Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter

Christophe Planas
Luxembourg[1]

October 15, 1997

[1]Tel: +352 423113469, email: planas@wsel.lu. This textbook was written within the framework of a study on Seasonal Adjustment methods, on request of Eurostat (contract nr. 66630002). The ideas expressed here are the author's ones and do not necessarily reflect the position of Eurostat. Thanks are due to B. Fisher for the providing of examples.

Preface

The general purpose of this textbook is to provide analysts in statistical institutes with a unified view of applied analysis of time series as can be conducted in the framework of linear stochastic models of the ARIMA-type. The issues discussed are modelling and forecasting, filtering, signal extraction and Unobserved Components analysis, and regression in time series models. The main concern is to help the readers in understanding some important tools that progress in statistical theory has made available. Emphasis is thus put on practical aspects, and the readers will find implementations of the techniques described in softwares like SEATS-TRAMO (see Gomez and Maravall, 1996) and X12-ARIMA (see Findley and al., 1996).

Contents

Part I

Introduction to Time Series Analysis

Chapter 1

Modelling Time Series

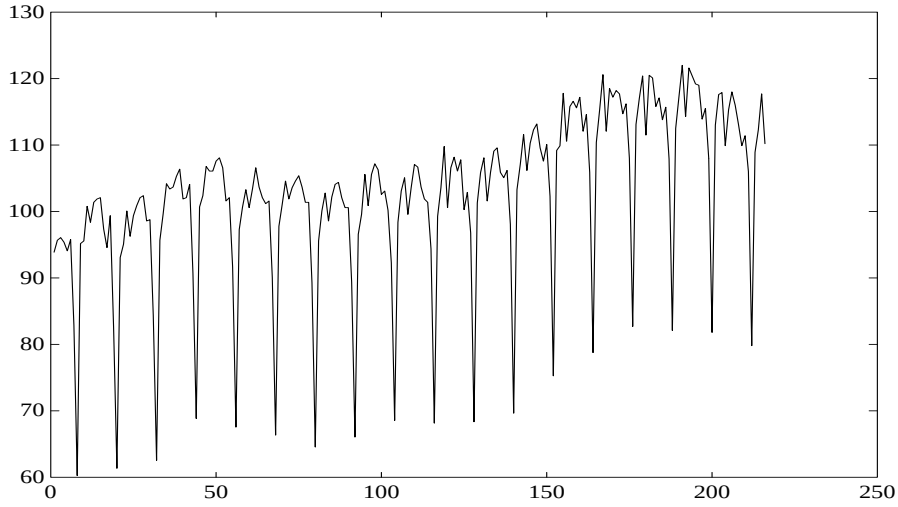
1.1 Introduction

A time series is a set of observations on a variable, say y, recorded for consecutive time intervals. If $t = 1, \cdots, T$ denotes time, then y_t represents the realisation of variable y at time t. The set of observations from 1 to T is written $[y_1, \cdots, y_T]$ or $\{y_t\}_{t=1}^{T}$, and T represents the sample length.

An example of time series is given on figure 1.1. The series plotted is the French total industry production (excluding construction, production index in constant prices). The observations are monthly, starting in January 1976 and ending in December 1993, that is 216 observations. We can see that it is characterized by a trend, that is a long-term movement, and by some large seasonal fluctuations. Figure 1.2 displays the series of Portugal employment in manufacture of bricks, tiles and construction products from january 1985 to february 1994, that is a sample of 110 observations. It shows a shaped downward long-term behavior upon which some irregularities and short-term cyclical fluctuations seem superposed. Time series analysis is concerned with two aims: the description of the salient features of the series and the prediction of the future values of the series. Typical features of economic time series are:

- The trend, or long-term movements of the series;
- The seasonal fluctuations;
- The cyclical fluctuations;

Figure 1.1: French Total Industry (1976-1 1993-12)

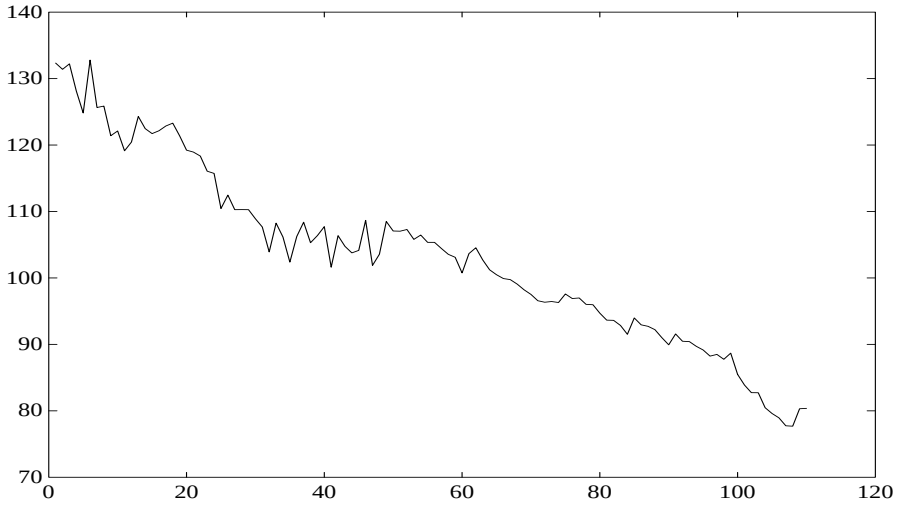


- The irregular component, which represents the non-systematic movements of the series.

For the simple reason that they are never directly observed, the variables catching these four movement-types are also termed unobserved components. The four unobserved components listed above have been explicitly stated in Persons (1919,1923). To characterize them, there are mainly two ways of proceeding. The first directly yields an estimator of these components without considering a statistical model for the series under analysis. It is thus model-free, and a popular example of a model-free procedure is given by X-11. However, in general, any model-free procedure admits a close approximation which is optimal for a particular statistical model. This model may thus be seen as implicit. A more informative approach makes instead the underlying assumptions explicit through the specification of a model. This is the concern of a second set of procedures called model-based procedures. Emphasis will be put on this approach, and in this first part, we discuss the set up of statistical models for time series.

The advantage of considering a model is not only that the assumptions are clearly

Figure 1.2: Portugal Employment in Manufacture of Bricks (1985-1 1994 2)

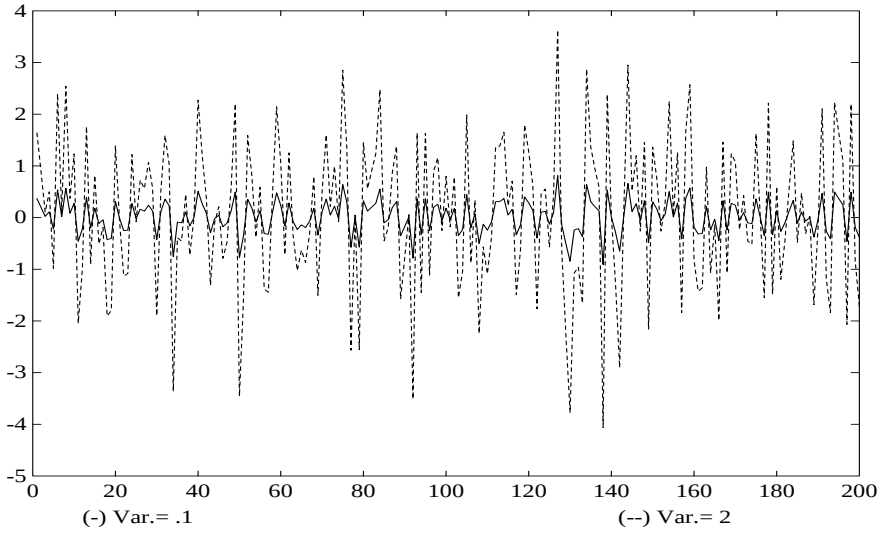


formulated, but also that a flexible model-type is likely to represent adequately the movements of a large number of time series with very different patterns. Furthermore, it is reasonable to expect that a model which does not provide a good description of the series will not yield satisfactory forecasts. However, it has to be underlined that a time series model is not an economic model and that it does not aim to represent the underlying data generation process. Its aim is to describe the behavior of the series analysed, and for that it is expected to be as flexible as possible. This is to be considered at the modelling stage, and the alternative we consider first is deterministic versus stochastic linear models.

1.2 Deterministic vs. Stochastic Linear Processes

Stochasticity will usually be introduced through a *white noise* variable. It is the process most used to represent an irregular component. A white noise is defined as a sequence of zero-mean random variables with constant variance and null autocorrelations. White noise variables may also be assumed to be normally distributed, in which case the absence of autocorrelations implies that the white noise variables are independent.

Figure 1.3: Examples of White Noises



Independent white noises are not predictable, and are sometimes called innovations. Figure 1.3 display two white noise processes, one with variance .1, the other with variance 2. It clearly appears that the more erratic process is associated with the largest variance. On the other hand, letting the variance going to zero will make the white noise closer to the zero-line.

Modelling time series relies heavily on this type of variable. The largest the innovations variance, the more erratic the process will be. Conversely, small innovations variance will in general be associated with stable processes.

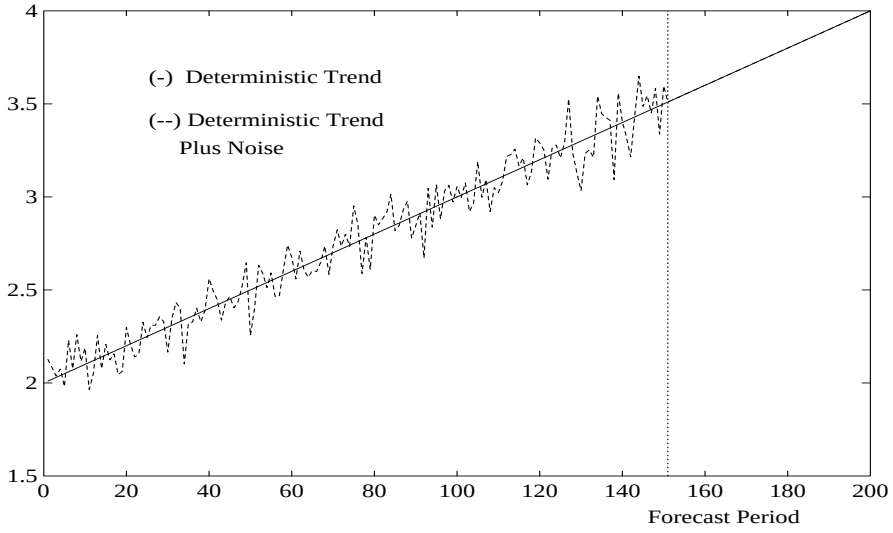
1.2.1 Modelling a trend

Suppose we are interested in modelling a series y_t whose prominent feature is a trend. The most obvious idea would be to consider a polynomial in time $f(t)$:

$$y_t = f(t) = a + bt, \tag{1.1}$$

where a and b are real. The plot of the function $f(t)$ is presented on figure 1.4. Such

Figure 1.4: Trends



models are called deterministic. It is the kind of model used in the regression approach, which was may be the first model-based approach for trend analysis. Here, it is seen that the series is described as moving in one single direction. Upturn and downturn points may be generated by increasing the order of the time-polynomial, but in any cases the movements of the series are very systematic.

As seen on figure 1.4, such models are very constraining. They are highly unrealistic: no actual series behaves in that way. This conclusion was already drawn by Macaulay in 1931 (p. 38, quoted in Bell and Hillmer (1984)), who noted that trends are evolving over time and "not necessarily representable through their length by any simple mathematical equation". In a first step, it may seem more reasonable to consider that the series z_t can be described by the sum of the deterministic function $f(t)$ plus a white noise disturbance e_t with mean zero and constant variance $V(e_t)$:

$$y_t = a + bt + e_t. \tag{1.2}$$

This model is quite close to the previous one: if $\sigma^2 = 0$, then $e_t = 0$ and the models are identical. On the other hand, as e_t is not correlated, it is not forecastable, so the

forecasts of y_t in (1.2) resume to the deterministic function $f(t)$ and are thus identical to the ones yielded by (1.1).

Suppose we take the first difference of (1.2); we would get:

$$\Delta y_t = b + e_t - \theta e_{t-1}. \tag{1.3}$$

with $\theta = 1$. We may here introduce some flexibility by allowing the parameter θ to take a value in $[-1, 1]$. Equation (1.3) gives an example of a stochastic linear process: y_t evolves over time according to a probability law. When $\theta = 1$, this process reduces to (1.2), which in turn becomes closer to the deterministic path displayed in (1.1) as $V(e_t)$ goes to zero.

The deterministic process (1.1) is thus a particular case of a stochastic process. As displayed in (1.3), y_t is said to follow an Integrated-Moving Average of order (1,1). If we set θ to 0, then we get a random walk plus drift, the drift being b. Setting b to 0 yields the random walk model $\Delta z_t = e_t$: it is an evolutionary process in the sense that the current level shifts each time there is a shock on the system.

1.2.2 Modelling cycles

We are now interested in modelling a cyclical behavior. Simple tools are given by trigonometric functions of the type $y = \cos x$, where x is an angle measured in radians. Since there are 2π radians in a circle, y takes all possible values when x moves from 0 to 2π. The pattern is then repeated when x moves to the successive interval $[2k\pi, 2(k+1)\pi]$. To introduce the time index t in the determination of y_t, we may substitute x by wt, where w is measured in radians and is refered to as the *frequency*. We have now: $y_t = \cos wt$. This pattern is now repeated every τ periods, τ being given by: $\cos wt = \cos w(t + \tau)$, which implies that $\tau = 2\pi/w$. A pattern which is repeated every ten periods will then have a frequency of $2\pi/10$.

Notice that the movements of y are bounded within $[-1, 1]$. We may modify this interval by multiplying the trigonometric function by a scalar ρ known as *amplitude*, so as to obtain variation of y_t within $[-\rho, \rho]$. Next, y_t is set here in such a way that it reaches its maximum for $w = 0$. We may shift this peak by introducing a *phase* parameter θ according to:

$$y_t = \rho \cos(wt + \theta) \tag{1.4}$$

Developing (1.4) using standard trigonometric formulas, it is readily obtained that y_t may also be written as:

$$y_t = \alpha \cos wt + \beta \sin wt, \tag{1.5}$$

with $\alpha^2 + \beta^2 = \rho^2$ and $\tan(\beta/\alpha) = \theta$.

The trigonometric formulation (1.4) may be put into a difference equation form. Saying for convenience $\theta = 0$, y_t may be written as: $y_t = .5\rho(e^{iwt} + e^{-iwt}) = .5\rho((e^{iw})^t + (e^{-iw})^t)$. This expression has the form of the solution of a difference equation, e^{iw} and e^{-iw} being the roots of the characteristic equation:

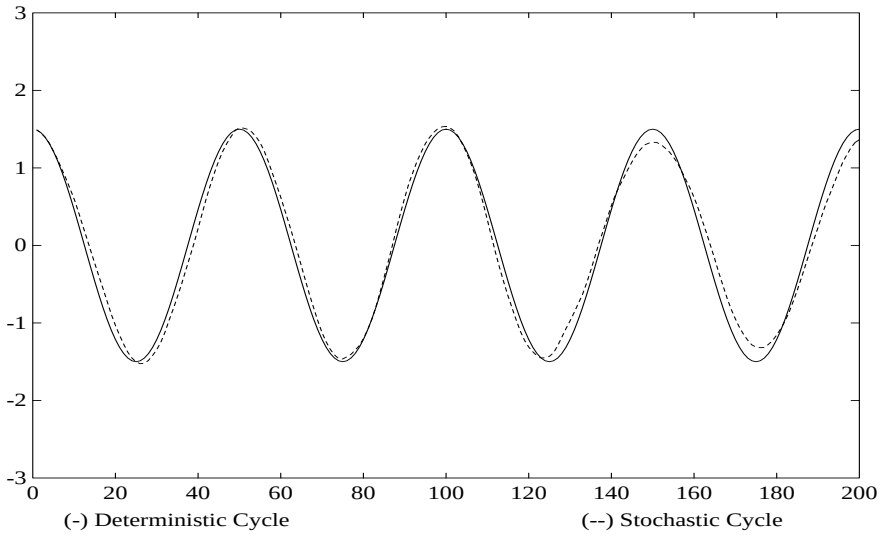
$$r^2 - 2\cos wr + 1 = 0.$$

which is associated with the second order difference equation: $y_t - 2\cos w\ y_{t-1} + y_{t-2} = 0$. The parameters ρ and θ in (1.4) are given by the initial conditions y_1 and y_2.

A deterministic cycle generated by (1.4) with $\rho = 1.5$, $\theta = 0$, and $w = 2 * \pi/50$ is displayed on figure 1.5. This specification for a cyclical pattern shares the problem of the deterministic modelling: unless the phenomenon obeys these exact laws, it is quite unlikely to provide a satisfactory representation. The exactness nature of the function does not able the movement to evolve over time. The forecasts are the exact replication of the past. Some flexibility may be introduced by considering some random disturbances along the waving path, while some generality may be obtained by letting the process being governed by two coefficients ϕ_1 and ϕ_2. We have then: $y_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} = e_t$.

This process is another example of a linear stochastic process. It is termed AutoRegressive of order 2, AR(2) in short. Again, setting $V(e_t)$ to zero and $\phi_1 = -2\cos w$, $\phi_2 = 1$ makes the deterministic function (1.4) recovered. For $V(e_t)$ as small as .0001, we can see on figure 1.5 that, though fluctuating, the stochastic cycle is close enough to the deterministic one.

Figure 1.5: Cycles



1.2.3 Modelling seasonality

Suppose now the series is observed s times a year. Patterns related to the s seasons may be modelled in a naive way using dummy variables:

$$y_t = D_{1t}\gamma_1 + D_{2t}\gamma_2 + \cdots + D_{st}\gamma_s \tag{1.6}$$

where D_{jt} takes the value 1 if $t = j, j+s, j+2s, ...$, zero otherwise, and the γ-variables represent the size of the seasonal fluctuations. This modelling implies that $y_t = y_{t+s}$: the seasonality is constant over the years. This in turn implies that the sum of the seasonal fluctuations over s consecutive time periods is constant: $\sum_{i=0}^{s-1} y_{t-i} = \sum_{i=0}^{s-1} y_{t-s-i} = c$. An important requirement is that this sum is null: $c = 0$. If it would not, then the constant c would be discarded from the seasonal process and incorporated into the level of the series. We have now a purely seasonal series represented as:

$$\sum_{i=0}^{s-1} y_{t-i} = 0 \tag{1.7}$$

which is another way of expressing (1.6). Yet another way would be to use trigonometric forms to characterize the seasonal patterns. Suppose the series is observed 3 times a year: then, we have $y_t + y_{t-1} + y_{t-2} = 0$, with the associated characteristic equation: $r^2 + r + 1 = 0$. Solving, we get the pair of complex roots $r_1 = e^{i2\pi/3}$ and $r_2 = e^{-i2\pi/3}$. Thus the solution of the differential equation yielded by the requirement that the seasonality sums to zero over a year is in that case $y_t = r_1^t + r_2^t$, which can be also written:

$$y_t = \cos(2\pi t/3). \tag{1.8}$$

This example illustrates some equivalence between the dummy variable modelling, the differential form, and the use of trigonometric functions. All capture the periodic nature of the seasonality, but the fluctuations are excessively restricted. The pattern is clearly enforced to repeat itself every year, and this limitation due to the deterministic character of the representation is very strong: in the 20's, that is more than 70 years ago, there was a general agreement of the idea that the seasonal components of time series change in time (for a review of the historical developments see Bell and Hillmer (1984)).

Small deviations from this strict model specification may be allowed by making the relationship subject to a random shock in each period:

$$\sum_{i=0}^{s-1} y_{t-i} = e_t$$

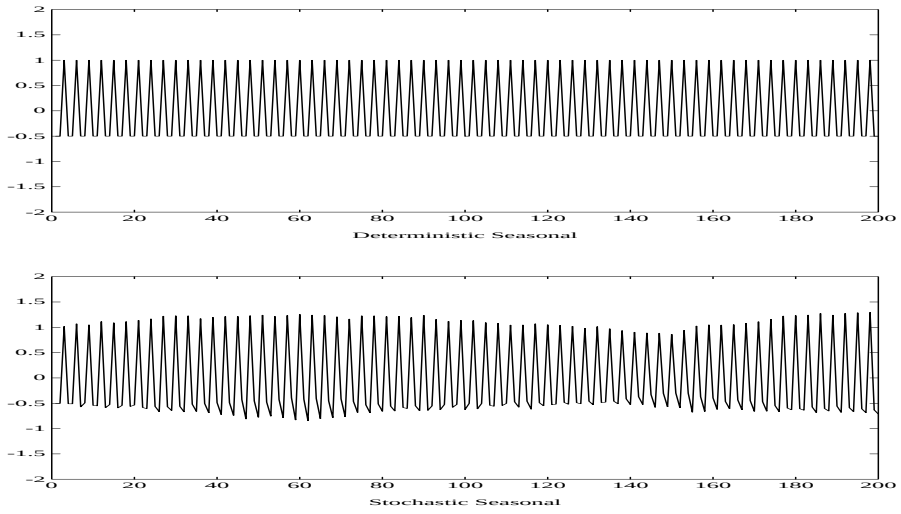
This stochastic process reduces to the deterministic specification (1.7) when $V(e_t) = 0$. Figure (1.6) displays a seasonal process generated by (1.8) and its stochastic extension. Although the noise variance has been set very low (.000625), we can see that the seasonal movements are much less systematic, evolving instead in time.

More generally, we can allow the deviation from zero to be correlated and consider :

$$\sum_{i=0}^{s-1} y_{t-i} = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}. \tag{1.9}$$

The right hand side of (1.9) is called Moving Average process of order q and denoted

Figure 1.6: Seasonal Fluctuations



MA(q).

Representations involving stochastic models may thus be seen as nesting the deterministic formulations. There are much more flexible, since the processes are not constrained to strictly follow a specific path but are allowed to evolve over time according to some random disturbances. Deterministic models are thus a particular case of stochastic modelling.

1.3 Tools and Concepts for Time Series Modelling

1.3.1 Time series operators

A very used tool in time series analysis is given by the backward operator B. It is defined according to:

$$By_t = y_{t-1}.$$

Applied to y_{t-1}, the backward operator yields y_{t-2}, so that in general we will have: $B^\tau y_t = y_{t-\tau}$. This definition is completed by defining B^0 such that $B^0 y_t = y_t$. Negative

powers imply forward shifts, so that we associate to B the forward operator F defined as $F = B^{-1}$ and $Fy_t = y_{t+1}$.

A polynomial in the backward operator takes the form: $\phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$. This polynomial is said to be of order p, and solving $\phi(z) = 0$, p roots are obtained. For example, the root of $\phi(z) = 1 + \phi z = 0$ is $z = -1/\phi$. A distinction will be made between the roots which are *outside* the unit circle, that is the roots of modulus greater then 1, and the roots on the unit which have a modulus of 1. Here, the root of $1 + \phi z$ lies outside the unit circle when $\mid \phi \mid < 1$.

A special kind of polynomial is given by the differencing operator Δ, defined by: $\Delta = 1 - B$. We have thus: $\Delta y_t = y_t - y_{t-1}$. Powers may also be taken: $\Delta^d = (1-B)^d$, and we will have for example: $\Delta^2 y_t = \Delta(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2} = (1 - 2B + B^2)y_t = (1-B)^2 y_t$. A subindex may be used to indicate the differencing *order*: $\Delta_d = 1 - B^d$ and $\Delta_d y_d = y_t - y_{t-d}$.

1.3.2 Stationarity

Time series modelling is thus concerned with characterizing in a satisfactory way the behavior of the series under analysis. We have seen that it is convenient to take the observations as the output of a probabilistic function: $y_t \sim f_t(y_t)$. However, there would be no way to approximate the function $f_t(.)$ if each observation were related to a single function $f_t()$. We thus have to introduce the concept of stationarity:

A process is said to be strictly stationary if the joint probability distribution of a set of r observations $y_1, \cdots, y_r$ is equal to the joint distribution of a set $y_{k+1}, ..., y_{k+r}$ for all k.

Then, the marginal distribution of the y_t will be such that: $f(y_t) = f(y_{t+k})$, and for any couple (y_t, y_{t+k}), the joint distribution will be independent of t: $f(y_t, y_{t+k}) = f(y_{t+l}, y_{t+l+k})$. In practice, it is enough to consider the less restrictive concept of weak stationarity which only requires the following conditions to be satisfied:

$$E(y_t) = \mu$$
$$E[(y_t - \mu)^2] = Var(y_t) = \gamma(0)$$

$$E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma(k)$$

for $k = 1, 2, ...$ Stationary processes thus present constant mean μ and variance γ_0, while the *autocovariances* γ_k are independent of the time index t; they are only related to the time-distance k. Given that normal distributions are fully characterized by the first two moments, a weakly stationary process which is normally distributed will also be a strictly stationary process.

1.3.3 The Wold decomposition

The Wold decomposition theorem states that any second-order stationary process y_t may be expressed as the sum of a deterministic function plus an infinite sequence of uncorrelated random variables:

$$\begin{aligned} y_t &= c(t) + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots \\ &= c(t) + \sum_{i=0}^{\infty} \psi_i e_{t-i} \\ &= c(t) + \psi(B) e_t \end{aligned}$$

with $\psi_0 = 1$ and $\sum_{j=0}^{\infty} \psi_i^2 < \infty$. Hence, any second-order stationary process may be expressed as the sum of a deterministic component plus a linear stochastic process. The deterministic part usually corresponds to the mean of the series, and the stochastic part is an infinite moving average process.

Chapter 2

Linear Time Series Models

2.1 AutoRegressive Integrated Moving Average Models

Given the straightforward availability of the mean of any stationary process, modelling the series y_t may be seen as the problem of finding an adequate representation of the process $\psi(B)e_t$. As the polynomial $\psi(B)$ is infinite, it would not be possible to estimate all the coefficients ψ_i. An approximation must be built in order to limit the number of parameters to estimate. The class of ARMA models provides a useful tool for approximating the stochastic part in the Wold decomposition with relatively few parameters. Without loss of generality, in the remainder of the discussion we will assume that the mean of the observed series has been removed.

2.1.1 Autoregressive models

Suppose the ψ-coefficients are such that: $\psi_i = \psi^i$ and $\mid \psi \mid < 1$. Then, y_t is given by:

$$\begin{aligned} y_t &= (1 + \psi B + \psi^2 B^2 + \cdots + \psi^n B^n)e_t = \\ &= \frac{1 - \psi^{n+1} B^{n+1}}{1 - \psi B} e_t = \\ &= \frac{1}{1 - \psi B} e_t, \end{aligned}$$

since $\lim_{n\to\infty} \psi^n = 0$. Rearranging, we get the AR(1) process:

$$(1 - \psi B)y_t = e_t.$$

or

$$y_t = \psi y_{t-1} + e_t \tag{2.1}$$

The variance of the AR(1) process is given by:

$$\begin{aligned} E[y_t^2] &= E[(e_t + \psi e_{t-1} + \cdots)^2] = \\ &= E[(e_t^2 + \psi^2 e_{t-1}^2 + \cdots)] = \\ &= \sigma^2 \frac{1}{1-\psi^2} = \\ &= \gamma(0), \end{aligned}$$

where use has been made of the fact that the e_t's are uncorrelated. For the covariances $\gamma(k)$ between y_t and y_{t-k}, $k = 1, 2, ...$, multiplying both sides of (2.1) by y_{t-k}, we have:

$$E[y_t y_{t-k}] = \psi E[y_{t-1} y_{t-k}] + E[e_t y_{t-k}],$$

so that we obtain: $\gamma(k) = \psi\gamma(k-1)$ and eventually $\gamma(k) = \psi^k\gamma(0)$.

The AR(1) model is thus able to describe a process whose innovation representation has coefficients which decay exponentially. This model may be extended to describe more general patterns: for example, discussing possible models describing a cyclical behavior, we have introduced the AR(2) model. More generally, an autoregressive process of order p is written as:

$$y_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} = e_t \tag{2.2}$$

and is denoted AR(p).

2.1.2 Moving Average models

Assume now that in (2.2) the ϕ_i-coefficients are such that $\phi_i = \theta^i$ with $\mid \theta \mid < 1$. Then, a similar reasoning than in the previous subsection yields:

$$\begin{aligned} e_t &= (1 + \theta B + \theta^2 B^2 + \cdots) y_t \\ &= \frac{1}{1 - \theta B} y_t, \end{aligned}$$

since $\lim_{n \to \infty} \theta^n = 0$. Rewriting, we obtain the MA(1) process:

$$y_t = (1 - \theta B) e_t.$$

The MA(1) is thus able to provide a short representation of an infinite but convergent AR process. However, that requires $\mid \theta \mid < 1$, otherwise the MA(1) cannot have a convergent AR equivalence. The condition $\mid \theta \mid < 1$ is known as the invertibility condition.

The variance of the MA(1) process is obtained as:

$$\begin{aligned} E[y_t^2] &= E[(e_t - \theta e_{t-1})^2] = \\ &= (1 + \theta^2)\sigma^2 = \\ &= \gamma(0). \end{aligned}$$

For the autocovariances, we have:

$$E[y_t y_{t-k}] = \gamma(k) = E[(e_t - \theta e_{t-1})(e_{t-k} - \theta e_{t-k-1}],$$

so that $\gamma(1) = -\theta\sigma^2$ and $\gamma(k) = 0$ for $k > 1$.

Generalizing the MA(1) process, a moving average of order q is written as:

$$y_t = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q},$$

and is denoted MA(q).

2.1.3 Mixed processes

Both AR(1) and MA(1) models are able to represent in a parsimonious way some particular processes. It is possible to extend their field of application by combining them, so as to obtain an ARMA(1,1) model defined as:

$$(1 + \phi_1 B)y_t = (1 + \theta_1 B)e_t. \tag{2.3}$$

The ARMA(1,1) model may be further generalized to the ARMA(p,q) model by combining an AR(p) with a MA(q) process:

$$y_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}. \tag{2.4}$$

Defining the polynomials $\phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_p B^p$, then (2.4) can be written in the compact form:

$$\phi(B)y_t = \theta(B)e_t. \tag{2.5}$$

Notice that an AR(p) process could also be written as an ARMA(p,0), in the same way that MA(q) process could be denoted ARMA(0,q).

2.1.4 Non Stationarity and Integrated Processes

Condition for nonstationarity

The Wold decomposition which provides the ground for the ARMA modelling requires that the observed series is stationary. In practice, very few economic time series are stationary. Nonstationarity may be due to a nonconstant mean as for example in the deterministic trend plus noise model:

$$E[y_t] = E[a + bt + e_t] = a + bt, \tag{2.6}$$

or to a divergent variance, as in the random walk specification: writing

$$\begin{aligned} y_t &= y_{t-1} + e_t = \\ &= y_{t-2} + e_{t-1} + e_t = \\ &= y_0 + e_t + e_{t-1} + \cdots + e_1 \end{aligned} \tag{2.7}$$

the variance of the random walk process is obtained as:

$$V[y_t] = t\sigma^2.$$

In an ARMA model, nonstationarity is equivalent to having the roots of the AR polynomial lying on the unit circle. Consider for example the AR(1) process $(1 - \phi B)y_t = e_t$. The AR polynomial is $\phi(B) = 1 - \phi B$, and the root of $\phi(z) = 0$ is readily obtained as $z = 1/\phi$. So the process will be nonstationary when $\mid \phi \mid = 1$. That case corresponds to the one yielding an infinite variance in (2.7), while $\mid \phi \mid < 1$ corresponds to the AR(1) process (2.1) whose variance is given by $\sigma^2/(1 - \phi^2)$.

In general, an AR(p) process will be stationary when the p roots of $\phi(B) = 0$ are of modulus greater than one, while roots of modulus one imply nonstationarity.

Nonstationarity is not influenced by the MA part of the processes. MA processes are only concerned by the invertibility condition which insures that they are equivalent to an infinite but convergent AR representation. The invertibility condition is stated similarly to the stationarity condition, that is a process is invertible if the q roots of the MA polynomial $\theta(B)$ satisfying $\theta(B) = 0$ are of modulus greater than one.

Stationarity inducing transformations

Consider for example the random walk process: it is seen in (2.7) that it is a nonstationary process. However, defining $z_t = y_t - y_{t-1} = \Delta y_t$, it is readily obtained that z_t is a white noise process which is stationary. First order differencing has thus made stationary the random walk y_t. Processes made stationary by first order differencing are said to be *integrated* of order 1, and denoted I(1). The term *integrated* may be understood by observing that in (2.1) the random walk variable y_t is made up of a starting condition plus a cumulative sequence of random variables.

Processes characterized by a time-trend have been seen in (2.6) to be also nonstationary. However, taking the first difference of $y_t = a + bt$ yields a constant: $\Delta y_t = b$:

hence the difference operator is able to achieve stationarity for deterministic functions. Also, the process y_t may need to be differenced d-times before yielding a stationary process. In this case y_t will be said to be integrated of order d and denoted $y_t \sim I(d)$. It is readily seen that differencing d times would reduce to a constant a polynomial in t of order $d+1$.

Differencing may also be needed at lags higher that one. Consider for example the deterministic process that has been discussed in (1.8) as a representation of the seasonal behavior of a series observed 3 times a year:

$$y_t = \cos 2\pi t/3. \tag{2.8}$$

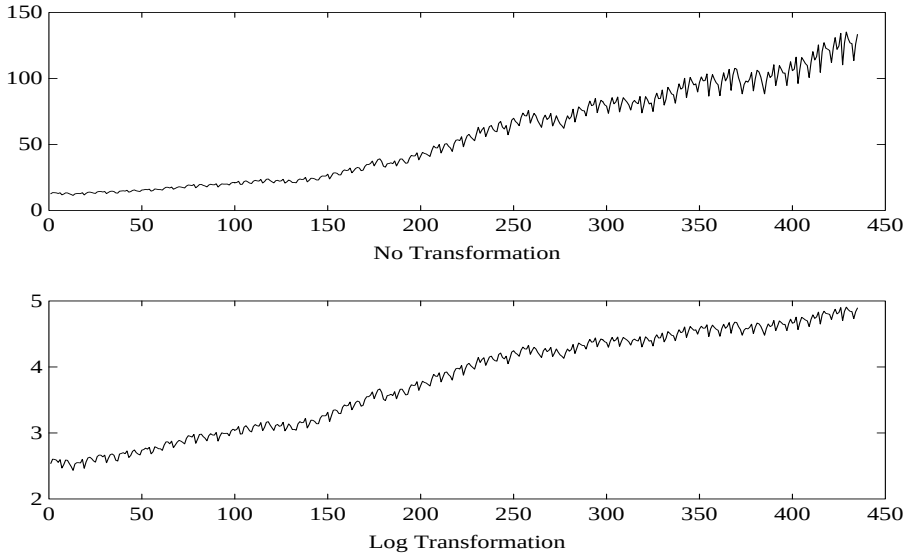
This series is clearly not stationary: $E(y_t)$ depends on t. Taking $z_t = y_t - y_{t-3} = \Delta_3 y_t$, we get:

$$\begin{aligned} \Delta_3 y_t &= \cos 2\pi t/3 - \cos\left[2\pi(t-3)/3\right] = \\ &= \cos 2\pi t/3 - \cos\left[2\pi t/3 - 2\pi\right] = 0. \end{aligned}$$

Differencing may also be considered after the series has been log-transformed. In that case, instead of $z_t = \Delta y_t$, one would consider $z_t = \Delta \log y_t$. This last expression gives a close approximation to the growth rates of the series: $\Delta \log y_t \simeq dy_t/y_t$. Thus, the use of logs together with differencing means that while the magnitude of the changes in the series depend on time, the series growth rate is stable. This is particularly adequate for series measured in current prices, which are sensitive to inflation in a geometrical way. Typically, these series present an increasing level over time, with fluctuations which are increasing as the level increases. An example is given by the series of the monthly US turnover index (in values; 1960-1 1995-3, 435 observations) displayed on figure 2.1. What is happening in these cases can be easily understood with the following example taken from Bell (1995) where y_t is such that $y_t = (1+i)^t y_0$. Then, it is readily checked that:

$$\begin{aligned} y_t - y_{t-1} &= (1+i)^{t-1} i y_0 \\ \log y_t - \log y_{t-1} &= log(1+i) \simeq i \end{aligned}$$

Figure 2.1: US Turnover Index (1960-1 1995-3)



If the changes in y_t are characterized by an exponential increase, those in $\log y_t$ are instead constant over time. Figure 2.1 displays also the log of the series: it is seen that the fluctuations of the series around its local level are more regular. Looking to changes in the series and to the series growth rates in figures 2.2 clearly indicates that the first is affected by an exponential increases in time while the second is much more stable.

