorrelation Function** and how we may use this to perform **hypothesis tests** on the sign and magnitude of correlations at different lags of the true underlying process which generated our data.

2. Estimation of ARMA type models in particular **OLS estimation of AR(p) models**. Also methods with which to test hypotheses on the coefficients of ARMA models.[3]

3. Methods to test that our model is dynamically complete, namely that the model we specify has errors which are uncorrelated (White Noise) and as such our specified models captures all the dynamics so that what remains is random variation.

4. Model selection betwee methods in ARMA models, namely using information criterion methods **Akaike and Bayesian Information Criterion (AIC & BIC).**

[3] We will not be concerned with proving the results on estimation and inference below, but more that you understand the intuition of the results and are comfortable in applying such techniques in practise.

1.1 *Sample Auto-Correlation Function & Testing*

The dependence properties of Time series can be characterised by their ACF, examples of which were given in lectures 1 and 2. For example an MA(1) process could allow a non-zero correlation at lag 1 and no correlation elsewhere. Hence if the data we observe came from an MA(1) process then the correlations between observations in the data should also satisfy this property for sufficiently large data sets.

It seems natural to then use sample values to estimate the populations correlations of the underlying process $\rho(k)$. Rememberer for stationary series the correlation between Y_t and Y_{t-k} (see notes from lecture 1) is

$$\rho(k) = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)}$$

where $Cov(Y_t, Y_{t-k})$ and $Var(Y_t)$ are the covariance and variance in the whole population (and hence unknown to us). However we have a sample of observations $\{y_t\}_{t=1}^{T}$ and we can replace the population averages with the sample averages.

Definition 1.1: Sample Auto-Correlation

The Sample Autocorrelation $\hat{\rho}_T(k)$ is defined

$$\hat{\rho}_T(k) = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

where $\bar{y} := \frac{1}{T}\sum_{t=1}^{t} y_t$ is the sample mean, i.e the estimator of $\mu = \mathbb{E}[Y_t]$ the population means (assumed constant due to stationarity). The **Sample Autocorrelation Function** is the Sample Autocorrelation at at all lags $k = 1, .., K$ for some k where $K < T$ (see discussion below).

Remarks

- The subscript 'T' in $\hat{\rho}_T(k)$ denotes this estimator is from a sample of size T where the properties of $\hat{\rho}_T(k)$ will depend on the sample size.

- We may re-write $\hat{\rho}_T(k)$

$$\hat{\rho}_T(k) = \frac{\frac{1}{T}\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{1.1}$$

  The numerator of $\hat{\rho}_T(k)$ is an estimate of

$$Cov[Y_t, Y_{t-k}] = \mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_{t-k} - \mathbb{E}[Y_{t-k}])] \tag{1.2}$$

  as it takes the average of $(y_t - \bar{y})(y_{t-k} - \bar{y})$ in the sample, noting that average is taken from $k+1$ to T as $t = k+1$ is the first observations t for which the pair (y_t, y_{t-k}) is available in the sample. As an implication $K < T$ which makes sense, to estimate how y_t and y_{t-k} move together we need to observe observations at least k periods ago.

- In practise we would only estimate sample autocorrelations up to some number K which we would select be much smaller than T to form reliable estimates. [4]

- The numerator is an estimator of the variance $Var[Y_t] = \mathbb{E}[(Y_t - \mathbb{E}[Y_t])^2]$ replacing the population average with the sample average.

- We can see the stationary assumptions is key here. For example we've used the fact the variance is constant over time so that the variance in the sample is reflective of the variance in any time period. This is also the case for estimating the mean and covariance which under stationarity do not change over time

- Intuitively the larger is the sample, the closer $\hat{\rho}_T(k)$ is to $\rho(k)$ though they never equal each other and will differ by sampling variation.[5] Namely $\hat{\rho}_T(k)$ is the realisation of a random variable and we must characterise this randomness (i.e it's distribution) when forming hypotheses on $\rho(k)$.

[4] For example if we observe only 10 years of UK GDP Data then for example we only have one data point with which to estimate the correlation at lag 9, namely observation 1 and 10. This is clearly going to be a very unreliable estimate of the correlation of UK GDP with 9 periods ago..

[5] For example suppose a stationary process has variance 5 (population variance). If we have a sample $T = 20$ in our realisation we may find the sample variance of this of 5.5, if we observed another realisation of $T = 20$ we may find variance in this sample of say 5.2 and so on.

As with the Population Auto-Correlation Function we may plot the Sample Auto-Correlations for different lags to form the **Sample Auto-Correlation Function**, also known as the **Correlogram**.

By matching the properties of the Correlogram with the theoretical autocorrelation properties for stationary $ARMA$ processes, an initial indication can be given of models that may be appropriate for the data. In doing so, however, it needs to be remembered that sample data will not precisely reproduce the properties of the underlying process, particularly when the available sample size is of moderate size.

In order to make inference on the form of the dependence in the true series we require methods to test certain hypothesis on the true unobserved autocorrelation function. To do this we must take in to account the sampling error in our estimate $\hat{\rho}_T(k)$ in estimating $\rho(k)$

1.1.1 *Sampling Properties of the Sample Autocorrelation (Function)*

The correlation structure of the true process $\rho(k)$ are unknown to us. We have an estimate $\hat{\rho}_T(k)$ which for T large should give a good approximation to $\rho(k)$. However this concept is vague when it comes to making inference. If for example in our sample we find $\hat{\rho}_T(1) = 0.09$ then the sample correlation at lag 1 is small, which may lead us to suspect that the true correlation at lag 1 is $\rho(1) = 0$ (or very small). However $\hat{\rho}_T(k)$ is a realisation from some random variable, for every different realisation we'd find a different estimate of the correlation at lag 1. As such we require methods with which to formalise the sampling uncertainty in our estimator to formally test hypotheses on $\rho(k)$ based on the sample data that formed $\hat{\rho}_T(k)$.

In order to see the sampling uncertainty in the sample autocorrelation we now consider a simulation experiment.

Suppose the true Process is $\boxed{Y_t = 0.5Y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim \text{WN}(1)}$. We know the ACF is $\rho(k) = 0.5^k$ which is plotted in Figure 1.2 below.

