Lecture Notes on Univariate Time Series Analysis and Box Jenkins Forecasting

John Frain
Economic Analysis, Research and Publications

April 1992
(reprinted with revisions)
January 1999

Abstract

These are the notes of lectures on univariate time series analysis and Box Jenkins forecasting given in April, 1992. The notes do not contain any practical forecasting examples as these are well covered in several of the textbooks listed in Appendix A. Their emphasis is on the intuition and the theory of the Box-Jenkins methodology. These and the algebra involved are set out in greater detail here than in the more advanced textbooks. The notes, thus may serve as an introduction to these texts and make their contents more accessible.

The notes were originally prepared with the scientific word processor Chiwriter which is no longer in general use. The reprinted version was prepared with the LaTeX version of Donald Knuth's TeX mathematical typesetting system. Some version of TeX is now the obligatory standard for submission of articles to many mathematical and scientific journals. While MS WORD is currently acceptable to many economic journals TeX has been requested and is sometimes very much preferred. Many books are now prepared with TeX. TeX is also a standard method for preparing mathematical material for the internet. TeX is free and the only significant cost of using it is that of learning how to use it.

It is often held that TeX systems are to difficult to use. On the other hand, it would have been impossible to produce this document in, for example, WORD 6.0a and WINDOWS 3.1x. I would not suggest that TeX be used for ordinary office work. A standard WYSIWYG word processor such as WORD would complete this work much better. For preparing material such as these notes TeX is better and should be considered.

An implementation of TeX for Windows is available from me on diskettes. TeX and LaTeX are freeware. A distribution (gTeX) is available from me on request. I can also provide some printed installation instructions if anyone wishes to install it on their own computer. While gTeX is designed to work with Windows its installation and operation requires some knowledge of MS/DOS. I am not in a position to support any TeX installation. For a knowledge of LaTeX please see Lapont (1994), "LaTeX document preparation system - User's Guide and Reference Manual", Addison-Wesley Publishing Company, ISBN 0-201-52983-1.

Contents

Chapter 1

Introduction

Univariate time series

Forecasting or seeing the future has always been popular. The ancient Greeks and Romans had their priests examine the entrails to determine the likely outcome of a battle before they attacked. To-day, I hope, entrails are not used to any extent in forecasting. Rather scientific forecasts are based on sound (economic) theory and statistical methods. Many people have mixed opinions about the value of scientific forecasting as they may have often found that such forecasts are often wrong.

This opinion is due to a basic missunderstanding of the nature of scientific forecasting. Scientific forecasting can achieve two ends

- provide a likely or expected value for some outcome – say the value of the CPI at some point in the future
- reduce the uncertainty about the range of values that may result from a future event

The essence of any risky decision is that one can not know with certainty what the result of the decision will be. Risk is basically a lack of knowledge about the future. With perfect foresight there is no risk. Scientific forecasting increases our knowledge of the future and thus reduces risk. Forecasting can not and will never remove all risk. One may purchase insurance or even financial derivatives to hedge or remove ones own risk, at a cost. This action is only a transfer of risk from one person or agency to another who is willing to bear the risk for reward.

Forecasting and economic modelling are one aspect of risk assessment. It relies on what can be learned from the past. The problem is that relying solely on the past will cause problems if the future contains events that are not similar to those that occurred in the past. Could events such as the October 1987 stock market crash, the 1982/3 ERM crisis, the far-east and Russian problems of 1998 have been predicted, in advance, from history. A minority of prophets may have predicted them in advance – some through luck and perhaps others

through genuine insight, but to the majority they were unexpected. The failure to predict such events should not be seen as a failure of forecasting methodology. One of the major assumptions behind any forecast is that no unlikely disaster will occur during the period of the forecast.

This does not imply that policy makers should not take possible disasters in deciding on policy. On the contrary, they should examine and make contingency plans where appropriate. This type of analysis is known as scenario analysis. For this type of analyses one selects a series of scenarios corresponding to various disasters and examines the effect of each scenario on the economy. This is a form of disaster planning. One then evaluates the likelihood of the scenario and its effects and sees what steps can be taken to mitigate the disaster. The analysis of scenarios is a much more difficult problem than univariate time series modelling. For an economy, scenario analysis will require extentions to an econometric model or a large computable general equilibrium model. Such procedures requires considerable resources and their implementation involves technical analyses beyond the scope of these notes. This does not take from the the effectiveness of a properly implemented univariate forecasting methodology which is valuable on its own account.

On the topic of scenario analysis one may ask what kind of disasters we should consider for scenario analysis. I can think of many disasters that might hit the financial system. For a central bank to consider many of these might give rise to a suspicion that the central bank thought that such a disaster might occur. There will always be concern in such cases that this may lead to stresses in the financial system. There is a problem here that is bound up with central bank credibility.

These notes are not intended as a full course in univariate time-series analysis. I have not included any practical forecasting examples. Many of the books in the annotated bibliography provide numerous practical examples of the use of univariate forecasting. Other books listed there provide all the theory that is required but at an advanced level. My emphasis is more on the intuition behind the theory. The algebra is given in more detail than in the theoretical texts. Some may find the number of equations somewhat offputting but this is the consequence of including more detail. A little extra effort will mean that the more advanced books will be more accessible.

These notes and a thorough knowledge of the material in the books in the references are no substitute for practical forecasting experience. The good forecaster will have considerable practical experience with actual data and actual forecasting. Likewise a knowledge of the data without the requisite statistical knowledge is a recipe for future problems. Anyone can forecast well in times of calm. The good forecaster must also be able to predict storms and turning points and this is more difficult.

When a forecast turns out bad one must find out why. This is not an exercise aimed to attach blame to the forecaster. An unfullfilled forecast may be an early warning of an event such as a downturn in the economy. It may indicate tha some structural change has taken place. There may be a large number of perfectly valid reasons why a forecast did not turn out true. It is important that these reasons be determined and acted on.

An unfullfilled forecast may be very good news. If the original forecast was for trouble ahead and persuaded the powers that be to take remedial policy action. If the policy changes produced a favourable outcome then one would appreciate the early warning provided by the forecast. In effect policy changes may invalidate many forecasts. In particular all forecasts not based on structural models are not robust with respect to policy changes. The construction of structural models which are invariant with respect to policy changes is an order of magnitude more difficult than building univariate forecasts

These notes deal with the forecasting and analysis of univariate time series. We look at an individual time series to fin out how an observations at one particular time is related to those at other times. In particular we would like to determine how a future value of the series is related to past values. It might appear that we are not making good use of available information by ignoring other time series which might be related to the series of interest. To some extent the gains from the rich dynamic structures that can be modelled in an univariate system outweigh the costs of working with more complicated multivariate systems. If sufficiend data are available recent reduction in the cost of and other advances in computer hardware/software have made some multivariate systems a practical possibility. Structural multivariate macroeconometric models may have better long-run properties but their poorer dynamic properties may result in poorer short-run forecasts.

Practical experience has shown that the analysis of individual series in this way often gives very good results. Statistical theory has shown that the method is often better than one would expect, at first sight. The methods described here have been been applied to analysis and forecasting such diverse series as:

- Telephone installations
- Company sales
- International Airline Passenger sales
- Sunspot numbers
- IBM common stock prices
- Money Demand
- Unemployment
- Housing starts
- etc. ...

The progress of these notes is as follows. Chapter 2 deals with the statistical properties of univariate time series. I include an account of the most common stationary (white noise, AR, MA, $ARMA$) processes, their autocorrelations, and impulse response functions. I then deal with integrated processes and tests for non-stationarity. Chapter 3 uses the theory set out in the previous chapter to explain the identification, estimation, forecasting cycle that is involved in the seasonal and non-seasonal Box-Jenkins methodology. Chapter 4 reviews a selection of software that has been used in the Bank for this type of work. The

exclusion of any item of software from this list is not to be taken as an indication of its relative value. It has been excluded simply because I have not used it. If any producer of econometric software for PCs feels that his software is superior and would like me to include an account of it in a future version of these notes I would be glad to receive an evaluation copy and time permitting I will include an account of it in the next version of these notes.

Chapter 2

Theory of Univariate Time Series

2.1 Basic Definitions

We start with some basic definitions. The elements of our time series are denoted by

$$X_1, X_2, \ldots, X_t, \ldots$$

The mean and variance of the observation at time t are given by

$$\begin{aligned} \mu_t &= E[X_t] \\ \sigma_t^2 &= E[(X_t - \mu_t)^2] \end{aligned}$$

respectively and the covariance of X_t, X_s by

$$cov(X_t, X_s) = E[(X_t - \mu_t)(X_s - \mu_s)] = \lambda_{ts}$$

In this system there is obviously too little information to estimate μ_t, σ_t^2, and λ_{ts} as we only have one observation for each time period. To proceed we need two properties — stationarity and ergodicity.

A series is second order stationary if:

$$\begin{aligned} \mu_t &= \mu, \quad t = 1, 2, \ldots \\ \sigma_t^2 &= \sigma^2, \quad t = 1, 2, \ldots \\ \lambda_{t,s} &= \lambda_{t-s}, \quad t \neq s, \ldots \end{aligned}$$

i.e. the mean, variance and covariances are independent of time.

A series is strictly stationary if the joint distribution of $(X_1, X_2, \ldots, X_t)$ is the same as that of $(X_{1+\tau}, X_{2+\tau}, \ldots, X_{t+\tau})$ for all t and τ. If a series has a multivariate normal distribution then second order stationarity implies strict stationarity. Strict stationarity implies second order stationarity if the mean

and variance exist and are finite. Be warned that text books have not adopted a uniform nomenclature for the various types of stationarity

In a sense we would like all our series to be stationary. In the real world this is not possible as much of the real world is subject to fundamental changes. For a nonstationary series we may try to proceed in the following way:

- Find a transformation or some operation that makes the series stationary
- estimate parameters
- reverse the transformation or operation.

This use of a single measurement at each time to estimate values of the unknown parameters is only valid if the process is *ergodic*. Ergodicity is a mathematical concept. In essence it means that observations which are sufficiently far apart in time are uncorrelated so that adding new observations gives extra information. We assume that all series under consideration have this property.

We often use autocorrelations rather than covariances. The autocorrelation at lag τ, ρ_τ is defined as:

$$\rho_\tau = \frac{\lambda_{t,t+\tau}}{\lambda_0} = \frac{\lambda_\tau}{\lambda_0} = \frac{E[(X_t - \mu)((X_{t+\tau} - \mu)}{E[(X_t - \mu)((X_t - \mu)]}$$

A plot of ρ_τ against τ is know as the autocorrelogram or auto-correlation function and is often a good guide to the properties of the series. In summary second order stationarity implies that mean, variance and the autocorrelogram are independent of time.