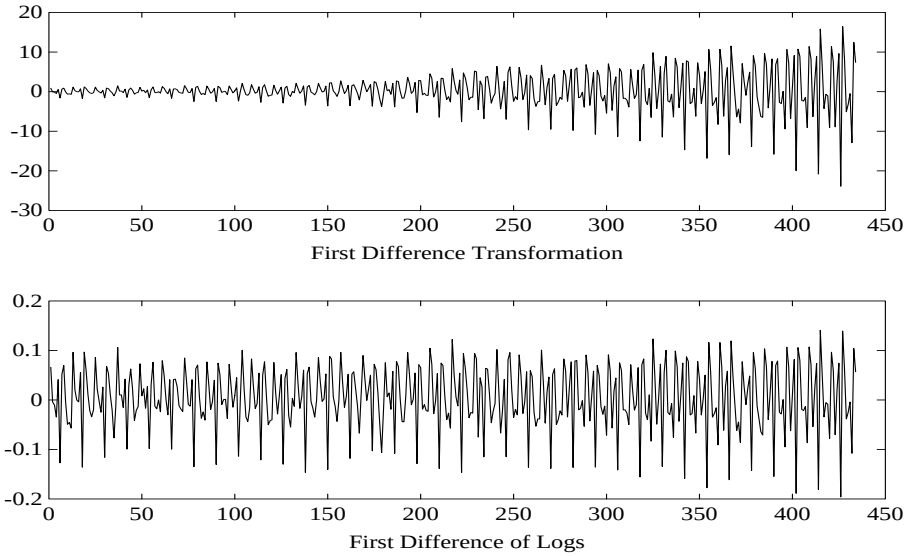
2.1.5 ARIMA models

It is now possible to set the general class of AutoRegressive Integrated Moving Average models, which consists of a mixed ARMA process for the series made stationary by differencing. For nonseasonal time series, these models are specified as:

$$\phi(B)\Delta^d y_t = \theta(B)e_t,$$

where $\phi(B)$ and $\theta(B)$ are polynomials satisfying the stationary and invertibility conditions respectively, $\Delta = 1 - B$ is the difference operator, d denotes the minimum

Figure 2.2: US Turnover Index (1960-1 1995-3)



number of differences required to render the series stationary and e_t is a white noise variable. If the polynomials $\phi(B)$ and $\theta(B)$ are respectively of order p and q, then y_t is said to follow an ARIMA(p,d,q) model.

Box and Jenkins (1970) extended the ARIMA models to cover seasonal time series. They started from the point that if a time series is observed with a frequency of s observations per year, then observations which are s periods apart should be similar. For example, if y_t represents a monthly time series, then it is expected that observations for the same month in successive years are related. An ARIMA model relating the observation y_t to the previous y_{t-s}, y_{t-2s}, $\cdots$, can simply be written as:

$$\Phi(B^s)\Delta_s^D y_t = \Theta(B^s)\alpha_t,$$

where $\Phi(B^s)$ and $\Theta(B^s)$ are finite polynomials in B^s, of order respectively P and Q, which satisfy the stationarity and invertibility condition, and $\Delta_s = 1 - B^s$ is the seasonal differencing operator. This nonstationary operator has its roots $e^{i2k\pi/s}$, $k = 0, 1, \cdots, s-1$, evenly spaced on the unit circle. The parameter D represents the minimum number of differences required to make the series stationary. It is usually

assumed that the relationship between the same month in successive years is common to all months, so the parameters of the polynomials $\Phi(B^s)$ and $\Theta(B^s)$ are constant.

Beside this relationship, for a monthly time series for example, a relationship is expected to occur between successive months in the same year. For this reason, the variable α_t will not be uncorrelated. Box and Jenkins account for the relationship between successive observations in a natural way, assuming that α_t itself follows the nonseasonal model:

$$\phi(B)\Delta^d \alpha_t = \theta(B) a_t.$$

It then comes out that the series y_t follow a multiplicative model specified as:

$$\phi(B)\Phi(B^s)\Delta^d \Delta_s^D y_t = \theta(B)\Theta(B^s) a_t.$$

where a_t is a normally distributed white noise. This ARIMA model is said to be of order $(p,d,q)(P,D,Q)_s$. In practice, this representation has the advantage of involving relatively few parameters and has proved to adequately approximate many seasonal time series. Multiplicative seasonal ARIMA models have been extensively used in the statistical literature, for applied research and for theoretical investigations.

2.2 Properties of ARIMA models

2.2.1 Time domain analysis

Given that the series is modelled as a linear combination of a normal variable, the stochastic properties of the series are fully described by the first two moments. The first moment is given by the mean of the process, while the second moments may be easily obtained from the AutoCovariance Generating Function (ACGF). This function will be denoted $\gamma(B)$:

$$\gamma(B) = \sum_{i=-\infty}^{\infty} \gamma_i B^i,$$

where γ_0 corresponds to the variance of the process while the γ_i, $i \neq 0$, gives the lag-i autocovariances. Dividing them by γ_0 yields the lag-i autocorrelations $\rho_i = \gamma_i/\gamma_0$.

Notice that the ACGF is symmetric, so it can be written as:

$$\gamma(B) = \gamma_0 + \sum_{i=1}^{\infty} \gamma_i (B^i + F^i)$$

For a stationary ARMA process, the ACGF is obtained as:

$$\gamma(B) = V_a \frac{\theta(B)\theta(F)}{\phi(B)\phi(F)}. \tag{2.9}$$

Example: ARMA(1,1)

Consider the following model:

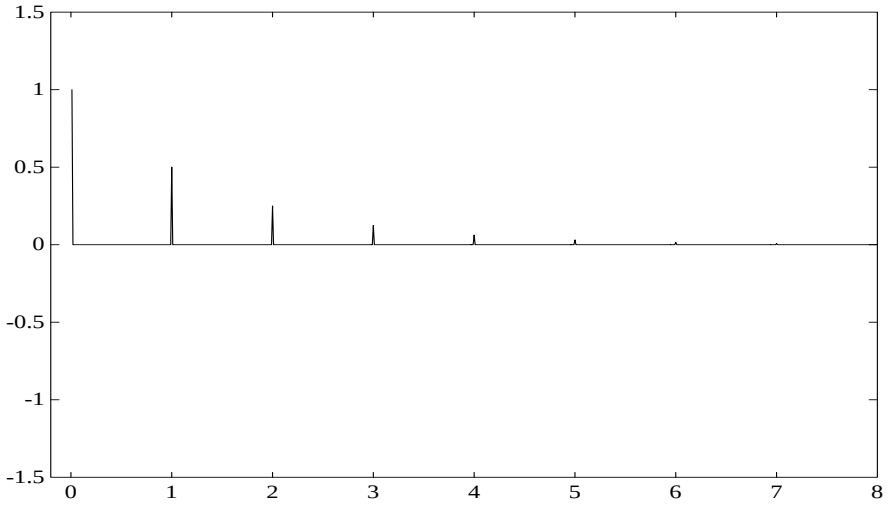
$$(1 - \phi B) y_t = (1 - \theta B) a_t \tag{2.10}$$

which reduces to the AR(1) if θ is set to zero, to a MA(1) if $\phi = 0$. The ACGF of model (2.10) can be derived according to:

$$\begin{aligned}
\gamma(B) &= V_a \frac{(1 - \theta B)(1 - \theta F)}{(1 - \phi B)(1 - \phi F)} = \\
&= V_a [1 + \theta^2 - \theta(B + F)](1 + \phi B + \phi^2 B^2 + \cdots). \\
&\quad .(1 + \phi F + \phi^2 F^2 + \cdots) = \\
&= V_a (1 + \theta^2 - \theta(B + F)) \frac{1}{1 - \phi^2} [1 + \phi(B + F) + \\
&\quad + \phi^2 (B^2 + F^2) + \cdots] = \\
&= V_a \frac{1}{1 - \phi^2} [1 + \theta^2 - 2\phi\theta + \{(1 + \theta^2)\phi - \theta(1 + \phi^2)\}. \\
&\quad .\{(B + F) + \phi(B^2 + F^2) + \cdots + \phi^{n-1}(B^n + F^n)\}]
\end{aligned}$$

Setting $\theta = 0$ in this last equation directly yields the ACGF of the AR(1) as found in subsection 2.1.1: $\gamma(k) = V_a \phi^k / (1 - \phi^2)$. The autocorrelations are then obtained as:

Figure 2.3: ACGF of an AR(1) process, $\phi = .5$



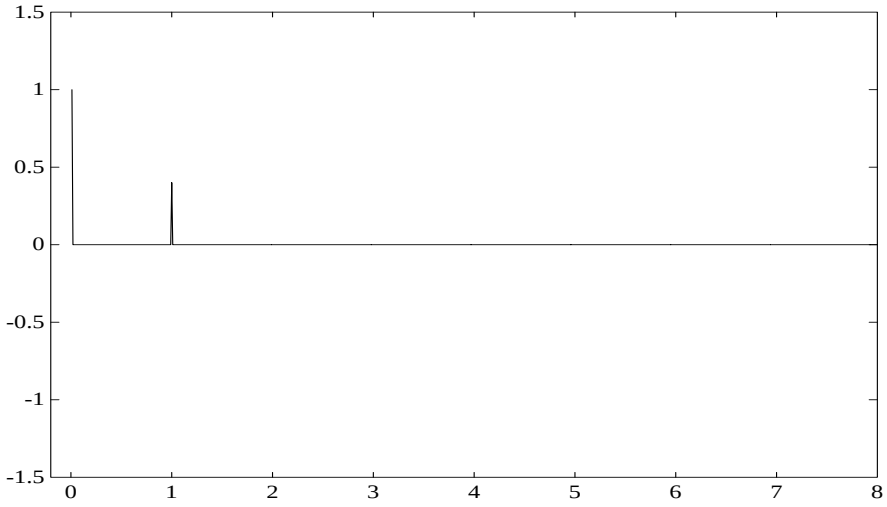
$\rho(k) = \gamma(k)/\gamma(0) = \phi^k$. Figure 2.3 shows the first 8 autocorrelations for an AR(1) process with $\phi = .5$. A fast convergence to zero as the time-lag increases is clearly appearent.

Similarly, setting $\phi = 0$ in the ACGF of the ARMA(1,1) leads to the covariance structure of an MA(1) found in 2.1.2. For the autocorelations, it is readily seen that $\rho(1) = -\theta/(1-\theta^2)$ while $\rho(k) = 0$ for $k > 1$. The first 8 autocorrelations for the MA(1) with $\theta = -.5$ are displayed on figure 2.4.

2.2.2 Frequency domain analysis

The study of the autocorrelations of a process gives information about the dynamic of a time series: it resumes the relationship between y_t to its past. However, we have previously seen that another feature of interest of time series is the regularity in the movements displayed by the series. For example, the process $y_t = \cos 2\pi t/5$ repeats itself every 5 periods; that is, it evolves with a frequency of $2\pi/5$. Because we are dealing with stochastic processes, interpretation will not be so straightforward. A convenient tool for analysing time series in the frequency domain is given by the *spectrum*. For a stationary stochastic process, the power spectrum is defined as:

Figure 2.4: ACGF of an MA(1) process, $\theta = -.5$



$$f(w) = \frac{1}{2\pi} \sum_{\lambda=-\infty}^{\infty} \gamma_\lambda e^{-i\lambda w}, \tag{2.11}$$

where $w \in [-\pi, \pi]$ is a frequency expressed in radians, i is the complex verifying $\sqrt{i} = -1$, and γ_λ is the lag-λ autocorrelation. Given that $\gamma_\lambda = \gamma_{-\lambda}$ in our framework, another writing of (2.11) may be:

$$f(w) = \frac{1}{2\pi}[\gamma_0 + 2\sum_{\lambda=1}^{\infty} \gamma_\lambda \cos \lambda w]. \tag{2.12}$$

Hence, for the applications we are concerned with, $f(w)$ will be symmetric around zero, and so it will be enough to consider the frequencies within $[0, \pi]$. Sometimes the power spectrum is divided by γ_0, defining the spectral density in a similar way than the correlations are obtained from the covariances.

All the different movements in the series are described over $[-\pi, \pi]$. The spectrum $f(w)$ of a time series describes the contribution of the movements with frequency w to the variance of the series. Summing these contributions, we obtain thus:

$$\int_{-\pi}^{\pi} f(w)dw = \gamma_0. \tag{2.13}$$

Consider the white noise process: not being correlated, its power spectrum is given by $\sigma^2/2\pi$. Result (2.13) is then immediately checked in that case. The lag-k covariances may also be recovered from the spectrum $f(w)$, and in particular:

$$\int_{-\pi}^{\pi} e^{ikw} f(w)dw = \gamma_k. \tag{2.14}$$

The *spectral generating function* (sgf in short) will prove to be more useful for our analysis. It is obtained by replacing the covariances $\gamma(\lambda)$ by the ones yielded by the ACGF $\gamma(B)$. Hence the sgf is simply defined by $\gamma(e^{-iw})$, the lag operator B being replaced by e^{-iw}, and from (2.9) we have for a stationary ARMA model:

$$g(w) = \gamma(e^{-iw}) = V_a \frac{\theta(e^{-iw})\theta(e^{iw})}{\phi(e^{-iw})\phi(e^{iw})}. \tag{2.15}$$

The power spectrum and the sgf are related through: $2\pi f(w) = g(w)$. Hence, once the autocovariance function of a given model has been derived, the computation of the power spectrum or of the sgf is trivial. It is the *Fourier transform* which relates the power spectrum to the autocovariance function (see 2.11), while the inverse *Fourier transform* of the power spectrum gives the autocovariances (see 2.14).

Example: Consider the ARMA(1,1) model:

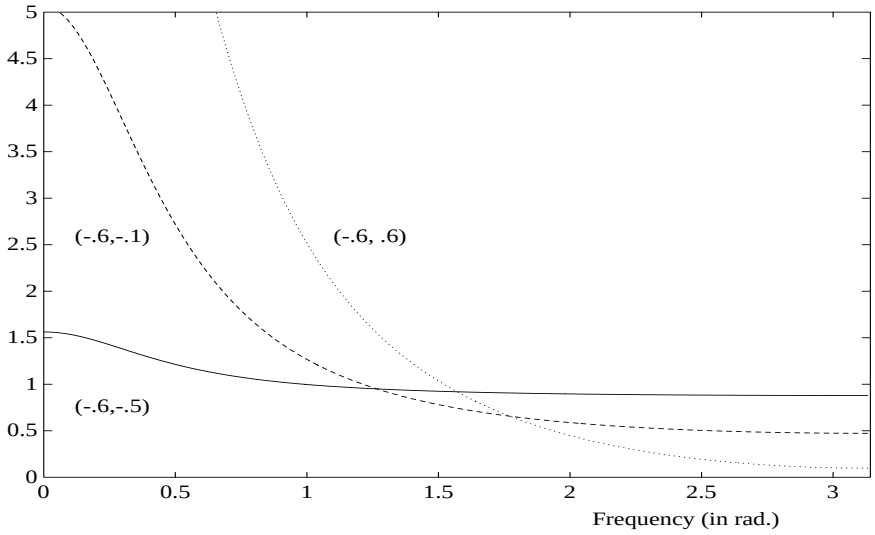
$$y_t + \phi y_{t-1} = a_t + \theta a_{t-1}. \tag{2.16}$$

The sgf is obtained as:

$$g(w) = V_a \frac{1 + \theta^2 + 2\theta \cos w}{1 + \phi^2 + 2\phi \cos w}. \tag{2.17}$$

This sgf is displayed for $w \in [0, \pi]$ on figure 1.11 for $(\phi, \theta) = (-.6, -.5)$, $(-.6, -.1)$ and $(-.6, .6)$. In the first case, due to the near cancellation of the factors in the AR

Figure 2.5: Spectra of an ARMA(1,1) for diff. values of (ϕ, θ)



and MA part, model (2.16) is close to a white noise process, and so the spectrum is flat. In the second and third cases, we observe a peak at the low-frequencies, which correspond to some long-term movement. This peak is clearly larger for the parameters (-.6,.6) than for the parameters (-.6,-.1). This reflects some more variability at the low-frequencies, and thus may be interpreted a sign of unstability in the long-term movements generated by an ARMA(1,1) with $(\phi, \theta) = (-.6, .6)$; conversely, the long-term movements generated by the ARMA(1,1) with parameter value (-.6,-.1) are more stable. In general, the width of the spectrum peaks is an indication of unstability.

Next, let us suppose that the process 2.16 is noninvertible, that is $\theta = 1$. The spectrum of s_t would thus be given by:

$$g(w) = V_a \frac{2 + 2\cos w}{1 + \phi^2 + 2\phi \cos w},$$

and it is easily seen that $g(w)$ displays a zero at the π-frequency. In general, noninvertible processes yield spectra with zero values in the range $[0, \pi]$.

Chapter 3

Building Linear Stochastic Models

3.1 Introduction

We now examine how the ARIMA models may be used to describe the behavior of time series. A general methodology for selecting a model has been developed by Box and Jenkins (1970), still refered to as the Box-Jenkins methodology. It is made up of three successive steps. First, the order of differencing necessary to render the series stationary is chosen, together with the order of the autoregressive part and of the moving average part. This is known as the *identification* stage, and it yields a tentative model. This tentative model is estimated in the second phase. To be satisfying, no underlying structure in the data must be left. This suppose that the residuals do not contain any systematic part, that is that they are random. Test of residuals randomness are performed in the third stage. If the diagnostic checks indicates a failure of the tentative model to capture the main features of the series under analysis, then the initial model specification is modified, a new model is estimated and evaluated. This procedure is repeated until a satisfying model is found. We now detail each step in this model selection procedure.

3.2 Identification

The main tools for identifying a model are provided by the AutoCorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). Using the quantities

$$\hat{\gamma}(k) = T^{-1} \sum_{t=k+1}^{T} (y_t - \overline{y})(y_{t-k} - \overline{y}) \tag{3.1}$$

where k=0,1,2,..., and $\overline{y}$ is the sample mean evaluated as $\overline{y} = T^{-1} \sum_{t=1}^{T} y_t$. This yields the sample variance $\hat{\gamma}(0)$ and the sample autocovariances $\hat{\gamma}(k)$, k=1,2... . The sample autocorrelations are then obtained as: $r(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$.

The PACF is given by the set of the k coefficients ϕ_{11}, ϕ_{22}, ..., ϕ_{kk} in the k autoregressions of order 1,2,...,k. For example ϕ_{11} is the first lag coefficient of the AR(1) model fitted to y_t, ϕ_{22} the second lag coefficient of the AR(2) model fitted to y_t, and so on. These coefficients may be obtained by fitting successively AR(j) models, $j = 1, ..., k$ and picking up every time the last coefficient. Another procedure consists in using the Yule-Walker equations which are derived as follows.

Fitting an AR(1) model to y_t, we have: $y_t = \phi_{11} y_{t-1} + e_t$; multiplying by y_{t-1} and taking expectation, we get:

$$\gamma_1 = \phi_{11} \gamma_0$$

By the same way, fitting the AR(2) process to y_t: $y_t = \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + e_t$, multiplying by y_{t-1}, by y_{t-2}, and taking expectation in the two cases, we get:

$$\begin{aligned} \gamma_1 &= \phi_{21}\gamma_0 + \phi_{22}\gamma_1 \\ \gamma_2 &= \phi_{21}\gamma_1 + \phi_{22}\gamma_0. \end{aligned}$$

Solving these two equations yields ϕ_{21} and the variable of interest ϕ_{22}. Repeating this operation k-times, we obtain the set of the *Yule-Walker* equation:

$$\begin{aligned} \gamma_1 &= \phi_{k1}\gamma_0 + \phi_{k2}\gamma_1 + \cdots + \phi_{kk}\gamma_{k-1} \\ \gamma_2 &= \phi_{k1}\gamma_1 + \phi_{k2}\gamma_0 + \cdots + \phi_{kk}\gamma_{k-2} \\ &\vdots \\ \gamma_k &= \phi_{k1}\gamma_{k-1} + \phi_{k2}\gamma_{k-2} + \cdots + \phi_{kk}\gamma_0 \end{aligned}$$

Dividing every equation by γ_0, the Yule-Walker equations may also be expressed in terms of the autocorrelations ρ_k. In pratice, the computations are done using the estimated autocorrelations $r(k)$ to produce the partial autocorrelations estimators $\hat{\phi}_{kk}$.

The first step in determining the order of a possible ARIMA(p,d,q) model is to identify d, the number of times the process must be differenced to become stationary. Although some tests have been constructed for that (see Dickey and Fuller (1979), Phillips (1987)), the ACF may also be used. Consider the random walk for which we had $y_t = y_0 + e_t + e_{t-1} + \cdots + e_1$. It is readily seen that $cov(y_t, y_{t-k}) = (t-k)V(e_t)$. Thus the theoretical autocorrelations will be $\gamma(k) = (t-k)/t$: the autocorrelations fall off slowly as k increases. In practice, the sample autocorrelations tend to follow the behavior of the theoretical autocorrelations, and so failure of the autocorrelations to die out quickly is a strong indication of nonstationarity. When this pattern is observed, the series is differenced until the resulting autocorrelations do converge rapidly. In pratice, the number of differences required to render a series stationary is either $d = 0, 1$, or 2.

Having identified d, it is possible to look for the orders p and q on the differenced series. The idea is to select a theoretical model whose autocorrelation pattern match the one estimated on the series. The main result in use for that can be stated as follow:

- The ACF of a MA(q) process shows a cut-off after lag q;
- The PACF of an AR(p) process exibits a cut-off after lag p.

and in a symmetric way:

- The PACF of a MA(q) process gradually tails off to zero;
- The ACF of a AR(p) process gradually goes to zero.

In practice, p and q take the value either 0,1,2, or 3. The identification procedure may become complicated if a mixed process is dealt with. Proceeding according to the scheme below may simplify the selection of a model specification:

Identification procedure

- Picking d: Plot the ACF of y_t. If it dies out rapidly $d = 0$. If not, plot the ACF of Δy_t. If it converges rapidly, $d = 1$. If not, check the ACF of $\Delta^2 y_t$. If it converges

rapidly, then take $d = 2$. Otherwise, check the series to see eventual problems before considering $d = 3$ which is very rarely met in practice.

- Picking p and q: Compute and plot the ACF and PACF of $\Delta^d y_t$. If the ACF truncates after some small q lags ($q \leq 3$), select an MA(q). If the PACF truncates after some small p lags ($p \leq 3$), choose an AR(p). If neither the ACF nor the PACF truncates after a small number of lags, a mixed model is to be considered. In that case it is better to keep looking for simple models, that is for model with low-order polynomials.

Because the sample autocorrelations and partial autocorrelations are estimated, they may show some departures from what they should theoretically be. For that reason, it is useful to consider confidence interval for these estimators. According to a result derived by Bartlett (1946), these estimators are asymptotically normally distributed with a variance which can be approximated by $(1/T)$. Thus, to test whether the sample autocorrelation of a process are zero for lags greater than q so as to identify an MA(q) process, it is enough the check whether the r_k, $k = q + 1, ...$, fall outside the confidence interval $[-2/\sqrt{T}, 2/\sqrt{T}]$. Similar reasoning is valid to test on the partial autocorrelations ϕ_{kk}, $k = p + 1, ...$, whether a process can be identified as an AR(p). Notice however that these bounds are obtained for a size of 5%, which means that for a white noise process, one over twenty autocorrelations may lie outside the confidence interval.

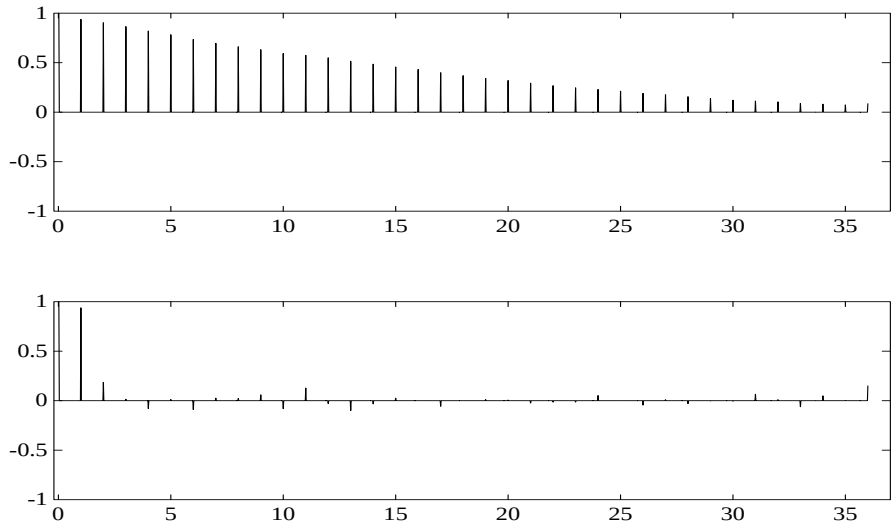
Identification of nonseasonal models

Example 1: Series POEM2640

The series POEM2640 of the employment in the portugal manufacture of bricks, tiles and construction products has been plotted on figure 1.2. Figure 3.1 shows the sample ACF and PACF computed on the raw observations (say z_t). The slow decay of the sample autocorrelations is very apparent. The series needs thus differencing to become stationary. In the same time, the PACF shows a very large first-lag partial autocorrelation, suggesting one single difference should be enough to obtain stationarity.

Figure 3.2 shows now the sample ACF and PACF of Δz_t, together with their confidence interval. Differencing z_t has completely modified the aspect of the sample ACF, which is now much more stable. However, the first lag of the ACF lies quite away of the

Figure 3.1: POEM2640: ACF and PACF of z_t

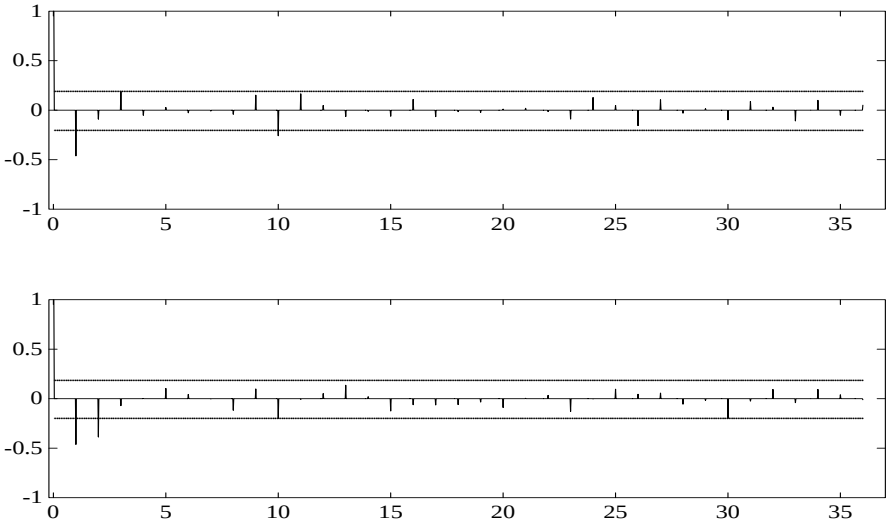


confidence interval. Depending on the pattern displayed the successive autocorrelations, this is an indication of either an AR process or of a MA process. To isolate one sensible specification, the PACF is of much help. Two significant first-lags partial autocorrelations suggest an AR(2) process for Δz_t. The model (2,1,0) is thus a candidate model.

Identification of seasonal models

The procedure described above can be straightforwardly applied to identify seasonal time series models presented in section 2.1.5. The attention is naturally reported to the autocorrelations at the seasonal lags: for example for a monthly time series, the lag-12 autocorrelation is of interest. Also, because the autocorrelation lag-distance for seasonal models is relatively large with respect to the number of observations available in standard applications, seasonal models are specified as simply as possible: the parameters P, D, Q characterizing the AR, the difference operator and the MA part of the seasonal model should be less or equal to one. Models for the seasonal part of time series most often belong to the class of the $(1,1,1)_s$, $(0,1,1)_s$, $(1,0,1)_s$ or $(1,1,0)_s$ models. One class of model which has proved to successfully represent a large number

Figure 3.2: POEM2640: ACF and PACF of Δz_t



of series is the so-called *airline* model:

$$\Delta\Delta_s y_t = (1 + \theta_1 B)(1 + \theta_s B^s)a_t \tag{3.2}$$

Assuming that y_t is a monthly series so that $s = 12$, then the ACGF of the stationary transformation of y_t is given by:

$$\begin{aligned}
\gamma(B) &= V_a(1 + \theta_1 B)(1 + \theta_{12} B)(1 + \theta_1 F)(1 + \theta_{12} F) \\
&= V_a(1 + \theta_1 B + \theta_{12} B^{12} + \theta_1\theta_{12} B^{13})(1 + \theta_1 F + \theta_{12} F^{12} + \theta_1\theta_{12} F^{13})
\end{aligned}$$

which gives:

$$\begin{aligned}
\gamma_0 &= V_a(1 + \theta_1^2 + \theta_{12}^2 + \theta_1^2\theta_{12}^2) \\
\gamma_1 &= V_a(\theta_1 + \theta_1\theta_{12}^2)
\end{aligned}$$

$$\gamma_i = 0 \quad i = 2, \cdots, 10$$

$$\gamma_{11} = V_a \theta_1 \theta_{12}$$

$$\gamma_{12} = V_a(\theta_{12} + \theta_1^2 \theta_{12})$$

$$\gamma_{13} = V_a \theta_1 \theta_{12}$$

The autocovariances are zero at lags greater than 13. Notice that the multiplicative seasonal model (3.2) implies non zero autocorrelations at lags 11 and 13, that is around the seasonal lag of 12. Furthermore the first lag autocorrelation does not depend on the seasonal part of the model.

In practice, the identification of seasonal time series models can be performed according to the following scheme:

Identification of models for seasonal time series

- Picking D and d: Plot the ACF of y_t. If large and persistent autocorrelations at lags 12, 24, 36 can be seen, differenciate y_t: $w_t = \Delta_{12} y_t$. Otherwise $D = 0$. Next look for d, the number of regular differenciation, as for nonseasonal models: that is the ACF of w_t is plotted, and if failure to cancel rapidly is detected, a firt difference is taken. Next, if the ACF of $w_t = \Delta\Delta_{12} y_t$ still shows a slow convergence of the autocorrelations, then $d = 2$. The transformed series $w_t = \Delta^d \Delta_{12}^D y_t$ is obtained.

- Picking p, P, q, Q: the ACF and PACF may be used in the same way than for the nonseasonal time series. For example, if the ACF of w_t shows a large autocorrelation at lag 12 and no significant correlations a the successive seasonal lags 24 and 36, take Q=1. If it is the PACF which shows a large autocorrelation at lag 12 and no significant correlations at lags 24 and 36, take P=1. For the regular part of the model (p and q), the identification is performed as described earlier.

Example 2: Series FRPDB101

The monthly series FRPDB101 of the french total industry production (excluding construction) has been plotted in figure 1.1. The ACF and PACF of y_t are displayed on figure 3.3.

Figure 3.3: FRPDB101: ACF and PACF of y_t

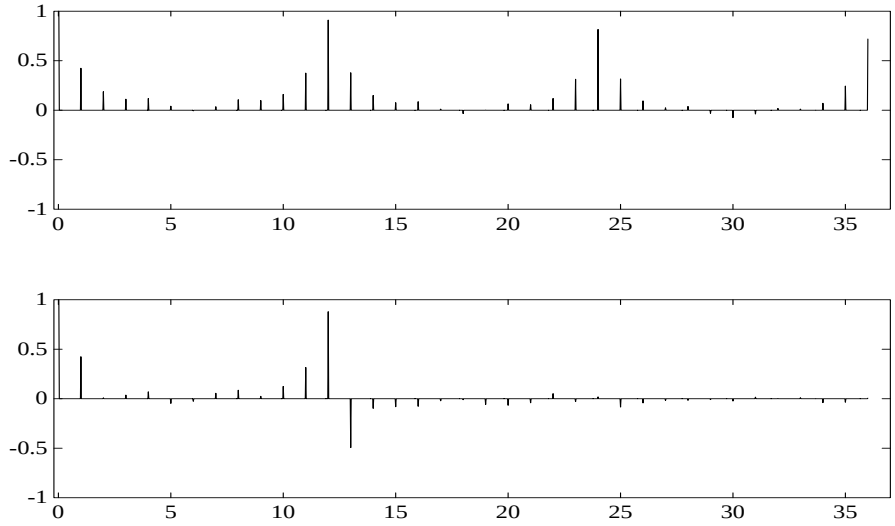


Figure 3.4: FRPDB101: ACF and PACF of $\Delta_{12} y_t$

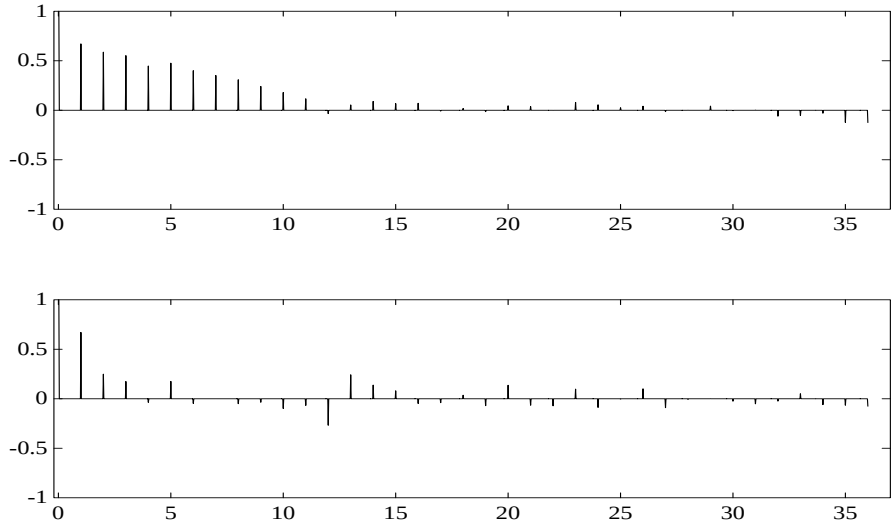
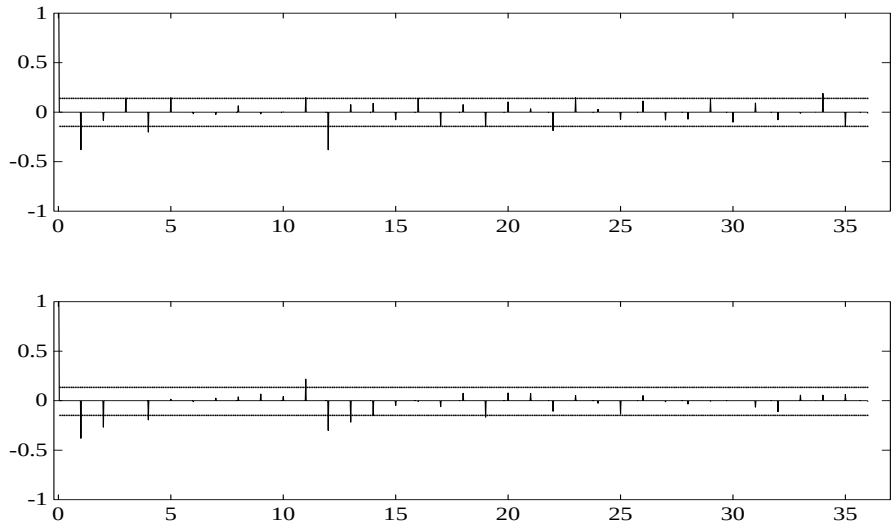


Figure 3.5: FRPDB101: ACF and PACF of $\Delta\Delta_{12}y_t$

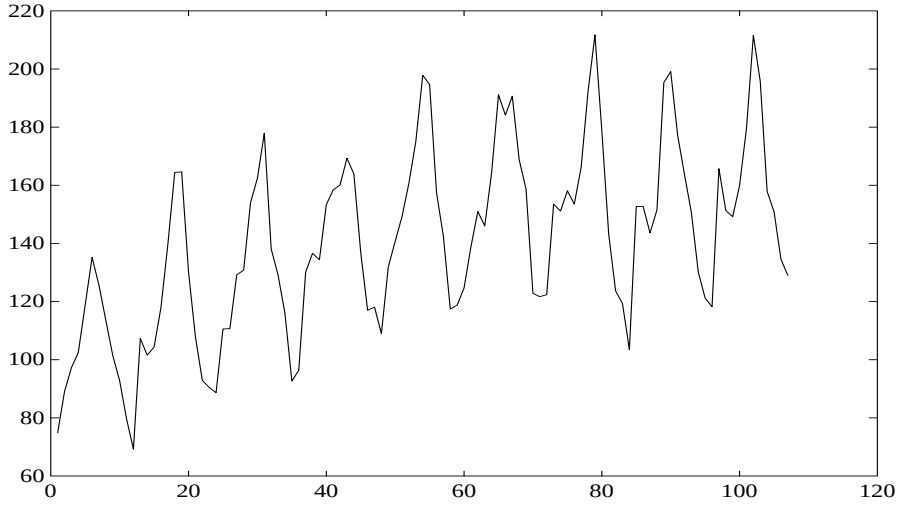


The ACF shows a large lag-12 autocorrelation which is still present at lags 24 and 36. This is an evidence of $D = 1$, which is further confirmed by the PACF which shows a close-to-one peak at lag 12. Taking thus D=1 and computing the resulting ACF and PACF, we can see on figure 3.4 that the ACF of $\Delta_{12}y_t$ shows a slow convergence, while the PACF display a large first partial autocorrelation. This suggests to consider $d = 1$. Figure 3.4 shows the ACF and PACF of $\Delta\Delta_{12}y_t$: no need for further differencing appear. The ACF shows significant correlations at lags 1 and 12. The partial correlations at lags 1 and 12 are also significant, but they are followed by significant autocorrelations at lags 2 and 13, 14. This information leads to consider the airline model $(0,1,1)(0,1,1)_{12}$ as a possible candidate for describing the autocorrelation structure of this series.

Example 3: Series ITPDB428

The series ITPDB428 represents the monthly italian production of soft drinks. It extends from january 1985 to november 1993, that is along a sample of 107 observations. It is plotted on figure 3.6. Figure 3.7 shows the sample ACF and PACF of the series. A seasonal first difference is clearly adequate. The ACF and PACF of the resulting transformed series can be seen on figure 3.8. Although the first autocorrelation is not very high, the failure of the ACF to converge rapidly suggest to consider also a

Figure 3.6: Series ITPDB428



non seasonal first-difference. Figure 3.9 displays the ACF and PACF of $\Delta\Delta_{12}y_t$: the sample ACF shows a long sequence of significant autocorrelations, while the partial autocorrelations are significant at lags 1, 2 and around the seasonal lag 12. This suggest to consider p=2, but to evaluate the relevance of this transformation, the two autoregressive parameters need now to be estimated. The results will allow us to pursue the search of an adequate model.