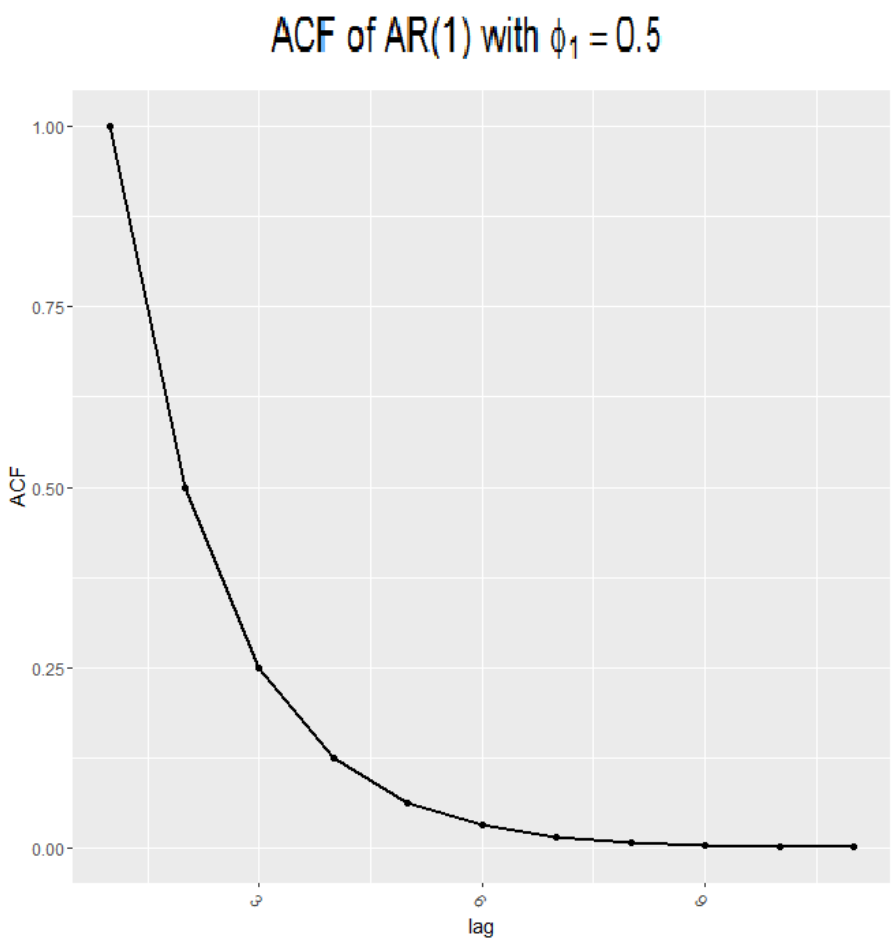


Figure 1.2: **ACF of an AR(1)** $Y_t = 0.5Y_{t-1} + \varepsilon_t$

Of course when observe our sample data in practise we do not know that the true process was an AR(1) above and that the ACF looked as in Figure 1.2. We would however plot the Sample Autcorrelation Function (the correlation patterns in our data) as our best guess of what the true ACF looked like. But in a simulation we can set the true process, generate data and then see the properties of our different estimation methods.

The graphs below are an animated figure which perform the following simulations for $T = \{30, 100, 1000, 10000\}$

1. Generate a sample $\{y_1^b, .., y_T^b\}$ from $Y_t = 0.5Y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim N(0,1)$
2. Calculate the Sample ACF up to lag 5 $\hat{\rho}_T^b(1)$,..,$\hat{\rho}_T^b(1)$ for each sample b
3. Plot Sample ACF $\hat{\rho}_T^b(1)$,..,$\hat{\rho}_T^b(5)$ for $b = \{1, .., 100\}$.

Take $T = 30$. As you click forward in the figure below a sample draw is taken from an AR(1) as in Step 1 and plotted, repeatedly.

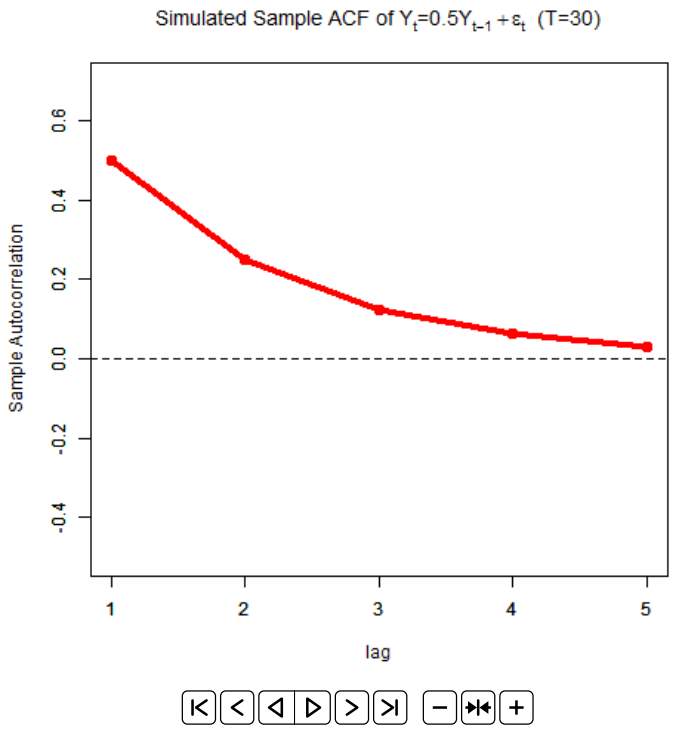


Figure 1.3: **Animated Simulation Sample Autocorrelation Function(T=30) of AR(1)** $Y_t = 0.5Y_{t-1} + \varepsilon_t$.

We generate 100 samples of size $T = 100$ *from* $Y_t = 0.5Yt - 1 + \varepsilon_t$ *setting* $Y_0 = 0$ *evaluated where* ε_t *is drawn from an i.i.d* $N(0,1)$ *distribution and plot the Sample ACF up to lag 5 for each sample. The line in red shows the population ACF* $\rho(k) = 0.5^k$.

What would we expect to happen to the sample realisations of the Sample ACF as the sample size T increases? Figure 1.4 repeats the above simulation for $T = 100$.

Figure 1.5 and Figure 1.6 repeats the above simulation again for $T = 1000, 10000$, again what do we expect to see happen?

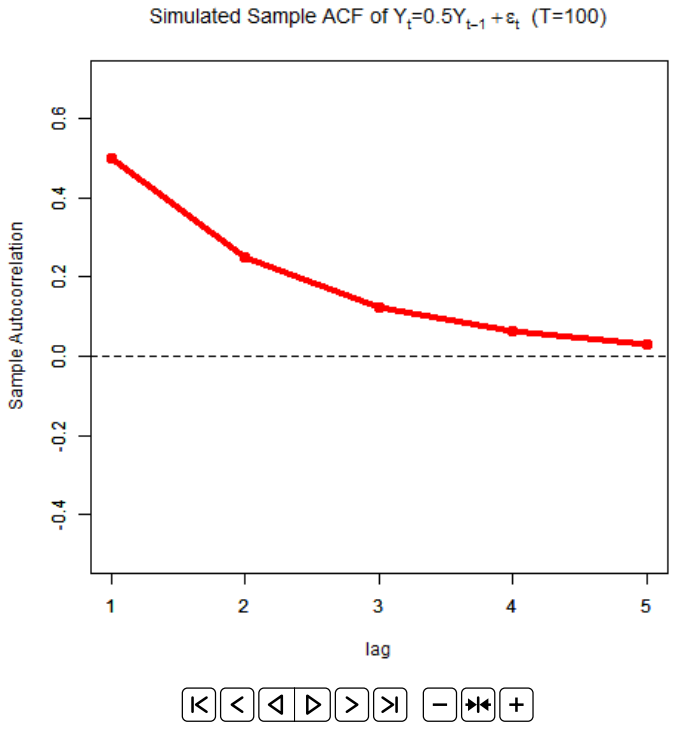


Figure 1.4: **Animated Simulation Sample Autocorrelation Function(T=30) of AR(1)** $Y_t = 0.5Y_{t-1} + \varepsilon_t$.