Examples of Time series Processes

2.1.1 Normal (Gaussian) White Noise

If ε_t are independent normally distributed random variables with zero mean and variance σ_ε^2 then it is said to be Normal (Gaussian) White Noise.

$$\begin{aligned} \mu &= E[\varepsilon_t] \\ &= 0 \\ Var[\varepsilon_t] &= \sigma_\varepsilon^2 \\ \rho_0 &= 1 \\ \rho_\tau &= E[\varepsilon_t \varepsilon_{t+\tau}] \\ &= 0 \text{ if } \tau \neq 0 \text{ (independence)} \end{aligned}$$

Normal White Noise is second order stationary as its mean variance and autocorrelations are independent of time. Because it is also normal it is also strictly stationary.

2.1.2 White Noise

The term *white noise* was originally an engineering term and there are subtle, but important differences in the way it is defined in various econometric texts. Here we define white noise as a series of un-correlated random variables with zero mean and uniform variance ($\sigma^2 > 0$). If it is necessary to make the stronger assumptions of independence or normality this will be made clear in the context and we will refer to independent white noise or normal or Gaussian white noise. Be careful of various definitions and of terms like weak, strong and strict white noise

The argument above for second order stationarity of Normal white noise follows for white noise. White noise need not be strictly stationary.

2.1.3 $AR(1)$ Process

Let ε_t be White Noise. X_t is an $AR(1)$ Process if

$$\begin{aligned}
X_t &= \alpha X_{t-1} + \varepsilon_t \quad (|\alpha| < 1) \\
X_t &= \varepsilon_t + \alpha(\alpha X_{t-2} + \varepsilon_{t-1}) \\
&= \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2 X_{t-2} \\
&= \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2(\alpha X_{t-3} + \varepsilon_{t-2}) \\
&= \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2)} + \alpha^3 X_{t-3} \\
&\quad \cdots \ \cdots \\
&= \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \alpha_{t-3}^3\varepsilon_{t-3} + \ldots \\
E[X_t] &= 0 \\
&\quad \text{(which is independent of } t\text{)}
\end{aligned}$$

Autocovariance is given by

$$\begin{aligned}
\lambda_k &= E[X_t X_{t+k}] \\
&= E\left[\sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i}\right]\left[\sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i}\right] \\
&= \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} \sigma_\varepsilon^2 \\
&= \alpha^{k+i}\sigma_\varepsilon^2 \sum_{i=0}^{\inf} \alpha^{2i} \\
&= \sigma_\varepsilon^2 \frac{\alpha^k}{1-\alpha^2}\rho_k \\
\rho_k &= \frac{\gamma_k}{\gamma_0} \\
&= \alpha^k k = 0, 1, \ldots \\
&= \alpha^{|k|} k = 0, \pm 1, \pm 2, \ldots \\
&\quad \text{(which is independent of } t\text{)}
\end{aligned}$$

We have shown that an $AR(1)$ Process is stationary. As an exercise you should now draw the autocorrelogram of a white noise and several AR processes and note how these change for values of α between -1 and 1

In most of the theoretical models, that we describe, we have excluded an intercept for ease of exposition. Including an intercept makes the expected value of the series non-zero but otherwise it does not effect our results.

Note that for our $AR(1)$ Process process we included a stipulation that $\alpha < 1$. This is required in order that various infinite series converge. If we allowed $\alpha \geq 1$ sums would diverge and the series would not be stationary.

2.1.4 Random Walk

We now consider the case $\alpha = 1$. Again ε_t is white noise. X_t is a random walk if

$$X_t = X_{t-1} + \varepsilon_t$$

There is a sense that errors or shocks in this model persist. Confirm this as follows. Let the process start at time $t = 0$ with $X_0 = 0$. By substitution:

$$X_t = \varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_1 + X_0$$

Clearly the effect of past ε's remain in X_t.

$$E[X_t] = X_0$$

but

$$var[X_t] = t\sigma_\varepsilon^2$$

Therefore the series is not stationary. as the variance is not constant but increases with t.

2.2 Lag Operators - Notation

Let $X_1, \ldots, X_t$ be a time series. We define the lag operator L by:

$$LX_t = X_{t-1}$$

if

$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p$$

An $AR(p)$ process is defined as

$$X_t = \alpha_1 x_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t$$

where ε_t is white noise. In terms of the lag operator this may be written:

$$X_t = \alpha_1 L x_t + \alpha_2 L^2 X_t + \cdots + \alpha_p L^p X_t + \varepsilon_t$$

$$(1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_p L^p) X_t = \varepsilon_t$$

$$\alpha(L) X_t = \varepsilon_t$$

The lag operator is manipulated using the ordinary rules of algebra. Further information on the lag operator is available in the references quoted at the end of these notes and in particular in Dhrymes(1976)

In terms of the lag operator the $AR(1)$ process may be written:

$$\begin{aligned}
(1 - \alpha L) X_t &= \varepsilon_t, \quad |\alpha| < 1 \\
X_t &= \left(\frac{1}{1 - \alpha L} \right) \varepsilon_t \\
&= (1 + \alpha^1 L + \alpha^2 L^2 + \cdots - \varepsilon_t \\
&= \varepsilon_t + \alpha \varepsilon_{t-1} + \ldots \\
&\quad \text{as before}
\end{aligned}$$

2.3 AR(2) Process

The $AR(2)$ process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

may be written in terms of the lag operator as

$$(1 - \phi_1 L - \phi_2 L^2) X_t = \varepsilon_t$$

We may write the process as

$$\begin{aligned}
X_t &= \psi(L) \varepsilon_t \\
&= (1 + \psi_1 L + \psi_2 L^2 + \ldots) \varepsilon_t \\
&\quad where \\
(1 - \phi_1 L - \phi_2 L^2)^{-1} &\equiv (1 + \psi_1 L + \psi_2 L^2 + \ldots)
\end{aligned}$$

or equivalently

$$(1 - \phi_1 L - \phi_2 L^2)(1 + \psi_1 L + \psi_2 L^2 + \ldots) \equiv 1$$

Equating coefficients we get:

$$\begin{aligned}
L^1 &: \quad -\phi_1 + \psi_1 = 0 && \psi_1 = \phi_1 \\
L^2 &: \quad -\phi_2 + \phi_1 \psi_1 + \psi_2 = 0 && \psi_2 = \phi_1^2 + \phi_2 \\
L^3 &: \quad -\phi_2 + \phi_1 \psi_1 + \psi_2 = 0 && \psi_3 = \phi_1^3 + 2\phi_1 \phi_2 \\
&\quad \cdots \\
L^j &: \quad \psi_j = \phi_1 \psi_{j-1} + \psi_2 \psi_{j-2}
\end{aligned}$$

and all weights can be determined recursively.

The $AR(1)$ process was stationary if $|\alpha| < 1$. What conditions should we impose on the $AR(2)$ process

$$(1 - \phi_1 L - \phi_2 L^2) X_t = \varepsilon_t$$

in order that it be stationary? Consider the reciprocals (say g_1 and g_2) of the roots of

$$(1 - \phi_1 L - \phi_2 L^2) = 0$$

Then the equation may be written

$$(1 - g_1 L)(1 - g_2 L) = 0$$

The process is stationary if $|g_1| < 1$ and $|g_2| < 1$. These roots may be real or complex. (It is usually said that $|g_1|^{-1}$ and $|g_2|^{-1}$ lie outside the unit circle). These restrictions impose the following conditions on ϕ_1 and ϕ_2.

$$\begin{aligned} \phi_1 + \phi_2 &< 1 \\ -\phi_1 + \phi_2 &< 1 \\ -1 < \phi_2 &< 1 \end{aligned}$$

The ACF (autocorrelation function) of a stationary $AR(2)$ process may be derived as follows: Multiply the basic equation

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \varepsilon_t$$

by X_{t-k} and take expectations

$$E[X_t X_{t-k}] - \phi_1 E[X_{t-1} X_{t-k}] - \phi_2 E[\varepsilon_t X_{t-k}] = E[X_{t-k} \varepsilon_t]$$

$$\gamma_k - \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} = E[\varepsilon_t X_{t-k}]$$

$$E[X_{t-2} X_{t-k}] = \begin{cases} \sigma_\varepsilon^2 & for \quad k = 0 \\ 0 & for \quad k = 1, 2, \ldots \end{cases}$$

$$\gamma_0 - \phi_1 \gamma_{-1} - \phi_2 \gamma_{-2} = \sigma_\varepsilon^2 = \gamma_0 - \phi_1 \gamma_1 - \phi_2 \gamma_2$$

$$\gamma_k - \phi_1 \gamma_{k-1} - \phi_2 \gamma_{k-2} = 0 \quad k = 1, 2 \ldots$$

or in terms of autocorrelations,

$$\rho_k - \phi_1 \rho_{k-1} - \phi_2 \rho_{k-2} = 0 \quad k = 1, 2 \ldots$$

The observant reader will notice that the autocorrelations obey the same difference equation as the time series apart from the missing random term (the

corresponding homogeneous difference equation) and the initial conditions ($\rho_0 = 1$, $\rho_{-1} = \rho_1$) We can solve this problem by direct substitution.

For k = 1

$$\rho_1 - \phi_1\rho_0 - \phi_2\rho_{-1} = 0$$

$$\rho_0 = 1$$

$$\rho_1 = \rho_{-1}$$

$$\rho_1 = \frac{\phi_1}{1-\phi_2}$$

For k = 2

$$\rho_2 = \phi_1\rho_1 + \phi_2\rho_0 = \frac{\phi_1^2}{1-\phi_2} + \phi_2$$

and all other values may be derived from the recursion and may be seen to be time independent.

We now work out the variance of an $AR(2)$ system.

Put $k = 0$ in recursion for covariances:

$$\gamma_0 - \phi_1\gamma_{-1} - \phi_2\gamma_{-2} = \sigma_\varepsilon^2$$

$$\gamma_0(1 - \phi_1\rho_1 - \phi_2\rho_2) = \sigma_\varepsilon^2$$

$$\gamma_0\left(1 - \frac{\phi_1^2}{1-\phi_2} - \frac{\phi_1^2\phi_2}{1-\phi_2} - \phi_2^2\right) = \sigma_\varepsilon^2$$

$$\gamma_0\left(\frac{1-\phi_2-\phi_1^2-\phi_1^2\phi_2-\phi_1^2(1-\phi_2)}{1-\phi_2}\right) = \sigma_\varepsilon^2$$

$$\gamma_0\left((1+\phi_2)\left(1-\phi_2-\phi_2^2\right) - \phi_2\left(1+\phi_2\right) - \phi_1^2\left(1+\phi_2\right)\right) = \left(1-\phi_2\right)\sigma_\varepsilon^2$$

$$\gamma_0\left(1 - 2\phi_2 + \phi_2^2 - \phi_1^2\right) = \frac{1-\phi_2}{1+\phi_2}\sigma_\varepsilon^2$$

$$\gamma_0 = \frac{1-\phi_2}{1+\phi_2}\frac{\sigma_\varepsilon^2}{\left(1-\phi_2-\phi_1\right)\left(1-\phi_1+\phi_1\right)}$$

which is independent of t. The conditions on g_1 and g_2, given earlier, ensure that $0 < \gamma_0 < \infty$.