3.3 Estimation

Once a tentative model has been selected, estimation may be computed either by maximum likelihood or by least-squares methods. There is a very abondant literature on ARMA models estimation; see for example Box and Jenkins, 1970, Brockwell and Davis, 1987, Harvey, 1989, among others. We present an outline of standard methods in use, but as the computational details are far beyond our scope, interested readers are refered to the literature above.

The maximum likelihood method consists in finding out the parameters $\psi = \{\phi_1, \cdots, \phi_p,$

Figure 3.7: ITPDB428: ACF and PACF of y_t

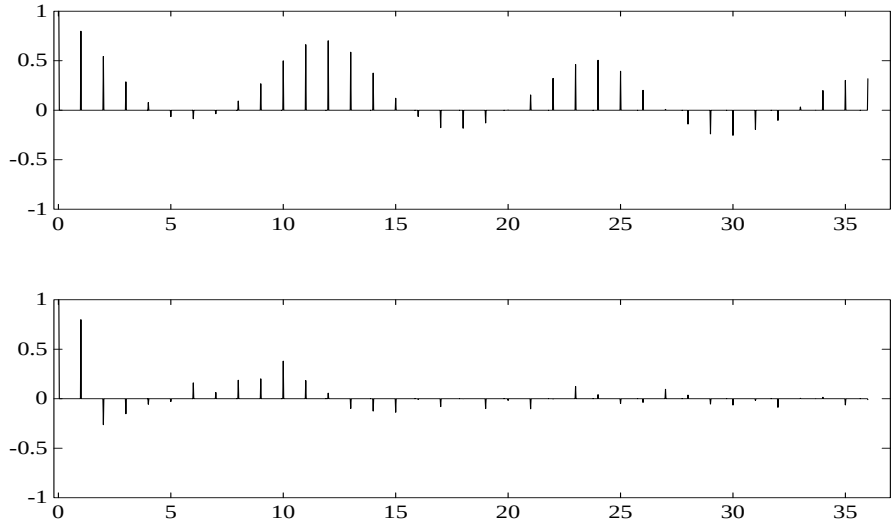


Figure 3.8: ITPDB428: ACF and PACF of $\Delta_{12} y_t$

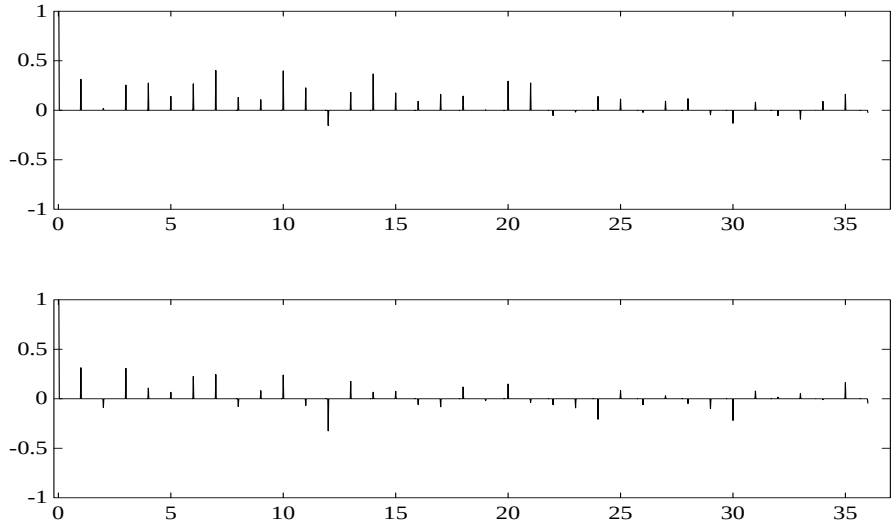
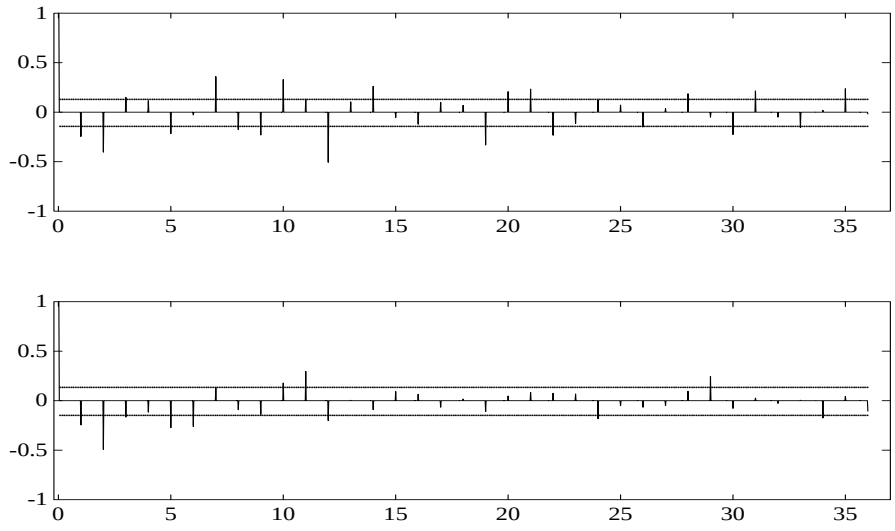


Figure 3.9: ITPDB428: ACF and PACF of $\Delta\Delta_{12}y_t$



$\theta_1, \cdots, \theta_q\}$ maximizing the joint probability function of the stationary transformation of the data. The joint probability function is built assuming that the innovations a_t are independent and normally distributed with variance σ^2. In the case of a stochastic process, say y_t, whose stationary transformation $w_t = \delta(B)y_t$ can be modelled as a process of the ARMA(p,q)-type, we have:

$$w_t = \phi_1 w_{t-1} + \cdots + \phi_p w_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \tag{3.3}$$

The polynomials $\phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ have distinct roots, all outside the unit circle. Further, the polynomial $\delta(B)$ is assumed to be of order d, so that if $n + d$ realisations of z_t are observed, equation (3.3) holds for $t = 1, \cdots, n$. Equation (3.3) involves the starting values $\mathbf{w}^*$=$\{w_0, \cdots, w_{1-p}, a_0, \cdots, a_{1-q}\}$. Assuming first the initial values known, and given that ψ denotes the set of parameters involved in (3.3), then the log-likelihood function of the ARMA(p,q) model conditional on $\mathbf{w}^*$ is given by

$$logL(\psi, \sigma^2) = -n \log \sigma^2 - \frac{S(\psi)}{2\sigma^2}, \tag{3.4}$$

where a constant has been omitted and $S(\psi)$ represents the sum of square function

$$S(\psi) = \sum_{i=1}^{n} \hat{a}_i^2(\psi).$$

The residuals $\hat{a}_i(\psi)$ are obtained from (3.3) as

$$\hat{a}_t(\psi) = w_t - \phi_1 w_{t-1} - \cdots - \phi_p w_{t-p} - \theta_1 \hat{a}_{t-1} - \cdots - \theta_q \hat{a}_{t-q}. \tag{3.5}$$

The parameter σ^2 may be *concentrated out* of the likelihood function using $\hat{\sigma}^2 = n^{-1}S(\psi)$, so that the function to maximize becomes:

$$logL(\psi, \sigma^2) = -n \log \frac{S(\psi)}{n} - \frac{n}{2}. \tag{3.6}$$

Maximizing the conditional likelihood function or minimizing the sum of squares is then equivalent. In practice however, the initial conditions $\mathbf{w}^*$ are not known, and computing the function $S(\psi)$ requires to set the values w_0, w_{-1},..., w_{1-p} and $a_0, ..., a_{1-q}$ in order to have available the first $\hat{a}_i$'s. A simple solution to this problem would consist in considering the unconditional expectations of the innovations and of the w_t's. Given that both are stationary processes with zero mean, these expectations are zero. This estimation procedure is known to perform poorly if the AR polynomial has roots close to the nonstationarity region. The distortion effects due to non vanishing transient may be avoided by considering instead the truncated function:

$$S(\psi) = \sum_{i=p+1}^{n} \hat{a}_i^2(\psi), \tag{3.7}$$

with $a_p = a_{p-1} = \cdots = a_{p-q+1} = 0$. Minimizing (3.7) with the starting conditions set in that way yields the so-called *constrained least squares* estimators. Setting the innovations at zero has a vanishing effect only if the MA polynomial $\theta(B)$ is invertible. Otherwise, the presence of unit root(s) in $\theta(B)$ makes persistent the effect of the starting conditions, and the estimators which results will be severely distorted. Invertibility of the MA polynomial is thus the requirement for constrained least squares estimators to be satisfying.

Alternatively, the starting conditions can be estimated from the data. A possibility developed by Box and Jenkins consists in reverting the model (3.3) by replacing B by B^{-1} in order to back forecast the values $w_0, w_{-1}, \cdots, w_{1-p}$. These values are then used in the recursion (3.3) for obtaining the residuals a_i, $i = 1-q, \cdots, 0, 1, \cdots$. This yields unconditional least square estimates.

Finally, completing the conditional likelihood (3.4) with the joint distribution of the starting values yields the *exact maximum likelihood* function. The difference with the unconditional sum of squares function is that the determinant of the initial values covariance matrix is considered. When an estimation procedure has been selected, maximisation can be computed using convergence algorithm like for example Newton-Raphson, Gauss-Newton, among others (see Brokwell and Davis, 1987). These algorithms maximize the objective function by trying parameters values whose choice is based the first two derivatives of the objective function evaluated at the previous parameter value.

Example 1: Series POEM2640

The identification stage pointed a (2,1,0) model as a possible model: $(1 + \phi_1 B + \phi_2 B^2)\Delta y_t = a_t$. Estimating the autoregressive polynomial by maximum likelihood, we get the following result:

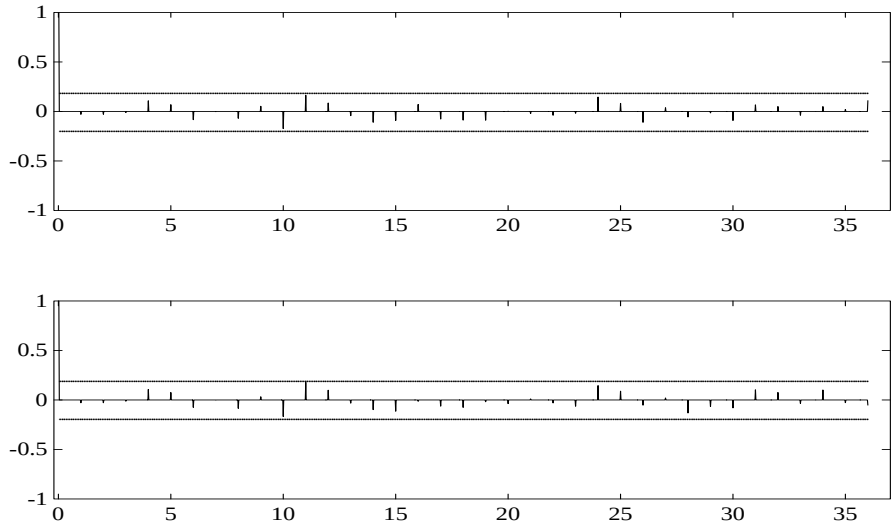
$$\begin{aligned} \phi_1 &= .494(.025) \\ \phi_2 &= .291(.018) \\ \sqrt{V_a} &= .019 \end{aligned}$$

where the standard errors of the parameters estimates are given in parenthesis. The mean of the differenced series has been obtained at $\hat{\mu} = -.005$, with a standard deviation of .001, hence largely significant. The model (2,1,0) fitted to the series yields residuals ACF and PACF displayed on figure 3.10. No autocorrelations are left in the residuals: the model selected pass this first examination stage.

Example 2: Series FRPDB101

The identification stage indicated that the airline model $(0,1,1)(0,1,1)_{12}$ could be a possible model. Estimating the MA parameters in $\Delta\Delta_{12} y_t = (1 + \theta_1 B)(1 + \theta_{12} B)a_t$. we get the following result:

Figure 3.10: POEM: ACF and PACF of residuals of a (2,1,0) model



$$\begin{aligned} \theta_1 &= -.561(.058) \\ \theta_{12} &= -.487(.061) \\ \sqrt{V_a} &= .017 \end{aligned}$$

The residuals ACF and PACF displayed on figure 3.11 does not show any autocorrelation left.

Example 3: Series ITPDB429

For this third example, the examination of the ACF and PACF of different transformation of the series yielded a $(2,1,0)(0,1,0)_{12}$ as a candidate model. Estimating the autoregressive polynomial by maximum likelihood, we get the following result:

$$\begin{aligned} \phi_1 &= .401(.133) \\ \phi_2 &= .502(.083) \\ \sqrt{V_a} &= .088 \end{aligned}$$

The ACF and PACF of the residuals of this model are displayed on figure 3.12. A large

Figure 3.11: FRPDB101: ACF and PACF of airline model residuals

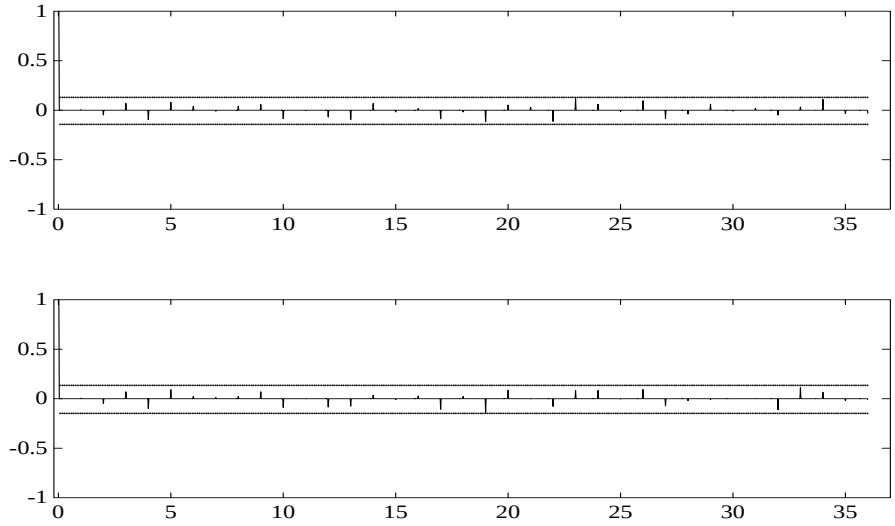


Figure 3.12: ITPDB428: ACF and PACF of residuals from $(2,1,0)(0,1,0)$

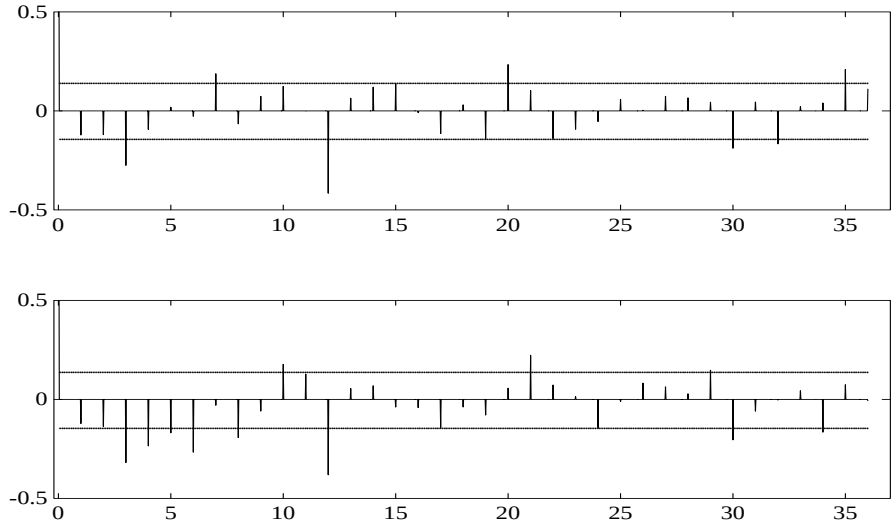
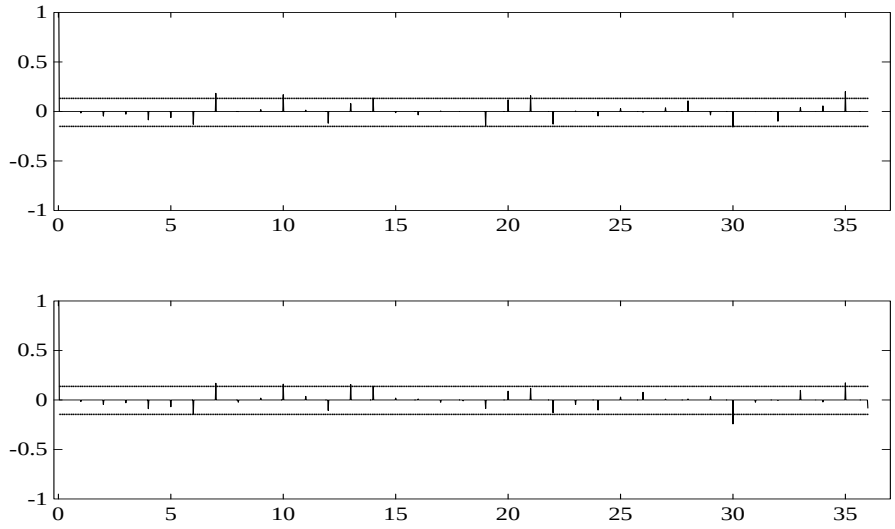


Figure 3.13: ITPDB428: ACF and PACF of residuals from $(2,1,0)(0,1,0)$



autocorrelation is seen at lag 12, pointing the need of considering Q=1. On the other hand, there are several large partial autocorrelations at the first lags which suggests to set q to 1. Fitting now the model $(2,1,1)(0,1,1)_{12}$, we obtain the following result:

$$\begin{aligned} \phi_1 &= -.085(.126) \\ \phi_2 &= .286(.109) \\ \theta_1 &= .-.750(.102) \\ \theta_{12} &= -.640(.090) \\ V_a &= .067 \end{aligned}$$

The ACF and PACF of the residuals of this model are displayed on figure 3.13: no left autocorrelations may be seen.

3.4 Hypothesis Testing and Confidence Interval

Test of significance of the parameters estimated may be carried out as t-tests. The variance-covariance matrix Ω of the parameters estimates may be obtained from the

second derivative of the likelihood function according to:

$$\Omega \simeq [\frac{-dL(\psi)}{d\psi d\psi'} \mid_{\psi=\hat{\psi}}]^{-1} \tag{3.8}$$

In practice, the matrix of variance-covariance is directly given by the algorithm used to maximized the likelihood function. The square root of the i-th element of the matrix Ω gives the standard error of the i-th parameter estimator. Confidence interval may then be computed given that, under certains conditions, the estimators are asymptotically distributed as $N(\psi_0, \Omega)$, where ψ_0 is the vector of true parameters.

Notice however that significance tests may only help in distinguishing between nested models. For example, they are useful in evaluating the relevance of an AR(1) against an ARMA(1,1). Distinguishing between non-nested models cannot be drawn in that way: this procedure cannot be applied to test for an AR(1) against a MA(1). The Bayesian Information Criterion (BIC), proposed by Akaike and discussed for example in Hannan (1980) (see also Sneek, 1984) is more useful in this respect. Denoting k the number of estimated parameter (for an ARIMA$(p, d, q), k = p + q + 1$), the BIC is obtained as:

$$BIC = -2 \log L(\psi, \sigma^2 | w) + k + k \log n \tag{3.9}$$

The BIC can be used to compare different models fitted to a time series. The preferred model is the one which minimizes the BIC accross the range of the different models.

3.5 Diagnostic Checking

At this stage, an evaluation of the temptative model based on the estimated residual properties is to be performed. For the model to provide a good description of the series, no underlying structure might be left in the residuals. So first of all it could be useful to examine the residual plot. For example, changes in the variance over time would suggest to modify the initial transformation of the data. But structures like remaining residual autocorrelation may be difficult to discern, and for that a more formal test could be useful. The Bartlett's result about the estimated autocorrelation

(see section 3.2) which states that the autocorrelations estimates are asymptotically normally distributed with variance $1/T$ suggests to consider the following *portmanteau test statistics*:

$$Q_m = T \sum_{i=1}^{m} r^2(i). \tag{3.10}$$

This statistics tests the significance of the first m-autocorrelations. The choice of m is abitrary; for example for monthly time series, one may consider $m = 24$. Since the $r(.)$ are normally and independently distributed, the Q_m statistics will have a χ^2_m distribution for white noise data. When the data are obtained as residuals of a fitted ARIMA model, two corrections must be made. First, Ljung and Box (1978) showed that in small samples the expression $(T-i)/T(T+2)$ approximates the variance of $r(i)$ more accurately than $1/T$. Hence, the corrected form of the Q-statistics is:

$$Q_m = T(T+2) \sum_{i=1}^{m} r^2(i)/(T-i). \tag{3.11}$$

Second, when the $r(.)$'s are computed from ARMA residuals, a better approximation of the limiting distribution is given by χ^2 with $(m-p-q)$ degrees of freedom.

The Q-statistics may also be modified so as to consider specifically seasonal lags. For example, for a monthly time series, the significance of the autocorrelations of the residuals at lags 12 and 24 may be tested using the statistics:

$$Q_s = T(T+2)[r^2(12)/(T-12) + r^2(24)/(T-24)]. \tag{3.12}$$

Pierce (1978) showed that Q_s can be roughly approximated by a χ^2_2 distribution.

Residuals which do not present significant autocorrelations can be seen as embodying no linear structure. However, it is possible that the residuals series a_t are related in some way to the a_t^3's. In other words, the residuals could be white noises but not independent because of some nonlinear underlying structure. The linearity assumption can be checked also with the Q-statistics. Maravall (1983) showed that if a series z_t is linear, then the lag-k autocorrelation is such that: $\rho_k(z_t^2) = [\rho_k(z_t)^2]$. Further,

McLeod and Li (1983) proved that $Q_m(a_t)$ and $Q_m(a_t^2)$ have the same distribution. So, computing the Q-statistics for both the residuals and the squared residuals, an increase in the Q-value for squared residuals is an indication of nonlinearity. Similarly, a test for nonlinearity at seasonal lags may be performed by comparing $Q_s(a_t)$ and $Q_s(a_t^2)$. Locating the nonlinearities present in the series at some particular lags may be useful to improve the modelling of the series under analysis (see for example Fiorentini and Maravall (1996)).

Yet another hypothesis to test is the residual normality. The linear models that we have presented are able to describe the autocorrelations behavior of time series, that is the second-order moment properties of the series. Through the Wold decomposition, these models can be seen simply as linear combination of a white-noise variable. Whether the white-noise variable is normally distributed or not determines the ability of the representation to catch the higher order moment of the series. In particular, for a normal white noise variable, the moment of order α:

$$m_\alpha = T^{-1} \sum_{i=1}^{T} e_t^\alpha \quad \alpha = 2, 3, 4, \tag{3.13}$$

is such that:

$$\sqrt{T} m_\alpha \sim N(\mu_\alpha, \alpha! \sigma^{2\alpha}) \quad \alpha = 2, 3, 4, \tag{3.14}$$

For example, it is possible to check whether the distribution of the estimated residuals is symmetric as the normal distribution is. This is done through the skewness test, which is computed on the on the third moment:

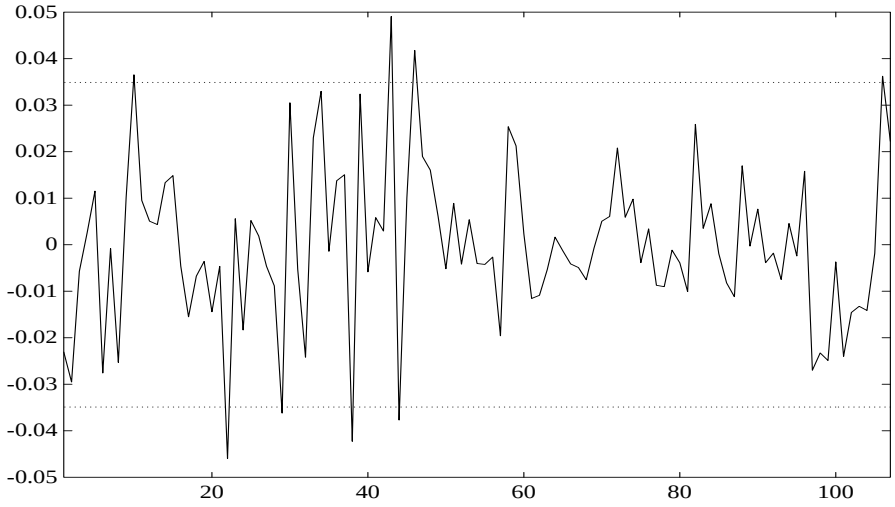
$$S = \sqrt{T} \frac{m_3}{\hat{\sigma}^3} \sim N(0, 6) \tag{3.15}$$

Also, a test on the fourth-moment can point out too large tails of the residual distribution: the kurtosis test, computed as:

$$K = \sqrt{T} \frac{m_4 - 3}{\hat{\sigma}^4} \sim N(0, 24) \tag{3.16}$$

becomes significant when too many residuals take large absolute values. It can thus be

Figure 3.14: POEM2640: Residuals from $(2,1,0)(0,0,0)$



used to check the presence of outliers in the residuals. Finally, the statistics S and K can be combined to form a test for normality N, obtained as: $N = S^2 + K^2$, and, given the independency of the S and K estimates, the statistics N is distributed as a χ^2_2.

Examples

The residuals of the final models fitted to the three series are plotted on figures 3.14, 3.15 and 3.16, together with their respective confidence interval.

For the POEM2640 residuals, a reduction in the variance may be seen in the second half of the sample. For the two other series, no particular pattern can be seen: the residuals seem to be white noises. A deeper evaluation of the randomness hypothesis requires to compute the statistics previously described. The table 3.1 sums up the results:

Figure 3.15: FRPDB101: Residuals from $(0,1,1)(0,1,1)_{12}$

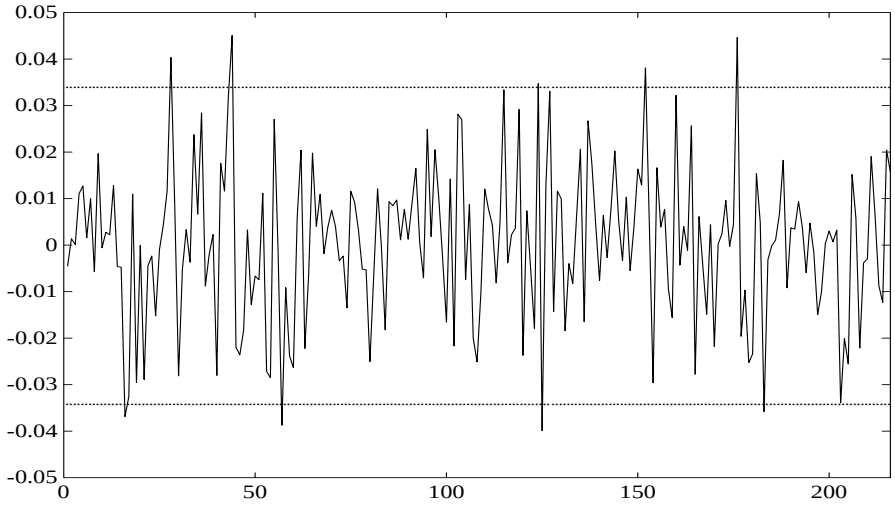


Figure 3.16: ITPDB428: Residuals from $(2,1,1)(0,1,1)_{12}$

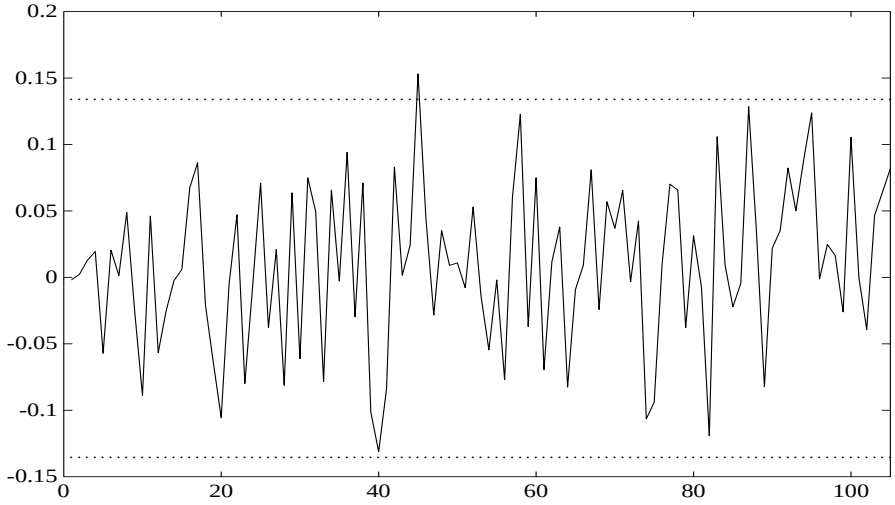


Table 3.1			
Diagnosis: Q-Statistics			
Series	POEM2640	FRPDB101	ITPDB428
$Q(a_t)$	$Q_{24} = 23.99$	$Q_{24} = 24.60$	$Q_{24} = 24.13$
$Q_s(a_t)$	$Q_2 = 2.44$	$Q_2 = 1.84$	$Q_2 = 1.83$
$Q(a_t^2)$	$Q_{24} = 26.06$	$Q_{24} = 15.91$	$Q_{24} = 20.99$
$Q_s(a_t^2)$	$Q_2 = 3.72$	$Q_2 = 2.22$	$Q_2 = .65$

There are thus no remaining correlations in the residuals. Furthermore, the linearity hypothesis is checked by the data. Looking at the tests of the normality hypothesis on table 3.2 confirms this result, and shows that in the three cases, the residuals come out normally distributed.

Table 3.2				
Diagnosis: Normality Statistics				
Series	Model	N	S	K
POEM2640	$(2,1,0)(0,0,0)$	1.79	.16 (.24)	3.55 (.47)
FRPDB101	$(0,1,1)(0,1,1)_{12}$	.12	-.03 (.17)	3.10 (.33)
ITPDB428	$(2,1,1)(0,1,1)_{12}$	1.44	-.16 (.24)	2.52 (.48)

3.6 Predictions in ARIMA models

3.6.1 Minimum Mean Square Error Forecasts

Once a model describing the sample behavior of the data in a satisfactory way has been identified and estimated, it can be used as a basis for forecasting. In this section, the forecasting procedure in stochastic linear models is presented. In a first step, the parameters of the models are assumed to be known.

Suppose the sample available is $y_1, \cdots, y_T$, and that we are interested in a forecast of y_{T+l}. We would like the forecast value, say $y^*_{T+l/T}$, to be as accurate as possible. Forecast accuracy is usually measured with a Mean Squared Error (MSE) criterion, which is defined as:

$$MSE(y^*_{T+l/T}) = E[(y_{T+l} - y^*_{T+l/T})^2].$$

It can be checked that the forecast with minimum MSE is given by the expectation of y_{T+l} conditional on the information set $y_1, \cdots, y_T$ available at time t:

$$\hat{y}_{T+l/T} = E(y_{T+l}|y_1, \cdots, y_T).$$

We shall also denote $E_T(y_{T+l})$ this expectation. Under the normality hypothesis, then the solution of the conditional expectation above will be a linear combination of the observed $y_1, \cdots, y_T$. This linear combination can be easily derived from the ARMA(p,q) process at time $T + l$:

$$\begin{aligned} y_{T+l} &= \phi_1 y_{T+l-1} + \cdots + \phi_p y_{T+l-p} + \\ &+ \theta_1 a_{T+l-1} + \cdots + \theta_q a_{T+l-q}. \end{aligned}$$

The expectation of y_{T+l} conditional on the information available at time T is given by:

$$\begin{aligned} \hat{y}_{T+l/T} &= E_T(\phi_1 y_{T+l-1}) + \cdots + E_T(\phi_p y_{T+l-p}) + \\ &+ E_T(\theta_1 a_{T+l-1}) + \cdots + E_T(\theta_q a_{T+l-q}). \end{aligned}$$

Given that the a_t's are assumed to be independent white noises, they do not depend on the past: the conditional expectation of the future innovations is then equal to their expected value, that is $E_T(a_t) = 0$ if $t > T$. On the other hand, given that past innovations are available by linear combination of the observed data, we have: $E_T(a_t) = a_t$ if $t \leq T$. Next, the expected value of y_{T+l} is simply obtained from the recursion

$$\begin{aligned} \hat{y}_{T+l/T} &= \phi_1 \hat{y}_{T+l-1/T} + \cdots + \phi_p \hat{y}_{T+l-p/T} + \\ &+ E_T(\theta_1 a_{T+l-1}) + \cdots + E_T(\theta_q a_{T+l-q}). \end{aligned}$$

with $\hat{y}_{T+j/T} = y_{T+j}$ if $j \leq 0$. For example, consider the AR(1) process $y_t = \phi y_{t-1} + a_t$: the forecast of y_{T+1} at time T is given by

$$\hat{y}_{T+1/T} = \phi y_T + E_T(a_{T+1}) = \phi y_T.$$

The two-step ahead forecast is instead:

$$\hat{y}_{T+2/T} = \phi \hat{y}_{T+1/T} + E_T(a_{T+2}) = \phi^2 y_T,$$

and in general it is easily obtained that for an AR(1) process,

$$\hat{y}_{T+l/T} = \phi^l y_T.$$

The forecast function of an AR(1)is thus a exponential function, decaying towards 0 since $|\phi| < 1$. Consider now an AR(p) process: it is readily seen that the one step-ahead forecast is given by

$$\begin{aligned} \hat{y}_{T+1/T} &= \phi_1 y_T + \cdots + \phi_p y_{T-p+1} + E_T(a_{T+1}) = \\ &= \phi_1 y_T + \cdots + \phi_p y_{T-p+1}. \end{aligned}$$

For the next forecast,

$$\begin{aligned} \hat{y}_{T+2/T} &= \phi_1 \hat{y}_{T+1/T} + \cdots + \phi_p y_{T-p+1} + E_T(a_{T+2}) = \\ &= \phi_1 \hat{y}_{T+1/T} + \cdots + \phi_p y_{T-p+1}, \end{aligned}$$

and forecasts beyond p-step ahead will verify the homogenous p-order linear difference equation:

$$\hat{y}_{T+p+i/T} = \phi_1 \hat{y}_{T+p+i-1/T} + \cdots + \phi_p \hat{y}_{T+i/T},$$

for $i \geq 0$ and starting value $\hat{y}_{T/T} = y_T$.

Forecasting a pure moving average process can be conducted in the following way. Let us consider a MA(1) process, $y_t = a_t + \theta a_{t-1}$. Then, taking the conditional expectation of y_{T+1} given $y_1, ..., y_T$, we get:

$$\hat{y}_{T+1/T} = E_T(a_{T+1}) + \theta E_T(a_T) = \theta a_T.$$

For the next step, it is readily obtained that $\hat{y}_{T+2/T} = 0$, and so will be the successive forecasts. In general, for a MA(q) process, the forecasting function is given by:

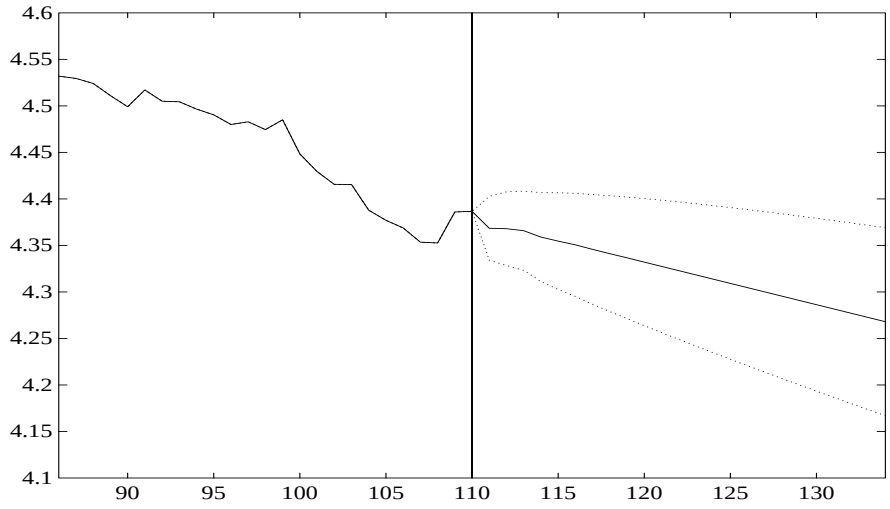
$$\begin{aligned} \hat{y}_{T+l/T} &= \theta_l a_T + \cdots + \theta_q a_{T+l-q} \quad \text{for } l \leq q \\ \hat{y}_{T+l/T} &= 0 \quad \text{otherwise.} \end{aligned}$$

The forecasts of an ARIMA(p, d, q) process is then derived according to the same principles. Suppose $\delta(B)$ is the polynomial containing the roots which are on the unit circle, so that the ARIMA model can be written $\phi(B)\delta(B)y_t = \theta(B)a_t$, where the polynomials may include seasonal lags or not. Defining $\alpha(B)$ the polynomial obtained as from the multiplication $\phi(B)\delta(B)$, then the ARIMA model may be rewritten $\alpha(B)y_t = \theta(B)a_t$ and the results on forecasting an ARMA apply. In particular,

$$\begin{aligned} \hat{y}_{T+1/T} &= \alpha_1 y_T + \cdots + \alpha_{p+d} y_{T-p-d+1} + \\ &+ \theta_1 a_T + \cdots + \theta_q a_{T+1-q} \\ \hat{y}_{T+2/T} &= \alpha_1 \hat{y}_{T+1/T} + \cdots + \alpha_{p+d} y_{T-p-d+2} + \\ &+ \theta_2 a_T + \cdots + \theta_q a_{T+2-q} \\ &\vdots \\ \hat{y}_{T+q/T} &= \alpha_1 \hat{y}_{T+q-1/T} + \cdots + \alpha_{p+d} \hat{y}_{T+q-p-d/T} + \theta_q a_T \\ \hat{y}_{T+l/T} &= \alpha_1 \hat{y}_{T+l-1/T} + \cdots + \alpha_{p+d} \hat{y}_{T+l-p-d/T} \quad l > q \end{aligned} \tag{3.17}$$

The forecast function of an ARIMA model turns out to have the same form than that of an AR process after a certain number of periods. However, it has to be noticed that

Figure 3.17: Series POEM264: Forecasts



Forecast — 95% Confidence Interval ···

while the forecasts of a pure AR process tend to zero, those of an ARIMA process do not present the same convergence pattern. This is due to the presence of unit roots. For example, the forecasts of the random walk $y_t = y_{t-1} + a_t$ are simply $\hat{y}_{T+l/T} = y_T, \forall l > 0$. Suppose then that a mean has been fitted. The model is thus $y_t = y_{t-1} + \mu + a_t$, and the forecasts are easily seen to be $\hat{y}_{T+l/T} = y_T + l\mu$, that is a first-order time polynomial. That pattern is illustrated on figure 3.17 where the forecast function given by the $(2, 1, 0)$ model chosen to describe the behavior of the series POEM2640 is plotted.