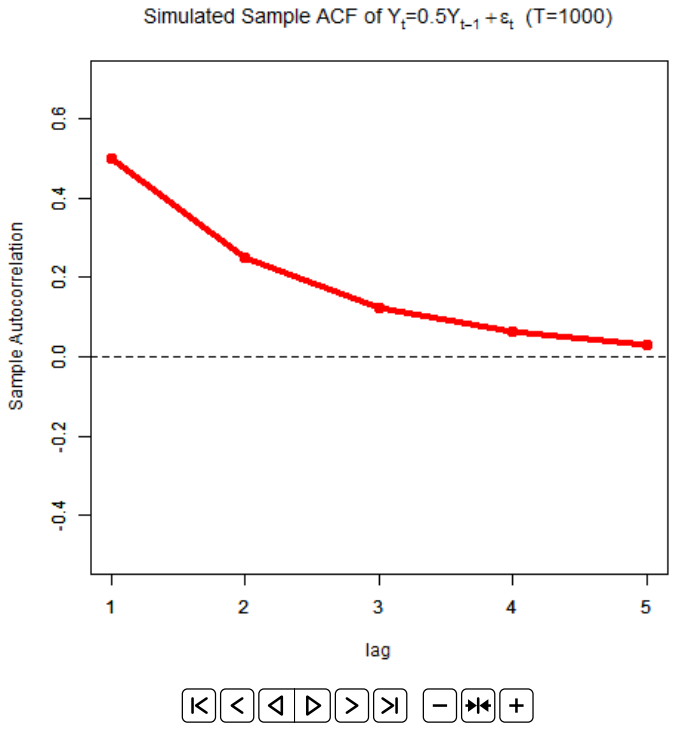


Figure 1.5: **Animated Simulation Sample Autocorrelation Function(T=30) of AR(1)** $Y_t = 0.5Y_{t-1} + \varepsilon_t$.

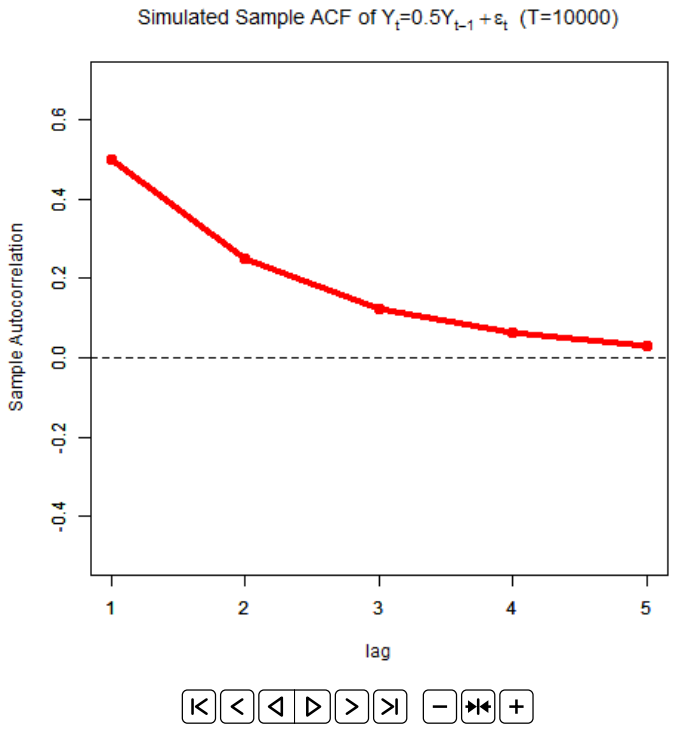


Figure 1.6: **Animated Simulation Sample Autocorrelation Function(T=30) of AR(1)** $Y_t = 0.5Y_{t-1} + \varepsilon_t$.

The animations above give a visual depiction of the sampling variation in the Sample ACF. We can see as the sample size increases the Sample ACF better approximates the true ACF of the process which generated our data. Again this makes intuitive sense, the data is generated from a process with a certain correlation pattern, and hence our data should reflect this in large enough samples.

We can also say more about the distribution or the sampling variation in the Sample Autocorrelations. To see this we perform a simulation below which shows the sampling distribution of $\hat{\rho}_T(1)$ for $T = 2, ..200$. Namely we plot the probabilty the Sample Autocorrelation at lag 1 takes on particular values. [6]

[6] For brevity we consider only the Sample ACF at lag 1, though we the idea holds for the Sample correlation at any lag.

The animated figure below performs the following simulation

1. Generate 1000 samples $\{y_1^b, .., y_T^b\}$ from $Y_t = 0.5Y_{t-1} + \varepsilon_t$ where $\varepsilon_t \sim N(0, 1)$
2. Calculate $\hat{\rho}_T^b(1)$ for each sample $b = \{1, .., 1000\}$
3. Plot probability distribution of $\hat{\rho}_T^b(1)$ $b = \{1, .., 1000\}$ for $T = 2, .200$.

The top graph in Figure **??** plots the sampling distribution of $\hat{\rho}_T(1)$ and the bottom graph the mean and standard deviation of $\hat{\rho}_T^(1)$ for $T = 2, .., 200$. What do we expect to happen to

A The form of the distribution of $\hat{\rho}_T^(1)$ as T increases?

B The mean of $\hat{\rho}_T(1)$ as T increases?

C The standard deviation (sd) of $\hat{\rho}_T^(1)$ as T increases?

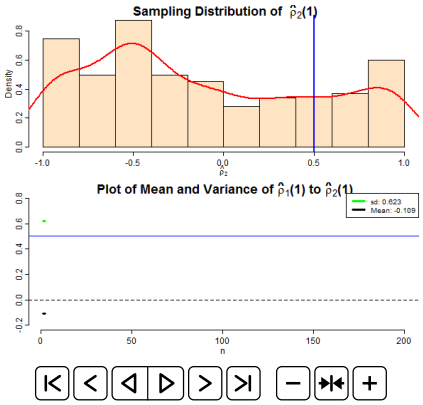


Figure 1.7: Animated Simulation of Sampling Distribution of $\hat{\rho}_T^b(1)$ T=2,..200.

We can see from Figures 1.2 to Figure **??** and the evidence in Figure **??** that

A The distribution $\hat{\rho}_T^(1)$ converges to a Normal Distribution as $T \to \infty$.

B The average value (mean) of $\hat{\rho}_T(1)$ converges to the population autocorrelation $\rho(1)$?

C The standard deviation (sd) of $\hat{\rho}_T(1)$ converges to zero.

All of these results verified in the simulations are formalsised in Section (**??**) more generally.