Thus an $AR\left(2\right)$ process is stationary.

The properties of the Autocorrelation function may be derived from the general solution of the difference equation

$$\rho_k - \phi_1\rho_{k-1} - \phi_2\rho_{k-2} = 0$$

which is of the form

$$\rho_k = Ag_1^k + Bg_2^k$$

where A and B are constants determined by initial conditions $\rho_0 = 1$ and $\rho_{-1} = \rho_{+1}$. If g_1 and g_2 are real the autocorrelogram is a mixture of two damped exponentials (i.e. both die out exponentially). This is similar to a weighted sum of two $AR(1)$ processes.

If g_1 and g_2 are complex the ACF is a damped sine wave.

If $g_1 = g_2$ the general solution is given by

$$\rho_k = (A_1 + A_2 k)\, g^k$$

2.4 AR(p) Process

An $AR(p)$ process is defined by one of the following expressions

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = \varepsilon_t$$

or

$$(1 - \phi_1 L - \cdots - \phi_p L^p)\, x_t = \varepsilon_t$$

or

$$\Phi(L)\, x_t = \varepsilon_t$$

where

$$\Phi(L) = 1 - \phi_1 L - \ldots\ldots - \phi_p L^p$$

For an $AR(p)$ process the stationarity conditions may be set out as follows: Write

$$\Phi(L) = (1 - g_1 L)(1 - g_2 L) \ldots (1 - g_p L)$$

Stationarity conditions require

$$|g_i| < 1 \quad \text{for } i = 1 \ldots p$$

or alternatively

g_i^{-1} all lie outside the unit circle.

We may derive variances and correlations using a similar but more complicated version of the analysis of an $AR(2)$ process. The autocorrelations will follow a difference equation of the form

$$\Phi(L)\rho_k = 0 \quad k = 1, \ldots$$

This has a solution of the form

$$\rho_k = A_1 g_1^k + A_2 g_2^k + \cdots + A_p g_p^k$$

The ACF is a mixture of damped exponential and sine terms. These will in general die out exponentially.

2.5 Partial Autocorrelation Function PACF

Considering all orders of AR processes, eventually, die out exponentially is there any way we can identify the order of the process. To do this we need a new concept—the Partial Autocorrelation function.

Consider the autocorrelation at lag 2. Observation 1 effects observation 2. Observation 1 affects observation 3 through two channels, i.e. directly and indirectly through its effect on observation 2 and observations $2's$ effect on observation 3. The autocorrelation measures both effects. The partial autocorrelation measures only the direct effect.

In the case of the k^{th} order the correlation between x_t and x_{t-k} can in part be due to the correlation these observations have with the intervening lags x_{t-1}, x_{t-2}, $\ldots..x_{t-k+1}$. To adjust for this correlation the partial autocorrelations are calculated.

We may set out this procedure as follows -

Estimate the following sequence of models

$$\begin{aligned}
x_t &= a_{11}x_{t-1} + \varepsilon_1 \\
x_t &= a_{21}x_{t-1} + a_{22}x_{t-2} + \varepsilon_2 \\
x_t &= a_{31}x_{t-1} + a_{32}x_{t-2} + a_{33}x_{t-3} + \varepsilon_3 \\
&\cdots \\
x_t &= a_{k1}x_{t-1} + \cdots + a_{kk}x_{t-k} + \varepsilon_k
\end{aligned}$$

The sequence a_{11}, a_{22}, a_{33}, $\ldots$, a_{kk}, $\ldots$ are the partial autocorrelations. In practice they are not derived in this manner but from the autocorrelations as follows.

Multiply the final equation above by x_{t-k}, take expectations and divide by the variance of x. Do the same operation with x_{t-1}, x_{t-2}, $x_{t-3} \ldots x_{t-k}$ successively to get the following set of k equations (Yule-Walker).

$$\begin{aligned}
\rho_1 &= a_{k1} + a_{k2}\rho_1 + \ldots a_{kk}\rho_{k-1} \\
\rho_2 &= a_{k1}\rho_1 + a_{k2} + \ldots a_{kk}\rho_{k-2} \\
&\cdots \\
\rho_k &= a_{k1}\rho_{k-1} + a_{k2}\rho_{k-2} + a_{kk}
\end{aligned}$$

Use Cramer's rule to solve for a_{kk} to get

$$a_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

It follows from the definition of a_{kk} that the partial autocorrelations of autoregressive processes have a particular form.

$AR(1)$	$a_{11} = \rho_1 = \alpha$			$a_{kk} = 0$	$k > 1$
$AR(2)$	$a_{11} = \rho_1$	$a_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$		$a_{kk} = 0$	$k > 2$
$AR(p)$	$a_{11} \neq 0$	$a_{22} \neq 0$	$a_{pp} \neq 0$	$a_{kk} = 0$	$k > p$

Hence for an AR process

- Autocorrelations consist of damped exponentials and/or sine waves.
- The Partial autocorrelation is zero for lags greater than the order of the process.

2.6 MA Process

An $MA(1)$ process is defined by:

$$X_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

where ε_t is white noise

$$\begin{aligned} E[X_t] &= 0 \\ var[X_t] &= E[(\varepsilon_t + \theta\varepsilon_{t-1})^2] \\ &= E[\varepsilon_t^2 + \theta^2 E[\varepsilon_{t-1}^2] \quad \text{(independence)} \\ &= (1+\theta^2)\sigma_\varepsilon^2 \\ \lambda_1 &= E[x_t x_{t-1}] \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] \\ &= \theta E[\varepsilon_{t-1}^2] \\ &= \theta\sigma_\varepsilon^2 \end{aligned}$$

therefore

$$\begin{aligned} \rho_1 &= \frac{\theta}{1+\theta^2} \\ \lambda_2 &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-2} + \theta\varepsilon_{t-3})] \\ &= 0 \end{aligned}$$

Clearly $\lambda_j = 0$ for $j \geq 2$. Thus an $MA(1)$ process is stationary (regardless of the value of θ).

An $MA(q)$ process is defined as follows. ε_t is as usual a Gaussian White noise.

$$\begin{aligned}
X_t &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \\
E[X_t] &= 0 \\
var[X_t] &= (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma_\varepsilon^2 \\
\lambda_k &= \mathrm{Cov}[X_t X_{t-k}] \quad = \\
&= E[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \ldots \theta_k\varepsilon_{t-k} + \\
&\qquad \theta_{k+1}\varepsilon_{t-k-1} + \cdots + \theta_q\varepsilon_{t-q}) \\
&\qquad (\varepsilon_{t-k} + \theta_1\varepsilon_{t-k-1} + \cdots + \theta_{q-k}\varepsilon_{t-q} + \ldots) \\
&= (\theta_k + \theta_{k+1}\theta_1 + \cdots + \theta_q\theta_{q-k})\sigma_\varepsilon^2
\end{aligned}$$

and

$$\rho_k = \frac{\lambda_k}{var[X_t]}$$

It is clear that an MA process is stationary regardless of the values of the $\theta's$.

$$\rho_k = \begin{cases} \frac{\sum_{i=0}^{n-k}(\theta_i\theta_{i+k})}{(1+\theta^2+\cdots+\theta_q^2)} & , \quad k \leq q \\ 0 & , \quad k > q \end{cases}$$

The important point to note is that the autocorrelation function for an $MA(q)$ process is zero for lags greater than q.

The duality between AR and MA processes is even more complete. The derivation of an expression for the partial autocorrelation function of an MA process is too complicated to give here. One would find that the partial autocorrelation function of an MA process has the same general form as the autocorrelation function of an AR process.

2.7 Invertibility

A property required on occasion in the analysis of such time series is that of invertibility. Recall that the $AR(1)$ process

$$(1 - \alpha L)\, x_t = \varepsilon_t$$

is stationary if $|\alpha| < 1$. In such cases the $AR(1)$ process has an $MA(\infty)$ representation.

$$\begin{aligned} x_t &= (1-\alpha L)^{-1}\varepsilon_t \\ &= (1+\alpha L+\alpha^2 L^2+\dots)\varepsilon_t \\ &= \varepsilon_t+\alpha\varepsilon_{t-1}+\alpha^2\varepsilon_{t-2}+\dots \end{aligned}$$

and this series converges due to stationarity conditions.

Consider the $MA(1)$ process with $|\theta|<1 \quad \left[|\theta|^{-1}>1\right]$

$$\begin{aligned} x_t &= (1-\theta L)\varepsilon_t \\ (1-\theta L)^{-1}x_t &= \varepsilon_t \\ (1+\theta L+\theta^2L^2-\dots)x_t &= \varepsilon_t \\ x_t+\theta x_{t-1}+\theta^2 x_{t-2}+\dots &= \varepsilon_t \end{aligned}$$

The left hand side converges if $|\theta|<1$. In such cases $MA(1)$ process has an $AR(\infty)$ representation and the process is said to be invertible. If the $MA(q)$

process $x_t=\Theta(L)\varepsilon_t$ is invertible the roots of $\Theta(L)=0$ are outside the unit circle.

The methodology that we are developing (i.e deriving properties of a series from its estimated autocorrelogram) depends on a unique relationship between the autocorrelogram and the series). It may be shown that this unique relationship holds for stationary $AR(p)$ and invertible $MA(q)$ processes.