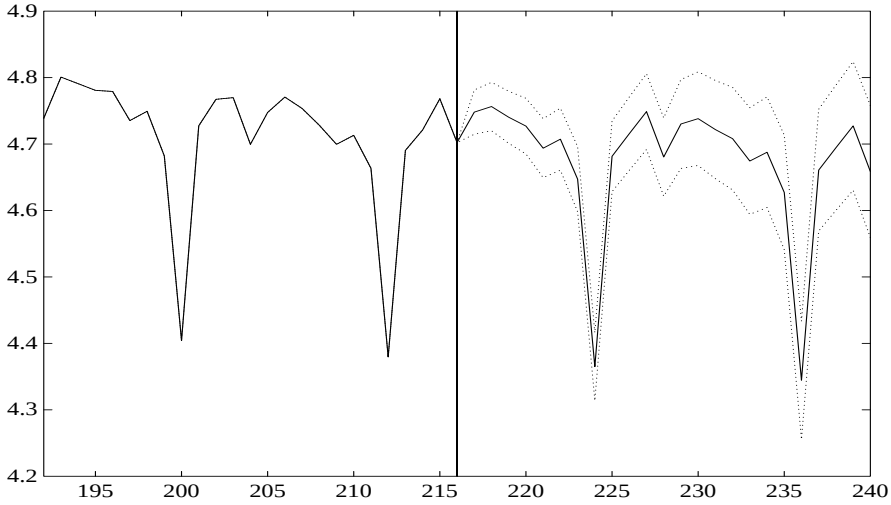
Next, consider also the *airline* model $\Delta\Delta_{12} y_t = (1 + \theta B)(1 + \theta_{12} B^{12}) a_t$. Beyond 12 periods, the forecast function reduces to:

$$\hat{y}_{T+l/T} = \hat{y}_{T+l-1/T} + \hat{y}_{T+l-12/T} - \hat{y}_{T+l-13/T},$$

$l > 12$, whose solution follows a fixed seasonal pattern of period 12 which is superposed on a linear trend. That pattern may be clearly observed on figures 3.18 and 3.19.

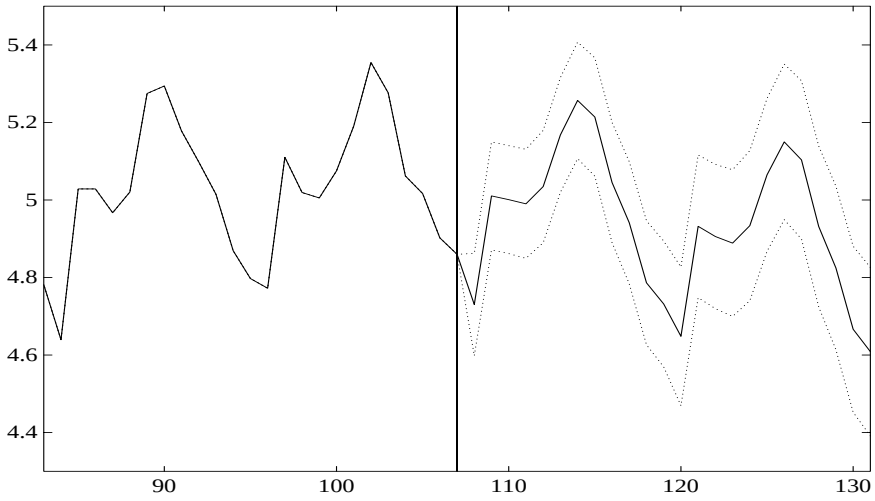
Finally, the forecast function of an ARIMA(0,1,1) presents a particular interest. Let us write that process as $\Delta y_t = (1 - \theta B) a_t$, $\mid \theta \mid < 1$. The one-step ahead forecast is

Figure 3.18: Series FPDB101: Forecasts



Forecast — 95% Confidence Interval · · ·

Figure 3.19: Series ITPDB428: Forecasts



Forecast — 95% Confidence Interval · · ·

readily seen to be such that:

$$\hat{y}_{T+1/T} = y_T - \theta a_T.$$

The innovation a_T may be also written as a linear combination of present and past values of y_t by inverting the MA polynomial $(1 - \theta B)$:

$$\begin{aligned} a_T &= (1 - \theta B)^{-1}(1 - B)y_T = \\ &= (1 + \theta B + \cdots + \theta^n B^n)(1 - B)y_T = \\ &= y_T - (1 - \theta)[y_{T-1} + \theta y_{T-2} + \cdots + \theta^{n-1} y_{T-n}] \end{aligned}$$

Replacing a_T by this linear combination of the observed process in the one-step ahead forecast equation directly yields:

$$\hat{y}_{T+1/T} = (1 - \theta)[y_T + \theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^n y_{T-n}]$$

The weights of this linear combination are given by $(1 - \theta)\theta^i$. Thus, they decline exponentially, and they present also the property of summing at 1. In fact, such linear combinations are termed *exponential weighted moving averages*. Notice that since $\hat{y}_{T/T-1} = (1 - \theta)\sum_{i=0}^{T-1} \theta^i y_{T-1-i}$, then we can also write:

$$\hat{y}_{T+1/T} = (1 - \theta)y_T + \theta \hat{y}_{T/T-1}.$$

The parameter θ weigths the importance given to the recent realizations: if $\theta = 0$, then no information on the past of the series is used in forecasting, only the current realization is considered, while $\theta = 1$ implies that past observations are heavily involved in the forecasting procedure. Quite often, values of the θ-parameter have been *a priori* assigned and the IMA(1,1) implicit specification *a priori* assumed. The absence of identification, estimation and formal diagnostic checks may have been justified because of the computational limitations, but the current availability of computional power has made obsolete the *exponential weighted moving average* techniques.

3.6.2 Confidence Interval for Forecasts

The building of confidence interval around the forecasts is useful for inference drawing. This requires knowledge of the MSE on the forecasts. In a first stage, we will consider the simple case where the model parameters are known. The effect of dealing with estimated parameters will be analysed next on the basis of an example.

Consider for example a MA(1) process: the one-step ahead forecast error is easily obtained as:

$$y_{T+1} - \hat{y}_{T+1/T} = a_{T+1},$$

and thus the MSE is: $\text{MSE}(\hat{y}_{T+1/T}) = E(a_{T+1}^2) = V_a$. The next forecasts of the MA(1) process are simply null, so the forecasting error is such that

$$y_{T+l} - \hat{y}_{T+l/T} = a_{T+l} + \theta a_{T+l-1},$$

for $l > 1$. It follows that for the MA(1) process, $\text{MSE}(\hat{y}_{T+l/T}) = (1+\theta^2)V_a$. It is not difficult to check that for a MA(q) process, then for $l > q$, $\text{MSE}(\hat{y}_{T+l/T}) = V_a \sum_{i=0}^{q} \theta_i^2$, where $\theta_0 = 1$, while the sum truncates at l if $l \leq q$.

The case of AR or mix ARMA models is less obvious. However, a simplification is reached by inverting the AR polynomial so as to write the model in the MA form:

$$y_t = \sum_{i=0}^{\infty} \psi_i B^i a_t \tag{3.18}$$

where the polynomial $\psi(B) = \psi_0 + \psi_1 B + \cdots + \psi_n B^n$ is obtained from $\psi(B) = \alpha^{-1}(B)\theta(B)$. An easy way to derive $\psi(B)$ consists in equating the coefficients of B^i in $(1 + \alpha_1 B + \cdots + \alpha_{p+d} B^{p+d})(1 + \psi_1 B + \cdots) = 1 + \theta_1 B + \cdots + \theta_q B^q$. Notice that this always yields $\psi_0 = 1$. Taking conditional expectation in (3.18) directly yields:

$$\hat{y}_{T+l/T} = \sum_{i=0}^{\infty} \psi_{i+l} a_{T-i}; \tag{3.19}$$

the predictions obtained with (3.19) and with (3.17) are identical. Then,

$$y_{T+l} - \hat{y}_{T+l/T} = \sum_{i=0}^{l-1} \psi_i a_{T+l-i};$$

It is of interest to notice that whatever is the model considered, the one-step-ahead forecast error is always a_{T+1}: it is for that reason that the a_t's have been called *innovations*, since they represent the part in the new observation y_{T+1} which is not predictable. Since the innovations are independent with mean zero and variance V_a, the MSE on the l-step ahead forecast is given by

$$V(l) = MSE(\hat{y}_{T+l/T}) = V_a \sum_{i=0}^{l-1} \psi_i^2. \tag{3.20}$$

Hence, $MSE(\hat{y}_{T+l/T})$ increases as the forecast horizon increases. Next, normality of the innovations imply that a 95% confidence interval around $\hat{y}_{T+l/T}$ can be built as $[\hat{y}_{T+l/T} - 1.96\sqrt{V(l)}, \hat{y}_{T+l/T} + 1.96\sqrt{V(l)}]$. These confidence intervals are also displayed on figures 3.17-3.19.

Let us consider an AR(1) process $(1 - \phi B)y_t = a_t$. It is readily seen that $\psi(B) = 1 + \phi B + \cdots + \phi^n B^n$, so $\psi_i = \phi^i$. Predicting y_{T+l} using (3.19) yields

$$\hat{y}_{T+l/T} = \sum_{i=0}^{\infty} \phi^{l+i} a_{T-i} = \phi^l \sum_{i=0}^{\infty} \phi^i a_{T-i} = \phi^l y_T,$$

so that expression for the l-step-ahead forecasts in an AR(1) is recovered. The forecast error is given by:

$$y_{T+l} - \hat{y}_{T+l/T} = \sum_{i=0}^{\infty} \phi^i a_{T+l-i} - \phi^l \sum_{i=0}^{\infty} \phi^i a_{T-i} = \sum_{i=0}^{l-1} \phi^i a_{T+l-i},$$

with associated variance $V(l) = V_a \sum_{i=0}^{l-1} \phi^{2i}$. This expression however supposes the parameter ϕ to be known, while in practice only its estimator is available. Taking into account the error in the parameter estimates and denoting $\hat{y}^*_{T+l/T}$ and $\hat{y}_{T+l/T}$ the forecasts derived with an estimated parameter and a known parameter, respectively, then the previous expression can be reformulated as:

$$y_{T+l} - \hat{y}^*_{T+l/T} = (y_{T+l} - \hat{y}_{T+l/T}) + (\hat{y}_{T+l/T} - \hat{y}^*_{T+l/T}).$$

In the case of an AR(1), this expression simplifies to:

$$y_{T+l} - \hat{y}^*_{T+l/T} = (y_{T+l} - \hat{y}_{T+l/T}) + y_T(\phi^l - \hat{\phi}^l),$$

with associated MSE:

$$MSE(\hat{y}^*_{T+l/T}) = MSE(\hat{y}_{T+l/T}) + y_T^2 E[(\phi^l - \hat{\phi}^l)^2].$$

Therefore the forecast error variance can be seen as made up of an error when ϕ is known plus an error due to the estimation approximation. To evalute the magnitude of this last term, let us consider the simple case $l = 1$. The least-squares estimate of ϕ is given by

$$\hat{\phi} = \sum_{t=2}^{T} y_t y_{t-1} / \sum_{t=2}^{T} y_{t-1}^2,$$

and it can be checked that

$$E[(\hat{\phi} - \phi)^2] = V_a / \sum_{t=2}^{T} y_{t-1}^2.$$

Hence, a confidence interval taking into account the error in the estimator must embody the additive term $V_a y_T^2 / \sum_{t=2}^{T} y_{t-1}^2$. In large samples however, the denominator is large enough and verifies: $\sum_{t=2}^{T} y_{t-1}^2 \simeq T V_a/(1-\phi^2)$, so the correcting term is negligeable for large enough sample sizes. In practice, in most applications, this term is neglected and the confidence interval given are underestimated. The bias, however, should not be important unless in small samples (see also Harvey, 1981, 1989; Box and Jenkins,1970; and Fuller (1976)).

Finally, the following concluding remarks about general properties of forecasts in ARIMA(p, d, q) models are of interest.

- Whether a time series is stationary or not determines the general pattern that the forecast function and the confidence interval display when the forecast horizon becomes large. In particular, for stationary time series ($d = 0$), as the forecast horizon tends to infinity, the prediction tends to the mean of the series while the forecast error variance tends to the variance of the series: as $l \to \infty$, $\hat{y}_{T+l/T} \to E(y_t)$, and $V(l) \to V(y_t)$.

  For nonstationary time series ($d > 0$), as $l \to \infty$, the prediction follows a pattern given by a polynomial of order d with d-th coefficient given by the mean of the stationary transformation of the series. For example, forecasts of an $I(1)$ variable is simply a constant if the series first difference has a zero-mean, a straight line otherwise; see figure 3.17. Also, forecasts of an $I(2)$ variable follows a straight line if the mean of the differenced process is zero, a second-order time polynomial otherwise. On the other hand, the variance of the forecast error becomes infinite as l increases, so the forecasts lose their relevance as the forecast horizon increases; see figures 3.17-3.19.

- When the series under analysis has been log-transformed ($y_t = \log Y_t$), it may be of interest to obtain forecasts and MSE in the original scale. Since the conditional distribution of Y_{T+l} given $Y_1, \cdots, Y_T$ is lognormal, the conditional mean is given by $E(Y_{T+l}|Y_1, \cdots, Y_T) = exp(\hat{y}_{T+l/T} + .5V(l))$. In practice however, the approximation $\hat{Y}_{T+l/T} = exp(\hat{y}_{T+l/T})$ is used, the loss in accuracy being rather small. The confidence interval can be obtained as $exp(\hat{y}_{T+l/T} +/- 1.96\sqrt{V(l)}) \simeq \hat{Y}_{T+l/T}(1 +/- 1.96V(l))$.

Part II

Decomposing Time Series into Unobserved Components

Chapter 4

ARIMA-Model-Based Decomposition of Time Series

4.1 Introduction

We have seen in the first part of this book how time series can be represented by stochastic linear models of the ARIMA-type. The standard methodology for selecting a particular model, estimating it, and checking its capabilities for describing the behavior of the data has been discussed. The forecasting procedure has also been presented. We now turn to the problem of decomposing an observed process into unobserved components, where every component catches a particular underlying pattern of the series. We shall see how the decomposition can be closely linked to the description of the series as operated in part I, so that modelling and decomposing time series will come out as an integrated procedure.

4.2 The General Framework

4.2.1 Model Specification and Assumptions

The ARIMA-model-based approach to decompose time series assumes that an observed process x_t is made up of Unobserved Components (UC), typically a seasonal component s_t, a trend p_t, a cycle c_t and an irregular component u_t. The UC and the observed series are assumed to be related according to:

$$x_t = s_t + p_t + c_t + u_t,$$

where the additivity may be obtained after some suitable transformation of the observed series. For example, a prior log-transformation of the series implies that the initial relationship is multiplicative. Some component may be discarded if it is not needed. In the remaining of the discussion, we shall consider the more general two-components decomposition into a signal s_t and a nonsignal n_t, such that

$$x_t = n_t + s_t, \tag{4.1}$$

the signal representing the component of interest, the nonsignal resuming the rest of the series. For example, in seasonal adjustment problems, n_t would represent the nonseasonal part of the series such that $n_t = p_t + c_t + u_t$.

The ARIMA-model-based procedure, as originally developed by Box, Hillmer and Tiao (1978), Burman (1980), and Hillmer and Tiao (1982), considers the three following assumptions on the Unobserved Components.

Assumption 1: The Unobserved Components are uncorrelated.

This assumption may appear to be somewhat restrictive; we shall see that it is not required in order to obtain estimates of the UC. Some decompositions used in the literature consider correlated components (see for example Watson (1986)). Probably the most popular example of correlated components is given by the Beveridge-Nelson (1981) decomposition of I(1) series into a temporary and a permanent component. In this procedure both components turn out to be defined as linear combinations of the observed series x_t (see Maravall (1993b)). Since x_t is stochastic, the Beveridge-Nelson decomposition implicitly assumes that the components share the same innovation, which is a strong assumption.

Assuming instead independent components is a simplification which has some intuitive appeal. It is justified by the idea that the evolution of the different components is driven by separate forces. A typical illustration of the applicability of this represen-

tation is provided by the monetary control problem. This arises because central banks often rely on seasonally adjusted money demand estimators to take decision about money supply in the next period. In this case, the orthogonality hypothesis amounts to considering the seasonal and long-term evolution of the monetary aggregates as being driven by different causes: the long-term path would be related to the economic fundamentals, while the seasonal variations would be related to events such as holidays timing or the Christmas period. This seems reasonable and thus supports the use of seasonally adjusted series for policy making. In general, the orthogonality hypothesis is standard in practical applications such as the short-term monitoring of the economy.

The next assumptions concern the stochastic properties of the components.

Assumption 2:

The correlation structure of the Unobserved Components is supposed to be well described by ARIMA *models of the type:*

$$\begin{aligned} \phi_n(B)n_t &= \theta_n(B)a_{nt}, \\ \phi_s(B)s_t &= \theta_s(B)a_{st}, \end{aligned} \tag{4.2}$$

where the variables a_{nt} and a_{st} are normally distributed white noise with variances V_n and V_s. The models are not reducible; that is each pair of polynomials $\{\phi_n(B), \theta_n(B)\}$, $\{\phi_s(B), \theta_s(B)\}$, are prime. Furthermore, the polynomials $\phi_n(B)$, $\phi_s(B)$, $\theta_n(B)$, and $\theta_s(B)$, of order respectively p_n, p_s, q_n, and q_s, may have their roots on or outside the unit circle, but $\phi_n(B)n_t$ and $\phi_s(B)s_t$ are required to be stationary. In the cases where an irregular component is considered separately, then the irregular u_t will be a white noise with variance V_u.

Notice that, as implied by assumption 1, the innovations a_{nt} and a_{st} are independent. The specification of ARIMA models for Unobserved Components can be found in Cleveland and Tiao (1976), Box, Hillmer and Tiao (1981), Pierce (1978), Burman (1980) and Hillmer and Tiao (1982) for the early references. A restriction must however be considered:

Assumption 3: The AR polynomials $\phi_n(B)$ and $\phi_s(B)$ do not share any common roots.

Assumption 3 implies that the spectra of the UC do not have peaks at the same frequencies. Since the movements displayed by different components are related to different spectral peaks, this is a reasonable feature of the decomposition. From assumptions 1, 2, and 3, we obtain:

$$\begin{aligned} \phi_n(B)\phi_s(B)x_t &= \phi_s(B)\phi_n(B)n_t + \phi_n(B)\phi_s(B)s_t = \\ &= \phi_s(B)\theta_n(B)a_{nt} + \phi_n(B)\theta_s(B)a_{st}. \end{aligned}$$

so the observed series x_t follows a ARIMA model of the type:

$$\phi_x(B)x_t = \theta_x(B)a_t,$$

where the polynomials $\phi_x(B)$ and $\theta_x(B)$ are respectively of order p_x and q_x. The polynomial $\phi_x(B)$ is such that: $\phi_x(B) = \phi_n(B)\phi_s(B)$, no common root between the polynomials $\phi_n(B)$ and $\phi_s(B)$ being allowed. Thus $p_x = p_n + p_s$. The repartition of the different roots of $\phi_x(B)$ between the polynomials $\phi_n(B)$ and $\phi_s(B)$ depends on the behavior that the components are expected to display. For example, a unit root $B = +1$ implies an infinite peak in the spectrum of x_t at the zero frequency. Given that the low frequencies movements are associated with the long-term evolution of the observed series, this unit root would be assigned to the trend component. The MA process $\theta_x(B)a_t$ verifies:

$$\theta_x(B)a_t = \phi_s(B)\theta_n(B)a_{nt} + \phi_n(B)\theta_s(B)a_{st}, \tag{4.3}$$

where a_t is a normally distributed white noise with variance V_a. We set $V_a = 1$ so that all other variances will be expressed as a fraction of V_a. Further, we shall also assume without loss of generality that the MA process $\theta_x(B)a_t$ is invertible. Equation (4.3) also implies that the order of the MA polynomial is constrained by:

$$q_x \leq max(p_n + q_s, p_s + q_n). \tag{4.4}$$

Equations (4.1), (4.2) and assumptions 1, 2, and 3 constitute a model which will be refered to as model (A). It is the general model, valid for linear processes with normal innovations, that will be discussed in this second part. Since our discussion will focus on the characterization and estimation of the components, and since the model for observed series can be consistently estimated, we shall retain the following assumption:

Assumption 4: The model for the observed series is known.

In other words, the polynomials $\phi_x(B)$, $\theta_x(B)$, and the innovation variance V_a are known. As discussed in the previous chapters, the knowledge of the model followed by the observed series is reached after estimation using Box-Jenkins techniques.

We need some notations about the Auto-Covariance Generating Function (ACGF) and about the spectra of the observed series and of the components. Throughout the remaining of the discussion, we will denote by A_i, $i = x, n, s$ the ACGF of respectively x_t, n_t, s_t. These are defined under the hypothesis of stationarity as:

$$A_i = V_i \frac{\theta_i(B)\theta_i(F)}{\phi_i(B)\phi_i(F)}. \tag{4.5}$$

Using the Fourier transform $B = e^{-iw}$ in (4.5), w denoting frequency in radians such that $w \in [-\pi, \pi]$, the equation above also defines the spectra $g_i(w)$. When one or both components are nonstationary, neither the spectra nor the ACGF of the nonstationary components and of the observed series are strictly defined: the presence of a unit root in a AR polynomial $\phi_i(B)$ implies an infinite peak in $g_i(w)$ and thus an infinite variance. However, the definitions of $g_i(w)$ and A_i provided in (4.5) may be extended to cover nonstationary cases, as in Hillmer and Tiao (1982), Bell and Hillmer (1984), and Harvey (1989), who refer to them as pseudo-spectrum and pseudo-ACGF. Since we do not make any assumptions about order of integration of the observed series, we shall refer to the functions $g_i(w)$ and A_i, $i = x, n, s$, simply as the spectrum and the ACGF, whether the components are stationary or not in order to simplify the presentation.

4.2.2 Characterization of the Components: some examples

Model (A) is quite general. It embodies many possible applications, the most important of which are possibly detrending of time series, seasonal adjustment, cycle analysis and noise extraction. These applications involve the components discussed by Persons (1919,1923), namely the trend, the seasonal, the cycle and the irregular component, and introduced in chapter 1. We briefly present the models most commonly used to characterize these components.

Trend Component

The general form for a stochastic linear trend can be written as:

$$\Delta^d s_t = \psi_s(B) a_{st},$$

where $0 \leq d \leq 3$, and $\psi_s(B)a_{st}$ is a low order ARMA processes. In the ARIMA-model-based approach, trends are often specified as IMA(2,2) models. Other model specifications used for example by Harrisson and Steven (1976), by Harvey and Todd (1983), and by Ng and Young (1990), and which are commonly encountered in the Structural Time Series approach, consider "second order" random walks processes such that:

$$\Delta s_t = \mu_t + a_{st},$$

where the drift is itself a random walk:

$$\Delta \mu_t = a_{\mu t},$$

where $a_{\mu t}$ is a white-noise variable with variance V_μ. Taking the second difference of s_t, it is readily seen that this model can be expressed as an IMA(2,1). Thus the second order random walks that STS models typically consider are a particular case of the IMA(2,2) models for the trend encountered in ARIMA-model-based approach. Notice that if $a_{\mu t}$ is a null constant, then the second order random walk model reduces to a simple random walk plus drift, which is commonly used in applied econometrics.

The above formulation may be easily interpreted as a stochastic extension of linear deterministic trends. Setting V_μ and V_s to zero, so that $\Delta \mu_t$ is constant, the corre-

sponding deterministic trend function is trivially obtained as a quadratic polynomial of time if $a_{\mu t}$ has a non-zero mean, as a linear function of time otherwise. With the IMA(2,2) modelling, this is equivalent to having roots of 1 in the MA polynomial. Allowing for stochasticity instead of deterministic expressions enables the trend component to evolve over time, which is an expected feature of the specification. Models for stochastic linear trends have been exhaustively discussed in Maravall (1993a).

Seasonal Component

As introduced in the first chapter, models for seasonal components could also be interpretated as stochastic extensions of deterministic models. The aim is still to allow the component to display a slowly changing behavior. Starting with a deterministic seasonality for which $s_t = s_{t-m}$, m representing the number of observations per year, then the sum of m consecutive seasonal components will be zero, or:

$$U(B)s_t = 0,$$

where $U(B) = 1 + B + \cdots + B^{m-1}$. The periodic nature of the seasonal component is captured here, but the seasonal fluctations are excessively restricted. Small deviations from this strict model specification may be allowed by making the relationship subject to a random shock in each period:

$$U(B)s_t = a_{st}.$$

This type of stochastic seasonal model is considered for example in the Gersch and Kitagawa (1983) and Harvey and Todd (1983) approaches. More generally, we can allow the deviation from zero of $U(B)s_t$ to be correlated and consider:

$$U(B)s_t = \theta_s(B)a_{st},$$

which is mostly used in the ARIMA-model-based approach with the MA polynomial $\theta_s(B)$ of order $m-1$. The spectrum of the corresponding component will present peaks centered on the seasonal frequencies $2k\pi/m$, $k = 1, 2, \cdots, m-1$.

Departures from this type of model specification may be found in the statistical literature. For example, Pierce (1978) considered both stochastic and deterministic seasonality. Probably the most common departure found in earlier model-based approaches was to model the seasonal component with Δ_m in its AR part. This should be avoided because this polynomial contains the root $(1 - B)$ which is related with low frequencies movements and should thus be assigned to the trend. This point is also treated in Maravall (1989) where the seasonal component model specification is thoroughly discussed.

Cyclical Component

The cyclical component can be handled in two different ways. The first approach designates the "cycle" to be the residual of the detrending of a nonseasonal series. This approach is quite common in macroeconomics, in particular in business cycle analysis where the "cycle" usually describes the nonseasonal deviations from the long term evolution of time series. With this representation, the cycle corresponds thus to the stationary variations of the series. In general, it is well described by an ARMA process.

The second approach explicitly models the cyclical component. It involves models which are able to generate periodicities longer than a year. For example, consider the model:

$$s_t + \phi_{s1} s_{t-1} + \phi_{s2} s_{t-2} = 0,$$

with $\phi_{s1}^2 < 4\phi_{s2}$. It is easily checked that a signal s_t so-defined will display a deterministic periodic behavior with frequency $w = \arccos(-\phi_{s1}/2\sqrt{\phi}_{s2})$. When this frequency is lower than the fundamental frequency $2\pi/m$, then the behavior of s_t will show a period longer than a year. As for the previous cases, small deviations from a strictly deterministic behavior are allowed by considering:

$$s_t + \phi_{s1} s_{t-1} + \phi_{s2} s_{t-2} = \theta_s(B) a_{st},$$

where $\theta_s(B) a_{st}$ is a low order moving average. Jenkins (1979) and Harvey (1985), among others, have used such "periodic cycles" models.

Irregular Component

The irregular component corresponds to the noisy part of the series. It is typically modelled as a stationary low-order ARMA process and most often, in the model-based approach, it is a pure white noise process. This component is of interest for example when the observations are known to be contaminated by some noise, and the user desires to recover the original signal. Such situations occur for instance in communications engineering.

It is worth noticing that in these four examples, the differences between the models for the different components come basically from the AR polynomial. However, it is also important to look at the MA polynomials and at the components innovations variances. We now return to the general model (A) to examine this point.

4.2.3 Admissible Decompositions.

Since the AR polynomials are identified directly from the factorization of $\phi_x(B)$, the unknowns of the model consist of the coefficients of the polynomials $\theta_s(B)$, $\theta_n(B)$, and of the innovation variances V_s and V_n. In model (A), information on the stochastic structure of the components is brought by the observed series and by the overall relationship (4.3). This relationship implies a system of $\max(p_s+q_n, p_n+q_s)+1$ covariance equations while the number of unknowns is q_n+q_s+2. So when $\max(p_s+q_n, p_n+q_n) < q_n+q_s+1$, and in the absence of an additional assumption, there exists an infinite number of ways to decompose the series x_t. Any decomposition consistent with the overall model for the observed series and insuring non negative components spectra will be called an 'admissible decomposition'. All admissible decompositions are of course observationally equivalent.

The underidentification of UC models is illustrated with the following example:

Example: Trend plus Cycle decomposition.

Consider the following decomposition:

$$\begin{aligned} x_t &= n_t + s_t \quad \text{with,} \\ \Delta s_t &= a_{st}, \\ (1-\phi B) n_t &= (1+\theta_n B) a_{nt}, \end{aligned} \tag{4.6}$$

where we shall set $\phi = -.7$. The assumptions 1-3 are supposed to hold. Equations (4.6) represents a simple model designed to decompose a time series into a trend (s_t) and a cycle component (n_t). The trend is modelled as a random walk and the cycle as a stationary ARMA(1,1), with period 2. It is a particular case of the model used in Stock and Watson (1993) to analyze the business cycle and to forecast recessions. This specification implies that the observed series x_t follows an ARIMA(1,1,2) model specified as:

$$(1 - \phi B)\Delta x_t = (1 - \theta_1 B - \theta_2 B^2)a_t,$$

with $a_t \sim$ sc Niid(0,1) and the MA polynomial is invertible given that $g_s(w)$ is always positive. Since the components must sum to the observed series, we have:

$$(1 - \theta_1 B - \theta_2 B^2)a_t = (1 - \phi B)a_{st} + (1 - B)(1 + \theta_n B)a_{nt},$$

from where we obtain the following system of covariance equations:

$$\begin{aligned} 1 + \theta_1^2 + \theta_2^2 &= (1 + \phi^2)V_s + (1 + (-1 + \theta_n)^2 + \theta_n^2)V_n \\ -\theta_1(1 - \theta_2) &= -\phi V_s - (1 - \theta_n)^2 V_n \\ -\theta_2 &= -\theta_n V_n. \end{aligned}$$

Taking $\theta_n = .3$, $V_n = .129$, and $V_s = 5V_n = .645$, the model for the observed series is then obtained as:

$$(1 + .7B)\Delta x_t = (1 + .404B - .039B^2)a_t.$$

The series is simulated on figure 4.1 and the components s_t and n_t on figure 4.2. On subsequent figures, this model will be refered to as "model 1".

Figure 4.1: Observed Series in Trend plus Cycle example

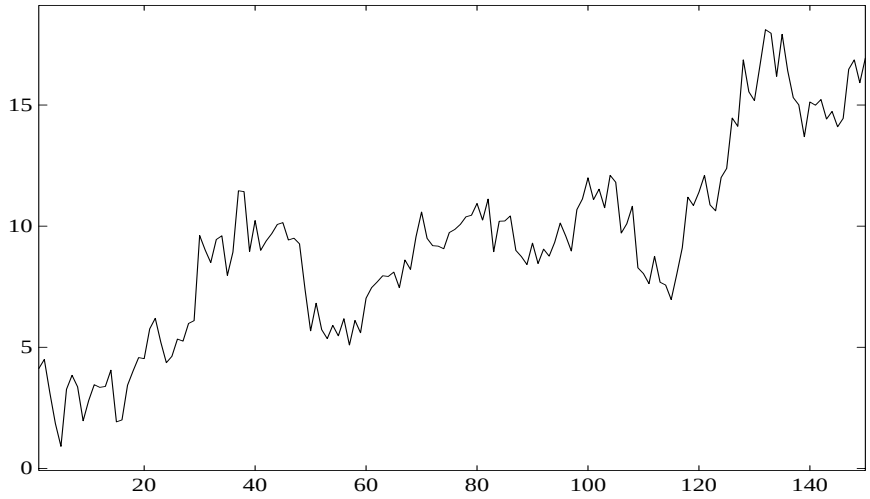
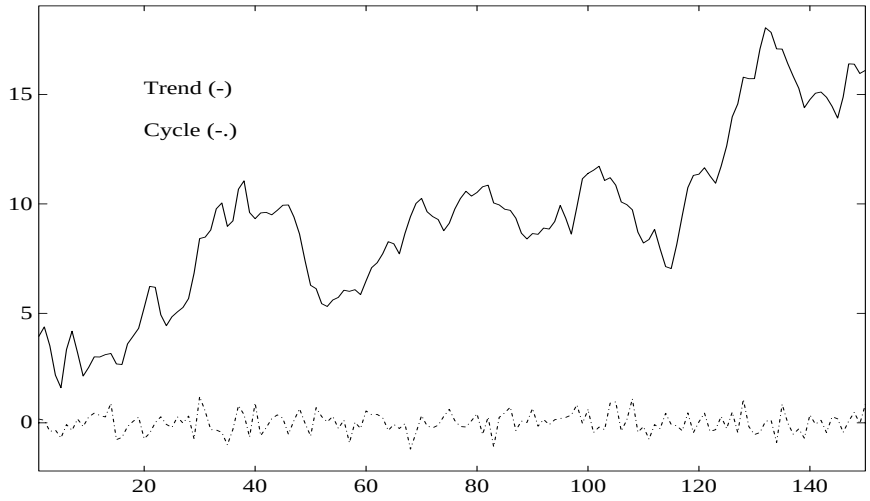


Figure 4.2: Components in Trend plus Cycle example



Now we change the specification of the components' models and consider an IMA(1,1) for the trend:

$$\begin{aligned} \Delta s_t &= (1+\theta_s B)a_{st} \\ (1-\phi B)n_t &= (1+\theta_n B)a_{nt}. \end{aligned}$$

The new system of covariance equations that this model specification implies is given by:

$$\begin{aligned} 1+\theta_1^2+\theta_2^2 &= [1+(\theta_s-\phi)^2+\phi^2\theta_s^2]V_s + [1+(-1+\theta_n)^2+\theta_n^2]V_n \\ -\theta_1(1-\theta_2) &= (\theta_s-\phi)(1-\phi\theta_s)V_s - (1-\theta_n)^2 V_n \\ -\theta_2 &= -\phi\theta_s V_s - \theta_n V_n. \end{aligned}$$

It can easily be checked that for $\theta_n = .434$, $V_n = .220$, and $V_s = .470$, and $\theta_s = .172$, the MA parameters θ_1 and θ_2 remain unchanged. This last decomposition, referred to as "model 2", is thus consistent with the overall model that (4.6) had generated. Two model specifications have generated the same observed series. They correspond thus to two admissible decompositions which are observationally equivalent.