1.1.2 *Large Sample Properties of Sample ACF and Hypothesis Testing*

In this section we overview the statistical properties of the Sample Autocorrelations of a White Noise Process. We do this as often we wish to test a process is white noise, or more specifically whether or not the correlation at some lag is zero or not. [7]

If Y_t is a White Noise Process (i.e $\rho(k) = 0$ for all $k = 1, 2, ..$) then it can be shown the following results hold.

Theorem 1.1.1: Consistency

If Y_t is a White Noise Process

$$\hat{\rho}_T(k) \xrightarrow{p} 0 \quad \text{all } k = 1, 2, ...$$

This result says that the sample ACF converges to the population ACF with probability converging to 1. We saw this above for the case $\hat{\rho}_T(1)$ where $\rho(1) = 0.5$, namely as T increased $\hat{\rho}_T(1)$ approached 0.5 with increasing probability.[8]

[7] We can provide more general results, that look very similar to those in this section though allowing the case the population ACF isn't zero.

[8] More generally when Y_t isn't White Noise we can show $\hat{\rho}_T(k) \xrightarrow{p} \rho(k)$ for all $k = 1, 2, ...$

Theorem 1.1.2: Central Limit Theorem

If Y_t is a White Noise Process

$$\sqrt{T}\hat{\rho}_T(k) \xrightarrow{d} N(0,1) \quad \text{all } k = 1, 2, ...$$

The results says that if Y_t were White Noise the variable $\sqrt{T}\hat{\rho}_T(k)$ for T large enough when is distributed as a Normal random variable with mean zero and variance 1 (formally that this statistic Converges in Distribution to a $N(0,1)$.[9]

If Y_t is White Noise then $\rho(k) = 0$ for any $k = 1, 2, ...$ This allows us to form tests a process has correlation at any lag, and hence is not White Noise, but a dependent series

Commonly we wish to test that there is no correlation at some lag k. The null hypothesis is then

$$H_0 : \rho(k) = 0$$
$$H_A : \rho(k) \neq 0$$

where the alternative H_A is that this correlation is non-zero.

Under the Null Hypothesis then $\rho(k) = 0$ (and also that Y_t is WN) the result above says that

$$\sqrt{T}\hat{\rho}_T(k) \xrightarrow{d} N(0,1) \tag{1.3}$$

hence if the null H_0 is true then $\sqrt{T}\hat{\rho}_T(k)$ is approximately normally distributed, where $\Pr\{-c < \sqrt{T}\hat{\rho}_T(k) < c\} = \Pr\{-c < Z < c\}$ where Z is the standard normal distribution. We can then use this result to perform a test H_0 is true. Take $c = 1.96$ we know by the properties of a $N(0,1)$ random variable that $\Pr\{-1.96 < Z < 1.96\} = 0.95$. Hence under H_0 we know $Pr\{-1.96 < \sqrt{T}\hat{\rho}_T(k) < 1.96\} = 0.95$ so that there is 95% of the time the statistic $\sqrt{T}\hat{\rho}_T(k)$ lies in the region $(-1.96, 1.96)$. If this statistic lies outside this range we know there's only a 5% probability of observing this outcome under H_0. Hence if $|\sqrt{T}\hat{\rho}_T(k)| > 1.96$ we would reject the null with 95% certainty and if $|\sqrt{T}\hat{\rho}_T(k)| < 1.96$ we can not find evidence to reject the null hypothesis at the 5% significance level.

This was a quick recap of the notion of a hypothesis test applied to the case of testing whether a correlation at lag k is zero, i.e if $\rho(k) = 0$.

1.1.3 Q-Test

In addition to testing that individual correlations at particular lags are zero, we often wish to perform a joint test that all correlations up to some point are zero. Namely we may wish to test

$$\begin{aligned} H_0 &: \rho(k) = 0 \text{ all } k = 1, 2, ..., L \\ H_A &: \text{At least one } \rho(k) \neq 0 \text{ for } k = 1, 2, ..., L \end{aligned} \tag{1.4}$$

for some L. E.g if $L = 2$ the null is the correlation at the first two lags is zero, against the alternative that at least one of these correlations are non-zero.

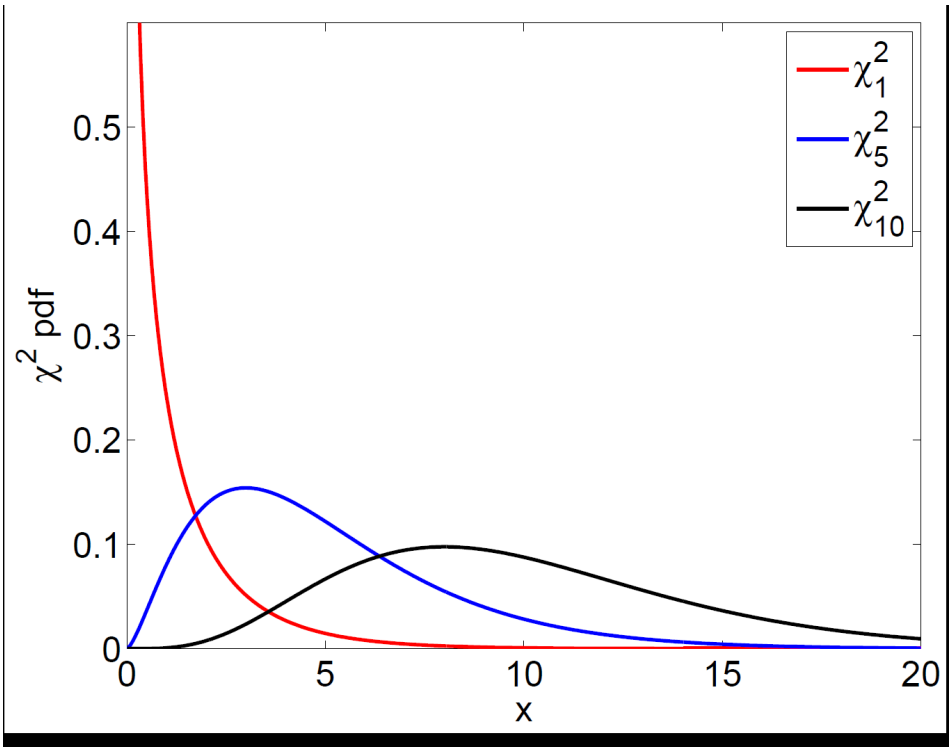
[9] This is just the Central Limit Theorem applied to the case of sample autocorrelations. It is well beyond the scope of this course to prove this result. However it is key to understand how to use this result to perform hypothesis tests on $\rho(k)$.

Definition 1.2: Chi Squared Distribution

For k independent random variables where $X_1, ..., X_k \sim N(0,1)$, then

$$\sum_{j=1}^{k} X_j^2 \sim \chi_k^2.$$

That is, the sum of squares of k independent standard normal variables is a χ^2 variable with k degrees of freedom. Below is the plot of the distribution of χ_k^2 for $k = 1, 5, 10$.