2.8 Examples

Example 1. Determine the ACF of the process

$$y_t=\varepsilon_t+0.6\varepsilon_{t-1}-0.3\varepsilon_{t-2}$$

where ε_t is White noise with variance σ^2

Solution

$$\begin{aligned} Ey_t &= 0 \\ Var(y_t) &= (1+(0.6)^2+(0.3)^2)\sigma^2=1.45\sigma^2 \\ E(y_ty_{t-1}) &= E(\varepsilon_t+0.6\varepsilon_{t-1}-0.3\varepsilon_{t-2})(\varepsilon_{t-1}+0.6\varepsilon_{t-2}-0.3\varepsilon_{t-3}) \\ &= \sigma^2(0.6-0.18) \\ &= 0.42\sigma^2 \\ \rho_1 &= 0.30 \\ E(y_ty_{t-2}) &= E(\varepsilon_t+0.6\varepsilon_{t-1}-0.3\varepsilon_{t-2})(\varepsilon_{t-2}+\dots) \\ &= -0.30\sigma^2 \\ \rho_2 &= 0.21 \\ \rho_3 &= \rho_4=\dots=0 \end{aligned}$$

Example 2. Calculate and plot the autocorrelations of the process $y_t = \varepsilon_t - 1.1\varepsilon_{t-1} + 0.28\varepsilon_{t-2}$ where ε_t is White Noise. Comment on the shape of the partial autocorrelation function of this process

Example 3 Calculate and plot the autocorrelation function of this process $y_t = 0.7y_{t-1} + \varepsilon_t$ where ε_t is White noise with variance σ^2

2.9 Autocorrelations for a random walk

Strictly speaking these do not exist but if we are given a sample from a random walk we can estimate the sample autocorrelation function. Will the shape of this be significantly different from that of the processes we have already examined? The random walk is given by

$$X_t = X_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \text{ is White Noise}$$

Let x_1, x_2, ... ,x_n be a sample of size n from such a process The sample autocovariance is given by

$$c_\tau = \frac{1}{n}\sum_{t=\tau+1}^{n}(x_t - \bar{x})(x_{t+\tau} - \bar{x})$$

where

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{j=t} x_j$$

As ε_t is stationary its autocovariances will tend to zero. We may write

$$\begin{aligned}
&\frac{1}{n}\sum_{t=\tau+1}^{n}\varepsilon_t\varepsilon_{t-\tau} \\
&= \frac{1}{n}\sum_{t=\tau+1}^{n}(x_t - x_{t-1})(x_{t-\tau} - x_{t-\tau-1}) \\
&= \frac{1}{n}\sum_{t=\tau+1}^{n}[(x_t - \bar{x}) - (x_{t-1} - \bar{x})][(x_{t-\tau} - \bar{x}) - (x_{t-\tau-1} - \bar{x})] \\
&= \frac{1}{n}\sum_{t=\tau+1}^{n}[(x_t - \bar{x})(x_{t-\tau} - \bar{x}) + (x_{t-1} - \bar{x})(x_{t-\tau-1} - \bar{x}) \\
&\quad -(x_t - \bar{x})(x_{t-\tau-1} - \bar{x}) - (x_{t-1} - \bar{x})(x_{t-\tau} - \bar{x})]
\end{aligned}$$

In this expression -

LHS		$\rightarrow$	0
RHS	1^{st} term	$\rightarrow$	c_τ
	2^{nd} term	$\rightarrow$	c_τ
	3^{rd} term	$\rightarrow$	$c_{\tau+1}$
	4^{th} term	$\rightarrow$	$c_{\tau-1}$

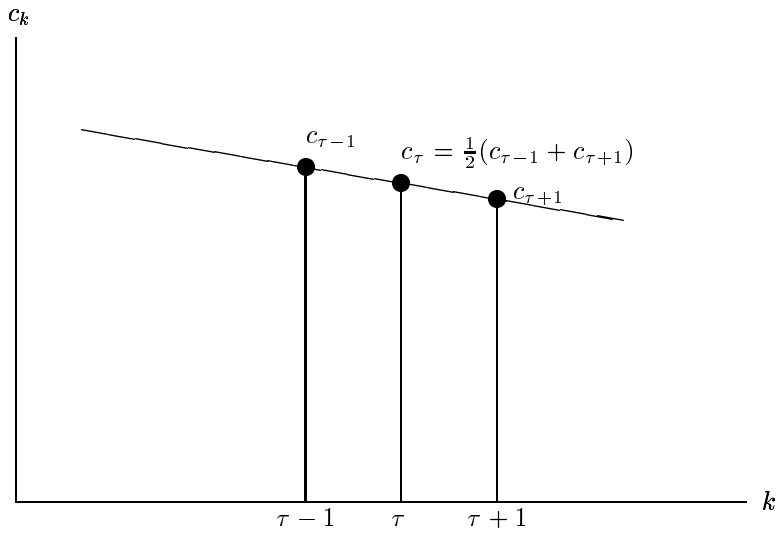

Figure 2.1: Sample autocorrelations of a Random Walk.

Thus for sufficiently large t we have $0 = 2c_\tau - c_{\tau+1} - c_{\tau-1}$ Thus $2c_\tau = c_{\tau+1} - c_{\tau-1}$. This is illustrated in Figure 2.1

Sample autocorrelations behave as a linear function and do not die out exponentially. This indicates that the series is not stationary. Note that the sample autocorrelations for a random walk are very similar to the theoretical autocorrelations of an $AR(1)$ process with ϕ close to 1. The theoretical autocorrelations for a random walk are all equal to 1. We will later look at a statistical test which is applicable in this case. Differencing may make a series stationary (see earlier comments on the random walk).

2.10 The $ARMA(p, q)$ Process

We mow consider (Mixed) $ARMA(p, q)$ processes:

Again let ε_t be white noise. X_t is a (mixed) Autoregressive Moving Average process of order p, q, denoted $ARMA(p, q)$ if

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) X_t = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) \varepsilon_t$$

or

$$\Phi(L) X_t = \Theta(L) \varepsilon_t$$

where Φ and Θ are polynomials of degree p and q respectively in L.

The conditions for stationarity are the same as those for an $AR(p)$ process. i.e. $\Phi(L) = 0$ has all its roots outside the unit circle. The conditions for invertibility are the same as those for an $MA(q)$ process. i.e $\Theta(L) = 0$ has all its roots outside the unit circle. The autocorrelogram of an $ARMA(p,q)$ process is determined at greater lags by the $AR(p)$ part of the process as the effect of the MA part dies out. Thus eventually the ACF consists of mixed damped exponentials and sine terms. Similarly the partial autocorrelogram of an ARMA (p,q) process is determined at greater lags by the $MA(q)$ part of the process. Thus eventually the partial autocorrelation function will also consist of a mixture of damped exponentials and sine waves.

There is a one to one relationship between process and autocorrelation function. for a stationary and invertible $ARMA(p,q)$ process

We have looked, at great length, into the properties of stationary $AR(p)$, $MA(q)$ and $ARMA(p,q)$ processes. How general are these processes? Wald in 1938 proved the following result (see Priestly).

Any stationary process X_t can be expressed in the form

$$X_t = U_t + V_t$$

where

1. U_t and V_t are uncorrelated
2. U_t has a representation $U_t = \sum_{i=0}^{\infty} g_i \varepsilon_{t-i}$ with $g_0 = 1 \quad \sum g_i^2 < \infty$ and ε_t white noise uncorrelated with V_t. (i.e. $E(\varepsilon_t, V_s) = 0$ all t, s). The sequence g_i are uniquely defined.
3. V_t can be exactly predicted from its past values.

Thus apart from a deterministic term any stationary process can be represented by an $MA(\infty)$ process.

We try to approximate the infinite polynomial

$$1 + g_1 L + g_2 L^2 + \cdots$$

by the ratio of two finite polynomials

$$\frac{\Theta(L)}{\Phi(L)}$$

. It may be shown that such an approximation can be achieved to any preassigned degree of accuracy.

2.11 Impulse Response Sequence

Any stationary and invertible $ARMA(p,q)$ may be represented as

$$\Phi(L) X_t = \Theta(L) \varepsilon_t$$

where

$$\begin{aligned}\Phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p \\ \Theta(L) &= 1 + \theta_1 L - \cdots - \theta_q L^q\end{aligned}$$

or by it autocorrelations.

In these conditions it may also be represented as

$$\begin{aligned}X_t &= \Psi(L)\varepsilon_t \\ &= \textstyle\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\end{aligned}$$

The sequence $\{\psi_j\}$ is known as the impulse response sequence for reasons which will become clear below. In linear systems theory the sequence $\{\varepsilon_j\}$is known as the input sequence and $\{X_j\}$ as the output sequence. A system is linear if when inputs $\{u_j^1\}$, $\{u_j^2\}$ produce outputs $\{y_j^1\}$, $\{y_j^2\}$, respectively, inputs $\{u_j^1 + u_j^2\}$ produces $\{y_j^1 + y_j^2\}$. Note the absence of a constant[1] in the definition of the system.

Let u_t, $-\infty \leq t \leq \infty$, be the input to a system. How does the output change if the input at $t = 0$ is increased by unity. By linearity the change is the same as the respons of a system with $\psi_t = 0$ for all t except for $t = 0$ when $\psi_0 = 0$. The effect of this shock is given by

Delay	effect of shock
0	1
1	ψ_1
2	ψ_2
$\vdots$	$\vdots$

The effect of the shock at a delay of t is to add ψ_t to the output at time t. For this reason $\{\psi_t\}$ is known as the impulse response sequence.

2.12 Integrated processes

Most of the processes encountered in economics are not stationary. Common sense will confirm this in many cases and elaborate statistical tests may not be required. Many economic series behave as random walks and taking first differences will make the series stationary. i.e. x_t is not stationary but $z_t = x_t - x_{t-1} = \Delta x_t$ is stationary. Such a series is said to be integrated of order 1, denoted $I(1)$.

On occasion a series must be differenced d times before it can be made stationary (It is not stationary if differenced $1, 2 \ldots d-1$ times). Such a series is said to be integrated of order d, denoted $I(d)$. If differencing a series d times makes it into a stationary $ARMA(p,q)$ the series is said to be an autoregressive integrated moving average process, denoted $ARIMA(p,d,q)$ and may be written

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t$$

[1] In linear systems theory a constant can be included in the initial conditions attached to the system (initial energy storage)

where $\Phi(L)$ is a polynomial of order p, $\Theta(L)$ of order q and Φ and Θ obey the relevant stationarity and invertibility conditions. In this expression the right-hand side has a unit root in the operator $\Phi(L)(1-L)^d$ Testing for stationarity is the same as looking for, and not finding, unit roots in this representation of the series In economics with monthly, quarterly or annual time series d will not be more than two.

If the presence of a unit root is not obvious it may become obvious from an examination of the sample autocorrelogram and indeed this tool was used for many years to indicate their presence. In recent years Dickey Fuller tests have been designed to test for a unit root in these circumstances.

If x_t has a unit root and we estimate the regression

$$x_t = \rho x_{t-1} + \varepsilon_t$$

we would expect a value of ρ close to one. Alteratively if we run the regression

$$\Delta x_t = \lambda x_{t-1} + \varepsilon_t$$

we would expect a value of λ close to zero. If we calculate the t-statistic for zero λ we should be able to base a test of $\lambda = 0$ (or the existence of a unit root) on this statistic. However the distribution of this statistic does not follow the usual t-statistic but follows a distribution originally tabulated by Fuller (1976).

We test

H_o	$\lambda = 0$	(unit root)
against		
H_1	$\lambda < 0$	(stationarity)

and reject the unit root for sufficiently small values of the t-statistic.