What is the difference between the two decompositions ? It is more convenient to answer the question in the frequency domain. The spectra of the components for the two decompositions are plotted on figure 4.3 and 4.4. It can be seen that for each component, the spectra obtained from the two model specifications differ only by a constant. This constant can be interpreted as the size of an orthogonal white noise which has been interchanged between the two components. To isolate it, let us look at the spectra minima. For the first model, the trend spectrum has a minimum at the π frequency equals to: $g_s(\pi) = V_s/4 = .161$. In the second case, this minimum becomes: $V_s/8 = .081$. Therefore, a white noise variable of variance $V_s/8$ has been removed from the trend component spectrum in model 2. Since the components must aggregate to the observed series, this noise has been assigned to the cycle which embodies more

Figure 4.3: Spectra for the cycle component

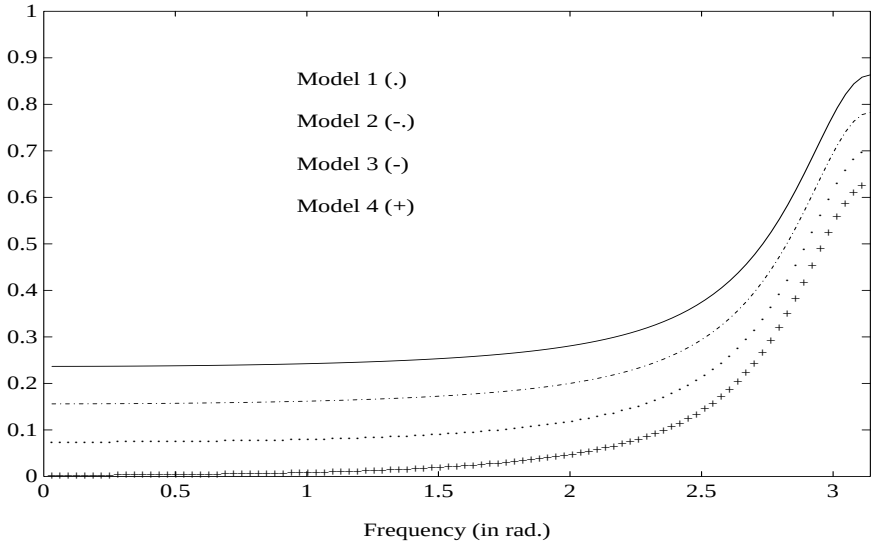
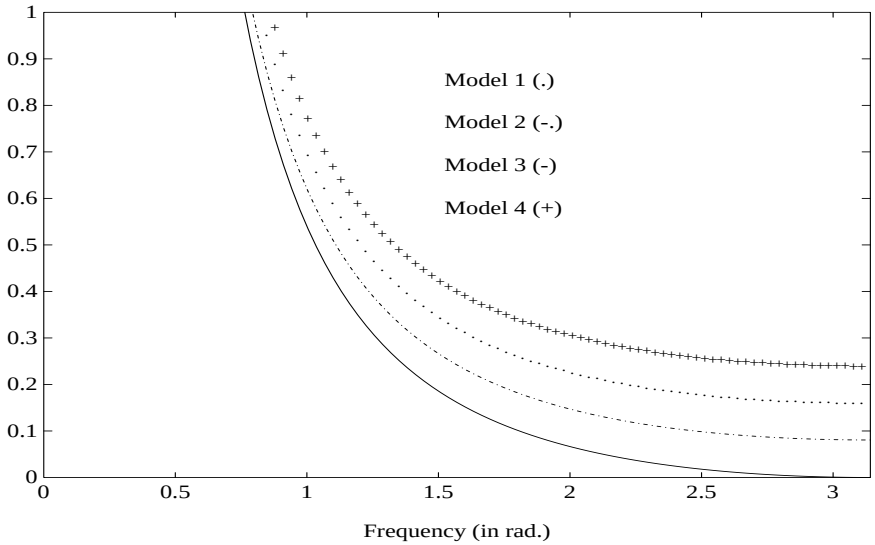


Figure 4.4: Spectra for the trend component



stochasticity. At the extreme, we could remove all the noise from the trend component and assign it to the cycle. The decomposition that we would obtain is found to be:

$$\begin{aligned} \Delta s_t &= (1+B)a_{st} \qquad V_s = .161 \\ (1+.7B)n_t &= (1+.496B)a_{nt} \qquad V_n = .306, \end{aligned} \tag{4.7}$$

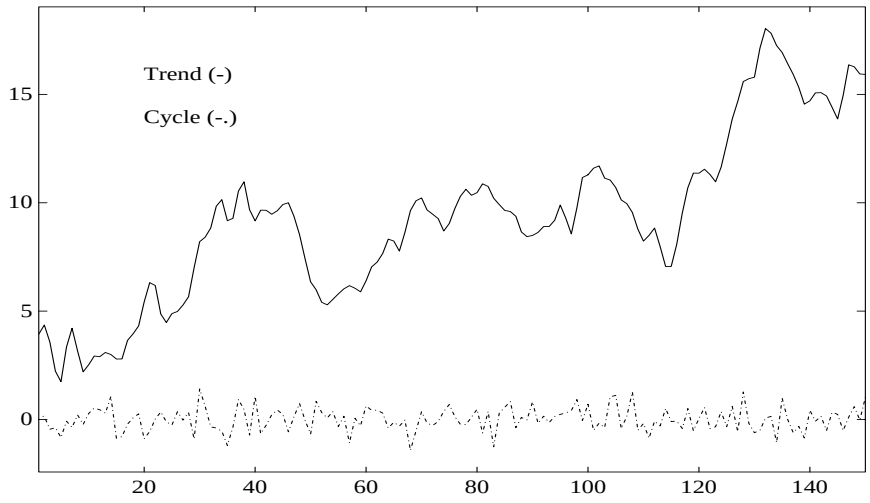
and the trend is now noninvertible, the zero in the spectrum at the π-frequency meaning that it does not embody any orthogonal noise. The spectrum for these components are also seen on figures 4.3 and 4.4. If we look on figure 4.5 at the plot of the components so-obtained and compare it with the components of the first decomposition, it can be clearly seen that the trend is smoother and the 'cycle' noisier. Decompositions where one component is noninvertible while the other concentrates all the noise of the model are called 'canonical' (see Hillmer and Tiao (1982)). Alternatively, one may be interested by the cycle analysis. In this case, it is possible to assign all the noise of the model to the trend component, and to remove it from the spectrum of the cycle. We thus obtain a second canonical decomposition, denoted "model 4" on figures 4.3 and 4.4, where the canonical component is now the cycle.

As shown on figures 4.3 and 4.4, the identification problem can be thus seen as the problem of determining the spectrum of the component within the range delimited below by the spectrum of the component free of noise and above by the spectrum of the component concentrating all the noise of the model. The admissible decompositions are thus characterized by a particular noise allocation. The problem is to select a decomposition among the admissible ones.

4.3 Identification of Unobserved Components Models.

We have just seen that in the "Trend plus Cycle" example, the decomposition was not unique. We had a system of 3 equations for 4 unknowns which thus could not be uniquely solved. Each set of parameters consistent with this system and insuring non negative components spectra defined an admissible decomposition. The difference between admissible decompositions could also be interpreted in terms of different noise repartitions. In this section, we discuss the identification problem for the general

Figure 4.5: Trend plus Cycle example: canonical trend and noisy cycle



model (A) under the two possible perspectives: as a parametric problem and as a noise repartition problem. Two identification criteria used in the statistical literature which correspond to these two perspectives are presented. We first need some essential concepts about identifiability.

4.3.1 The Concept of Identifiability

The definition of identifiability involves the distinction between a *structure* and a *model*. According to Rothenberg (1971) and Hotta (1989), a structure and a model may be defined in the following way:

Definition 4.1 *A structure S is a set of hypothesis which implies a unique distribution function of the observable Y, say $P[Y/S]$. The set denoted $\mathcal{S}$ of all a priori possible structures is called a model.*

A structure thus specifies a set of parameters for a given distribution. That distribution is in turn characterized by the model. Considering model (A), each structure is formed by a particular set of parameters $S_A=\{\theta_{n1}, \cdots, \theta_{nq_n}, \theta_{s1}, \cdots, \theta_{sq_s}, V_n, V_s\}$ lying within

the admissible parameter space. Two structures will then be *observationally equivalent* if they imply the same distribution function, or, formally:

Definition 4.2 *Two structures S_1 and $S_2 \in \mathcal{S}$ are said to be observationally equivalent if $P[Y/S_1] = P[Y/S_2]$ almost everywhere.*

Identifiability may then be defined in the following way:

Definition 4.3 *A structure $S \in \mathcal{S}$ is said to be identifiable if there is no other structure in $\mathcal{S}$ which is observationally equivalent. A model $\mathcal{S}$ is identified if all the structures are identified.*

This is the standard econometric/statistical definition of identification. No confusion should be made with the concept of identification used in Box and Jenkins (1976) which is related to specification.

Model (A) would thus be identified if the distribution of the observable x_t would be generated by a unique set of parameters S_A. This is not the case, and in general unobserved component models are not identified. It is tempting to relate the underidentification of the components with the lack of a precise definition of the components. For example, suppose we are interested in removing the seasonal variations of a time series. The seasonality is related to movements which repeat themselves with a periodicity corresponding to year. Consequently a seasonal component can be defined by the spectral peaks located at the seasonal frequencies, as in Granger (1978). However, such a definition is not precise enough to imply a unique model for the seasonal component. Spectral peaks are generated by large roots in AR polynomial; nothing is said about what should be the MA polynomial and the component innovation variance. In the same way, since it represents the long-term movements, a trend component may be defined by an infinite spectral peak at the low frequencies. But from this definition several models for a trend component may be derived. In the trend plus cycle example, we had several possible trend components, all perfectly admissible.

In other words, unobserved components cannot be precisely defined, and as a consequence they are not identified. Identification problems are also encountered in simultaneous econometric models. But, if in these cases economic theory enables the analysts to overcome the underidentification of the model by setting at zero some parameters, in

statistics, as noticed in Maravall (1988b), no such a priori information is available. Any model-based approach must thus consider an arbitrary assumption on the components.

This gives rise to two important problems in UC models. Firstly, as recognized by statisticians (see Bell and Hillmer (1984)), it makes difficult to evaluate a unobserved component estimation procedure: it is not possible to compare methods estimating different signals. This point is particularly relevant in the seasonal adjustment context where several different methods are still in use. Secondly, given that one is interested in a signal, which model form should be chosen ? One of the most attractive feature of a model-based approach is that the arbitrary assumptions on the components models are made explicit, while they are somewhat hidden in empirical methods. We turn now to see the main assumptions adopted in model-based decompositions of time series.

4.3.2 Identification by zero-coefficients restrictions and the Structural form approach

Necessary and sufficient condition for identifiability

For a Gaussian model such as model (A), a structure reduces to a set of parameters consistent with the first and second moments of the stationary transformation of the observed series. When the first moment of the stationarised series is null, as in model (A), it is enough to use the autocovariance generating function or the spectrum of the stationarised series to check the identifiability of the underlying structures.

For the general model (A), the relationship (4.3) provides the following identity:

$$\theta_x(B)\theta_x(F) = \phi_s(B)\phi_s(F)\theta_n(B)\theta_n(F)V_n \quad + \quad \phi_n(B)\phi_n(F)\theta_s(B)\theta_s(F)V_s \tag{4.8}$$

which implies a set of covariances equations by equating the coefficient the coefficient of B^j. The right hand side of (4.8) contains $q_n + q_s + 2$ unknowns, which are θ_{n1}, $\cdots$, θ_{nq_n}, V_n, θ_{s1}, $\cdots$, θ_{sq_s}, and V_s, while the left hand side yields $q_x + 1$ covariances equations. So when $q_n + q_s + 1 > q_x$, the system is underidentified and instead of a unique decomposition, a set of observationally equivalent decompositions is obtained. Using $q_x = max(p_n + q_s, p_s + q_n)$, we can easily deduce that the necessary condition

for identification of model (A) is:

$$q_s < p_s \quad \text{and/or} \quad q_n < p_n. \tag{4.9}$$

Hotta (1989) has shown that (4.9) is also a sufficient condition. It thus possible to restrict the order of the components MA polynomial in order to identify the underlying structure of the model. This has been mostly used in 'structural models':

4.4 Example of identification by zero-parameters restriction: Structural Time Series models

The Structural Time Series (STS) models have been developed by Engle (1978) and Harvey and Todd (1983), and are applied mostly to the modelling of economic time series. They are usually designed for the purposes of extracting the trend or seasonally adjusting time series. The approach followed consists firstly of specifying *a priori* the models for the components, and then in estimating the parameters. Identification is obtained by reducing the order of one of the components MA polynomial, typically the trend or the seasonal component. Consider for example the Basic Structural Model which decomposes an economic time series into a trend and a 'nontrend' component. The trend component is typically modelled as a random walk with a drift, the drift itself being a random walk:

$$\begin{aligned} \Delta s_t &= \mu_t + u_t, \\ \Delta \mu_t &= v_t, \end{aligned}$$

where u_t and v_t are orthogonal white-noises. Maravall (1985) noted that this model is equivalent to specifying an IMA(2,1) model for the trend component, which makes explicit the identification restrictions in structural models. For the seasonal and the cycle components, they are usually specified as subject to an uncorrelated random shock in each period, so that no MA term is considered.

4.5 The Canonical Decomposition

Setting the order of the MA polynomials in order to identify the UC is however an *a priori* identification procedure that may not be justified by any extra consideration. In general, from any invertible component specified in this manner, there exists a particular amount of noise that can be removed, yielding a component balanced in the polynomials orders. Consider for example a trend specified as a random walk:

$$\Delta s_t = a_{st}.$$

The spectrum of s_t is given by: $g_s(w) = V_s/(2 - 2\cos w)$. It is immediately seen that the spectrum of the random walk has a minimum at the frequency π, of magnitude: $g_s(\pi) = V_s/4$. Removing a proportion $\alpha \in]0, 1]$ of this noise from $g_s(w)$ yields

$$g_s(w) - \alpha V_s/4 = \frac{V_s}{4}\frac{4 - 2\alpha + 2\alpha\cos w}{2 - 2\cos w} \tag{4.10}$$

which is the spectrum of an IMA(1,1) model.

Balanced components are more general in the sense that they allow different noise repartitions. The order of the MA polynomial for the remaining component will however be constrained by the overall model for the observed series. Clearly, with balanced components the necessary and sufficient condition (4.9) for identification of model (A) is no longer satisfied: an additional condition must be imposed. The canonical decomposition is an identification procedure which uses some extra considerations in order to obtain identification. It handles the identification problem explicitly as a noise repartition problem.

The hypothesis of independence of the components yields the following relationship:

$$g_x(w) = g_s(w) + g_n(w).$$

As in Burman (1980), we write : $\varepsilon_s = \min_w g_s(w)$ and $\varepsilon_n = \min_w g_n(w)$. The quantity $\varepsilon_s + \varepsilon_n$ can be seen as the variance of a pure noise component embodied in the spectrum of the observed series which can be attributed arbitrarily. It is clear that the

identification problem arises because we do not know which amount of noise must be assigned to the components : a fraction of ε_s and of ε_n could be removed from every component spectrum, and assigned to the other component. If we remove as much noise as possible from s_t and attribute it to n_t, then we obtain : $g_s^0(w) = g_s(w) - \varepsilon_s$ and $g_n^0(w) = g_n(w) + \varepsilon_s$ the spectra of a noninvertible signal and of a nonsignal which concentrates all the noise of the model. This decomposition is known as *canonical*. In the random walk example, taking $\alpha = 1$ in (4.10) yields:

$$g_s^0(w) = g_s(w) - V_s/4 = \frac{V_s}{4}\frac{2 + 2\cos w}{2 - 2\cos w},$$

which corresponds to the noninvertible model:

$$\Delta s_t = (1 + B)a_{st}^*,$$

with $Var(a_{st}^*) = V_s/4$.

The canonical decomposition was first proposed by Box, Hillmer and Tiao (1978) and Pierce (1978). As illustrated above, the approach consists in specifying a component as clean of noise as possible, and a canonical signal has a zero in its spectrum which corresponds to a unit root in the MA polynomial. That is, a canonical signal is noninvertible. An interesting property of canonical decompositions is that the admissible models for a signal can always be written as the sum of the canonical one plus an orthogonal white noise. Furthermore, Hillmer and Tiao (1982) showed that the canonical decomposition minimizes the variance of the signal innovation. Next, where an irregular component is to be isolated, then its variance is maximized when the other components are set canonical. The canonical decomposition is widely used in the ARIMA-model-based approach. Additional properties of the canonical decomposition will be given in chapter 5.

Chapter 5

Estimation of the Components by Ad Hoc Filtering

5.1 Introduction

The previous chapter has shown how models for the components may be set up. Since the components are never observed, the estimation of the components is then a necessary step. This chapter presents some general principles of estimation by linear filtering. It will be seen how these principles are involved in standard decomposition procedures like the ones developed in the softwares like X12-ARIMA (see Findley and al., 1996). Some drawbacks of this type of approach will be underlined.

5.2 Linear Filters

A linear time invariant filter can be expressed as:

$$a(B) = \sum_{k=-m}^{r} a_k B^k, \tag{5.1}$$

where the weights a_i are real, do not depend on time, and satisfy $\sum a_i^2 < \infty$. We shall focus on filters which, like $a(B)$, are time invariant. When the a_i's weights sum to unity, $a(B)$ is also called a *moving average filter*. Care must be taken to not make confusion with a moving average process. Filtering a series x_t with $a(B)$ yields a process y_t such that

$$y_t = a(B)x_t = \sum_{k=-m}^{r} a_k x_{t-k}.$$

Assuming that x_t is a stochastic stationary process, the stochastic properties of the series x_t are transmitted to the series y_t by means of the filter $a(B)$. It is convenient to draw the interpretation of this filtering operation in the frequency domain. Taking the Fourier transform $B = e^{-iw}$ in (5.1), then the *transfer function* $\Gamma(w)$ associated with the filter $a(B)$ can be defined according to:

$$\Gamma(w) = a(e^{-iw}) = \sum_{k=-m}^{r} a_k e^{-ikw}. \tag{5.2}$$

The transfer function enables us to understand the effect of a linear filter on an input time series: it shows that a selection of the movements in the input series x_t which are passed to y_t is performed on the basis of the frequency of the different movements. For example, suppose that $a(B)$ is such that its transfer function verifies $\Gamma(w) = 0$ for $w \in [2\pi/5, 2\pi/10]$. If the x_t's were annual data, the movements of periods from 5 to 10 years in the x_t's would be suppressed by the filter $a(B)$, so that y_t would not display any fluctuations in that band.

Notice that, in general, the transfer function $\Gamma(w)$ takes complex values. If the real and imaginary parts are expressed as $\Gamma(w) = \Gamma_R + i\Gamma_I$, another writing of $\Gamma(w)$ is then

$$\Gamma(w) = \gamma(w)e^{-iPh(w)},$$

where $\gamma(w)$ and $Ph(w)$ are respectively the modulus and the argument of the transfer function. Standard algebra yields

$$\gamma(w) = \mid \Gamma(w) \mid = \sqrt{\Gamma_R^2 + \Gamma_I^2}$$

and

$$Ph(w) = Arg(\Gamma(w)) = \tan^{-1}(-\Gamma_I/\Gamma_R).$$

The quantities $\gamma(w)$ and $Ph(w)$ are of particular relevance for the analysis of the filter $a(B)$, and are usually refered to as *gain* and *phase*, respectively. Gain and phase characterize the filter $a(B)$ in different ways. The gain is related to the amplitude of the movements, and if we look at the relationship between the spectra of the input and output, respectively denoted $f_x(w)$ and $f_y(w)$, we have:

$$f_y(w) = a(e^{-iw})a(e^{iw})f_x(w) = \gamma(w)^2 f_x(w).$$

The squared gain controls the extent in which a movement of particular amplitude at a frequency w is delivered to the output series. For example, a gain of zero in $[w_1, w_2]$ corresponding to a transfer function vanishing in this band will make the output series free of movements in this range of frequencies. Consider for example the transformation:

$$y_t = x_t - x_{t-5}.$$

The filter $a(B)$ is thus such that: $a(B) = 1 - B^5$, with response function given by:

$$a(e^{-iw}) = 1 - e^{-i5w} = 1 - \cos 5w + i \sin 5w,$$

while the squared gain is

$$\gamma^2(w) = \mid a(e^{-iw}) \mid^2 = (1 - e^{-i5w})(1 - e^{i5w}) = 2 - 2\cos 5w.$$

The squared gain of the filter is null at the frequency $2\pi/5$: if x_t is observed with a yearly periodicity, then every movement with 5-years periodicity in x_t would be cancelled by the application of the filter Δ_5, and the resulting series y_t will not display any 5-years periodicity fluctuations. In general, the Δ_d operator cancels the movements associated with the $k2\pi/d$ frequencies, $k = 1, 2, \cdots$. On the opposite, a gain higher than one would lead to an increase in variability in y_t with respect to x_t at the associated frequencies: trivially, if y_t is such that $y_t = a_0 x_t$, $a_0 > 1$, then y_t will display a larger variability than x_t all along the frequency range $[0, \pi]$.

Now let us consider the simple filter $a(B) = B^5$, for which an input series x_t yields $y_t = x_{t-5}$. Applying (5.2), the transfer function of this filter is given by

$$\Gamma(w) = e^{-5iw} = \cos 5w - i \sin 5w,$$

whose argument is readily seen to be $Ph(w) = 5w$. The lateral shift of five periods in time corresponds to a shift of $5w$ in the frequency domain. It is this shift in the frequency domain which is represented by the phase. In general, for periodic movements associated with the frequency w, y lags x by $Ph(w)/w$.

Equation (5.2) implies that the phase will be null when $a_i = a_{-i}$. This is a reason why symmetric filters are often preferred: they imply no-phase shift in the filtered series. This requirement is of importance, as it would be particularly inconvenient for data analysis to have the series and the filter output showing swings with ups and downs at different timings. A simple example of symmetric filter is given by

$$y_t = x_{t-1} + x_t + x_{t+1}$$

so that $a(B) = B + 1 + F$. The transfer function associated with this filter is obtained as

$$\Gamma(w) = e^{-iw} + 1 + e^{iw} = 1 + 2\cos w,$$

and the squared gain is

$$\gamma(w)^2 = (e^{-iw} + 1 + e^{iw})(e^{iw} + 1 + e^{-iw}) = 3 + 4\cos w + 2\cos 2w.$$

It is readily seen that the gain is null at the frequencies $w = 2\pi/3$ and $w = 4\pi/3$. Thus movements in x_t occuring with a periodicity 3 are cancelled by the filter. Supposing x_t is observed three times a year, then $w = 2\pi/3$ corresponds to the *seasonal fundamental frequency*, and the output will not display this seasonal behavior.

It is thus possible to build linear filters in order to annihilate some movements of particular periodicities in time series. This is the approach which has been adopted in seasonal adjustment by empirical filtering as performed by X11.

5.3 Ad hoc filters: X11 Seasonal Adjustment filters

During the sixties, the Bureau of the Census has developed a method for seasonally adjusting time series and for estimating trends (see Shiskin and al., 1967), which is known as X11. This method and the sofware implementing it have been widely spreadt, and most national statistical offices and central banks are still using it. Several updates of the programs has been delivered, the last of them being called X12-RegArima (see Findley and al., 1996). Given the consideration that these programs have received in applied works, the filters that they implement are now discussed.

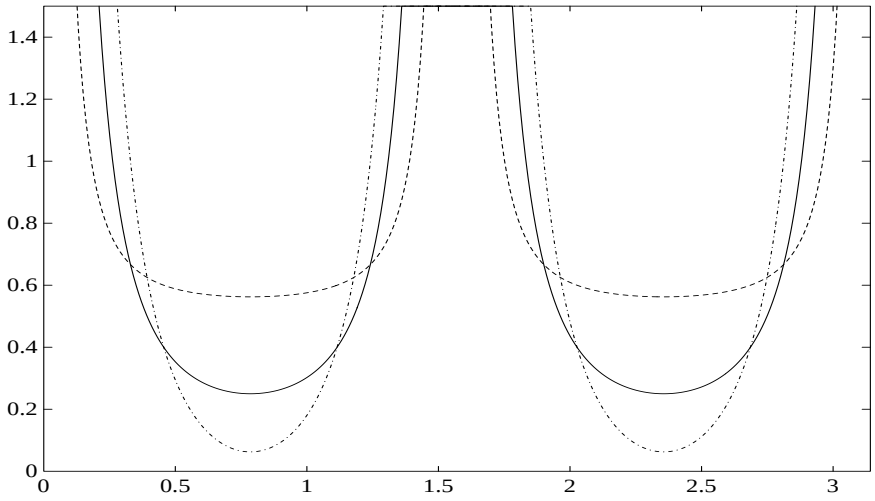
We have concentrated the discussion on additive or log-additive decompositions, as set since the beginning of section 4 by equation 4.1. However, in X11 (and its variants) the decomposition may also be multiplicative. While the filters used in additive and log-additive decompositions are linear and identical, since they apply either to the raw series or to the log-transform on it, the filters involved in multiplicative decompositions differ in some way. Young (1968) however showed that the linear filters can be seen as approximation of the multiplicative approach. Only some nonlinearities are missed by the linear approximation, which according to Young are in general not important.

The linear filters in X11 can be seen as convolutions of moving averages. Details of the procedure can be found in Wallis (1974, 1982). Bell and Monsell (1992) make very easy the reconstruction of the additive X11 filters: they give some explicit expressions for the filters involved in the convolutions. According to the filter chosen at each step, a different outcome is obtained. We illustrate the main lines of the procedure with the following example: consider a quarterly series whose behavior is described by an $I_4(1)$ model:

$$\Delta_4 x_t = (1 + \theta B) a_t. \tag{5.3}$$

The corresponding spectrum is displayed on figure 5.1 for different values of θ: namely, $\theta = -.5, 0, .5$. Clearly, more stable seasonal patterns are obtained when θ goes to -1. This spectrum is characterized by infinite peaks at the frequencies 0, $\pi/2$ and π, corresponding to movements occuring with an infinite period, with a period of 4 times a year, and twice a year which is the harmonic. The seasonal adjustment of this series consists in removing the movements associated with the intra-year periods 2 and 4.

Figure 5.1: Spectrum of an $I_4(1)$ process



$\theta = 0$ (—), $\theta = -.5$ (- -), $\theta = .5$ (- .)

The first seasonal moving average filter that would be applied in X11's default is a so-called 3x3 moving average filter:

$$\lambda_1(B) = (1/9)(F^4 + 1 + B^4)(F^4 + 1 + B^4),$$

where the terming 3x3 reflects the construction of the filter as a product of two filters of length 3. Developing $\lambda_1(B)$, we get:

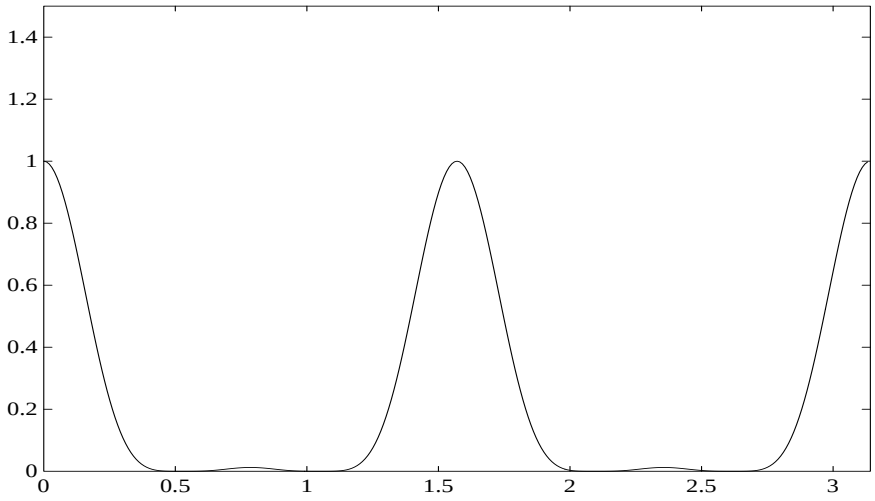
$$\lambda_1(B) = (1/9)(F^8 + 2F^4 + 3 + 2B^4 + B^8),$$

and using $B = e^{-iw}$, the transfer function $\lambda_1(w)$ is obtained as:

$$\lambda_1(w) = (1/9)(3 + 4\cos 4w + 2\cos 8w).$$

This function is of course real: the filter being symmetric, no phase shift appears. It is also symmetric around $\pi/2$. The squared gain is given by $\mid \lambda_1(w) \mid^2$, and is plotted

Figure 5.2: Squared Gain of $\lambda_1(B)$



on figure 5.2. It is seen that the filter has a gain of one at the 0-frequency, so that long-term movements are unaffected by the filter. The squared gain then decreases until it reaches a zero at the frequency $\pi/6$ and then at $w = 2\pi/6$: no movements at these frequencies are transmitted to the output series. Still moving toward the π-frequency, the gain reaches one for the seasonal frequency $\pi/2$. The same patterns are then repeated in $[\pi/2, \pi]$, so another peak at $w = \pi$ is observed.

The seasonality in x_t is related to the frequencies $w = \pi/2$ and $w = \pi$. Given that the filter $\lambda_1(w)$ preserves the frequencies $w = 0, \pi/2, \pi$, the long term behavior of the x_t must still be separated. This is done through the application of a 2x4 trend moving average filter of the type:

$$\begin{aligned} \mu(B) &= (1/8)(F^2 + F)(1 + B + B^2 + B^3) = \\ &= (1/8)(F^2 + 2F + 2 + 2B + B^2), \end{aligned}$$

with transfer function

$$\mu(w) = (1/8)(2 + 4\cos w + 2\cos 2w).$$

Figure 5.3: Squared Gain of $\mu(B)$)

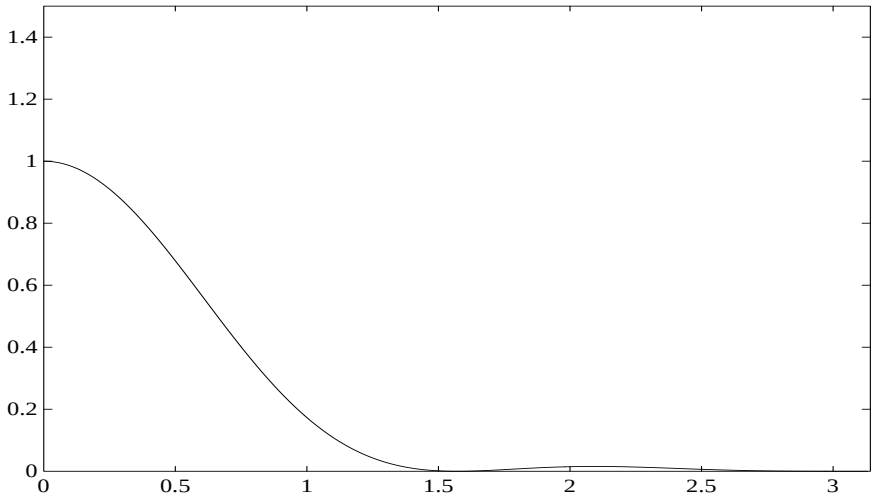
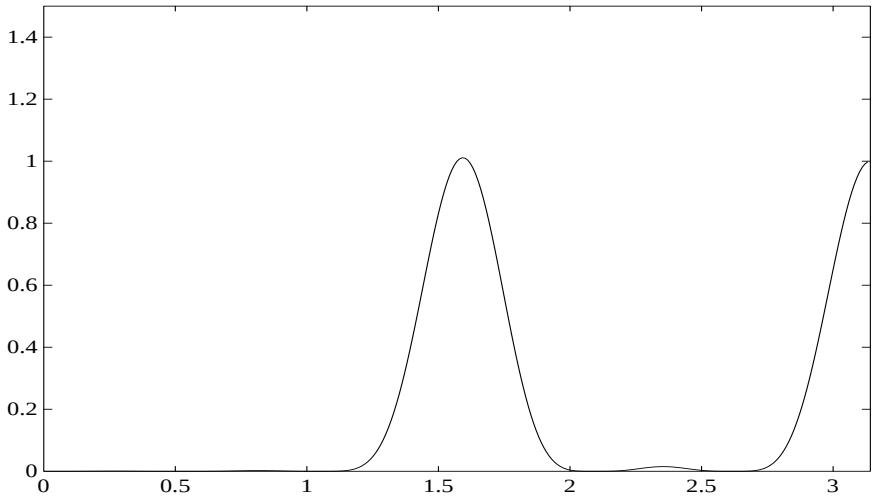


Figure 5.3 shows the corresponding squared gain: this filter operates a cut in the high-frequencies, preserving the lower ones. So the series $x_t - \mu(B)x_t = (1 - \mu(B)x_t$ will not display any trend behavior, and thus may be considered as input to $\lambda_1(B)$. This amounts to consider the filter $(1 - \mu(B))\lambda_1(w)$ whose corresponding gain is readily obtained since $\mid (1-\mu(w))\lambda_1(w) \mid = \mid (1-\mu(w)) \mid\mid \lambda_1(w) \mid$. Figure 5.4 plots it, and it is seen that the gain of the resulting filter has unit values concentrated on the seasonal frequencies $\pi/2$ and π, so that the output of the filtering operation will display mainly seasonal movements.

The process of convoluting filters is repeated several times in X11. It is illustrated here with a simple convolution of two filters, but in X11, four filters are involved in this process. These are the polynomials $\mu(B)$ and $\lambda_1(B)$, as described above for quarterly series, another seasonal moving average filter (say $\lambda_2(B)$) and a more refined trend filter known as the Henderson filter (say $H(B)$) that we shall discuss in the next section. The default for $\lambda_2(B)$ is a 3x5 moving average specified as:

$$\lambda_2(B) = (1/15)(F^4 + 1 + B^4)(F^8 + F^4 + 1 + B^4 + B^8)$$

Figure 5.4: Squared Gain of $(1-\mu(B))\lambda_1(B))$

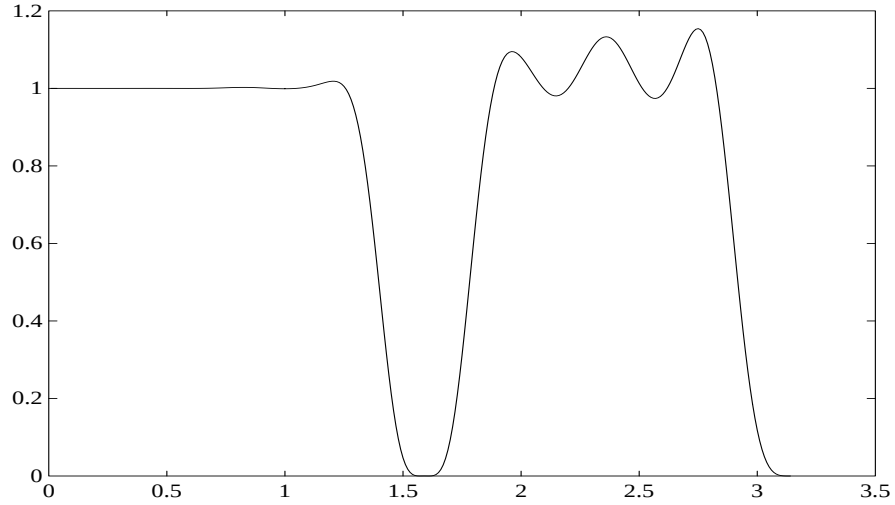


still for the quarterly case. The squared gain of the final seasonal filter and of the corresponding adjustment filter can be seen on figure 5.5. The band pass structure appears clearly: there is a dip in the adjustment filter around $\pi/2$ and π. The width of the dip is related to the stability of the seasonality which is supposed to be removed. If the input series is like (5.3) with $\theta = -.5$, then the default filter may be adequate. But for a series like (5.3) with $\theta = .5$, then figure 5.1 shows that the seasonality emobodied by this series would be too much unstable for that filter. The default adjustment filter of X11 will not succeed in removing all the seasonal fluctuations that the series will display. A remaining seasonality will be found in the output of the adjustment filter.

In general, different time series will display movements with very different characteristics. It is thus convenient to have available more than a single filter to catch possibly different features. X11 offers the choice between the default seasonal, the 3x3, the 3x5, and the 3x9 seasonal filters. The default involves a 3x3 for $\lambda_1(B)$ and a 3x5 for $\lambda_2(B)$, as illustrated above in the quarterly case. For the other options saying the seasonal filter is 3xl amounts to consider that both $\lambda_1(B)$ and $\lambda_2(B)$ are equal to

$$\lambda(B) = (3l)^{-1}(F^m + 1 + B^m)(F^{mk} + \cdots + F^m + 1 + B^m + \cdots + B^{mk})$$

Figure 5.5: Squared Gain of X11 Default Adjustment Filters, Quarterly



where m is the data periodicity and k is such that $l = 2 * k + 1$. Graphics of the X11 filters weights and squared gains can be found in numerous articles; the most complete may be Bell and Monsell (1992). For convenience, we reproduce the squared gains of the seasonal adjustment filters, when the filters are obtained with a 5-term and 13-term Henderson filter for the quarterly and monthly cases, respectively. Namely, the default, 3x3, 3x5 and 3x9 filters are presented for quarterly and monthly series.

These graphics give another illustration of how filtering works in the frequency domain. The seasonal component is designed to capture the movements in the series which occurs with a seasonal frequency. Thus the seasonal adjustment filters should annihilate the variability associated with the seasonal frequencies, and let the other unchanged. In agreement with that, the gain of the X11 adjustment filters presented on figures 5.6-5.7 displays this bandpass structure: they show a gain of 0 around the seasonal frequencies and a gain close to one in the other regions.