Again we can choose the relevant critical value to perform the test, which is done automatically in the Correlogram in Eviews.

Since we can show the sample autocorrelations $\hat{\rho}_T(j)$ where $\rho(j) = 0$ (assuming Y_t is WN) for $j = 1, 2, ...L$ are asymptotically independent where by eq. (**??**) ($\sqrt{T}\hat{\rho}_T(j)$ is distributed approximately $N(0,1)$ (so that $(\sqrt{T}\hat{\rho}_T(j))^2$) is approx. χ_1^2 distributed) so that) we can consequently state that

$$Q = \sum_{j=1}^{k} (\sqrt{T}\hat{\rho}_T(j))^2 = T\sum_{j=1}^{k} \hat{\rho}_T(j)^2 \xrightarrow{d} \chi_k^2$$

namely that Q is approximately distributed χ_k^2 if the null is true.

This Q-statistic therefore provides a means to test (**??**). If the Null is true then Q is approximately distributed χ_k^2 so if the Q statistic is less than the

critical value from the relevant χ^2_k distribution we accept the null, rejecting otherwise. This test is performed automatically in Eviews. We now move on to the Correlogram Output from Eviews and provide examples of these tests.

We now move on to discuss the Correlogram output in Eviews and give an example of performing hypothesis tests on auto-correlations.

1.1.4 *Correlogram output in Eviews*

The Correlogram of GDP growth (here denoted DLDGP) based on a data set of $T = 116$ observations from 1984Q1 to 2012Q4 is given below. This has been obtained using EViews (which we will see in the PC Lab in Week 5).

Correlogram of Quarterly UK GDP Growth

1. Sample: 1984Q1 2012Q4
Included observations: 116

Autocorrelation	Partial Correlatio		AC	PAC	Q-Stat	Prob
3.		1	0.582	0.582	40.333	0.000
		2	0.439	0.151	63.439	0.000
		3	0.289	0.024	73.582	0.000
		4	0.151	0.078	76.385	0.000
		5	0.111	0.037	77.893	0.000
		6	0.094	0.048	78.995	0.000
		7	0.033	0.060	79.129	0.000
		8	0.075	0.080	79.836	0.000
		9	0.104	0.073	81.214	0.000
		10	0.078	0.032	81.994	0.000
		11	0.084	0.005	82.923	0.000
		12	0.051	0.182	83.263	0.000
		13	0.033	0.067	83.406	0.000
		14	0.011	0.061	83.421	0.000
		15	0.032	0.035	83.560	0.000
		16	0.056	0.112	83.982	0.000

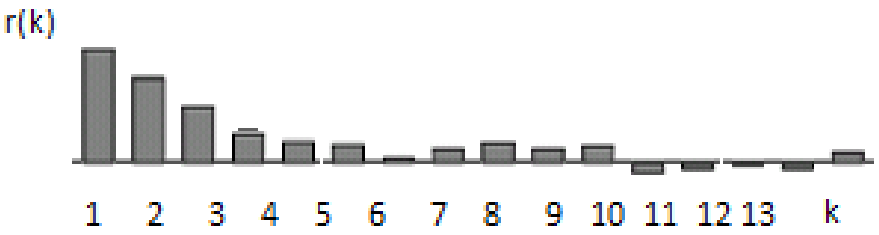
2. 4.

CORRELOGRAM EVIEWS OUTPUT: USER GUIDE

1. Gives the dates of the observation 1984Q1 to 2012Q4 with $T = 116$ observations.

2. Column indicating the lag, e.g at 1 all entries in that row refer to the sample correlation information pertaining to lag 1.

3. Vertical bar plot of $\hat{\rho}_T(k)$ for $k = 1, ..16$. It is often clearer to interpret plotted horizontally as done for the Population Auto-Correlation Function.



where the dotted lines have been removed. The column labelled AC gives the actual values of $\hat{\rho}_T(k)$ represented in the plot. For $k = 1$ here $\hat{\rho}_T(1) = 0.582$. The dotted lines in the plot in 1 give the rejection/acceptance region of a test that $\rho(k) = 0$ against the alternative that $\rho(k) \neq 0$. Remember we reject the null (at the 5% level) if $|\sqrt{T}\hat{\rho}_T(k)| > 1.96$ namely if $|\hat{\rho}_T(k)| > 1.96/\sqrt{T}$. The dotted lines are horizontal lines at $\pm 1.96/\sqrt{T}$. In this case at $\pm 1.96/\sqrt{116} = \pm 0.186$. Namely for any lag k is $\hat{\rho}_T(k)$ lies outside the banded region around 0 we have evidence at the 5% level that the correlation at lag k is not zero. For example we see $\hat{\rho}_T(1) = 0.582 > 0.182$ and lies outside this region so we would find evidence to reject the null that $\rho(1) = 0$ at the 5% level.

4. At any lag k the Q-Stat gives the test statistic corresponding to the Joint Hypothesis that $\rho(1) = ..\rho(k) = 0$ [namrly the null hypothesis is that the ACF up to lag k are all zero] against the alternative hypothesis that at least one of these correlations are non-zero. Namely it calculates $Q = \sum_{j=1}^{k} (\sqrt{T}\hat{\rho}_T(j))^2 = T \sum_{j=1}^{k} \hat{\rho}_T(j)^2$.

   The corresponding entry to the right in the column Prob is the p-value for this test, i.e the probability that this joint null hypothesis is true. Namely it is the probability that a χ^2_k takes a value at least as large as Q, i.e $\Pr\{\chi^2_k > Q\}$. If the null is true we know that Q was drawn from a χ^2_k and hence the probability of observing a value at least as large as Q corresponds to how likely it is the null holds. If this probability is very small it is unlikely to observe the statistic as large as Q if the null were true (as we know it came from a χ^2_k in this case).

5. The columns in red for the graph of the Partial Autocorrelations with the corresponding values provided PAC are not covered in this course and we can ignore these.

Hence the Correlogram Output from Eviews provides all the key information on the dependence properties of a particular time series variable and allows us to both visualise the form of the dependence in the true process, and perform formal hypothesis tests about its structure. We perform an example analysis for the Correlogram above.

Properties of UK GDP Growth From the Correlogram

- The Correlogram shows clear evidence of dependence in GDP growth. We can see the first three autocorrelations lie outside the critical region bounds and hence we'd reject the null that any of these correlations were zero. This is verified by the Q-Statistic where the joint hypothesis that the first three correlations are zero has a p-value of 0, i.e there is zero probability the null is true. After lag 4 we find no evidence of non-zero correlations, all the sample autocorrelations lie within the acceptance region.
- We see the first few correlations are positive and decline quickly. This provides evidence the true process may be an AR(1), as remember the theoretical ACF an AR(1) for a positive AR(1) coefficient had positive correlations decaying exponentially quickly.

Hence there is evidence the true process of UK GDP Growth may be of an AR(1) form. However this is not formally tested, so the true model may be of another form, however it is likely to be related to the AR(1) process.