In effect there are four such tests

	Test Regression	True Model
1.	$\Delta x_t = \lambda x_{t-1} + \varepsilon_t$	$\Delta x_t = \varepsilon_t$
2.	$\Delta x_t = \alpha_1 + \lambda x_{t-1} + \varepsilon_t$	$\Delta x_t = \varepsilon_t$
3.	$\Delta x_t = \alpha_1 + \lambda x_{t-1} + \varepsilon_t$	$\Delta x_t = \alpha_1 + \varepsilon_t$
4.	$\Delta x_t = \alpha_0 t + \alpha_1 + \lambda x_{t-1} + \varepsilon_t$	$\Delta x_t = \alpha_1 + \varepsilon_t$

The t statistics for $\lambda = 0$ in 1. 2. and 4. yield the test statistics that Fuller calls $\widehat{\tau}, \widehat{\tau}_\mu$ and $\widehat{\tau}_\tau$ respectively. These are referred to as the 'no constant', 'no trend', and 'with trend statistics'. Critical values for these statistic and the t-statistic are compared below.

Comparison of Critical Values

		$samplesize = 25$			$samplesize = 50$		
$size$	$t-stat$	$\hat{\tau}$	$\hat{\tau_\mu}$	$\hat{\tau_\tau}$	$\hat{\tau}$	$\hat{\tau_\mu}$	$\hat{\tau_\tau}$
1%	−2.33	−2.66	−3.75	−4.38	−2.62	−3.58	−4.15
5%	−1.65	−1.95	−3.00	−3.60	−1.95	−2.93	−3.50
10%	−1.28	−1.60	−2.62	−3.24	−1.61	−2.60	−3.18

The t-statistic in 3 has an asymptotic Normal distribution. This statistic is not, in my opinion, as important in econometrics. It has been suggested that, in finite samples, the Dickey-Fuller distributions may be a better approximation than the Normal distribution. In 1, 2 and 4 the joint distribution of α_0, α_1 and λ have non-standard distributions. It is possible to formulate joint hypotheses about α_0, α_1 and λ. Critical values are given in Dickey and Fuller (1981) and have been reproduced in several books

The Dickey Fuller critical values are not affected by the presence of heteroscedasticity in the error term. They must, however, be modified to allow for serial autocorrelation. The presence of autocorrelation in the may be thought of as implying that we are using the 'wrong' null and alternative hypotheses. Suppose that we assume that the first difference follows an $AR(p)$ process. Augmented Dickey-Fuller (ADF) are then appropriate. In an ADF test the regressions are supplemented by lags of ΔX_t.

	Test Regression	True Model
5.	$\Delta x_t = \lambda x_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t$	$\lambda = 0$
6.	$\Delta x_t = \alpha_1 + \lambda x_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t$	$\alpha_1 = \lambda = 0$
7.	$\Delta x_t = \alpha_1 + \lambda x_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t$	$\lambda = 0$
8.	$\Delta x_t = \alpha_0 t + \alpha_1 + \lambda x_{t-1} + \sum_{j=1}^{p} \phi_j \Delta X_{t-j} + \varepsilon_t$	$\alpha_1 = \lambda = 0$

In 5, 6, and 8 the t-statistics for $\lambda = 0$ have the same $\widehat{\tau}, \widehat{\tau}_\mu$ and $\widehat{\tau}_\tau$ distributions as those of the unaugmented regressions. The t-statistics for $\phi_j = 0$ have standard distributions in all cases. Note that the joint distributions of α_0, α_1 and λ may have non-standard distributions as in the unaugmented case.

The ADF test assumes that p the order of the AR process is known. In general this is not so and p must be estimated. It has been shown that if p is estimated using the Akaike (1969) AIC or Schwartz (1978) BIC criterion or using t-statistics to test significance of the ϕ_j statistics the confidence intervals remain valid. The ADF test may be extended to the ARMA family by using the ADF and AIC or BIC to insert an appropriate number of lags.

Philips (1987) and philips and Perron (1988) proposed an alternative method of dealing with autocorrelated variables. Their method is somewhat more general and may be considered an extention to testing within an $ARMA$ class of series. They calculate the same regressions as in the Dickey Fuller case but adjust the test statistics using non-parametric methods to take account of general autocorrelation and heteroscedasticity. Said and Dickey ADF tests also provide a valid test for general $ARMA$ processes.

The choice of test may appear somewhat confusing. In an ideal situation one would hope that the conclusions might be the same regardless of the test. In the type of forecasting exercise one would expect that the type of test used would be consistent with the model being estimated. Thus if an AR(3) model (in levels) were estimated one would choose an ADF test with two lags In small samples the power of unit root tests is low (i.e. it may accept the hypothesis of a unit root when there is no unit root). Thus care must be exercised in applying these tests.

Chapter 3

Box-Jenkins methodology

The Box-Jenkins methodology is a strategy for identifying, estimating and forecasting autoregressive integrated moving average models. The methodology consists of a three step iterative cycle of

1. Model Identification
2. Model Estimation
3. diagnostic checks on model adequacy

followed by forecasting

3.1 Model Identification

For the moment we will assume that our series is stationary. The initial model identification is carried out by estimating the sample autocorrelations and partial autocorrelations and comparing the resulting sample autocorrelograms and partial autocorrelograms with the theoretical ACF and PACF derived already. This leads to a tentative identification. The relevant properties are set out below.

	ACF	**PACF**
AR(p)	Consists of damped exponential or sine waves—dies out exponentially	Is zero after p lags
MA(q)	Is zero after q lags	Consists of mixtures of damped exponential or sine terms—dies out exponentially
ARMA(p,q)	Eventually dominated by AR(p) part— ... then dies out exponentially	Eventually dominated by MA(q) part— ... then dies out exponentially

This method involves a subjective element at the identification stage. This can be an advantage since it allows non-sample information to be taken into

account. Thus a range of models may be excluded for a particular time series. The subjective element and the tentative nature of the identification process make the methodology difficult for the non experienced forecaster.

In deciding which autocorrelations/partial autocorrelations are zero we need some standard error for the sample estimates of these quantities.

For an $MA(q)$ process the standard deviation of $\hat{\rho}_\tau$ (the estimate of the autocorrelation at lag τ) is given by

$$n^{-\frac{1}{2}} \left(1 + 2\left(\rho_1^2 + \cdots + \rho_q^2\right)\right)^{\frac{1}{2}} \text{ for } \tau > q$$

For an $AR(p)$ process the standard deviation of the sample partial autocorrelations $\hat{a}_{kk}$ is approximately $\frac{1}{\sqrt{n}}$ for $k > p$.

By appealing to asymptotic normality we can draw limits of ± 2 standard deviations about zero to assess whether the autocorrelations or partial autocorrelations are zero. This is intended as an indication only as the sample sizes in economics are, in general small. In particular the sample estimates of the autocorrelations of a stationary series are correlated in small samples - Thus invalidating many standard inferences.

The identification process is explained in Figures 3.1 and 3.2. It is assumed that the constant is zero in each illustrated system and this does not change the shape of the theoretical autocorrelations or partial autocorrelations.

Figure 3.1 gives the theoretical autocorrelations and partial autocorrelations for the $AR(1)$ process $X_t = \phi X_{t-1} + \varepsilon_t$ for $\phi = 0.4$, 0.6, 0.8 and 0.99. Note that the partial autocorrelation function is zero except for the first autocorrelation. This is the particular property of an $AR(1)$ process. Note that the autocorrelations die out exponentially. This process is slow when ϕ is close to one. In particular the theoretical autocorrelation function for the $AR(1)$ process with ϕ close to 1 is very similar to the shape of the sample autocorrelation function for a random walk.

Figure 3.2 plots the theoretical autocorrelations and partial autocorrelations for three $AR(2)$ processes. The first process

$$X_t = X_{t-1} - 0.24X_{t-1} + \varepsilon_t$$

which may be written

$$(1 - 0.6L)(1 - 0.04L)X_t = \varepsilon_t$$

The roots of the equation

$$(1 - 0.6L)(1 - 0.4L) = 0$$

are

$$L = 1.67 \text{ or } L = 2.5$$

, both of which are outside the unit circle (modulus or absolute value greater than one). Thus the process is stationary. The autocorrelogram is very similar to

those or the $AR(1)$ processes in Figure 3.1. What distinguishes the process and indentifies it as an $AR(2)$ process is the two non-zero partial autocorrelations.

From the second system

$$X_t = 0.6X_{t-1} - 0.25X_{t-1} + \varepsilon_t$$

the equation

$$1 - 0.6L + 0.25L^2 = 0$$

has roots

$$L = 1.2 \pm 1.6i$$

These roots are complex conjugates and their modulus is 2.5. Thus the roots are outside the unit circle and the process is stationary. In this case the autocorrelations oscillate about zero and die out exponentially. This oscillatory behaviour is a result of the complex roots that can occur in $AR(2)$ and higher order processes.

If ϕ is negative in an $AR(1)$ process the sign of the autocorrelations may alternate but they can not oscillate in the same way as those of an $AR(2)$ or higher order process. The PACF again shows the two non-zero values of the partial autocorrelations typical of an $AR(2)$ process.

Higher orders of AR processes show autocorrelations which are mixtures of those of $AR(2)$ and $AR(1)$ processes with the number of non-zero partial autocorrelations corresponding to the order of the process

We could generate similar diagrams for $MA(1)$, $MA(2)$ and higher order MA processes. Such diagrams would be very similar to those already generated for AR processes of similar order but with the autocorrelations and partial autocorrelations interchanged. The number of non-zero autocorrelations for an MA process corresponds to the order of the process. The partial autocorrelations for an MA process resemble the autocorrelations for an AR process.

Figure 3.3 shows an example of the autocorrelations and partial autocorrelations for an $ARMA(2,2)$ process. Note that the autocorrelations are similar to those of an AR process and the partial autocorrelations resemble those of an MA process.