The width of the region where the gain is null is related to the stability of the seasonal movements which are supposed to be removed: the 3x9 filter corresponds to a stable seasonality while the 3x3 seasonal would be adequate for a relatively unstable

Figure 5.6: Squared Gain of X11 Adjustment Filters, Quarterly

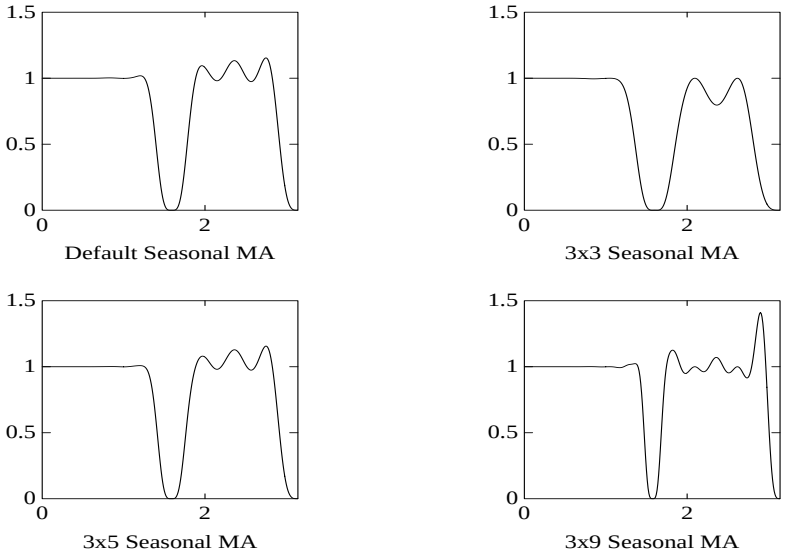
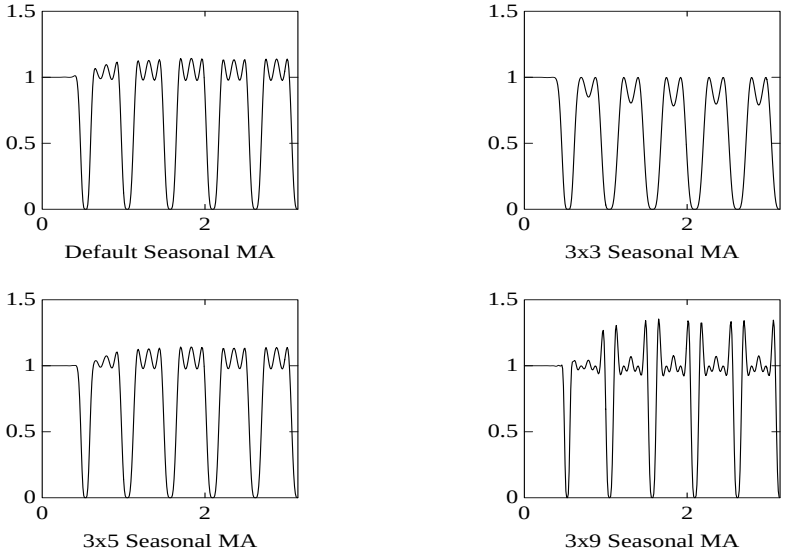


Figure 5.7: Squared Gain of X11 Adjustment Filters, Monthly



seasonality. For series displaying seasonal movements whose characteristics will be outside these two extreme patterns considered with the X11 filters, then either too much or not enough seasonality could be removed by simple application of these filters (see for example Fiorentini and Planas 1996a). In general, the choice of a particular filter must be conducted on the ground of information about the stochasticity of the movements that the filter should catch. At the limit, whether the series presents a seasonal behavior or not has to be checked in a prior stage.

It can be seen also on figure 5.6-5.7 that the gain of the central X11 adjustment filters is greater than 1 at frequencies between the fundamental seasonal frequency and the Nyquist frequency. The 3x3 adjustment filter is the only one which does not present this feature. These frequencies are related to short-term movements in the series. A gain greater than one implies that the seasonally adjusted series will present a larger short-term variability than the series itself (see Fiorentini and Planas, 1996b).

5.4 Henderson Trend Filter

X11 also embodies a filter designed to smooth time series so as to yield estimate of the trend. This filter, denoted $H(B)$, is known as the Henderson trend filter, and it is involved in the convolutions yielding the adjustment and the seasonal filters. Saying its length is n and denoting $l = (n-1)/2$, then $H(B)$ can be written as

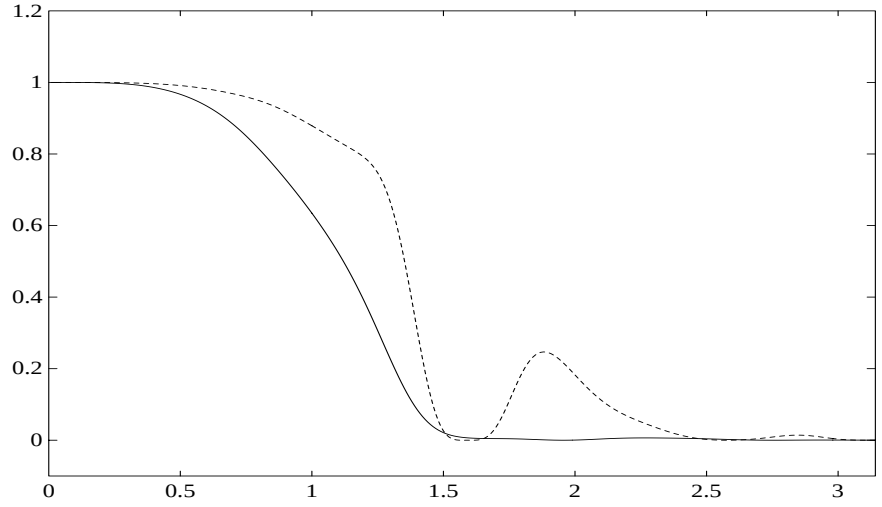
$$H(B) = \sum_{i=-l}^{l} h_i B^i.$$

The weights h_i can be obtained by setting $m = (n+3)/2$ from the formula:

$$h_i = 315 \frac{[(m-1)^2 - i^2][m^2 - i^2][(m+1)^2 - j^2][(3m^2 - 16) - 11j^2]}{8m(m^2-1)(4m^2-9)(4m^2-25)}.$$

This expression is given is Macaulay (1931), also reproduced in Dagum (1985) and Bell and Monsell (1992). Standard lengths of the filter are 9, 13, 17 or 23 terms for monthly time series, of 5 and 7 terms for quarterly series. In practice, the Henderson filter is not directly applied to the series under analysis but to the seasonally adjusted transformations. This prior procedure is needed because the Henderson filter does not

Figure 5.8: Squared Gain of X11 Trend Filters, Quarterly



7-term (—), 5-term (- -)

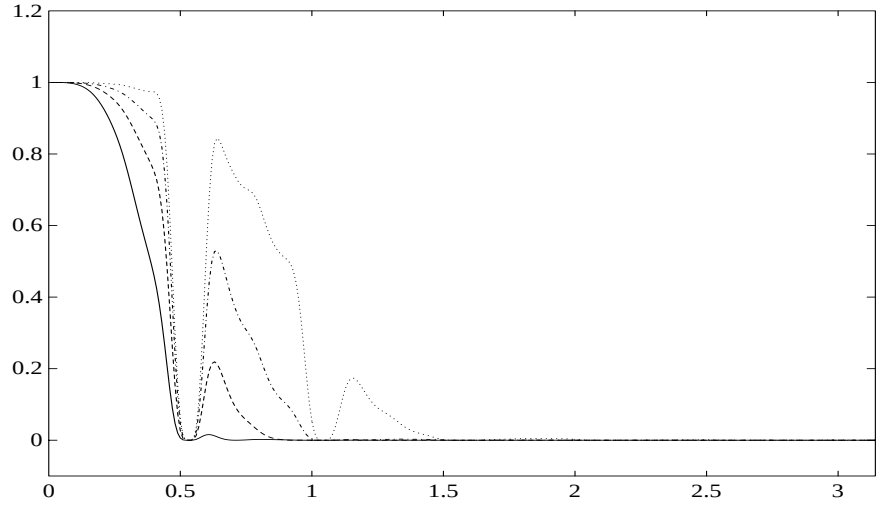
necessary display a zero-gain at the seasonal frequencies. In order to avoid the trend estimates to present seasonal fluctuations, the filter $H(B)$ is applied together with a seasonal adjustment filter of the form described in the previous section.

Figures 5.8 and 5.9 represent the squared gains associated to every standard filter, the seasonal adjustment filter being set as the default. It is seen that trend filters operate as low pass filters, the width of frequency range that are transmitted to the output depending on the filter length. Clearly, the longer the filter, the more narrow the frequency band for which movements in the input series are passed to the output, and the more stable will be the trend estimate.

5.5 Limits of Ad Hoc Filtering

The filters presented in this chapter are *ad hoc* in the sense that they do not depend on the stochastic properties of the series under analysis: they are available in the software X12-ARIMA, and it is up to the analysts to select the most adequate one given a

Figure 5.9: Squared Gain of X11 Trend Filters, Monthly



23-term (—), 17-term (- -), 13-term (- .), 9-term (..)

particular series. We have mentioned the problems may arise with series displaying extreme characteristics. Blinded use may also be dangerous, as illustrated in Maravall (1996) with a very explicit example: simple application of an empirical seasonal filter to a white noise variable yields a seasonal component ! It is true however that the X12-ARIMA software embodies many controls which inform the user that something may have gone wrong. But because of the empirical nature of ad hoc filtering procedures, these controls are mainly qualitative and no rigorous checks may be performed. Further, an exact measure of precision is not available. The building of confidence interval around the estimates is thus not possible. Also, Fiorentini and Planas (1996a) have pointed out the impossibility of such pre-defined filters to handle correctly time series characterised by very unstable patterns, like a fastly evolving seasonal component. On the other hand, they also underlined the overestimation of the short-term movements that the X11 filters may imply (see Fiorentini and Planas 1996b). We now turn to present a more refined decomposition procedure.

Chapter 6

Minimum Mean Squared Error Estimation of the Components

6.1 Introduction

This chapter presents an estimation procedure overcoming the limits of ad hoc filters. The procedure exploits the information given by the modelling of the series to build the filters estimating the components. The properties of the components estimators so obtained are discussed, and the estimation errors are analysed in relationship with an interpretation in the frequency domain of the procedure. Finally, some features of estimators of components identified through the canonical requirement are pointed out.

Writing $X_T = [x_1, \cdots, x_T]$, the optimal estimator of the signal will be given by the expectation of s_t conditional on X_T :

$$\hat{s}_{t/T} = E(s_t/X_T). \tag{6.1}$$

If $t = T$, the conditional expectation (6.1) yields the concurrent estimator of the signal. If $t < T$, (6.1) provides the preliminary estimator of a past realization of the signal, while for $t > T$, (6.1) corresponds to the $t - T$ period-ahead forecast of the signal. In the model-based approach framework, two methods of calculating this expectation can be used, each one having specific advantages. The Kalman Filter method (see,

for example, Anderson and Moore (1979)) proceeds firstly by writing the model in a state space format, then by setting some initial conditions, and finally by deriving the optimal estimator through recursive computations. Because of its computational tractability, it has been used in many applied works (for a general presentation, see for example Harvey (1989)), and in particular in the STS models. The Wiener-Kolmogorov (WK) filter has the benefit of being well suited for analytical discussion, and provides as an output a clear and precise information about the structure of the estimators. Its computational tractability has been improved in a decisive way by the T. Wilson algorithm presented in Burman (1980). It is most often involved in the ARIMA-model-based decomposition of time series, and the emphasis put on this approach throughout the discussion motivates the choice of focusing on the WK filter. The properties presented will also apply to estimators obtained with the Kalman Filter.

6.2 The Wiener-Kolmogorov filter

For completness, this section presents the construction of the WK filter. For stationary time series, this filter is derived as follow (from Whittle (1963)). First, the assumption that an infinite set of observations on the process x_t is available is needed. The estimator obtained for an infinite sample will be denoted $\hat{s}_t$, so that $\hat{s}_t = \hat{s}_{t/\infty}$. This assumption will be relaxed later.

The WK filter is a linear filter of the past and future realizations of the observed series, so the estimator $\hat{s}_t$ of the signal s_t may be expressed as:

$$\begin{aligned} \hat{s}_t &= \sum_{k=-\infty}^{\infty} \nu_k x_{t-k}, \\ &= \nu(B) x_t, \end{aligned}$$

where $\nu(B) = \sum_{k=-\infty}^{\infty} \nu_k B^k$. As the optimal estimator is defined as a conditional expectation, the WK filter is optimal in the sense that it minimizes the Mean Squared Errors on the estimator. So, by orthogonal projection of s_t on the x_{t-j}, $j = -\infty, \cdots, 0, \cdots, +\infty$, the estimator $\hat{s}_t$ must verify:

$$cov[s_t - \hat{s}_t, x_{t-j}] = 0.$$

Thus, for each j, $j = -\infty, \ \cdots, 0, \cdots, +\infty$, we have:

$$cov[s_t, x_{t-j}] - \sum_{k=-\infty}^{\infty} \nu_k cov[x_{t-k}, x_{t-j}] = 0.$$

Denoting w the frequency in radians and $g_{xs}(w)$ the cross-spectrum density function, this last expression can be translated in the frequency domain as:

$$\begin{aligned} 0 &= \int_{-\pi}^{\pi} e^{iwj} g_{xs}(w)dw - \int_{-\pi}^{\pi} e^{iwj} [\sum_{k=-\infty}^{\infty} \nu_k e^{-iwk}] g_x(w)dw, \\ &= \int_{-\pi}^{\pi} e^{iwj} [g_{xs}(w) - \nu(e^{-iw}) g_x(w)]dw, \end{aligned}$$

where the Fourier transform $B = e^{-iw}$ is used to write: $\nu(e^{-iw}) = \nu(B)$. This integral is finite since the observed series is supposed to have a finite variance (stationary case). Then, for all $j = -\infty, \cdots, 0, \cdots, +\infty$, we have:

$$g_{xs}(w) - \nu(e^{-iw}) g_x(w) = 0,$$

which leads to: $\nu(e^{-iw}) = g_{xs}(w)/g_x(w)$. When the components are assumed independent as in model (A), the filter $\nu(e^{-iw})$ may be written simply as:

$$\nu(e^{-iw}) = g_s(w)/g_x(w). \tag{6.2}$$

The WK filter was initially designed to deal with stationary time series. Under certain assumptions, Cleveland and Tiao (1976), Pierce (1979), Bell (1984), and Maravall (1988a) have shown that the filter yields a finite Mean Squared Error even if the processes are nonstationary, so the WK filter is still valid for nonstationary time series. Given that most of the series encountered in practice are nonstationary, this extension was of a great importance for the applicability of the WK filter. A similar extension has been developed for the Kalman Filter (see, for example, Kohn and Ansley (1987)).

6.3 Optimal estimators

6.3.1 MMSE Estimators

The WK filter was expressed in (6.2) as the ratio of the spectrum of the signal to the spectrum of the observed series. An appealing feature of ARIMA models is that they provide a convenient way to parametrize the spectrum of time series. Applying the expression (6.2) to model (A), and under the hypothesis of independence of the components, the estimators can be obtained as (see, for example, Hillmer and Tiao (1982)):

$$\begin{aligned} \hat{s}_t &= \nu_s(B)x_t, \\ \nu_s(B) &= \frac{A_s}{A_x} = V_s \frac{\theta_s(B)\theta_s(F)\phi_n(B)\phi_n(F)}{\theta_x(B)\theta_x(F)}, \end{aligned}$$

and for the nonsignal estimator :

$$\begin{aligned} \hat{n}_t &= \nu_n(B)x_t, \\ \nu_n(B) &= \frac{A_n}{A_x} = V_n \frac{\theta_n(B)\theta_n(F)\phi_s(B)\phi_s(F)}{\theta_x(B)\theta_x(F)}. \end{aligned} \tag{6.3}$$

The hypothesis of invertibility of the polynomial $\theta_x(B)$ insures the convergence of the filter. It is a symmetric filter, which depends on the polynomials of the models for the observed series and those for the components. This dependence allows the WK filter to adapt itself to the series under analysis, in contrast to the empirical filters which assume that a particular filter holds for set of time series. Since the WK filter is symmetric, it can be seen as the ACGF of a particular model, which is given in the following lemma:

Lemma 6.1 *The WK filter $\nu_s(B)$ corresponds to the* ACGF *of the model*

$$\theta_x(B)z_t = \phi_n(B)\theta_s(B)b_t$$

where b_t is a white noise with variance V_s. ∎

Hence, the derivation of the weights of the filter amounts to the computation of a simple autocovariance function.

Lemma 6.1 makes clear that the MMSE estimators are available only when all the polynomials of model (A) have been determined. Estimating the components requires the practitioners to first select an admissible decomposition. Different assumptions made on the models for the component will imply different properties of the estimators through the squared polynomial $\theta_i(B)\theta_i(F)$ and the innovation variance V_i, where $i = n, s$. However, there is not a strict correspondence between the stochastic properties of the components and those of the estimators. Indeed, some discrepancies do exist. This point has been often discussed in the statistical literature (see for example Bell and Hillmer (1984)). It can be easily understood from a study of the distribution followed by the estimators.

6.3.2 The distribution of the estimators

The distribution of the estimators will be studied both in the time domain and in the frequency domain. Starting in the time domain, if we replace x_t by $\phi^{-1}(B)\theta(B)a_t$ in (6.3), then the estimator $\hat{s}_t$ can be written:

$$\hat{s}_t = V_s \frac{\theta_s(B)\theta_s(F)\phi_n(F)}{\phi_s(B)\theta_x(F)} a_t. \tag{6.4}$$

Comparing the model for the estimator (6.4) and the model (4.2) for the theoretical signal, it is easily seen that they share the same AR and MA polynomials in B (see, for example, Maravall (1993b)). So, if a component is nonstationary, the component and its estimator will share the same stationarity inducing transformation.

However, the models for the theoretical components and the models for the estimators are structurally different. The difference is due to the presence of a polynomial in F: for example, the model for the signal estimator $\hat{s}_t$ contains $\theta_x(F)$ as AR polynomial and $\theta_s(F)\phi_n(F)$ as MA polynomial. Notice that the model for the MMSE estimator will be noninvertible when the model for the theoretical component is noninvertible or when the other component follows a nonstationary process. This latter case expresses the dependence of the estimator $\hat{s}_t$ on the model for the other component, the nonsignal n_t, through the AR polynomial $\phi_n(F)$. Yet another way to express (6.3) is:

$$\hat{s}_t = \frac{\theta_s(B)}{\phi_s(B)}\hat{a}_{st},$$

where $\hat{a}_{st}$ is such that:

$$\hat{a}_{st} = V_s \frac{\theta_s(F)\phi_n(F)}{\theta_x(F)} a_t. \tag{6.5}$$

It can be checked that $\hat{a}_{st}$ is the MMSE estimator of the pseudo-innovation in the signal a_{st} (see for example Harvey and Koopman, 1992). The differences between the structure of theoretical component and that of the estimator may thus be seen as due to the pseudo-innovations estimator: while the pseudo-innovations are assumed to be white noises, their MMSE estimators are correlated, following some ARMA models. For example, the irregular component is supposed to be white noise, but its estimator present an ARMA correlation structure. The pseudo-innovations estimators are however not forecastable, model (6.5) being time reversed so the current estimators depend on the future innovations on the series. The ACGF of the estimator (6.5), denoted $A(\hat{a}_{st})$, is given by:

$$A(\hat{a}_{st}) = V_s^2 \frac{\theta_s(B)\theta_s(F)\phi_n(B)\phi_n(F)}{\theta_x(B)\theta_x(F)},$$

and thus it verifies $A(\hat{a}_{st}) = V_s\nu_s(B)$: up to a scale factor, the covariances of the pseudo-innovations estimator are equivalent to the weigths of the WK filter designed for estimating the associated component.

A look at the way the filter works on the series, in the frequency domain, helps in understanding the differences between the final estimator and the component. Denoting $g_{\hat{s}}(w)$ the spectrum of the estimator $\hat{s}_t$, then from (6.3)

$$\begin{aligned} g_{\hat{s}}(w) &= \nu_s(e^{-iw})\nu_s(e^{iw})g_x(w) = \\ &= \left[\frac{g_s(w)}{g_x(w)}\right]^2 g_x(w). \end{aligned}$$

The gain of the filter is given by $\mid \nu_s(e^{-iw}) \mid = g_s(w)/g_x(w)$: since spectra are always non negative, gain and frequency response function of the WK filter (6.2) are equivalent. The WK filter is built in such a way that the frequency response function relates the spectrum of the estimator to the spectrum of the signal. This property is specific to the WK filter, and it can easily be understood from:

$$\begin{aligned} g_{\hat{s}}(w) &= \left[\frac{g_s(w)}{g_x(w)}\right]^2 g_x(w) = \\ &= \left[\frac{g_s(w)}{g_x(w)}\right] g_s(w), \end{aligned}$$

and the frequency response function may also be written as:

$$\frac{g_s(w)}{g_x(w)} = \frac{1}{1 + \frac{g_n(w)}{g_s(w)}}. \tag{6.6}$$

This property allows us to draw the following interpretation of the WK filter mechanism. When the relative contribution of the signal is high at a particular frequency w^*, then $g_n(w^*)/g_s(w^*) \simeq 0$. The frequency response function is thus close to 1, and $g_{\hat{s}}(w^*) \simeq g_s(w^*)$ holds. The component and the estimator will display movements of similar variability around the frequency w^*. Also, the gain of the filter at this frequency is close to 1, so we get: $g_{\hat{s}}(w^*) \simeq g_x(w^*)$: most of the observed series spectrum is used for the signal estimation.

Conversely, when the relative contribution is low at a particular frequency, the WK filter just ignores it for the signal estimation. For example, suppose that either the signal's spectrum has a zero at the frequency w^* or the nonsignal's spectrum admits an infinite peak at the frequency w^*, so that we have: $g_n(w^*)/g_s(w^*) \to \infty$. Then the spectrum of the signal estimator will display a zero: $g_{\hat{s}}(w^*) \simeq 0$, and the estimator will follow a noninvertible model. This conclusion was already obtained from a direct observation of the model for the estimator. Furthermore, it is easily deduced from (6.6) that $g_{\hat{s}}(w) \leq g_s(w)$ for every frequency, so the signal is always underestimated (see for example Burman (1980)). A straightforward consequence is that the estimator will always be more stable than the component.

6.4 Covariance between estimators.

Another important discrepancy between the properties of the components and those of the estimators is that, if the components are assumed independent, the MMSE estimators will always be covariated. This is a consequence of the estimation procedure which orthogonally projects both components on a space of dimension one defined by the observed series. The existence of covariances between the estimators even if the theoretical components were assumed independent has been subject of attention in the literature (see for example Nerlove (1964), Granger (1978), Garcia-Ferrer and Del Hoyo (1992)). A similar result has been discussed by Harvey and Koopman (1992) concerning the estimators of the 'pseudo-residuals' a_{nt} and a_{st}. Maravall and Planas (1996) also discussed the covariances between the UC estimators; most of the results presented in this section is taken from their article.

The covariances between the estimators are easily obtained from the following lemma:

Lemma 6.2 *If $C(\hat{n}, \hat{s})$ denotes the cross-covariance generating function between $\hat{n}_t$ and $\hat{s}_t$, then $C(\hat{n}, \hat{s})$ is equal to the ACGF of the model:*

$$\theta_x(B) z_t = \theta_s(B)\theta_n(B) b_t, \tag{6.7}$$

where b_t is a white noise with variance $V_n V_s$.

Proof: From (6.3), we have:

$$C(\hat{n}, \hat{s}) = A_n A_s / A_x. \tag{6.8}$$

Developing and simplifying, we get:

$$C(\hat{n}, \hat{s}) = V_n V_s \frac{\theta_n(B)\theta_n(F)\theta_s(B)\theta_s(F)}{\theta_x(B)\theta_x(F)}, \tag{6.9}$$

which is the ACGF of the model (6.7). ∎

So $C(\hat{n}, \hat{s})$ can be seen as the ACGF of an ARMA process with AR polynomial $\theta_x(B)$, MA polynomial $\theta_n(B)\theta_s(B)$, and with innovation variance $V_n V_s$. Notice the dependence of the covariances between the estimators to the admissible decompositions through the MA polynomial and the innovation variance of model (6.7). Using lemma 6.2, the following properties of the MMSE estimators can be derived.

Property 6.1 *The lag-0 covariance between the estimators is always positive.*

This is immediate since, from lemma 6.2, the lag-0 covariance between the estimators is equal to the variance of the process z_t: $cov[\hat{n}_t, \hat{s}_t] = var[z_t] > 0$. Since the estimators must sum to the observed series, the existence of a positive cross-covariance between the estimators reflects the loss in the component spectrum which is implied by the estimation procedure. As emphasized in the previous section, the estimator has a spectrum always below the one of the corresponding component, and this difference is then found in the estimators cross-spectrum which is always positive.

Property 6.2 *The covariances between the components estimators are symmetric, finite and converge to zero.*

Proof: The process z_t, with ACGF $C(\hat{n}, \hat{s})$, has $\theta_x(B)$ as AR polynomial. The model for the observed series being assumed invertible, z_t is stationary. So the covariances between the signal and the nonsignal estimators are finite and converge to zero, even if the estimators are nonstationary. ∎

For nonstationary time series, the UC estimators may diverge in time, but according to property 6.2, they will diverge together. This result holds independently of the components' order of integration. An interesting consequence is:

Property 6.3 *When the observed series x_t is nonstationary, the estimators $\hat{s}_t$ and $\hat{n}_t$ are uncorrelated whatever the selected admissible decomposition is.*

Proof: The correlations between the estimators are given by the ratio of the covariances to the product of the estimators standard deviations. As shown in property 6.2, the covariances are finite, while the sample standard deviation of at least estimator will not converge. So the ratio of the covariances to the standard deviations will tend to zero. ∎

So, when the observed series is nonstationary, the estimation procedure preserves the property of the theoretical components in terms of zero cross-correlations. Given that most of the economic time series encountered in practice are nonstationary, this result limits the relevance of the discussions about the non zero estimators cross-correlations.

We now focus on the errors in the estimators.

6.5 Estimation errors

The estimation errors can be decomposed into two types of errors: the *final estimation error* and the *revision error*. The first one corresponds to $s_t - \hat{s}_t$ or $n_t - \hat{n}_t$ and is obtained under the hypothesis of a complete realization of the observed series. Given that the WK filter is convergent, in practice, for large enough sample, the final estimation error concerns the estimators computed around the center of the series. The revision error is related to the impossibility to actually deal with infinite samples, and concerns in practice the estimators computed near the ends of the sample. The independence of both types of errors, as demonstrated in Pierce (1980), allows us to analyse them separately.

6.5.1 Final Estimation Error

The following lemma is due to Pierce (1980):

Lemma 6.3 *The* ACGF *of the final estimation error* $e_t = s_t - \hat{s}_t = \hat{n}_t - n_t$ *is equivalent to that of the* ARMA *process*

$$\theta_x(B)e_t = \theta_s(B)\theta_n(B)b_t,$$

where b_t *is a normally distributed independent white noise with variance* $V_n V_s$.

Proof: The error in the signal estimator can be written as:

$$\begin{aligned} e_t &= s_t - \hat{s}_t = \\ &= s_t - [A_s/A_x]x_t = \end{aligned}$$

$$= [1 - A_s/A_x]s_t - [A_s/A_x]n_t =$$

$$= [A_n/A_x]s_t - [A_s/A_x]n_t.$$

Developing the ACGF's and writing, in order to simplify the expressions, η for any polynomial $\eta(B)$ and $\overline{\eta}$ for $\eta(F)$, we have

$$e_t = \frac{[V_n\theta_n\overline{\theta}_n\overline{\phi}_s\phi_s s_t - V_s\theta_s\overline{\theta}_s\overline{\phi}_n\phi_n n_t]}{\theta_x\overline{\theta_x}}.$$

Using the models (4.2) for the components:

$$e_t = \frac{[V_n\theta_n\overline{\theta}_n\overline{\phi}_s\theta_s a_{st} - V_s\theta_s\overline{\theta}_s\overline{\phi}_n\theta_n a_{nt}]}{\theta_x\overline{\theta}_x}.$$

which can also be written as

$$e_t = \theta_n\theta_s[V_n\overline{\theta}_n\overline{\phi}_s a_{st} - V_s\overline{\theta}_s\overline{\phi}_n a_{nt}]/\theta_x\overline{\theta}_x.$$

The ACGF of the estimation error is then given by

$$A(e_t) = \theta_n\theta_s\overline{\theta}_n\overline{\theta}_s\frac{V_n^2V_s\theta_n\overline{\theta}_n\phi_s\overline{\phi}_s + V_s^2V_n\theta_s\overline{\theta}_s\phi_n\overline{\phi}_n}{\theta_x\overline{\theta}_x\theta_x\overline{\theta}_x},$$

The relationship (4.3) between the MA polynomials provides:

$$\theta_x\overline{\theta}_x = V_n\theta_n\overline{\theta}_n\phi_s\overline{\phi}_s + V_s\theta_s\overline{\theta}_s\phi_n\overline{\phi}_n. \tag{6.10}$$

Inserting (6.10) into the expression for $A(e_t)$, we get:

$$A(e_t) = V_nV_s\frac{\theta_n\overline{\theta}_n\theta_s\overline{\theta}_s}{\theta_x\overline{\theta}_x}. \tag{6.11}$$

which is the expected result. Working on n_t instead of s_t would lead to $z_t = \hat{n}_t - n_t$ which has the same ACGF than $s_t - \hat{s}_t$. ∎

The MA polynomial of the process for the final estimation error corresponds to the product of the MA polynomials of the models for the two components, while the AR polynomial is given by the MA polynomial of the model for the observed series. The observed series' process being invertible, the final estimation error follows a stationary model whose variance gives the MSE of the estimation. An immediate consequence of lemmas 6.2 and 6.3 is that (see Maravall and Planas, 1996):

Lemma 6.4 *For all admissible decompositions, the theoretical estimators cross-covariance generating function is equivalent to the* ACGF *of the final estimation error :*

$$C(\hat{n}, \hat{s}) = ACGF(\hat{n}_t - n_t) = ACGF(s_t - \hat{s}_t).$$

It is interesting to notice that both will depend on the identifying assumptions adopted. This dependence will be explored in the next section. For decompositions into more two components, then lemmas 6.2 and 6.3 still holds with s_t being the component of interest and n_t aggregating the other components.

Returning to the analysis of the final estimator in the frequency domain, it was noted in section 6.3.2 that the largest discrepancies between the spectrum of the signal and that of the estimator are observed at the frequencies where the ratio $g_n(w)/g_s(w)$ is large. Hence, the error is mainly related to the frequencies where the stochastic variability of the signal is relatively low. In general, unstable signals are more accurately estimated than stable signals.

6.5.2 Revision error

The hypothesis of having an infinite realization of the series x_t was needed because the WK filter in (6.3) goes to $-\infty$ to ∞. Since the filter is convergent, it can be safely truncated at some point. However, at the beginning or at the end of a sample, the computation of the estimator requires unknown past or future realizations of x_t. We shall focus on the distortion induced by the lack of future observations. Near the end of a finite sample, optimal preliminary estimates can be computed by replacing unknown future realizations by their forecasts (Cleveland and Tiao (1976)). The forecast errors imply that the preliminary estimates will be contaminated by an additional error,

termed 'revision error'. As new observations become available, forecasts are updated and eventually replaced by the observed values, and the preliminary estimator is revised. The total revision error in the concurrent estimate of s_t, that is the estimate of s_t computed at time t, is given by: $\hat{s}_t - E(\hat{s}_t/X_t)$. To simplify the presentation, we shall denote by $E_t\hat{s}_t$ the expectation of the estimate of s_t conditional on the information available at time t, so that: $E_t\hat{s}_t = E(\hat{s}_t/X_t)$. Writing $\psi(B) = \theta(B)/\phi(B)$, then the optimal estimator may be expressed in function of the innovations on the observed series as:

$$\begin{aligned} \hat{s}_t &= \nu_s(B)x_t = \\ &= \nu_s(B)\psi(B)a_t = \\ &= \xi_s(B)a_t, \end{aligned}$$

with $\xi_s(B) = \cdots + \xi_{s-1}B + \xi_{s0} + \xi_{s1}F + \cdots$. Taking the expectation of $\hat{s}_t$ conditional on the information available at time t, the concurrent estimator of s_t is then readily obtained as:

$$E_t\hat{s}_t = \sum_{i=-\infty}^{0} \xi_{si}a_{t+i}.$$

Hence, the total revision error in the concurrent estimator is given by:

$$\hat{s}_t - E_t\hat{s}_t = \sum_{i=1}^{\infty} \xi_{si}a_{t+i},$$

and, for the revisions in any preliminary estimate of s_t computed at time $t+k$, $k \neq 0$:

$$\hat{s}_t - E_{t+k}\hat{s}_t = \sum_{i=k+1}^{\infty} \xi_{si}a_{t+i}.$$

The revision errors are thus an infinite moving average process. As shown in Pierce (1980), the MA processes followed by the revision errors are stationary. Thus the variance of the revision error can be computed as:

$$V[\hat{s}_t - E_{t+k}\hat{s}_t] = \sum_{i=k+1}^{\infty} \xi_{si}^2. \tag{6.12}$$

The stationarity of the revisions can be easily understood by noticing that the polynomial $\xi_s(B)$ is convergent in F. A consequence is that $lim_{k\to\infty} V[\hat{s}_t - E_{t+k}\hat{s}_t] = 0$: in practice, the revisions become negligible after some number of periods. The polynomial $\xi_s(B)$ is obtained by polynomials convolution, which does not raise any particular problem. A convenient algorithm for computing the ξ-weights can be found in Maravall and Mathis (1994). Adding final estimation error and total revision error, we obtain the *total estimation error* in the concurrent estimates as $s_t - E_t\hat{s}_t$. As final estimation error and revision error are independent, the variance of the total estimation error is simply obtained as $V(s_t - E_t\hat{s}_t) = V(s_t - \hat{s}_t) + V(\hat{s}_t - E_t\hat{s}_t)$.

6.6 Properties of the canonical decomposition

All along the previous section, we have seen that the choice of a particular decomposition affects both the final estimation error and the revision errors through the MA polynomials $\theta_s(B)$, $\theta_n(B)$, and the innovation variances V_s and V_n. Maravall and Planas (1996) fully explored that dependency, and some of their results are reported here. Interested readers are refered to Maravall and Planas (1996) for the proofs of the results discussed in this section. Let us suppose that s_t and n_t are component specified in their canonical form, so that the decomposition actually considered is: $x_t = s_t + n_t + u_t$, where u_t is a white noise with variance maximized. The coefficients ν_{s0} and ν_{n0} thus denote the central coefficients of the WK filter designed to estimate the canonical signal and the canonical nonsignal. Then,

Lemma 6.5 *A canonical decomposition always minimizes the final estimation error. Which component must be specified canonical is determined by the following rule:*

- *set s_t canonical if $\nu_{s0} \leq \nu_{n0}$;*
- *set n_t canonical otherwise.*

Further, from lemma 6.3 the canonical decomposition minimizing the final estimation error of the components also minimizes the covariances between the estimators. Lemma 6.5 provides a simple procedure to design the components models in order to get this property: the canonical requirement must be applied to the component which has the smallest central coefficient in the WK filter designed to estimate its canonical form. On the other hand, all the noise of the model must be assigned to the component which is relatively more important. This relative importance is evaluated by comparing the central weight of the WK filters designed to estimate the components in their canonical forms. A very simple way to compute the central weight of the WK filter as the variance of a simple ARMA model has been described in section 6.1.

Given the model for the observed series, this canonical decomposition will yield the highest coherence between the spectrum of the signal and the spectrum of the estimator. Nerlove (1964) saw such a coherence as a desirable feature when he discussed in the seasonal adjustment context several spectral criteria that adjustment procedures should satisfy. Specifying the unobserved components models as suggested in lemma 6.2 will thus reduce the discrepancies between the model of the signal and the model of its estimator.

For preliminary estimator however, the canonical criterion has an unpleasant property: for a given model for the observed series it always maximizes the revision error among the set of admissible decompositions. This result has been source of confusion for many practitioners, and it seems thus worth to explain its correct meaning. First, it only concerns the set of admissible decompositions of a given time series model. It does not mean that another method of decomposition will perform better in terms of revision. Secondly, in the two-component decomposition, the other canonical specification may perfectly minimize the revisions and provide the most precise preliminary approximation of the final estimates. Thirdly, this feature is of secondary importance since if we consider the total estimation error in any preliminary estimate, lemma 6.5 can be generalised. Writing $\nu_{s0}^{t/t+k} \leq \nu_{n0}^{t/t+k}$ the coefficients of B^0 in the WK filter designed to estimate the canonical signal s_t and the canonical nonsignal n_t estimated at time $t+k$,

Lemma 6.6 *The variance of the Total Estimation Error on the concurrent signal estimator is always maximized at a canonical decomposition. The component to be specified canonical is*

- s_t *if* $\nu_{s0}^{t/t+k} \leq \nu_{n0}^{t/t+k}$;
- n_t *otherwise.*

Lemma 6.5 thus holds for any preliminary estimator.

The aggregation of the historical estimation error and of the revision error preserves the property of the canonical decompositions that they yield the most precise estimator. Whatever is the set of observations available on x_t and the period for which we are interested in the signal, it is always true that the minimum variance of the total estimation error is obtained by assigning all the noise of the model to the canonical component s_t or n_t whose estimator gives the most weight to the particular realization x_t, the other one being let canonical.

However, it is possible that while one canonical decomposition minimizes the error in the historical estimator, the other one minimizes the error in the concurrent estimator. In that case, the aggregation of the two types of errors makes the noise repartition minimizing the error switch to one bound of $[0, 1]$ to the opposite bound: for example when $\nu_{s0} > \nu_{n0}$ and $\nu_{s0}^{t/t} < \nu_{n0}^{t/t}$. Because the revision error variance is always maximized at one bound of $[0, 1]$, the switching of solutions means that the decomposition minimizing the final error variance is also the one which maximizes the revision error in the concurrent estimator. Notice that in the concurrent estimation case, the equality $\nu_s(B)(\theta(B)/\phi(B)) = \xi_s(B)$ implies $\nu_{s0} = \xi_{s0}$. The same holds of course for n_t, and in the concurrent case, the condition for having a canonical signal minimizing the estimation error simplifies to $\xi_{s0} < \xi_{n0}$.

The case of forecasts of the components is trivial. Since every admissible decomposition differ by some noise repartition, and that this noise is unforecastable, the variance of the forecast error on the signal is always minimized with the canonical specification. Adding an unpredictable white noise to a variable just increases the forecast error variance without changing the variable forecast. In general, noninvertible components are always best forecasted.

Chapter 7

Examples

7.1 Introduction

We now apply the previous results to the decomposition of time series. Both theoretical models and actual time series are treated. In this last case, the computations have been performed by the program SEATS developed by A. Maravall and V. Gomez (see Maravall and Gomez, 1996), so all the results commented may also be found in the SEATS output. This software also contains a facility to deal with theoretical models.