Often assessing the properties of the process from the Correlogram is a key step in guiding estimation which we now move on to. Though a good guide to the likely true form of the process, we still require formal methods with which to estimate the exact form of the process.[10]

[10] For example even if the true process for UK GDP growth was an AR(1), we still do not know the exact form of the process, just that it is $Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$ for some μ, ϕ_1. We now need a method to estimate the unknown coefficients μ, ϕ_1 which we now move on to.

1.2 *Estimation of ARMA(p,q) Models*

This section discusses how to estimate an ARMA(p,q) model given a data set $\{y_t\}_{t=1}^{T}$. We focus mostly on the AR(p) as we can use the OLS estimator you have seen before. Similar results holds the same for general ARMA(p,q) model. Though when a model has MA components a different method is used and details are given for the simple case of an MA(1).[11]

[11] Again you are not expected to prove any results below, but to understand the intuition and how to use and interpret the results in practise.

1.2.1 *AR(p) Models*

Say that, on whatever bases, it is considered that an $AR(p)$ may have generated the available data. For example for UK GDP growth by looking at the Correlogram we think an AR(1) may be the true underlying process. The next step then is how to estimate the AR(1) coefficients? Then the investigator uses the sample observations (y_1, y_2, ..., y_T) to estimate the parameters of

$$Y_t = \alpha + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t \tag{1.5}$$

$\varepsilon_t \sim WN(\sigma^2)$. We can see this is just a linear regression model, and using terminology from the ECON21001 course the dependent variable is Y_t and the independent variables are a constant $Y_{t-1}, .., Y_{t-p}$. We could then just perform an OLS regression to estimate $\alpha, \phi_1, .., \phi_p$. It turns out for OLS to correctly estimate these parameters in large samples we require that ε_t is uncorrelated with all past values, i.e that ε_t are serially uncorrelated which we have maintained by the White Noise assumption.

To see how we would perform this regressions in practise we can see in the following extract from an Excel worksheet:

t	y(t)	y(t-1)	y(t-2)	y(t-3)
1	55	NA	NA	NA
2	63	55	NA	NA
3	85	63	55	NA
4	79	85	63	55
5	66	79	85	63

where the first observations are 'lost" at the beginning of the sample in order to create the required lags; hence some lagged values are not available (indicated here as NA). For example, to estimate an $AR(2)$, then the initial observations for $t = 1, 2$ cannot be used in the $AR(2)$ regression as there are no corresponding observations available for y_{t-2}

Programs typically make this adjustment automatically. However, when comparing different models estimated for a single variable, it is good practice to ensure that all are estimated over the same sample period. The reason being we want to compare the models over the same time period. Thus, if you want to compare $AR(1)$, $AR(2)$, $AR(3)$ and $AR(4)$ models for some y_t, all should be estimated using the observations $t = 5, 6, ..., T$. To allow for such an adjustment, we denote the number of observations actually employed for estimation as $T^* = T - p - q$.

1.2.2 *Properties of OLS*

Let's assume that the investigator estimates an $AR(1)$ where the data were generated by an AR(1). Then the process is

$$Y_t = \alpha + \phi_1 Y_{t-1} + \varepsilon_t \tag{1.6}$$

and the OLS estimators are

$$\hat{\alpha} = \bar{y} - \hat{\phi}_1 \bar{y}_{-1}$$

$$\hat{\phi}_1 = \frac{\sum_{t=2}^{T} (y_t - \bar{y})(y_{t-1} - \bar{y}_{-1})}{\sum_{t=2}^{T} (y_{t-1} - \bar{y}_{-1})^2}$$

where $\bar{y}_{-1}$ is the mean of the observations for y_{t-1}. [Note again that estimation can start only from $t = 2$ as discussed above.] For an AR(2) or higher then we have a multivariate regression and Eviews automatically calculates the regression coefficients.

We can also derive the large sample properties of the OLS Estimator as we did for the Sample ACF.

Theorem 1.2.1: OLS Consistency

If $|\phi_1| < 1$ and $\varepsilon_t \sim WN(\sigma^2)$ then

$$\hat{\phi}_1 \xrightarrow{p} \phi_1$$

Theorem 1.2.2: OLS Central Limit Theorem

f $|\phi_1| < 1$ and $\varepsilon_t \sim WN(\sigma^2)$ then

$$\sqrt{T}\frac{(\hat{\phi}_1 - \phi_1)}{\text{sd}(\hat{\phi}_1)} \xrightarrow{d} N(0,1) \tag{1.7}$$

Crucial Assumptions

1. ε_t **Serially Uncorrelated** (i.e ε_t uncorrelated with all past values). If errors are serially correlated then the OLS assumption of exogeneity fails. For example if ε_t is correlated with ε_{t-1} then ε_t is correlated with Y_{t-1} as $Y_{t-1} = \mu + \phi_1 Y_{t-2} + \varepsilon_{t-1}$ where in this case Y_{t-1} is our dependent variable.
2. $|\phi_1| < 1$ **Stationarity**. IIF $|\phi_1| = 1$ (unit root) then the variance of Y_t is infinite and the properties of the OLS estimator are non-standard (covered in more detail in the final lecture.)

All these arguments for estimation in the $AR(1)$ extend to an $AR(p)$ with $p > 1$. There is one further comment, however. To apply a joint hypothesis test in the $AR(p)$, such as

$$\begin{aligned} H_0 &: \phi_1 = ... = \phi_p = 0 \\ H_A &: \text{at least one } \phi_i \neq 0,\ i = 1, ..., p \end{aligned} \tag{1.8}$$

we can perform an F-test of this joint hypothesis using the techniques from ECON21001. again noting that this test will work well for large sample sizes. Again you are not expected to prove any of this but to understand these tests are approximate. See your second notes on joint hypothesis tests on OLS to refresh your memory further.

All that can be said from a theoretical perspective is that a Central Limit Theorem result analogous to (**??**) allows a "large sample" test statistic to be derived (the F-Statistics) which has n F distribution with for the test of (**??**). This test is calculated and performed in Eviews and an example given below.

1.2.3 MA Models

The material in this subsection is not examinable but is here to give you the intuition on how MA processes are estimated. You are expected to be able to

perform tests on the MA coefficients based on the Eviews output, just as for the AR case but not to prove any of these results.

To highlight the intuition of how MA models in general can be estimated consider the $MA(1)$

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{1.9}$$

where we have set $\mu = 0$ for simplicity. The coefficient θ_1 cannot be estimated using linear regression: however in this case our regressor is ε_{t-1} and unlike Y_{t-1} there are no values are available to use for ε_{t-1} and hence it cannot be treated as a regressor. But there are ways around this problem.