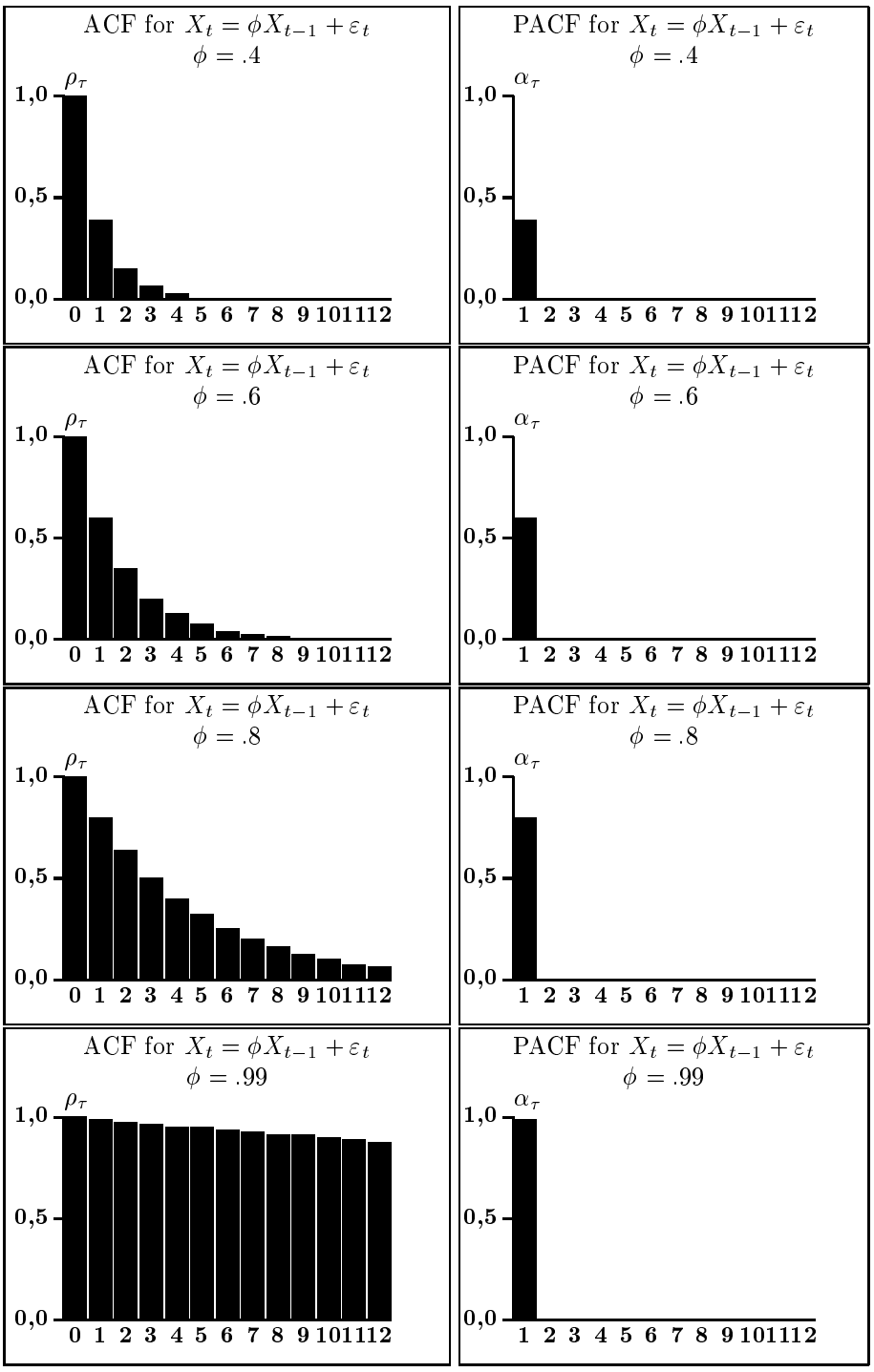

Figure 3.1: Autocorrelations and Partial Autocorrelations for various $AR(1)$ Processes

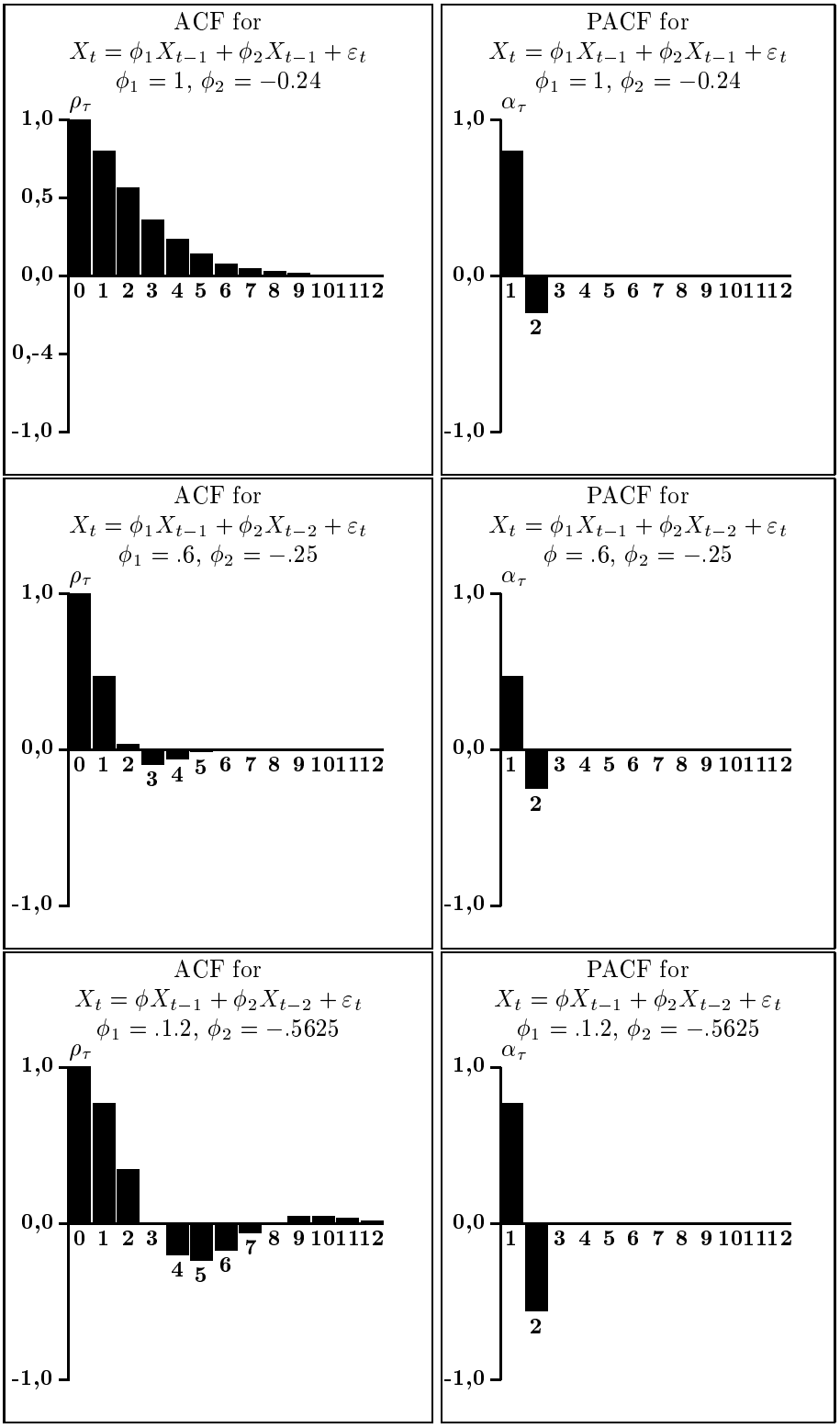

Figure 3.2: ACF and PACF for various $AR(2)$ Processes

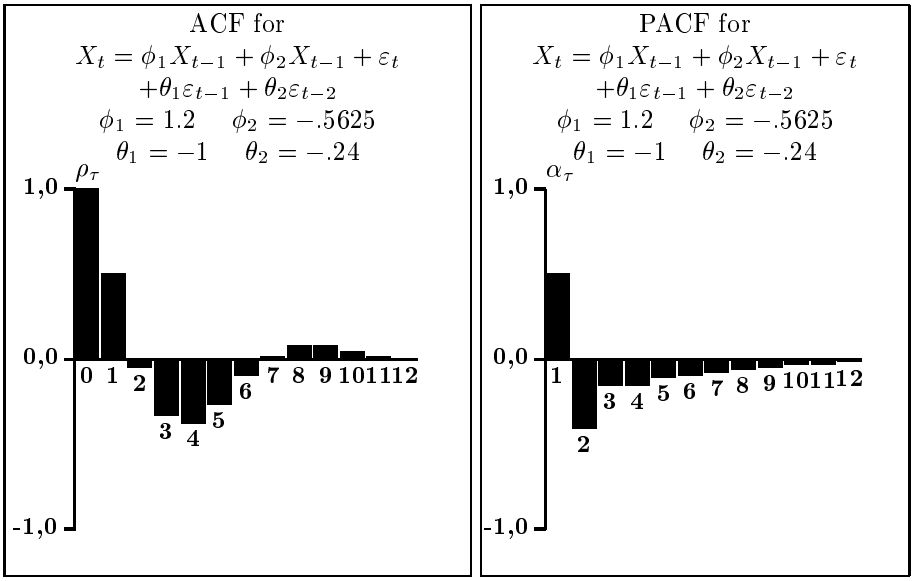

Figure 3.3: ACF and PACF for an $ARMA$ Process

If the series are not stationary we try to make it stationary by a process of preliminary transformations and/or differencing the data. Preliminary transformations are simple transformations which are intended to do two things

- Straighten out trends
- Reduce heteroscedasticity i.e. produce approximately uniform variability in the series over the sample range

In the second case we often find that the variance of x_t is proportional to x_t. In general one of the following will suffice

- Do nothing
- Take logarithms
- Take square roots

In deciding how to proceed bear the following in mind

- Do you think of the series in terms of growth rates (G.N.P., money, prices etc.)? If so take logs.
- If a percentage growth rate has no meaning for the series—do nothing or possibly take square roots if the series is more variable at higher values (e.g. some count data).

If the choice of transformation is not obvious then a transformation will probably make little or no difference to the forecast. In particular difficult cases some form of Box-Cox transformation may be used but this will not generally be required in economics.

Forecasting methodology is generally very sensitive to errors in differencing—particularly to underdifferencing. The Dickey-Fuller tests may be used to test the degree of differencing. The amount of differencing and the inclusion of a constant in the model determine the long-term behaviour of the model. The following table lists the implications of various combinations of differencing and the inclusion/exclusion of an intercept.

Differences	**Intercept**	**Behaviour**
0	Yes	Clusters around mean level (unemployment?)
1	No	Doesn't trend—Doesn't seek a level (interest rates)
1	Yes	Trends at a fairly constant rate (real G.D.P.)
2	No	Trends at a variable rate (price index)

A very important principle in this type of analysis is that of parsimony. Many stationary processes can be well fitted by a high order AR process

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

where p may be reasonably large. The possibility of using an ARMA process for approximation may allow us to achieve a good fit with many fewer parameters. In effect this more parsimonious model may forecast better. The smaller the data set the less parameters you can estimate and the more important judgment becomes.

Time series models should not be taken too seriously. They are designed to fit the serial correlation properties of the data and not to explain them. You should aim to find a model which fits the data well with as few parameters as possible.

The most carefully thought out model is worthless if it cannot be estimated using the available data. While it may be thought that four parameters can be estimated from thirty data points, experience has shown that if a three parameter model fits almost as well (even if the difference is statistically significant) then the smaller model will forecast better most of the time.