7.2 Regular ARIMA model

Suppose the observed series follows an IMA(2,1) process given by :

$$(1-B^2)x_t = (1+\theta B)a_t \quad \text{with} \quad \begin{array}{c|c} a_t & \sim NID(0,1). \\ \theta & |<1 \end{array}$$

This simple model, adequate for bi-annual data, has been the subject of a pedagogical discussion in Maravall and Pierce (1987), with the simplification $\theta = 0$.

Suppose we wish to decompose the observed series as: $x_t = n_t + s_t$, where n_t and s_t represent respectively the trend and the seasonal components of the observed series x_t. The AR polynomial $(1-B^2)$ has a root the zero-frequency, which is thus assigned to the trend component, and a root at the π-frequency, with period two, which characterises

the seasonal fluctuations of the series. Possible models for the components are then of the type:

$$\begin{aligned}(1+B)s_t &= (1-\theta_s B)a_{st},\\ (1-B)n_t &= (1-\theta_n B)a_{nt}.\end{aligned}$$

For these models, we have the overall relationship:

$$(1-\theta B)a_t = (1-B)(1-\theta_s B)a_{st} + (1+B)(1-\theta_n B)a_{nt},$$

which provides a system of 3 covariances equations with the four unknowns θ_s, V_s, θ_n, V_n. The system is thus not identified. We overcome this problem by imposing zero-coefficients restrictions on both components. The system of variance-covariances equations is then

$$\begin{aligned}1+\theta^2 &= 2V_s + 2V_n\\ -\theta &= -V_s + V_n.\end{aligned}$$

where it is easily obtained that $V_s = (1+\theta)^2/4$ and $V_n = (1-\theta)^2/4$. Alternatively, this solution may be derived in the frequency domain by partial fraction decomposition of the observed series spectrum. It is the approach favored in most computational implementations. Using $\cos 2w = 2\cos^2 w - 1$, the spectrum of x_t may be written as:

$$g_x(w) = \frac{1+\theta^2-2\theta cos w}{2-2cos2w} = \frac{1+\theta^2-2\theta cos w}{4-4cos^2 w}.$$

Taking then $z = \cos w$, this last expression can be factorised as:

$$\frac{1+\theta^2-2\theta z}{4-4z^2} = V_s\frac{1}{2+2z} + V_n\frac{1}{2-2z},$$

with $V_s = (1+\theta)^2/4$ and $V_n = (1-\theta)^2/4$, in agreement with the solution derived in the time domain. The amount of noise embodied in the spectra $g_s(w)$ and $g_n(w)$ is given by:

$$\epsilon_s = \min_w g_s(w) = g_s(0) = (1+\theta)^2/16,$$

$$\epsilon_n = \min_w g_n(w) = g_n(\pi) = (1-\theta)^2/16,$$

so the 'pure' noise part of the observed series is:

$$V_u = \epsilon_s + \epsilon_n = (1+\theta^2)/8.$$

Removing ϵ_s from the spectrum of s_t, the spectrum of a canonical seasonal component is obtained as:

$$\begin{aligned} g_s(w) &= V_s \frac{1}{2+2\cos w} - V_s/4 = \\ &= (V_s/4)\frac{4-(2+2\cos w)}{2+2\cos w} = \\ &= (V_s/4)\frac{2-2\cos w}{2+2\cos w}. \end{aligned} \tag{7.1}$$

It is easily seen that $g_s(w)$ characterizes the canonical seasonal specified as:

$$(1+B)s_t = (1-B)a_{st},$$

with $V_s = (1+\theta)^2/16$. The nonseasonal component concentrates all the noise of the model, and its spectrum is given by:

$$\begin{aligned} g_n(w) &= V_n \frac{1}{2-2\cos w} + V_s/4 = \\ &= (1/4)\frac{4V_n + 2V_s - 2V_s\cos w}{2-2\cos w} = \\ &= (1/16)\frac{4(1-\theta)^2 + 2(1+\theta)^2 - 2(1-\theta)^2\cos w}{2-2\cos w}. \end{aligned} \tag{7.2}$$

In order to get the model for the nonseasonal component, the spectrum numerator must be factorised. The standard procedure consists in writing $\cos w = (1/2)(z + z^{-1})$ and in solving in z which yields roots r and r^{-1}. Picking up the roots less than one in modulus gives then the factors composing the MA polynomial. The variance is simply obtained as a rescaling factor. In the case studied, the analytic expression of the corresponding model is a bit tedious. In the simple case where $\theta = 0$, it simplifies to:

$$\Delta n_t = (1 + (3 - 2\sqrt{2})B)a_{nt},$$

with $V_n = (3 + 2\sqrt{2})/16$. In practice, the models for the components are not needed to derive the optimal filter. For example, the WK filter designed to estimate the canonical seasonal component is given by:

$$\begin{aligned}
\nu_s(B) &= V_s \frac{(1-B)^2(1-F)^2}{(1-\theta B)(1-\theta F)} = \\
&= V_s \left[(6 - 4(B+F) + (B^2 + F^2)\right] \\
&\quad \cdot \left[1 + \theta B + \theta B^2 + \cdots\right] \left[1 + \theta F + \theta F^2 + \cdots\right] \\
&= \frac{V_s}{1-\theta^2} \left[(6 - 4(B+F) + (B^2 + F^2)\right] \\
&\cdot \quad \left[1 + \theta(B+F) + \theta^2(B^2 + F^2) + \cdots\right]
\end{aligned}$$

It is then easily seen that

$$\begin{aligned}
\nu_{s0} &= \frac{V_s}{1-\theta^2}(6 - 8\theta + 2\theta^2) \\
\nu_{s1} &= \frac{V_s}{1-\theta^2}[6\theta - 4(1+\theta^2) + \theta + \theta^3] \\
&\vdots \\
\nu_{si} &= \frac{V_s}{1-\theta^2}\theta^{i-2}[6\theta^2 - 4(\theta + \theta^3) + 1 + \theta^4]
\end{aligned}$$

Using the time domain equivalent of (7.1), (7.2) and lemma 6.3, the ACGF of the final estimation error e_t is given by the product

$$A(e_t) = \frac{V_s}{16}\frac{[2-(B+F)][4(1-\theta)^2+2(1+\theta)^2-(1-\theta)^2(B+F)]}{(1-\theta B)(1-\theta F)}$$

and looking at the central term:

$$\begin{aligned} V(e_t) &= \frac{V_s}{16(1-\theta^2)}[1+\theta(B+F)+\cdots]\cdot \\ &\cdot \quad [10(1-\theta)^2+4(1+\theta)^2-[2(1+\theta)^2+6(1-\theta)^2](B+F) \\ &+ \quad (1-\theta^2)(B^2+F^2)] \end{aligned}$$

Simplifying, we get:

$$\begin{aligned} V(e_t) &= \frac{V_s}{16(1-\theta^2)}[10(1-\theta)^2+4(1+\theta)^2- \\ &- \quad 2\theta[2(1+\theta)^2+6(1-\theta)^2]+ \\ &+ \quad 2\theta^2(1-\theta)^2] \end{aligned}$$

In the simple case where $\theta = 0$, we get: $V(e_t) = 7/128$.

Finally, the total revision error in concurrent estimates of the signal can be derived by writing the estimator $\hat{s}_t$ in function of the innovations a_t (see section 6.5):

$$\begin{aligned} \hat{s}_t &= \nu_s(B)x_t = \\ &= \nu_s(B)(1-\theta B)/(1-B^2)a_t = \\ &= \xi_s(B)a_t, \end{aligned}$$

so that we obtain :

$$\begin{aligned}
\xi_s(B) &= V_s \frac{(1-B)^2(1-F)^2}{(1-\theta B)(1-\theta F)} \frac{(1-\theta B)}{(1-B^2)} = \\
&= V_s \frac{1-B}{1+B} \frac{(1-F)^2}{1-\theta F} = \\
&= V_s[1-2B+2B^2+\cdots+2(-1)^n B^n] \cdot \\
&\cdot \quad [1+(-2+\theta)F+(\theta-1)^2F^2+\cdots+(\theta-1)^2\theta^{n-2}F^n]
\end{aligned}$$

Denoting $\xi_{s1},...,\xi_{si}$ the coefficient of F,...,F^i, then according to 6.12, the variance of the total revision is given by $\sum_{i=1}^{\infty} \xi_{si}^2$. The coefficients $\xi_{s1},...,\xi_{si}$ are easily obtained as

$$\begin{aligned}
\xi_{s1} &= V_s(-2+\theta) + V_s(\theta-1)^2 \frac{-2}{1+\theta} \\
\xi_{s2} &= V_s(\theta-1)^2 \frac{-1+\theta}{1+\theta} \\
&\vdots \\
\xi_{si} &= V_s(\theta-1)^2\theta^{i-2} \frac{-1+\theta}{1+\theta}
\end{aligned}$$

Summing the squared coefficients yields the variance of the total revisions in the concurrent estimates:

$$\begin{aligned}
V(r_t) &= \sum_{i=1}^{\infty} \xi_{si}^2 = \\
&= V_s^2(-2+\theta)^2 + 4V_s^2 \frac{(\theta-1)^4}{(1+\theta)^2} - 4V_s^2 \frac{(-2+\theta)(\theta-1)^2}{(1+\theta)^2} + \\
&+ V_s^2(\theta-1)^4 \frac{(1-\theta)^2}{(1+\theta)^2} \frac{1}{(1-\theta^2)}.
\end{aligned}$$

The variance of the total revisions in the concurrent estimates takes the value 17/256 if $\theta = 0$. Denoting $\hat{n}_t^0$ the concurrent estimator of the seasonally adjusted series,

then the final estimator of n_t lies with a probability of 95% in an interval given by $\hat{n}_t^0 + / - 2\sqrt{17}/16$. In order to build a confidence interval for the unobserved n_t, the final error variance must be taken in consideration; this yields $\hat{n}_t^0 + / - 2(\sqrt{17}+\sqrt{14})/16$ still for the case $\theta = 0$.

7.3 Trend plus Cycle example

We analyze the model discussed in section 4.1.3. To ease the readers understanding, we recall that the model for the observed series was:

$$(1 + .7B)\Delta x_t = (1 + .404B - .039B^2)a_t.$$

This series was the sum of a trend s_t and a cycle n_t component, and making the former canonical, we obtained :

$$\begin{aligned} \Delta s_t &= (1 + B)a_{st} \qquad V_s = .161 \\ (1 + .7B)n_t &= (1 + .496B)a_{nt} \qquad V_n = .306. \end{aligned}$$

The estimators of the trend and nontrend components are immediately derived as:

$$\hat{s}_t = V_s \frac{(1 + B)(1 + F)(1 + .7B)(1 + .7F)}{(1 + .404B - .039B^2)(1 + .404F - .039F^2)} x_t$$

$$\hat{n}_t = V_n \frac{(1 + .496B)(1 + .496F)(1 - B)(1 - F)}{(1 + .404B - .039B^2)(1 + .404F - .039F^2)} x_t$$

The WK filter then corresponds to the ACGF of the model,

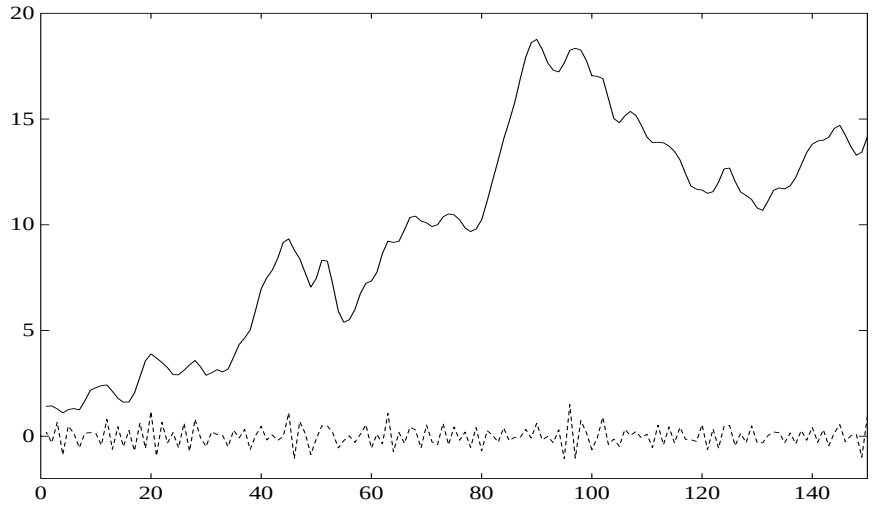
$$\theta(B)z_t = (1 + B)(1 + .496B)b_t,$$

with $V_b = V_s V_n = .049$. It is interesting to compare the estimators obtained in this canonical decomposition with the ones of the decomposition refered to as "model 1" in section 4.1.3:

$$\begin{aligned} \Delta s_t &= a_{st} \qquad V_s = .645 \\ (1 + .7B)n_t &= a_{nt} \qquad V_n = .129, \end{aligned}$$

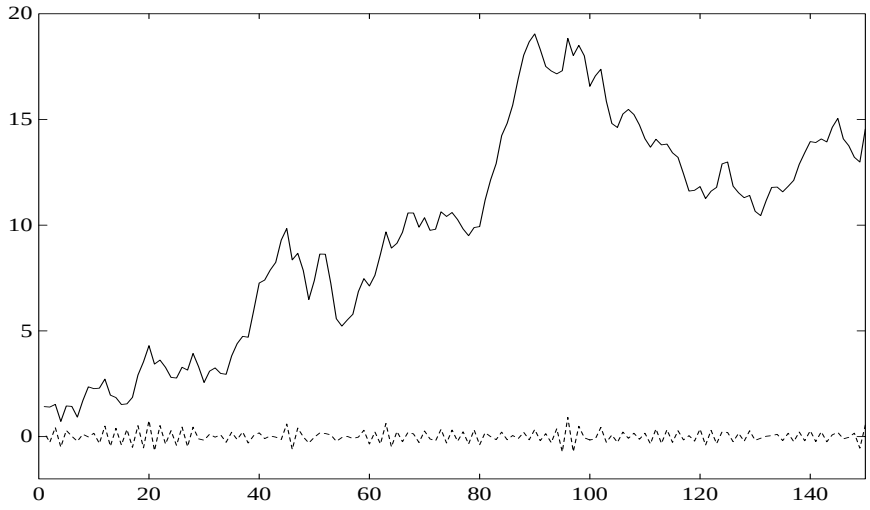
where the identification criterion is the one used in STS models. The corresponding estimators are plotted on figure 7.1 for the canonical decomposition, on figure 7.2 for this last model. It is clearly seen that the estimator of the trend is smoother when the trend is specified canonical than when it specified according to the STS criterion. But given that the cycle catches all the noise of the model in this canonical decomposition, the cycle so-obtained is in turn more unstable than the STS cycle.

Figure 7.1: Trend and Cycle Estimators in Canonical Decomposition



Trend — Cycle - -

Figure 7.2: Trend and Cycle Estimators in STS Decomposition



Trend — Cycle - -

7.4 Portugal Employment in Manufactured Bricks Production

This series has been showed on figure 1.1, and it has been analysed in section 3.5. A satisfying model for describing the stochastic behavior of the series was the (2,1,0) model with parameters

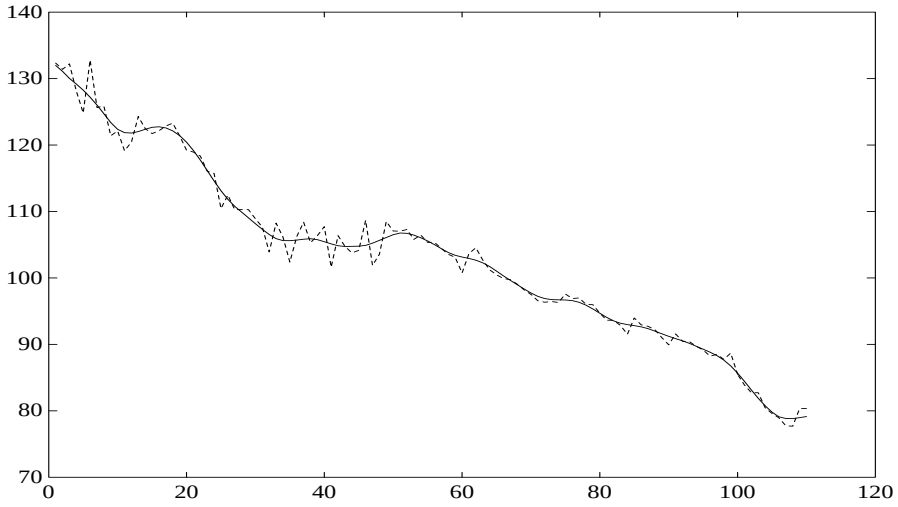
$$(1 + .494B + .291B^2)\Delta x_t = a_t,$$

where the standard deviation of the residual was $\sqrt{V_a} = .019$. According to the discussion held in section 3.5, these residuals seem to be randomly distributed. The autoregressive parameters of the models are significant, and the roots of the AR polynomial are complex conjugates roots of modulus .539 and argument 117 degrees. Thus, a pattern in the data of period close to three-times a year is caught by this model. If we assign this polynomial to a seasonal component, then this series POEM2640 can be decomposed as the sum of a trend, of a seasonal plus an irregular component according to:

$$\begin{aligned} \Delta p_t &= (1 + B)a_{pt} \quad V_p = .078, \\ (1 + .494B + .291B^2)s_t &= (1 - .473B - .567B^2)a_{st} \quad V_s = .098, \\ V_u &= .255 \end{aligned}$$

Notice that the seasonal component so-obtained is unusual: it is stationary, and represent movements which repeat themselves 3 time a year. The corresponding estimators are plotted on figures 7.3 and figures 7.4. The former plots the series together wih the estimated trend, while the second plots the two short-term movements. It is interesting to look at the spectra of the components and of the associated estimators on figure 7.5, 7.6 and 7.7. It is seen on the three pictures that the spectrum of the estimator is always below the one of the component. The larger difference occurs at the frequency where the contribution of the component is relatively low. Also, figure 7.7 illustrates the general result that the estimator of the irregular component is not a white noise. The differences between the components and their estimators are related to the estimation error.

Figure 7.3: Series POEM2640 (- -) and Trend (—)



The variance of the estimation errors in the final estimators of the canonical trend and the seasonal both specified canonical are obtained as $.099V_a$ and $.106V_a$, respectively. The final canonical trend is better estimated than the final canonical seasonal. This is in agreement with the values of the central coefficient of the WK filter designed to estimate these components, $\nu_{p0} = .308 < \nu_{s0} = .338$ (see lemma 6.5). This shows that the the final trend is more stable than the final seasonal, and for that reason all the noise of the series must be removed from the trend and assigned to the seasonal for having the decomposition the most accuratly estimated. For concurrent estimation, the opposite works since it is found that the variance of the total estimation error on the trend is at $.192V_a$ against $.136V_a$ for the seasonal. A look can the coefficient of B^0 in the ξ-weights confirms this gain in accuracy since we had: $\xi_{p0} = .478 > \xi_{s0} = .265$.

Figure 7.4: Seasonal and Irregular estimators in POEM2640

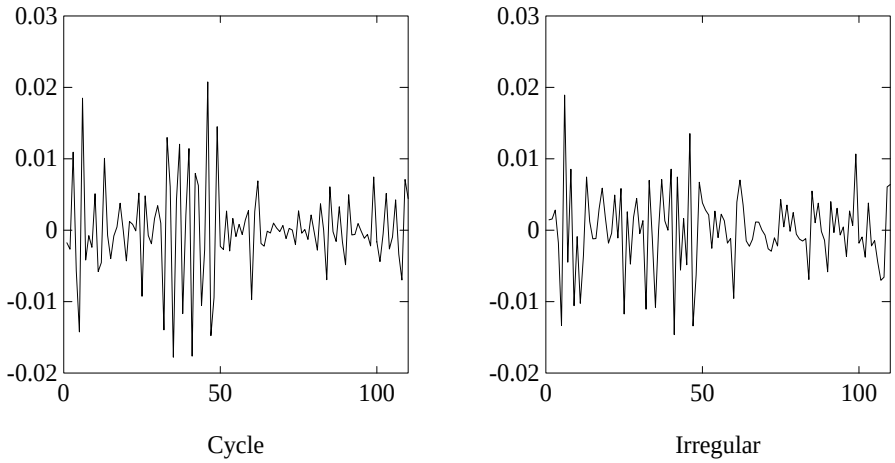
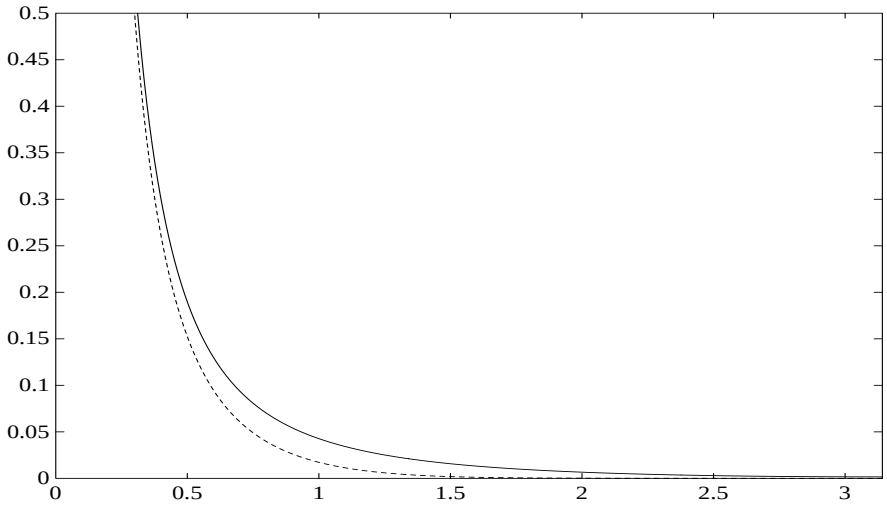
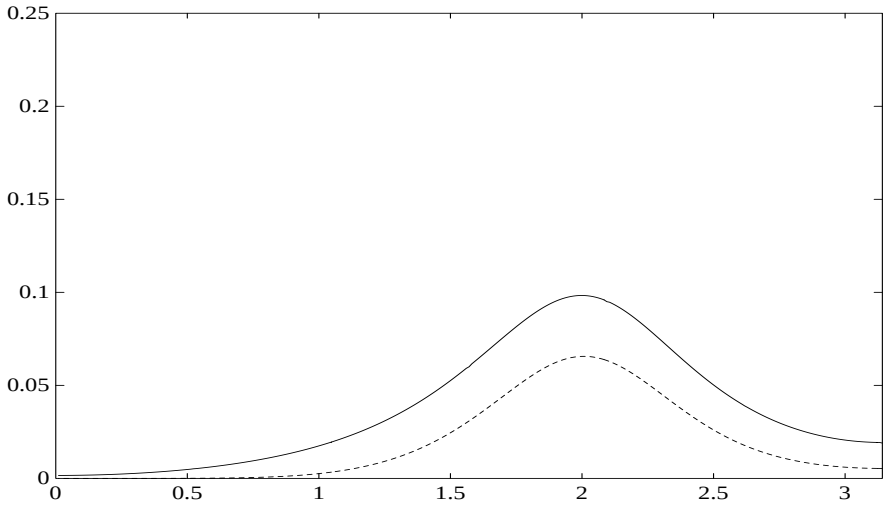


Figure 7.5: Trend Spectrum



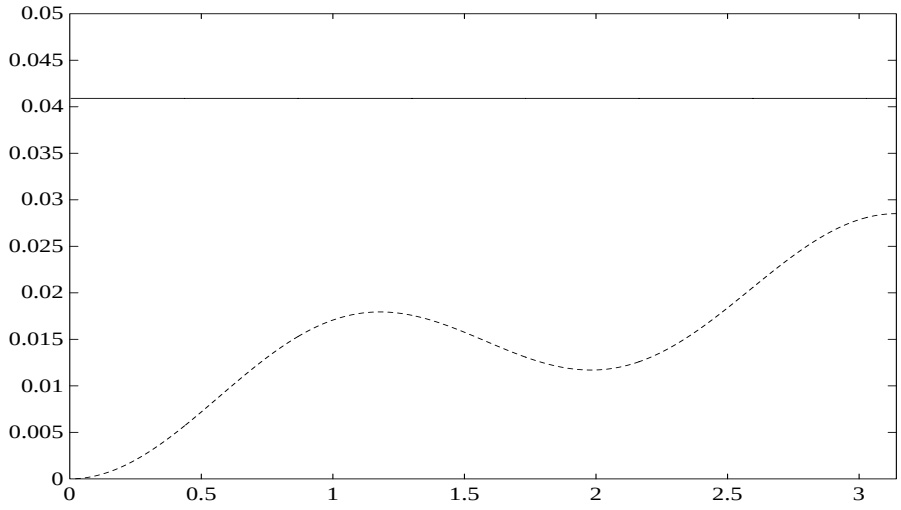
— Component - - Estimator

Figure 7.6: Seasonal Spectrum



— Component - - Estimator

Figure 7.7: Irregular Spectrum



— Component - - Estimator

7.5 French Total Industry Production

We now focus on the series FRPD101 already plotted in figure 1.1 This series, with another sample size, has been subject of discussion in Fiorentini and Planas (1996b). The behavior of this series is well described by an airline model such that

$$\Delta\Delta_{12}x_t = (1 - .561B)(1 - .488B^{12})a_t,$$

The MA parameters are significant, while the residual standard deviation is $\sqrt{V_a} = .017$. These residuals have been found randomly distributed in section 3.6. Decomposing the series into a nonseasonal plus a seasonal component specified as

$$\begin{aligned} U(B)s_t &= (1 + 1.150B + 1.031B^2 + .830B^3 + .570B^4 + .311B^5 + .073B^6 \\ &- 0.130B^7 - 0.278B^8 - 0.398B^9 - .462B^{10} - .645B^{11})a_{st} \\ \Delta n_t &= (1 - 1.517B + .542B^2)a_{nt} \end{aligned}$$

where the pseudo-innovations variances are $V_s = .0263$ and $V_n = .576$. The spectra of these two components, together with the spectrum of the observed series, are plotted in figure 7.8. It is seen that the series is characterised by very stable seasonal fluctuations. This can seen also in the magnitude of the pseudo innovations variances which is relatively low. The series seasonally adjusted by means of the WK filter can be seen in figure 7.9. In order to draw a comparison, the adjustment by X12 filtering is also reported on figure 7.9. The default adjustment filter of X12 with a 13-term Henderson filter has been found adequate for this seasonality. In fact the gain of the WK filter and of the X12 default adjustment filter are very close for this series. The output of both procedures are pretty similar, yet some more short term variability in the X12's output may be seen if one focus on the noise which is embodied in the SA series. Indeed, the variance of the stationary transformation of the seasonally adjusted series with X12 is 1.13 times larger that the one obtained with WK estimation. An explanation of this phenomemon can be found in Fiorentini and Planas (1996b); in particular, they show how the X12 adjustement filters may overestimate the short term movements in time series.

Figure 7.8: FRPDB101: Spectra of Series and Components

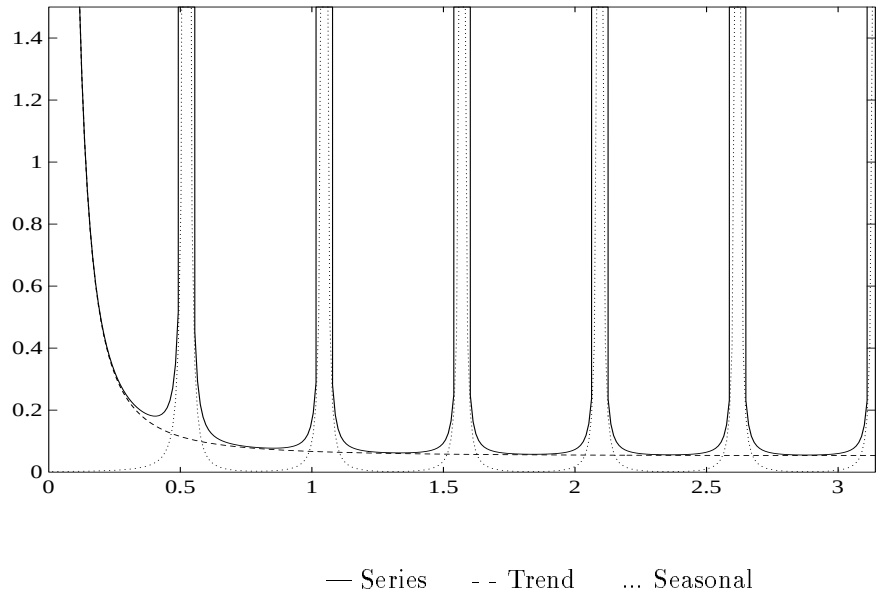
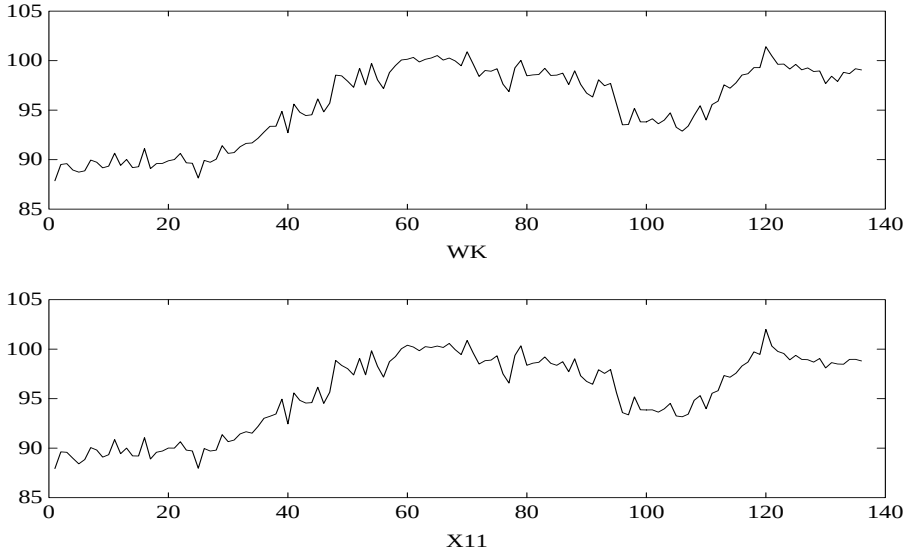


Figure 7.9: FRPDB101: Seasonally Adjusted Estimates



7.6 Italian Production of Soft Drinks

The Italian production of soft drinks has been presented in figure 3.6, section 3.5. It is characterized by a $(2,1,1)(0,1,1)_{12}$ model specified as:

$$(1 - .085B + .286B^2)\Delta\Delta_{12}x_t = (1 - .750B)(1 - .640B^{12})a_t,$$

with residuals standard error $\sqrt{V_a} = .067$. Every coefficient is significantly different from zero, and the residuals satisfied the diagnostic checks performed in section 3.6. The AR polynomial, denoted $\phi(B)$, has complex conjugate roots with modulus .54 and argument 85.42 degrees, that is close enough to the seasonal harmonic at 90 degrees which is associated with four-times a year movement. This root is thus assigned with the seasonal behavior of the series. The resulting model for the seasonal component is found to be:

$$\begin{aligned}\phi(B)U(B)s_t &= (1 + 1.341B + .846B^2 + .810B^3 + .737B^4 + .660B^5 + \\ &+ .629B^6 + .534B^7 + .544B^8 + .443B^9 + .529B^{10} + \\ &+ .4372B^{11} - .050B^{12} - .293B^{13})a_{st},\end{aligned}$$

while for the nonseasonal part of the series:

$$\Delta^2 n_t = (1 - 1.690B + .700B^2)a_{nt}.$$

The pseudo innovations variances are obtained as $V_s = .283$ and $V_n = .394$. The MMSE estimator of the seasonally adjusted series is displayed on figure 7.10. The final estimation error in historical estimator of the seasonally adjusted series is found at $.152V_a$, while the variance of the total revisions in concurrent estimate is of $.064V_a$. Hence, adding $+/-2\sqrt{.152}\sqrt{V_a}$ on the historical estimator gives the 95% confidence interval for the component, while 95% confidence interval around concurrent estimator is obtained by adding $+/-2\sqrt{.216}\sqrt{V_a}$ on the concurrent estimtes. Figure 7.11 displays the forecast of the series and of the trend together with their forecast confidence interval.

Figure 7.10: ITPDB428: Original Series (- -) and SA Series (—)

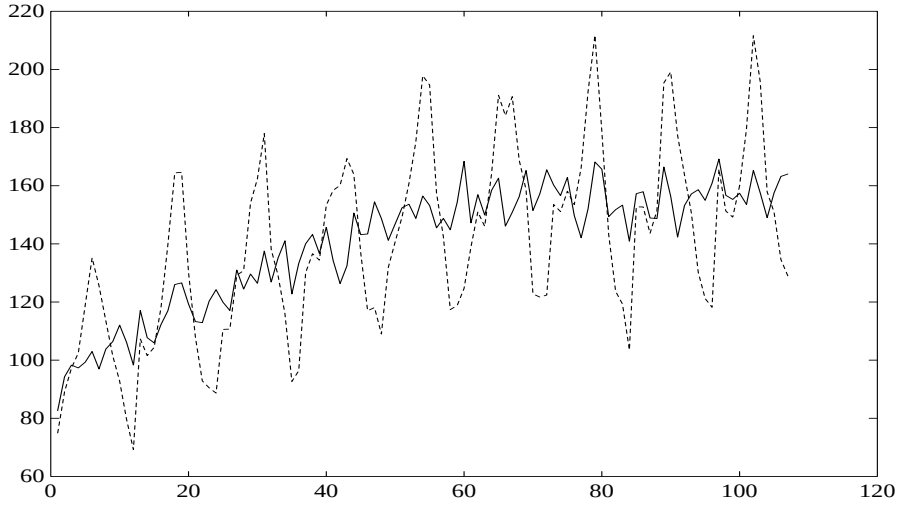
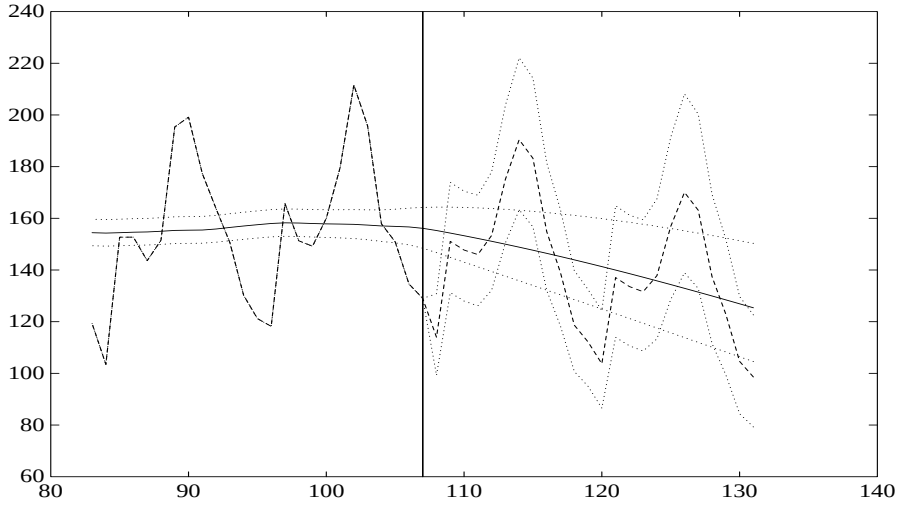


Figure 7.11: ITPDB428: Series and Forecast



- - Series Forecasts — Trend Forecasts ... Confidence Interval

The typical result that unobserved components are better forecasted than the observed series is clearly illustrated.

Part III

Regression in Time Series Analysis

Chapter 8

Linear Regression Models with ARIMA Errors

8.1 Model Specification

In the first chapter, we have seen that the Wold decomposition states that a covariance stationary time series can always be written as the sum of a deterministic plus a stochastic part. Up to now, we have focused on the modelling and on the extraction of signals in stochastic processes. We now turn to the modelling of some deterministic effects. First, suppose that an observed series Y_t is stationary; we write

$$Y_t = \mu_t + Z_t,$$

where μ_t represents the mean of the process such that :

$$\mu_t = E(Y_t) = X_t'\beta.$$

The variable X_t is a vector of r regression variables, $X_t' = (X_{1t}, \cdots, X_{rt})$, weighted by the vector of coefficients β, $\beta' = (\beta_1, \cdots, \beta_r)$. The Z_t's follow the general ARMA process, $\phi(B)Z_t = \theta(B)a_t$, where $\phi(B)$ and $\phi(B)$ satisfy the stationarity and invertibility conditions, respectively, while a_t is NID$(0, V_a)$ distributed. The variance the process Y_t is then given by

$$V(Y_t) = V(Z_t)$$

which is the standard covariance matrix of an ARMA process. In practice, stationarity is a property which most often requires a prior data transformation. Denoting $\delta(B)$ the stationarity inducing transformation of the data and $y_t = \delta(B)Y_t$, $z_t = \delta(B)Z_t$, we have:

$$y_t = \delta(B)X_t'\beta + z_t$$

where $\delta(B)X_t'$ stands for $\delta(B)X_t' = (\delta(B)X_{1t}, \cdots, \delta(B)X_{rt})$. The polynomials $\phi(B), \theta(B)$ and $\delta(B)$ may include or not seasonal lags. Another writing is

$$\phi(B)(\delta(B)Y_t - \delta(B)X_t'\beta) = \theta(B)a_t.$$

For notational simplicity, we shall write $x_t = \delta X_t$, so that:

$$\phi(B)(y_t - x_t'\beta) = \theta(B)a_t.$$

This type of models are known as linear regression models with ARIMA errors, REGARIMA in short. A good discussion of REGARIMA models can be found in Bell (1995). An appealing feature of REGARIMA models is that the first and second moments of the variable, that is the mean function and the variance-covariance matrix, are handled separatly.