The usual approach to estimation for an MA assumes $\varepsilon_0 = 0$ (corresponding to $t = 0$, the observation prior to the sample period). Also write (**??**) as

$$\varepsilon_t = Y_t - \theta_1 \varepsilon_{t-1}.$$

Now, say you have some initial "guess" for θ_1, let's call it $\widehat{\theta}_1^{(0)}$; this initial "guess" could be $\widehat{\theta}_1^{(0)} = 0$. Now you can recursively calculate a pseudo residual series using this last equation, together with $\varepsilon_0 = 0$, as

$$\begin{aligned}
e_1^{(0)} &= Y_1 \\
e_2^{(0)} &= Y_2 - \widehat{\theta}_1^{(0)} e_1^{(0)} \\
e_3^{(0)} &= Y_3 - \widehat{\theta}_1^{(0)} e_2^{(0)} \\
&\vdots \\
e_T^{(0)} &= Y_T - \widehat{\theta}_1^{(0)} e_{T-1}^{(0)}
\end{aligned}$$

The series $e_t^{(0)}$ then gives a proxy for the unobserved ε_{t-1}. So we discard $\widehat{\theta}_1^{(0)}$ and set up the regression model

$$Y_t = \theta_1 e_{t-1}^{(0)} + \varepsilon_t. \tag{1.10}$$

A new (hopefully better) estimate of θ_1 is obtained by applying OLS to this model. The process can then be repeated: the new estimate $\widehat{\theta}_1^{(1)}$ is used to compute a new residual series $e_t^{(1)}$, which in turn is used in the artificial regression model in place of $e_t^{(0)}$. This is known as iterative estimation: the process stops when two consecutive estimations deliver the same value for $\widehat{\theta}_1$. The idea generalises to any finite order process $MA(q)$, and also to $ARMA$ models.

There are other, more sophisticated, ways to estimate the coefficients of an MA or $ARMA$ model, but virtually all involve iteration. However, this is not a problem for modern computers. Despite requiring iteration for estimation, hypothesis testing is performed for MA (and $ARMA$) models as for AR models. In other words, the estimators are asymptotically normally distributed, with tests applied in practice using the t and F distributions for single and joint tests, respectively.

1.3 Model Specification

The theoretical properties of estimators of the coefficients of an $ARMA$ model assume that the model coincides with the DGP. For example, if an $AR(2)$ model is estimated, then the theoretical properties are for the situation where an $AR(2)$ process did, indeed, generate the data. Therefore, the theory of estimators and hypothesis testing assumes that the investigator knows the nature of the true process, but not its coefficients. In practice, of course, the investigator does not know that an $AR(2)$ generated the data. Hence s/he not only has to estimate the coefficients, but also deduce from the data what model or models should be estimated.

Historically, $ARMA$ model specification was based on the Correlogram, together with the estimated partial autocorrelation function. We do not study the latter because these methods are often difficult to apply in practice and do not clearly indicate the appropriate form of $ARMA$ model. Nevertheless, the Correlogram remains very useful as a summary of the dependence properties of the data.

The modern approach uses what are known as "information criteria". Many such criteria have been proposed, with two in widespread use. The first and oldest is known as the **Akaike Information Criterion**, usually abbreviated to **AIC.**

For the $ARMA\,(p,q)$ model (expressed succinctly in it's Lag Operator Form, see the end of notes for Lecture 2)

$$\phi\,(L)\,T_t = \alpha + \theta(L)\varepsilon_t$$

let $\hat{\sigma}^2$ be the variance estimator for a given choice of p,q so that

$$\hat{\sigma}^2 = \frac{RSS}{T^*}$$

where RSS is the residual sum of squares from the ARMA(p,q) regression and T^* is the number of observations used for estimating the model. Then, AIC is defined by

$$AIC = \log\hat{\sigma}^2 + \frac{2\,(p+q+1)}{T^*} \tag{1.11}$$

in which $p+q+1$ is the number of coefficients estimated, including the intercept. If no intercept is included, $p+q$ is used rather than $p+q+1$. Obviously, if no AR is estimated, $p=0$ and similarly $q=0$ if no MA coefficient is estimated. The $ARMA(0,0)$ model with intercept only has $p=q=0$.

The AIC it turns out gives a measure of how well a model fits to the data. The smaller the AIC roughly speaking the better the model fits and hence the more likely the true process takes that form. Hence we calculate AIC for a range of values, such as $0 \leqslant p \leqslant p_{\max}$ and $0 \leqslant q \leqslant q_{\max}$ and then choose the $(p,\,q)$ which gives the minimum value of AIC. This still requires the maximum lags, $p_{\max}$ and $q_{\max}$ to be specified. In practice, a subset of models may be considered based on information in the Correlogram For example, for UK

GDP growth, the declining pattern in the Correlogram appears to indicate the presence of an least an $AR(1)$ term, though it it not evident whether there is any MA.

In terms of Eviews output, there is an alternative, equivalent, definition in terms of a quantity called the "log-likelihood function". We will not define it, except to note that Eviews provides this automatically in its regression output as the "Log Likelihood". Denote by L this log likelihood. Then, the AIC is computed by Eviews as

$$AIC = -2\left(\frac{L}{T^*}\right) + \frac{2(p+q+1)}{T^*}.$$

Selecting the model that minimises this expression is equivalent to minimising (**??**).

An alternative is the **Schwarz Information Criterion**, abbreviated **SIC**[12]**:**

$$SIC = \log \hat{\sigma}^2 + \frac{(p+q+1)\log T^*}{T^*} \tag{1.12}$$

[12] This criterion is sometimes also known as the **Bayesian Information Criterion**, or BIC.

for which there is also an equivalent expression in terms of the log likelihood (and used by Eviews). As for AIC, a model is chosen by selecting the orders p and q that minimise SIC.

It can be seen from (**??**) and (**??**) that both AIC and SIC trade-off goodness of fit (measures by $\hat{\sigma}^2$) against the number of coefficients estimated in relation to the sample size (measured by $p+q+1$), although the latter is weighted differently by the two criteria. In general, increasing the model order improves goodness of fit, so the criteria ask whether the better fit (smaller $\hat{\sigma}^2$) from increasing p and/or q is outweighed by the additional parameters. Remember, that better fit is obtained even when the additional parameters are truly equal to zero. Because the criteria are different, there is no guarantee that the two criteria will select the same model orders as being appropriate.

The relative weight given to the number of coefficients estimated is 2 for AIC and $\log T^*$ for SIC; the latter is greater that the former unless T^* is very small. A consequence of this heavier "cost" that SIC places on estimating coefficients is that, if we are selecting the order for an AR model, then

$$p_{AIC} \geq p_{SIC} \tag{1.13}$$

and a similar relation applies for pure MA models. While they can imply the same p, in practice AIC often chooses a model with more coefficients than does SIC. Keeping with the AR case, and assuming that the true process is an $AR(p_0)$, it can also be shown that, provided $p_{\max} \geq p_0$

$$\begin{aligned} \text{as } T^* &\to \infty, p_{SIC} \to p_0 \\ \text{as } T^* &\to \infty, p_{AIC} \geq p_0 \end{aligned}$$

Thus, while SIC will asymptotically select the true p_0, AIC can (even asymptotically) over-specify the order. For this reason, some people prefer to use SIC.