3.2 Estimation

The class of models we have considered so far may be expressed as

$$\Phi(L)\nabla^d x_t = \alpha + \Theta(L)\varepsilon_t$$

where

$$\begin{aligned} \Phi(L) &= 1 - \phi_1 L - \cdots - \phi_p L^p \\ \Theta(L) &= 1 + \theta_1 L - \cdots - \theta_q L^p \\ \nabla &= 1 - L \end{aligned}$$

and we have inserted a constant α. If d is known we write $y_t = \nabla^d x_t$

If $\varepsilon_1 \ldots \varepsilon_t$ are independent normal we may write their joint density as

$$f(\varepsilon_1, \ldots, \varepsilon_n / \alpha, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{T}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{T}\varepsilon_i^2\right]$$

From this joint density we can derive the liklihood function. The calculations are not trivial as the ε are not observed. The procedure may be compared to a regression where the residual follows an $AR(1)$ process. Two posibilites are

- Cochrane-Orcutt – works by using an iterative process which is conditional on the first observation and
- the corresponding Maximum Liklihood which improves efficiency by including the first observation in the calculation of the liklihood.

In the estimation of an $ARMA$ model it is possible to estimate the liklihood conditional on the early observations. With modern software there is no need to do this and if you should use full Maximum Liklihood. The estimation of the liklihood can be achieved with many different software packages on a PC.

If the numerical optimization does not converge it is most likely that the model that is being estimated is not the right model. Check that the polynomials $\Phi(L)$ and $\Psi(L)$ do not have a common or near common factor (that is both are divisible or almost divisible by $(1-\phi L)$). In such cases reducing the order of Φ or Θ by one may make the process converge and result in a more parsimonious model that will forecast better.

3.3 Model testing: diagnostic checks for model adequacy

We will consider two types of diagnostic checks. In the first we fit extra coefficients and test for their significance. In the second we examine the residuals of the fitted model to determine if they are white noise (i.e. uncorrelated).

3.3.1 Fitting extra coefficients

Suppose we have tentatively identified and estimated an $ARMA(p,q)$ model. Consider the following $ARMA(p+q*, q+q*)$ model.

$$(1 - a_1 L - \cdots - a_p L^p - \cdots - a_{p+p*} L^{p+p*}) X_t = \\ (1 + b_1 L + \cdots + b_q L^q + \cdots + b_{q+q*} L^{q+q*}) \varepsilon_t$$

We can calculate a Lagrange Multiplier test of the restrictions

$$\begin{array}{ccccccc} a_{p+1} & = & a_{p+2} & = & \ldots & = & a_{p+p*} = 0 \\ b_{q+1} & = & b_{q+2} & = & \ldots & = & b_{q+q*} = 0 \end{array}$$

If the hypothesis is accepted we have evidence of the validity of the original model.

3.3.2 Tests on residuals of the estimated model.

If the model is correctly specified the estimated residuals should behave as white noise (be uncorrelated). If $e_t \quad t = 1, \ldots, T$ are the estimated residuals we estimate the sample autocorrelations.

$$r_\tau(e_t) = \frac{\sum_{t=\tau+1}^{T} e_t e_{t-\tau/}}{\sum_{t=1}^{T} e_t^2}$$

These sample autocorrelations should be close to zero. Their standard errors are functions of the unknown parameters of the model but may be estimated as $\frac{1}{\sqrt{T}}$. Thus a comparison with bounds of $\frac{\pm 2}{\sqrt{T}}$ will provide a crude check on model adequacy and point in the direction of particular inadequacies.

In addition to the test on individual autocorrelations we can use a joint test (portmanteau) known as the Q statistic

$$Q = n(n+2) \left(\sum_{i=1}^{M} (n-\tau)^{-1} r_\tau^2 \right)$$

M is arbitrary and is generally chosen as 10 to 20. Some programs produce a Q-statistic based on $M = \sqrt{T}$. The Q statistic is distributed as χ^2 with $M-p-q$ degrees of freedom. Model adequacy is rejected for large values of the Q-statistic. The Q-statistic has low power in the detection of specific departures from the assumed model. It is therefore unwise to rely exclusively on this test in checking for model adequacy.

If we find that the model is inadequate we must respecify our model, reestimate and re-test and perhaps continue this cycle until we are satisfied with the model

3.4 A digression on forecasting theory

We evaluate forecasts using both subjective and objective means.

The subjective examination looks for large errors and/or failures to detect turning points. The analyst may be able to explain such problems by unusual unforeseen or unprovided for events. Great care should be taken to avoid explaining too many of the errors by strikes etc.

In an objective evaluation of a forecast we may use various standard measures. If x_i is the actual datum for period i and f_i is the forecast then the error is defined as

$$e_i = x_i - f_i \tag{3.1}$$

The following measures may be considered

Mean Error	ME	=	$\frac{1}{n}\sum_{i=1}^{n} e_i$
Mean Absolute Error	MAE	=	$\frac{1}{n}\sum_{i=1}^{n} \lvert e_i \rvert$
Sum Squared Errors	SSE	=	$\sum_{i=1}^{n} e_i^2$
Mean Squared Error	MSE	=	$\frac{1}{n}\sum_{i=1}^{n} e_i^2$
Root Mean Square Error	RMS	=	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$

Alternatively consider a cost of error function $C(e)$ where e is the error and

$$\begin{aligned} C(0) &= 0 \\ C(e_i) &> C(e_j) \text{ if } |e_i| > |e_j| \end{aligned}$$

In many cases we also assume that $C(e) = C(-e)$. In some cases an expert or accountant may able to set up a form for $C(e)$. In much practical work we assume a cost function of the form

$$C(e) = ae^2 \quad \text{for } a > 0$$

This form of function is

1. not a priori unreasonable
2. mathematically tractable, and
3. has an obvious relationship to least squares criterion.

We can show that for this form of cost function the optimal forecast f_{nh} (h period ahead forecast of x_{n+h} given x_{n-j} for $j \geq 0$) is given by

$$f_{nh} = E(x_{n+h}/x_{n-j}, \quad j \geq 0)$$

This result may in effect be extended to more general cost functions.

Suppose we have two forecasting procedures yielding errors

$$e_t^{(1)} \quad e_t^{(2)}$$

, $t = 1 \ldots n$. If MSE is to be the criterion the procedure yielding the lower MSE will be judged superior. Can we say if it is statistically better? In general, we cannot use the usual F-test because the MSE's are probably not independent.

Suppose that $e_t^{(1)} \quad e_t^{(2)}$ is a random sample from a bivariate normal distribution with zero means and variances $\sigma_1^2\sigma_1^2$ and correlation coefficient ρ. Consider the pair of random variables $e_t^{(1)} + e_t^{(2)}$ and $e_t^{(1)} - e_t^{(2)}$

$$E(e^{(1)} + e^{(2)})(e^{(1)} - e^{(2)}) = \sigma_1^2 - \sigma_2^2$$

Thus the difference between the variances of the original variables will be zero if the transformed variables are uncorrelated. Thus the usual test for zero correlation based on the sample correlation coefficient

$$r = \frac{\sum_{t=1}^{n} \left((e_t^{(1)} + e_t^{(2)})(e_t^{(1)} - e_t^{(2)}) \right)}{\left[\sum_{t=1}^{n} \left((e_t^{(1)} + e_t^{(2)}) \right)^2 \sum_{t=1}^{n} \left((e_t^{(1)} - e_t^{(2)}) \right)^2 \right]^{\frac{1}{2}}}$$

can be applied to test equality of expected forecast errors. (This test is uniformly most powerful unbiased).

Theil proposed that a forecasting method be compared with that of a naive forecast and proposed the U-statistic which compared the RMS of the forecasting method with that derived from a random walk (the forecast of the next value is the current value). Thus

$$U = \frac{\frac{1}{n} \sum_{t=1}^{N} (f_t - X_t)^2}{\frac{1}{n} \sum_{t=1}^{N} (X_t - X_{t-1})^2}$$

Sometimes U is written

$$U = \frac{\frac{1}{n} \sum_{t=1}^{N} \left(\frac{f_t - X_t}{X_{t-1}} \right)^2}{\frac{1}{n} \sum_{t=1}^{N} \left(\frac{X_t - X_{t-1}}{X_{t-1}} \right)^2}$$

if $U > 1$ the naive forecast performs better than the forecasting method being examined.

Even if the value of U is very much less than one we may not have a very good forecasting methodology. The idea of a U statistic is very useful but today it is feasible to use a Box Jenkins forecast as our base line and to compare this with the proposed methodology.

3.5 Forecasting with $ARMA$ models

Let X_t follow the stationary ARMA model

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \quad [\theta_0 = 1]$$

At time t let f_{nh} be the forecast of X_{n+h} which has smallest expected squared error among the set of all possible forecasts which are linear in $X_{n-j}, (j \geq 0)$.

A recurrence relationship for the forecasts f_{nh} is obtained by replacing each element in the above equation by its forecast at time n, as follows

1. replace the unknown X_{n+k} by their forecast $f_{nk} \quad k > 0$
2. "forecasts" of X_{n+k} $(k \leq 0)$ are simply the known values
3. since ε_t is white noise the optimal forecast of ε_{n+k} $(k > 0)$ is simply zero
4. "forecasts" of ε_{n+k} $(k \leq 0)$ are the known values of the residuals

The process

$$\Phi(L) X_t = \Theta(L) \varepsilon_t$$

may be written

$$X_t = c(L) \varepsilon_t$$

where $c(L)$ is an infinite polynomial in L such that

$$\Phi(L) c(L) = \Theta(L)$$

write

$$c(L) = c_o + c_1 L + \cdots$$

where c_i may be evaluated by equating coefficients.

$$\begin{array}{lcll} X_{n+h} & = & c_0 \varepsilon_{n+h} + c_1 \varepsilon_{n+h-1} + \cdots & + c_h \varepsilon_n + c_{h+1} \varepsilon_{n-1} + \cdots \\ f_{nh} & = & & c_h \varepsilon_n + c_{k+1} \varepsilon_{n-1} \end{array}$$

Thus the forecast error is given by

$$\begin{aligned} e_{nh} &= X_{n+h} - f_{n,h} \\ &= c_0 \varepsilon_{n+h} + c_1 \varepsilon_{n+h-1} + \cdots + c_{h-1} \varepsilon_{n+1} \\ &= \sum_{j=0}^{h-1} c_j \varepsilon_{n+h-j} \end{aligned}$$

As the e_i are independent the variance of the forecast error is given by

$$V_h = E\left(e_{nh}^2\right) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} c_j^2$$

A similar method will be used for ARIMA processes. The computations will be completed by computer. These estimates of the forecast error variance will be used to compute confidence estimates for forecasts.

3.6 Seasonal Box Jenkins

So far the time series considered do not have a seasonal component. Consider for example a series giving monthly airline ticket sales. These sales will differ greatly from month to month with larger sales at Christmas and during the holiday season. In Ireland sales of cars are often put off until the new year in order to qualify for a new registration plate. We may think of many such examples. In Box-Jenkins methodology we proceed as follows.

If the seasonal properties repeat every s periods then X_t is said to be a seasonal time series with periodicity s. Thus $s = 4$ for quarterly data and $s = 12$ for monthly data and possibly $s = 5$ for daily data. We try to remove the seasonality from the series to produce a modified series which is non-seasonal, to which an ARIMA model could be fitted. Denote the nonseasonal series by u_t. Box Jenkins proposed the seasonal ARIMA filter.

$$\Phi_s(L^s)(1-L^s)^D X_t = \Theta_s(L^s)u_t$$

where

$$\begin{aligned}\Phi_s(L^s) &= 1 - \phi_{1s}L^s - \phi_{2s}L^{2s} - \cdots - \phi_{Ps}L^{Ps} \\ \Theta_s(L^s) &= 1 - \theta_{1s}L^s - \theta_{2s}L^{2s} - \cdots - \theta_{Qs}L^{Qs}\end{aligned}$$

u_t is then approximated using the usual ARIMA representation (notation as before)

$$\Phi(L)(1-L)^d u_t = \Theta(L)\varepsilon_t$$

and u_t is $ARIMA(p,d,q)$.

Substituting for u_t

$$\Phi(L)(1-L)^d\Phi_s(L^s)(1-L^s)^D = \Theta(L)\Theta_s(L^s)\varepsilon_t$$

This is known as a seasonal $ARIMA$ $(SARIMA)$ $(p,d,q)\times(P,D,Q)_s$ process.

In processing such a a series we follow the same cycle of

1. provisional identification

2. estimation

3. testing

and finally forecasting as in the non-seasonal model.

3.6.1 Identification

We now have six parameters $pdqPD$ and Q to identify.

Step 1: Identify a combination of d and D required to produce stationarity. If the series is seasonal the autocorrelogram will have spikes at the seasonal frequency. For example quarterly data will have high autocorrelations at lags 4, 8, 12 etc. Examining these will indicate the need for seasonal differencing. If seasonal differencing is required then the autocorrelogram must be reestimated for the seasonally differenced series. Identification of d proceeds similarly to the non seasonal case. An extension of the Dickey-Fuller tests due to Hylleberg, Engle, Granger and Yoo exists and may be used. These problems

****

Insert Examples

****

Step 2: Once d and D are selected we tentatively identify p, q, P and Q from the autocorrelation and partial autocorrelation functions in a somewhat similar way as in the non-seasonal model. P and Q are identified by looking at the correlation and partial autocorrelation at lags $s, 2s, 3s, \ldots$ (multiples of the seasonal frequency). In identifying p and q we ignore the seasonal spikes and proceed as in the nonseasonal case. The procedure is set out in the table below. AC and PAC are abbreviations for the autocorrelogram and partial autocorrelogram. SAC and SPAC are abbreviations for the AC and PAC at multiples of the seasonal frequency. Bear in mind that we are likely to have very few values of the SAC and SPAC. For quarterly data we may have lags 4 8 12 and probably 16. For monthly data we have 12 and 24 and possibly 36 (unless the series is very long). Identification of P and Q is very approximate. The need for parsimony must be borne in mind.

Examples of Identification

Properties	Inference
SAC dies down, SPAC has spikes at $L, 2L, \cdots, PL$ and cuts off afterPL	seasonal AR of order P
SAC has spikes at lags $L, 2L, \ldots, QL$ and SPAC dies down	seasonal MA
SAC has spikes at lags $L, 2L, \ldots, PL$ SPAC has spikes at lags $L, 2L, \ldots, QL$ and both die down	use either • seasonal MA of order Q or • seasonal AR of order P • (Fit MA first)
no seasonal spikes	$P = Q = 0$
SAC and SPAC die down	possible $P = Q = 1$

Important systems are

1. $X_t = (1 + \theta_1 L + \theta_2 L)(1 + \theta_{1s} L^s + \theta_{2s} L^{2s})\varepsilon_t$
2. $(1 - \phi_1 L)(1 - \phi_{1s} L^s)X_t = (1 + \theta_1 L)(1 + \theta_{1s} L^s)\varepsilon_t$
3. $x_t = (1 + \theta_1 L + \theta_s L^s + \theta_{s+1} L^{s+1})\varepsilon_t$

or

1. $(0, d, 2) \times (0, D, 2)_s$
2. $(1, d, 1) \times (1, D, 1)_s$
3. is strictly a non-seasonal $(0, d, s + 1)$ with restrictions on the coefficients.

3.7 Automatic Box Jenkins

The procedure outlined above requires considerable intervention from the statistician/economist completing the forecast. Various attempts have been made to automate the forecasts. The simplest of these fits a selection of models to the data, decides which is the "best" and then if the "best" is good enough uses that. Otherwise the forecast is referred back for "standard" analysis by the statistician/economist.

The selection will be based on a criterion such as the AIC (Akaike's Information Criterion), FPE (Forecast Prediction Error), HQ (Hannon Quinn Criterion), SC (Schwarz Criterion) or similar. The form of these statistics are given by

$$
\begin{aligned}
AIC &= \ln \hat{\sigma}^2 + \frac{2}{n} \\
HQ &= \ln \hat{\sigma}^2 + \frac{n \ln(\ln n)}{n} \\
SC &= \ln \hat{\sigma}^2 + \frac{\ln n}{n}
\end{aligned}
$$

The FPE can be shown to be equivalent to the yCI. $\hat{\sigma}^2$ is the estimate of the variance of the model under assessment. The chosen model is that which minimises the relevant criterion. Note that each criterion consists of two parts. The variance of the model will decrease as the number of parameters is increased (nested models) while the second term will increase. Thus each criterion provides a way of measuring the tradeoff between the improvement in variance and the penalty due to overfitting.

It should be noted that AIC may tend to overestimate the number of parameters to be estimated. This does not imply that models based on HQ and SC produce better forecasts. In effect it may be shown that asymptotically AIC minimizes 1-step forecast MSE.

Granger and Newbold (1986) claim that automatic model fitting procedures are inconsistent and tend to produce overly elaborate models. The methods provide a useful additional tool for the forecaster, but are not a fully satisfactory answer to all the problems that can arise.

The behaviour of the sample variances associated with different values of d can provide an indication of the appropriate level of differencing. Successive values of this variance will tend to decrease until a stationary series is found. For some series it will then increase once over-differencing occurs. However, this will not always occur (consider for example an $AR(1)$ process for various values of ϕ_1). The method should, therefore, only be used as an auxiliary method of determining the value of d.

ARIMA processes appear, at first sight, to involve only one variable and its own history. Our intuition tells us that any economic variable is dependent on many other variables. How then can we account for the relative success of the Box Jenkins methodology. Zellner and Palm (1974) argue " ... ... ARMA processes for individual variables are compatible with some, perhaps unknown joint process for a set of random variables and are thus not necessarily "naive", "ad hoc" alternative models". Thus there is an expectation that a univariate ARIMA model might out-perform a badly specified structural model.

The use of univariate forecasts may be important for several reasons:

1. In some cases we have a choice of modelling, say, the output of a large number of processes or of aggregate output, leaving the univariate model as the only feasible approach because of the sheer magnitude of the problem.

2. It may be difficult to find variables which are related to the variable being forecast, leaving the univariate model as the only means for forecasting.

3. Where multivariate methods are available the univariate method provides a yardstick against which the more sophisticated methods can be evaluated.

4. The presence of large residuals in a univariate model may correspond to abnormal events—strikes etc.

5. The study of univariate models can give useful information about trends long-term cycles, seasonal effects etc in the data.

6. Some form of univariate analysis may be a necessary prerequisite to multivariate analysis if spurious regressions and related problems are to be avoided.

While univariate models perform well in the short term they are likely to be outperformed by multivariate methods at longer lead terms if variables related to the variable being forecast fluctuate in ways which are different to their past behaviour.

Appendix A

REFERENCES

A.1 Elementary Books on Forecasting with sections on Box-Jenkins

(1) Bowerman, and O'Connell (1987): *Time Series Forecasting: Unified Concepts and Computer Implementation*, Duxbury. This is a good introduction and is elementary and non-mathematical

(2) Chatfield (1987), 1^{st} edition [(1999)? 4^{th} edition]: *Analysis of Time Series—Theory and Practice*, Chapman and Hall. This is a very good introduction to the theory of time series in general at a not too advanced level

(3) Makridakis, Wheelwright and McGee (1983): *Forecasting: Methods and Applications*, Wiley. A large (> 900 page textbook) that covers a wide range of forecasting techniques without getting too involved in their theoretical development. It is much much more comprehensive than either 1 or 2.

A.2 Econometric texts with good sections on Box-Jenkins

(4) Pindyck and Rubenfeld (1991): *Econometric Models and Economic Forecasting*, McGraw-Hill, (recent edition 1998) This is s very good introductory text. The new US edition contains a disk giving the data for all the problems in the model. It is a pity that this disk has not been included in the European version

(5) Judge, Hill, Griffiths, Lutkepohl and Lee (1988), *an introduction to the theory and practice of econometrics*, Wiley.

(6) Judge, Griffiths, Hill, Lutkepohl and Lee (1985): *The theory and practice of econometrics*, Wiley. (6) is a comprehensive survey of econometric theory

and is an excellent reference work for the practising econometrician—A new edition must be due shortly. (5) is an introduction to (6) and is very comprehensive ($> 1,000$ pages). It has a very good introduction to non-seasonal Box-Jenkins

A.3 Time-Series Books

(7) Box, Jenkins (1976): *Time Series Analysis – forecasting and Control*, Holden Day This covers both theory and practice very well but theory is advanced – very useful, if not essential, for practising forecasters

(8) Granger, Newbold (1986): *Forecasting Economic Time Series*, Academic Press A very good account of the interaction of standard econometric methods and time series methods. Some sections are difficult but much of the material will repay the effort involved in mastering it

(9) Priestly (1981): *Spectra Analysis and Time Series*, Academic Press, A comprehensive look on time series analysis

(10) Mills, T.C. (1990): *Time series techniques for economists*, Cambridge University Press Well described by title – intermediate level – recommended – written for economists

(11) Brockwell and Davis (1991) 2^{nd} edition, *Time Series: Theory and Methods*, Springer-Verlag An advanced book – probably the most advanced of those listed here

(12) Jenkins, G.M. (1979): *Practical Experiences with modelling and forecasting time series*, Gwilyn Jenkins and Partners (Overseas) Ltd, Jersey. The object of this book is to present, using a series of practical examples, an account of the models and model building methodology described in Box Jenkins (1976). It presents a very good mixture of theory and practice and large parts of the book should be accessible to non-technical readers