Nevertheless, it is also worth noting that in practice we do not have an "infinite" sample size available, and hence it is debatable how much weight should be given to the asymptotic result that p_{AIC} can over-specify the true p_0. Practitioners have found both criteria to work well in practice, and it is sensible to combine their use with hypothesis tests to check whether coefficients are significant and also with hypothesis testing for the presence of serial correlation in the residuals, to which we turn next.

1.4 Testing for Serial Correlation

1.4.1 Residual Serial Correlation

If the disturbances of a time series model are correlated, and hence are not white noise, then they are **serially correlated**. It is good practice to always apply a test for serial correlation after estimation of a time series model. If the model does not account for all the dependence in the data, hence leaving serial correlation in the residuals, then a higher order model is required.

To see this, consider an $AR(p)$ model with where the disturbances are $AR(m)$ rather than white noise:

$$\phi(L)Y_t = \alpha + u_t \tag{1.14}$$

with

$$\psi(L)u_t = \varepsilon_t \tag{1.15}$$

where $\psi(L) = 1 - \psi_1 L - ... - \psi_m L^m$ and ε_t is white noise. Substituting $u_t = \psi(L)^{-1}\varepsilon_t$ from (**??**) into (**??**) yields

$$\psi(L)\phi(L)Y_t = \psi(1)\alpha + \varepsilon_t$$

as the correct form of AR process for y_t. In other words, (**??**) implies that the correct process for Y_t is an $AR(p+m)$, not $AR(p)$. Clearly, if the disturbances u_t in (**??**) are an $AR(m)$ process, the white noise disturbance assumption is wrong. In this case, the estimated model is *misspecified* and a more general one, the higher order $AR(p+m)$, should be used.

Analogous considerations apply if u_t is an $MA(q)$, with

$$u_t = \theta(L)\varepsilon_t$$

then the appropriate model is $ARMA(p,q)$

$$\phi(L)Y_t = \alpha + \theta(L)\varepsilon_t.$$

In practice, however, it may be unclear whether the additional terms required should be of the AR or MA form.

The same considerations apply also in an MA or $ARMA$ model. That is, the presence of serial correlation in the residuals implies that a higher order model is required.

1.4.2 *Diagnostic Test*

The test that ε_t are not autocorrelated, against the possibility that they follow the $AR(m)$ of (**??**), can be carried out by estimating the auxiliary regression

$$y_t = \alpha + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \psi_1 e_{t-1} + ... + \psi_m e_{t-m} + \varepsilon_t, \qquad (1.16)$$

where e_t are the residuals obtained from estimating (**??**), and testing the joint null hypothesis

$$H_0 : \psi_1 = ... = \psi_m = 0$$

against the alternative

$$H_A : \text{ at least one } \psi_j \neq 0, j = 1, ..., m.$$

Strictly speaking, this test is only valid asymptotically, when the corresponding test statistic is asymptotically distributed as χ^2_m under the null hypothesis. However, although they are asymptotically equivalent, it has been shown that, in practice, an F-test for $\psi_j = 0$ ($j = 1, ..., m$) in (**??**) performs better than a χ^2 test.

Serial Correlation & Breusch Godfrey LM Test

From an examination of the form of (**??**) you might guess that the test is against the alternative that the disturbances are of an $MA(m)$ form. Indeed, this is the case, in that the test does not discriminate against alternative hypotheses that the disturbances are $AR(m)$ or $MA(m)$. In this sense the test is a general one for the presence of serial correlation, rather than specifically against an AR or MA alternative.

The test just outlined is a particular case of a *Lagrange multiplier* test. Here it is the Lagrange multiplier test for residual autocorrelation of order m in an $AR(p)$ model. However, the same principle can be used also to test for residual serial correlation in an estimated MA or $ARMA$ model. This test is often called the *Breusch-Godfrey test*, after the econometricians (Trevor Breusch and Les Godfrey) who independently developed this test in the 1970s.

The usual guideline is that the order m used for the test should be 4 for

quarterly data and 12 for monthly data, with the intention of testing for any remaining seasonality in the residuals. For annual data, $m = 1$ or 2 might be used, depending on the available size.

One important implication of the discussion is that the Durbin-Watson test for first order autocorrelation is NOT VALID in the context of an $ARMA$ process, because it does not take account of any relationship between the regressors and the lagged disturbances. However, by including the AR terms, in addition to the lagged residuals, (**??**) does take account of this. In practice, the Durbin-Watson test is effectively useless for checking serial correlation in a time series model and **should not be used**, or its value reported - despite the fact that it is automatically computed by many regression packages, including Eviews.

1.5 *Empirical Example*

For the GDP growth data over 1984Q1 to 2012Q4, say the investigator considers (for some reason) that an $MA(2)$ model may be appropriate. Estimating this model in Eviews gives

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.599462	0.097589	6.142706	0.0000
MA(1)	0.432890	0.088259	4.904768	0.0000
MA(2)	0.343070	0.089958	3.813683	0.0002

R-squared	0.314704	Mean dependent var	0.601666
Adjusted R-squared	0.302575	S.D. dependent var	0.710753
S.E. of regression	0.593564	Akaike info criterion	1.820178
Sum squared resid	39.81196	Schwarz criterion	1.891392
Log likelihood	-102.5703	Hannan-Quinn criter.	1.849087
F-statistic	25.94618	Durbin-Watson stat	1.796944
Prob(F-statistic)	0.000000		

and both estimated MA coefficients, together with the intercept, are highly significant. A test for serial correlation to lag $m = 4$ however yields

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	2.831882	Prob. F(4,109)	0.0280
Obs*R-squared	10.91930	Prob. Chi-Square(4)	0.0275

so that significant residual serial correlation is present at the 5% level.

The pattern of declining sample autocorrelations at short lags in the Correlogram suggests an $AR(1)$ may be appropriate for this GDP growth data. Estimation gives

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.238187	0.071080	3.350964	0.0011
DLGDP(-1)	0.591083	0.075976	7.779879	0.0000

R-squared	0.346804	Mean dependent var	0.601666
Adjusted R-squared	0.341074	S.D. dependent var	0.710753
S.E. of regression	0.576949	Akaike info criterion	1.754964
Sum squared resid	37.94714	Schwarz criterion	1.802439
Log likelihood	-99.78790	Hannan-Quinn criter.	1.774236
F-statistic	60.52652	Durbin-Watson stat	2.106444
Prob(F-statistic)	0.000000		

with both the constant and estimated $AR(1)$ coefficient being highly significant. The autocorrelation test for this model gives

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.639092	Prob. F(4,110)	0.6357
Obs*R-squared	2.634579	Prob. Chi-Square(4)	0.6207

which provides no evidence of significant residual serial correlation.

It is clear that the $AR(1)$ specification is preferable to the $MA(2)$: both AIC and SIC would select the $AR(1)$ model, which also has no evidence of residual serial correlation. The $MA(2)$ is not satisfactory in showing significant serial correlation in the residuals, suggesting further dynamics not being captured by the MA(2) model that the AR(1) seems to capture.