8.2 Identification and Estimation

The problem of identifying and estimating an ARIMA model is now to be considered together with that of estimating the regression coefficients. It is convenient to rewrite the REGARIMA equation in matrix form:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{z}$$

where $\mathbf{y} = (y_1, \cdots, y_T)'$, $\mathbf{x}$ is a Txr matrix with column i given by $(x_{i1}, \cdots, x_{iT})'$ and $\mathbf{z} = (z_1, \cdots, z_T)'$. In the simple case where the z_t are white noises, then the best linear unbiased estimate of β is the *Ordinary Least Squares* OLS estimator :

$$\hat{\beta}_{OLS} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}.$$

The estimator $\hat{\beta}_{OLS}$ is unbiased since $E(\hat{\beta}_{OLS}) = \beta$ and it is consistent since its variance is obtained as $V(\hat{\beta}_{OLS}) = V_a(\mathbf{x}'\mathbf{x})^{-1}$, which is minimum. Why is this variance a lower bound is explained in details in many general textbooks; see for example Spanos (1986). Further, since it is a linear combination of a normal variable, the estimator is normally distributed: $\hat{\beta}_{OLS} \sim N(\beta, V_a(\mathbf{x}'\mathbf{x})^{-1})$.

The ordinary least squares estimators are straightforwardly extended to the cases where the z_t are correlated. When $V(\mathbf{z}) = \Sigma$ is known, the *Generalized Least Squared* (GLS) estimator is obtained as:

$$\hat{\beta}_{GLS} = (\mathbf{x}'\Sigma^{-1}\mathbf{x})^{-1}\mathbf{x}'\Sigma^{-1}\mathbf{y}.$$

which is the minimum variance linear unbiased estimator in that case. It is unbiased since it is readily checked that $E(\hat{\beta}_{GLS}) = \beta$, and it is efficient since $V(\hat{\beta}_{GLS}) = \mathbf{x}'\Sigma^{-1}\mathbf{x}$, which in that case is the lower bound for the variance of any linear estimator. Under the assumption of white noise observations, $\Sigma = V_a I$, the GLS estimator reduces to the OLS one. On the other hand, computing an OLS estimation when the errors are correlated would lead to an unbiased estimator with variance given by:

$$V(\hat{\beta}_{OLS}) = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\Sigma^{-1}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}.$$

It can be checked that for every β_i-estimate, $i = 1, \cdots, r$, the inequality $V(\hat{\beta}_{iOLS}) > V(\hat{\beta}_{iGLS})$ holds. Thus, the OLS estimator of β is not efficient, and since the GLS estimation yields more precise results, it is the favored estimation procedure. The variance-covariance matrix needs however to be known. In chapter 3, we have seen how the variance-covariance matrix of a variable described by an ARIMA model may be derived simply as a function of the AR and MA parameters. Thus, conditional on the choice of a particular model, the knowledge of the matrix Σ only depends on

the estimators of the ϕ's, θ's and of the innovation variance. We make explicit that dependence by writing:

$$\Sigma = V_a C(\phi, \theta),$$

where $\phi = \{\phi_1, \cdots, \phi_p\}$ and $\theta = \{\theta_1, \cdots, \theta_q\}$. Now assuming that the observations are normally distributed, $\mathbf{y} \sim N(\mathbf{x}\beta, V_a C(\phi, \theta))$, then the likelihood function for the RegArima model specification is given by:

$$\begin{aligned} L(\beta, \phi, theta, V_a) &= (\sqrt{V_a 2\pi})^{-n} |C(\phi, \theta)|^{-1/2} \cdot \\ &\cdot \exp(-1/2V_a)(\mathbf{y} - \mathbf{x}\beta)' C(\phi, \theta)^{-1} (\mathbf{y} - \mathbf{x}\beta). \end{aligned}$$

The associated log-likelihood is obtained as

$$\begin{aligned} l(\beta, \phi, theta, V_a) &= (-n/2) log V_a - (1/2) log|C(\phi, \theta)| - \\ &- (1/2V_a)(\mathbf{y} - \mathbf{x}\beta)' C(\phi, \theta)^{-1} (\mathbf{y} - \mathbf{x}\beta). \end{aligned}$$

Maximizing this last expression with respect to β and V_a yields the Gls estimators:

$$\hat{\beta} = (\mathbf{x}' C(\phi, \theta)^{-1} \mathbf{x})^{-1} \mathbf{x}' C(\phi, \theta)^{-1} \mathbf{y}$$

$$\hat{V}_a = (1/n)(\mathbf{y} - \mathbf{x}\hat{\beta})' C(\phi, \theta)(\mathbf{y} - \mathbf{x}\hat{\beta})$$

Thus the maximisation of the log likelihood may be computed over the variables ϕ and θ, the optimal values for β and V_a resulting from the above formula. The maximisation of the likelihood of RegArima models with respect to the Ma and Ar parameters turns then out to be similar to that of an Arima model.

Furthermore, the estimators of the regression parameters have the property to be asymptotically uncorrelated with the Arma parameters estimators. Inference may

thus be drawn separatly, and considering the regression parameters, the GLS estimator is still normally distributed: $\hat{\beta}_{GLS} \sim N(\beta, (\mathbf{x}'\Sigma^{-1}\mathbf{x})^{-1})$. Replacing Σ by its estimator in function of $\hat{\phi}$ and $\hat{\theta}$, then the variance-covariance matrix of $\hat{\beta}$, say $\hat{\Omega}$, can be estimated. Hypothesis tests are then easily derived, under t-statistics forms $\hat{\beta}_i/\hat{\Omega}_{jj}$ or as χ^2-tests for testing joint significance of set of parameters using the appropriate partition of the matrix $\hat{\Omega}$. For example, a test of $\beta = \beta_0$ may simply be computed using: $(\hat{\beta} - \beta_0)'\hat{\Omega}^{-1}(\hat{\beta} - \beta_0) \sim \chi^2_r$.

It is now possible to turn to examine the identification procedure of REGARIMA models. The choice of the regressors involved in $\mathbf{x}$ depends on the information available on the series under analysis. For example, suppose one would deal with a monthly series of energy consumption. Then, a series representing the monthly temperature may bring some valuable information for understanding the behavior of the energy consumption series, and thus may be seen as a candidate regressor. Given that the choice of regressors is operated first on qualitative grounds, the problem then reduces to identifying a stochastic structure when exogenous regressors are present. This is similar to the identification problem in ARIMA models (see chapter 3), although some modification should be considered to deal with the regression variables. In agreement with Bell (1995), the following sheme can be adopted:

- The regressors $(X_{1t}, \cdots, X_{rt})$ are supposed known;
- Determine the minimum order of differencing needed to make the series stationary by plotting the ACF of Y_t, ΔY_t, $\Delta_m Y_t$, $\cdots$ (m stands for data periodicity); the polynomial $\delta(B)$ is then obtained, typically as $\delta(B) = \Delta^d \Delta^D_m$, $d = 0, 1, 2$ and $D = 0, 1$;
- Transform the series and regressors into $y_t = \delta(B)Y_t$ and $x_t = \delta(B)X_t$;
- Regress y_t on x_t and get the coefficients $\hat{\beta}_{OLS}$ and the residuals $\hat{z}_t$;
- Apply the standard identification methodology for ARIMA models to the residuals $\hat{z}_t$: that is, examine the ACF and PACF of $\hat{u}_t$;
- A model for Z_t is finally found using $\delta(B)Z_t = z_t$; this model may then be estimated together with a GLS reestimation of the regression coefficients β.

Notice that, in the second step, the differencing polynomial appropriate for the variable under analysis is derived without prior removing of the regression effects. This is because in general deterministic regressors do not obscure the identification of the differencing polynomial. The only case where some problems may be met is the one of a deterministic time polynomial taken as regressor, but it has been seen that such a specification should be better avoided. The third step involves OLS instead of GLS estimation for two main reasons: first, the ARMA strucutre is not known at that stage, and second because the OLS estimators are still consistent. Given that the aim is to identify a stochastic structure and that this involves computation of ACF and PACF, consistency of the sample estimates of ACF and PACF is what matters at that stage. The result that consistent estimates of the β imply consistency of the sample estimates of ACF and PACF can be found in Fuller (1976). Diagnostic checking of the ARIMA part may then be computed using the techniques presented in chapter 3 together with tests of the relevance of the regressors. We now turn to examine the regressors most often used in practice.

8.3 Special Regressors: Calendar Effects

8.3.1 Trading Days

If one would deal with daily series to measure business activity, an expected result would be that the activity varies over the different days of the week. Daily series are however rarely available, and most of the statistical treatment in statistical agencies like EUROSTAT involves monthly or quarterly series. Consider the case of monthly series: as a month embodies a varying number of Mondays, Tuesdays, ..., and Sundays, the business activity varies accordingly. A correction for this "trading day" effect may thus be needed. A simple procedure would be to build 7 dummy variables (one by day) $X_{1t}, \cdots, X_{7t}$, such that X_{1t} is the number of mondays in month t, ..., X_{it} the number of i-th day of the week, and to regress the observed series Y_t on the X_{it}'s so as to obtain:

$$Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_7 X_{7t} + Z_t$$

Of course, the number of mondays in month t depends also on the year, but to ease the presentation we omit here that dependence. In practice, the β's coefficients tend

to be highly correlated, and a reparametrization is needed. This may be done mainly in two different ways. The first consists in a reparametrisation of the regression above. Interpreting the β_i's as the mean activity in day i, then $\overline{\beta} = (1/7)\sum_{i=1}^{7} \beta_i$ represents the average daily activity. Let us denote m_t the length of month t, which verifies $m_t = \sum_{i=1}^{7} X_{it}$. Writing,

$$\begin{aligned}
\sum_{i=1}^{7} \beta_i X_{it} &= \sum_{i=1}^{7} (\beta_i - \overline{\beta}) X_{it} + \sum_{i=1}^{7} \overline{\beta} X_{it} = \\
&= \sum_{i=1}^{7} (\beta_i - \overline{\beta})(X_{it} - X_{7t}) + X_{7t} \sum_{i=1}^{7} (\beta_i - \overline{\beta}) + \sum_{i=1}^{7} \overline{\beta} X_{it} = \\
&= \sum_{i=1}^{7} b_i (X_{it} - X_{7t}) + \overline{\beta} m_t
\end{aligned}$$

There are thus six dummy variables used plus a length of month adjustment variable. Given that $b_i = \beta_i - \overline{\beta}$, in this reparametrization the b_i's represent the difference between the mean activity in day i and the average daily activity. However, the lost of one dummy may not be enough to decrease the correlations between the regressors, and so another modelling may be to consider some "working days" adjustment which involve a single regressor defined as X_t = #(Mondays,...,Fridays) - 5/2#(Saturdays,Sundays). It is justified by the idea that the pattern for working days and week-ends days are different. Again, a length of month adjustment may be introduced.

8.3.2 Easter Effects

Activity also varies around some special dates in the year, like Christmas or Easter, which are typically associated with sales increases. While the effect of Christmas on activity is always caught by the month of December, the effect of Easter may concern either March or April, according to the year. The date of Easter during the year thus implies some instabilities in the seasonal patterns related to the months of March and April. For that reason, the Easter effect requires a special attention.

To be incorporated to the set of regressors, the Easter effect must be translated into something quantitative. Suppose that the increase in the purchases related to Easter starts n-days (say n=10) before Easter and ends the Saturday before Easter. To represent this effect when one is dealing with monthly time series, a simple way is

to create a dummy variable, say $H(.)$. This variable is monthly, so we write $H(.,t) = 0$ for every month t. Next, to model the special effect on march, april, the number of days in march and april which are belonging to the interval of n-days is counted. For example, in 1996, the Easter date was 07/04: for $n = 10$, we have thus 4 days in march and 6 days in april. The regression variable $H(n,t)$ takes then the value .4 in March 1996 and .6 in April 1996.

8.3.3 Example: Series FRPDB101

The French Total Production series (except construction) gives a nice example of the importance of calendar effects in flow time series. The series plotted in figure 1.1 was in fact corrected for trading days and easter effect. The original series for the sample dates 1985-1 1996-11 (that is 143 observations) is given in figure 8.1: compared to figure 1.1, much less regularity is seen in the original series. Regressing the observed Y_t on six trading days plus a length of month effect, $TD_{1t}, \cdots, TD_{7t}$ as described in the previous section and incorporating an easter effect EE_t, then the airline model given in (8.1) is fitted for the remaining term, yielding the general model:

$$\begin{aligned} Y_t &= \sum_{i=1}^{7} b_i TD_{it} + \alpha EE_t + Z_t \\ \Delta\Delta_{12} Z_t &= (1+\theta B)(1+\theta_{12} B^{12}) a_t \end{aligned} \tag{8.1}$$

The parameters estimates are obtained as $\hat{\theta}_1 = -.426(.008)$, $\hat{\theta}_{12} = -.449(.009)$, and while the regressors estimates are given by:

Table 8.1			
Regressors in model (8.1)			
Parameter	Estimate	Std. Dev.	t-ratio
$\hat{b}_1$	0.858E-04	.002	0.04
$\hat{b}_2$	0.642E-02	.002	2.88
$\hat{b}_3$	0.338E-02	.002	1.53
$\hat{b}_4$	0.906E-02	.002	4.18
$\hat{b}_5$	0.311E-02	.002	1.43
$\hat{b}_6$	-.145E-01	.002	-6.70
$\hat{b}_7$	0.269E-01	.007	3.82
$\hat{\alpha}$	-.243E-01	.004	-5.44

According to subsection 8.3.1, the coefficient b_i, $i = 1, \cdots, 6$, refers to the difference between the mean activity in day i and the daily mean activity. Table 8.1 shows first that the production on the Mondays is not significantly different from the mean production over the 7 days of the week. Conversely, production is relatively low on the Saturdays, and so $\hat{b}_6$ takes a negative value at $-.0145$, with a very high t-value. The other days of the week show a relatively high activity. The coefficient b_7 refers to the influence of the length of month; it takes a positive value at .027, also very significant, which reflects the common sense expectation that the longer the month, the more output is produced. Finally the easter effect is significantly negative. This model has yielded a BIC value of -8.055. Reducing the number of trading days regressors to 1, and keeping the length of month adjustment, then fitting the corresponding model yields a BIC value of -8.11. Hence, for this series, the correction by a single trading day regressor seems preferable over a correction with 6 trading day regressors.

8.4 Outliers and Intervention Analysis

It is sometimes useful to remove data irregularities. The different types of data irregularities most often considered are additive outlier (AO), temporary change (TC) and level shift (LS). They can be presented in the following way. Suppose a model is fitted to the series y_t, so that the residuals e_t are available. Denoting $I_{t_0}(t)$ a dummy variable

Figure 8.1: FRPD101, Uncorrected Series

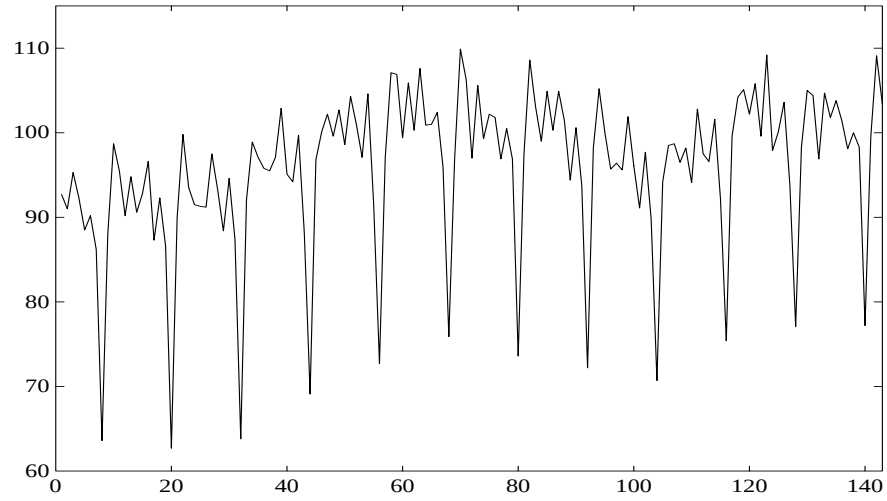
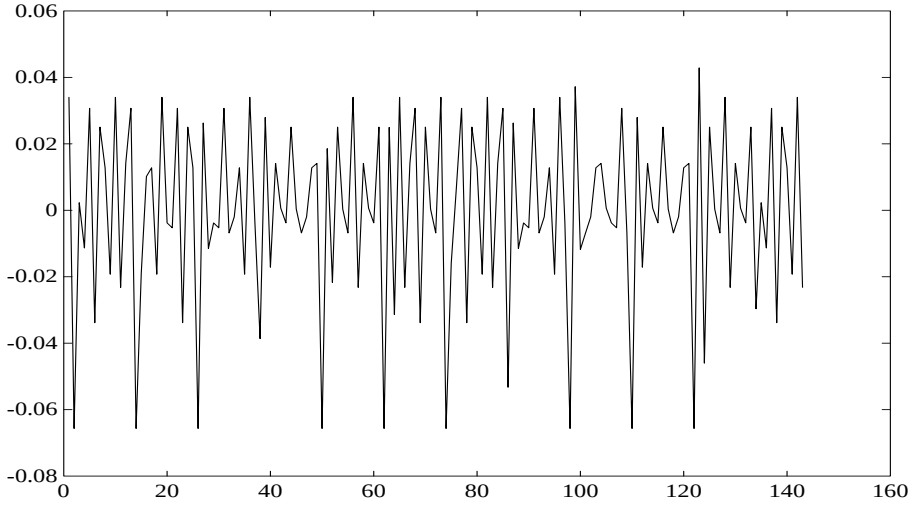


Figure 8.2: Trading days and Easter effect in FRPDB101



such that $I_{t_0}(t) = 1$ if $t = t_0$, 0 otherwise, then these three types of irregularities are defined by:

- AO: $e_t = a_t + w_A I_{t_0}(t)$;
- TC: $e_t = a_t + w_T/(1 - \eta B) I_{t_0}(t)$;
- LS: $e_t = a_t + w_L/(1 - B) I_{t_0}(t)$.

An additive outlier is thus able to catch a single point in the data, a temporary change a single point jump followed by a smooth return to the original path, and a level shift a permanent change in the level of the series. In the outlier analysis, two cases must be distinguished according to whether the irregularity in the data is known or not. For example, suppose a series representing the monthly traffic on a particular highway is analysed. Then, closure of the highway during a month would produce a zero in the data. This effect could be catched by the AO, and it would be possible to impose that pattern directly at the modelling stage. Other patterns like TC may be relevant when dealing with phenomena related for example to strikes, since return to a normal activity after the end of the strikes may take several periods. Finally, the patterns described by a LS may correspond for example to changes in nomenclature, since they typically make time series subject to a shift in the level. In all these cases, the event affecting the series is known together with the time of occurence. Imposing then an AO, TC or a LS as a regressor enables the analyst to quantify the significativity of the effect. This type of analysis is known as *intervention analysis* (see Box and Tiao, 1975).

On the other hand, no information may be available about some possible event affecting the series. The only way to find out such irregularities is thus to analyse the data. In practice, the analysis concentrates on the residuals of the fitted model. The removal of outliers is important because they may drastically distort the residuals sample ACF and PACF and, for example, an AO not taken into consideration may lead to overdifferencing (see for example Perron, 1986). The methodology for outlier detection, identification and estimation has been discussed in Chang, Chen and Tiao (1988), and further developed in Tsay (1984), Chen and Liu (1993) among others. It can be summarized as follow:

- A model is fitted to the series, and the residuals e_t are obtained;
- For every residual, estimators of w_A, w_T, w_L are computed together with their variance;
- Compute the t-values: when the t-value of one or some w_i's at some time t_0 exceeds a critical value C, then an outlier at time t_0 has been detected;
- To identify which type of outlier is dealt with, a comparison between the different t-values obtained is performed: the chosen outlier pattern is the one related with the greatest significativity.

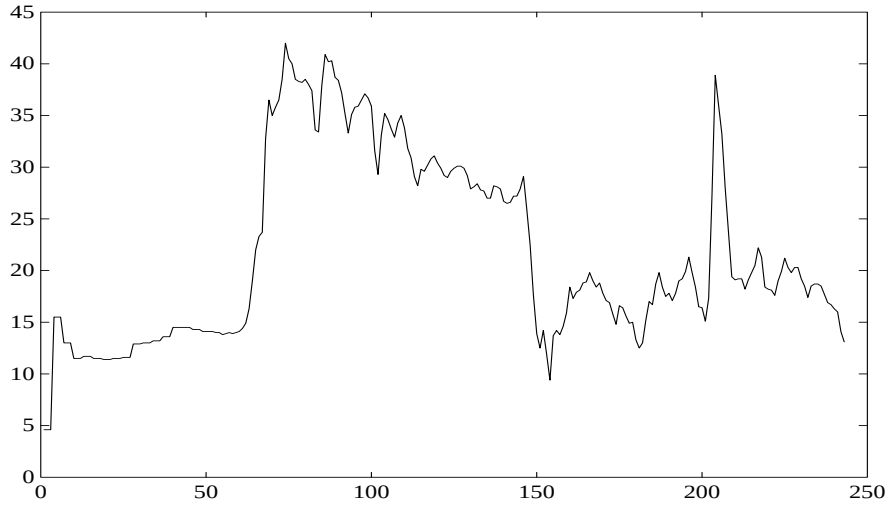
In practice, the critical size for outlier significativity is set at $C = 3.0, 3.5, 4.0$ for high, medium and low sensitivity, respectively, for series of length around 200. This critical size tends to decrease as the sample size increase. At that stage, a set of potential outliers has been obtained; it is then important to insure that the procedure did not produce spurious outliers. One possibility consists in returning to the general model for the observed series including the outliers previously found as regressors. Reestimating the RegArima model, the t-ratios of the regressors may be recomputed, and if the minimum t-ratio is less than C, then the corresponding outlier is deleted. The regression is recomputed with the updated regressor set and updated parameter estimators. This operation is repeated until no more outliers are found.

When an outlier have been detected, identified and estimated, then an important step of the analysis is to see if an interpretation can be given. The modelling and estimation tools have pointed out an irregularity in the data, and the practitioner should try to explain that irregularity. Consider for the example the Oil Price series (in \$ per barrel) which is displayed on figure 8.3. This series is monthly, starting in October 1975 and ending after 243 observations in October 1993. Many irregularities are readily seen in this time series. Despite of that, suppose that an analyst first fits a linear (2,1,1) model plus mean without allowing for outliers:

$$(1 + \phi_1 B + \phi_2 B^2)\Delta x_t = \mu + (1 + \theta_1 B)a_t, \tag{8.2}$$

the parameters estimators are found to be $\hat{\mu} = .003$ $(.005)$, $\hat{\phi}_1 = -1.044$ $(.177)$, $\hat{\phi}_2 = .213$ $(.063)$, $\hat{\theta}_1 = -.873$ $(.175)$, with residuals standard deviation $\sqrt{V_a} = .110$. A

Figure 8.3: Oil Price ($ per barrel), 1975-10 1993-10.



look at the residuals e_t, figure 8.4, clearly shows that several residuals are lying outside the confidence interval. Accordingly, the Kurtosis statistics is very large at 63.5. Proceeding according to the scheme described above and using C=4.0, 6 outliers are eventually found as:

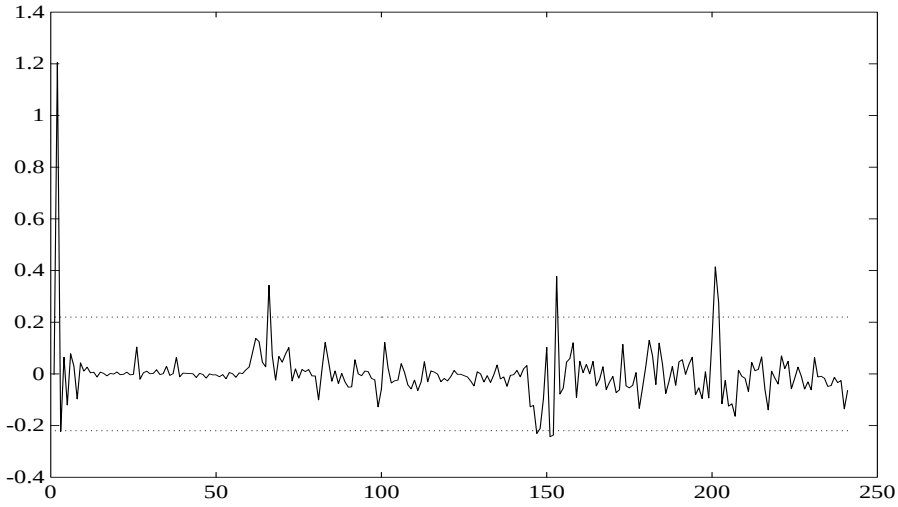
Table 8.2

Outliers in model (8.2)

Date	Type	Value	t-ratio
1 1974 (4)	LS	1.216	24.51
7 1986 (154)	AO	-.300	-9.92
5 1986 (152)	AO	.173	5.73
5 1979 (68)	LS	.279	5.62
9 1990 (204)	TC	.340	7.33
8 1990 (203)	LS	.392	7.41

while the reestimation of model (8.2) finally gives the estimates $\hat{\mu} = -.004 \quad (.005)$, $\hat{\phi_1} = -.388$ (.002), $\hat{\phi_2} = -.003$ (.001), $\hat{\theta_1} = .005$ (.006), with residuals standard deviation $\sqrt{V_a} = .054$. It is of interest to notice that removing the outliers have

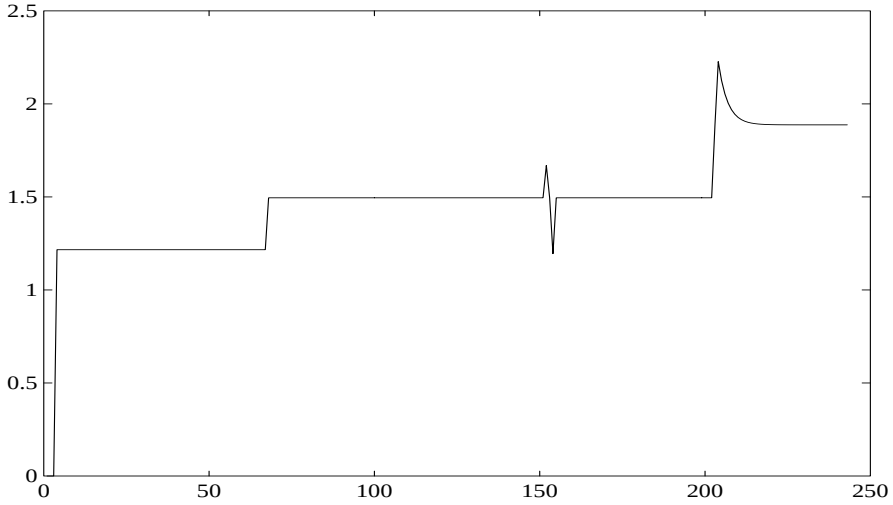
Figure 8.4: Oil Price Series: Residuals from (2,1,1) model



decreased by 50% the residuals standard deviation. The kurtosis statistics has been lowered to 4.43 (.31), which still shows some departures from normality but in much more limited extent than without the outlier correction. No autocorrelations seem left in the residuals since the Ljung-Box Q-statistics for the first 24 autocorrelations is 25.37, for 21 degress of freedom.

A plot of the outliers found can be seen on figure 8.5. It is of interest to relate these features to some events which have occured during the period. First, the level shift in Oil Prices detected in January 1974 obviously corresponds to the first oil crisis, where the price of a barrel jump from 4.6$ to 15.5 $ between December 1973 and January 1974. A second level shift is detected in May 1979, which is the date of the Shah fall in Iran. No doubt that this political event is reponsible for the rising of the barrel price from 23.7$ in April 1979 to 32.8$ in May 1979. Two additive outliers are then found in May and July 1986, the last one with a negative sign points a relatively large decrease in oil prices. These movements may be related to tensions within the OPEC, in particular between Iran-Iraq in one side and Saudi-Arabia in the other side. After a disagreement between these actors, the Saudi-Arabians flooded the oil market and in July 1986 the oil price were downsized from about 30$ to 9.4$. Finally, the level shift

Figure 8.5: Outliers in Oil Price Series



detected for August 1990 followed by a smooth return to a level close to the original one as the temporary change for September 1990 indicates can be related to the Iraqian invasion of Kuweit.

The Oil Price series has given an illustration of how outliers in time series may be interpretated. In some cases however, this task may be difficult, not only because of the lack of relevant information, but also because the occurence of outliers may be due to a nonlinear structure in the data not taken into account in the modelling. This point has been discussed in Fiorentini and Maravall (1995) and Planas (1996).

References

Aigner, D.J., Schneider, F., and Ghosh, D. (1988), "Me and My Shadow: Estimating the Size of the U.S. Hidden Economy From Time Series Data", in *Dynamic Econometric Modeling*, ed. Barnett, W., Berndt, E. and White, H., Cambridge: Cambridge University Press, 297-334.

Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, Wiley, New York.

Anderson, B. and Moore, J. (1979), *Optimal Filtering*, New Jersey: Prentice Hall.

Bartlett, M.S. (1946), "On the Theoretical Specification and Sampling Properties of Autocorrelated Time Series", *Journal of the Royal Statistical Society*, Ser. B,, 8, 27-41.

Bell, W.R. (1984), "Signal Extraction for Nonstationary Time Series", *The Annals of Statistics*, 12, 2, 646-664.

Bell, W.R. (1995), "X12-Time Series Modeling Course", Washington D.C.: U.S. Dept. of Commerce, Bureau of the Census.

Bell W.R., and Hillmer, S.C. (1984), "Issues Involved with the Seasonal Adjustment of Economic Time Series", *Journal of Business and Economic Statistics*, 2, 291-320.

Bell W.R., Hillmer S.C., and Tiao G.C. (1981), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series", *ASA-Census-NBER Conference on Applied Time Series Analysis of Economic Data.*

Bell, W.R. and Monsell, B.C. (1992), "X-11 Symmetric Linear Filters and their Transfer Functions", Bureau of the Census, Research Report n. RR 92/15, Washington.

Beveridge, W.H. (1921), "Wheather and Harvest Cycles", *Economic Journal*, 31,

429-452.

Beveridge, W.H. (1922), "Wheat Prices and Rainfall in Western Europe", *Journal of the Royal Statistical Society*, 85, 412-459.

Beveridge, S. and Nelson, C.R. (1981), "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'", *Journal of Monetary Economics*, 7, 151-174.

Box, G.E.P and Jenkins, G.M (1970): *Time Series Analysis: Forecasting and Control*, San Francisco: Holden Day.

Box, G.E.P, Hillmer S.C., and Tiao G.C. (1978), "Analysis and Modeling of Seasonal Time Series", in *Seasonal Analysis of Time Series*, ed. A. Zellner, Washington, D.C: U.S. Department of Commerce, Bureau of the Census, 309-334.

Box, G.E.P., Pierce, D.A., and Newbold, P., (1987), "Estimating Trend and Growth Rates in Seasonal Time Series", *Journal of the American Statistical Association*, 82, 276-282.

Box, G.E.P., and Tiao, G.C., (1975), "Intervention Analysis with Applications to Economic and Environmental Problems", *Journal of the American Statistical Association*, 70, 71-79.

Brockwell, P.J. and Davis, R.A. (1987): *Time Series: Theory and Methods*, New-York: Springer-Verlag.

Burman, J.P. (1980), "Seasonal Adjustment by Signal Extraction", *Journal of the Royal Statistical Society*, Ser. A, 143, 321-337.

Burridge, P. and Wallis, K.F. (1984), "Unobserved Components Models for Seasonal Adjustment Filters", *Journal of Business and Economics Statistics*, 2, 350-359.

Burridge, P. and Wallis, K.F., (1985), "Calculating the Variance of Seasonally Adjusted Series", *Journal of the American Statistical Association*, 80, 391, 541-552.

Buys Ballot, C, H. D. (1847), *Les Changements Périodiques de Température*, Utrecht:

Chang, I., Tiao, G.C., and Chen, C. (1988), "estimation of Time Series Models in the Presence of Outliers", *Technometrics*, 30, 2, 193-204.

Cleveland, W.P. (1972), "The Inverse Autocorrelations of a Time Series and their Applications", *Technometrics*, 14, 277-298.

Cleveland, W.P. and Tiao, G.C. (1976) "Decomposition of Seasonal Time Series: a Model for the X-11 Program", *Journal of the American Statistical Association*, 71, 581-587.

Cochrane, J.H. (1988), "How Big is the Random Walk in GNP", *Journal of Political Economy*, 96, 893-920.

Dagum, E.B. (1980), "The X11 ARIMA Seasonal Adjustment Model", Statistics Canada, Catalogue 12-564E.

Dickey, D.A, and Fuller, W.A. (1979), "Distribution of the Estimators for Autoregressive Time Series with an Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

Durbin, J. (1960), "The Fitting of Time-Series Models", *Review of the International Statistical Institute*, 28, 233-243.

Engle, R.F. (1978), "Estimating Structural Models of Seasonality", ed. A. Zellner, in *Seasonal Analysis of Economic Time Series*, Washington, D.C.: Dept. of Commerce, Bureau of the Census, 281-297.

Findley D.F. (1985), "An analysis and generalisation of M. Watson's minimax procedure for determining the component estimates of a seasonal time series", Washington, D.C.: Statistical Research Division, Bureau of the Census.

Findley, D.F., Monsell, B., Bell, W., Otto, M.C. and Chen, B.C. (1996), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program", Washington D.C.: Bureau of the Census.

Fiorentini, G. and Maravall, A. (1996), "Unobserved Components in ARCH Models:

An Application to Seasonal Adjustment", *Journal of Forecasting*, 15, 175-201.

Fiorentini, G. and Planas, C. (1996a), "Non-Admissible Decompositions in Unobserved Components Models", Cemfi, Working Paper n. 9613, Madrid.

Fiorentini, G. and Planas, C. (1996b), "Short-Term Variability in Seasonally Adjusted Series", mimeo, Eurostat, Luxembourg.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, New York: Wiley.

Garcia-Ferrer A., Del Hoyo J. (1992): "On Trend Extraction Models : Interpretation, Empirical Evidence and Forecasting Performances", *Journal of Forecasting*, 11, 645-665.

Gersch, W. and Kitagawa, G. (1983), "The Prediction of Time Series with Trends and Seasonalities", *Journal of Business and Economics Statistics*, 1, 253-264.

Gilbart, J.W. (1852), "On the Laws of the Currency in Ireland, as Exemplified in the Changes That Have Taken Place in the Amount of Bank Notes in Circulation in Ireland, Since the Passing of the Acts of 1845", *Journal of the Statistical Society of London*, 15, 307-325.

Gomez, V. and Maravall, A. (1996) "Programs TRAMO and SEATS, Instructions for the User", Bank of Spain, Working Paper n. 9628, Madrid.

Gourieroux, C. and Monfort, A. (1990), *Series Temporelles et Modèles dynamiques*, Paris: Economica.

Granger, C.W.J. (1978), "Seasonality: Causation, Interpretation and Implications", in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington D.C.: U.S. Dept. of Commerce, Bureau of the Census, 33-46.

Hannan, E.J. (1980), "The Estimation of the Order of ARMA Processes", *Annals of Statistics*, 8, 1071-1081.

Harrisson, P.J. and Steven, C.F. (1976), "Bayesian Forecasting", *Journal of the Royal Statistical Society*, Ser. B, 38, 205-247.

Harvey, A.C. (1981), *Time Series Models*, Oxford: Phillip Alan.

Harvey, A.C. (1985), "Trends and Cycles in Macroeconomic Time Series", *Journal of Business and Economic Statistics*, 3, 3, 216-227.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.

Harvey, A.C. and Koopman, S.J. (1992), "Diagnostic checking of Unobserved Components Time Series Models", *Journal of Business and Economic Statistics*, 10, 377-390.

Harvey, A.C., and Todd, P.H.J. (1983), "Forecasting Economic Time Series with Structural and Box-Jenkins Models : a Case Study", *Journal of Business and Economic Statistics*, 1, 299-307.

Haussman, J.A., and Watson, M.W. (1985), "Errors-in-Variables and Seasonal Adjustment Procedures", *Journal of the American Statistical Association*, 80, 531-540.

Hibbert Committee (1988), "Report of the Working Party on Seasonal Adjustment of the Monetary Aggregates", *CSO* 8.

Hillmer, S.C. (1985), "Measures of Variability for Model-Based Seasonal Adjustment Procedures", *Journal of Business and Economic Statistics*, 3, 1, 60-68.

Hillmer S.C. , and Tiao G.C. (1982) "An ARIMA-Model-Based Approach to Seasonal Adjustment", *Journal of the American Statistical Association*, 77, 63-70.

Hodrick, R. and Prescott, E. (1980), "Post-War U.S. Business Cycles: An Empirical Investigation", Carnegie Mellon University Manuscript.

Hotta, L.K. (1989), "Identification of Unobserved Components Models", *Journal of Time Series Analysis*, 10, 259-270.

Jenkins, G.M. (1979), "Practical Experience with Modeling and Forecasting Time Series", ed. Anderson, O.D., *Forecasting*, Amsterdam: North-Holland.

Joy, A. and Thomas, W. (1928), "The Use of Moving Averages in the Measurement of Seasonal Fluctuations", *Journal of the American Statistical Association*, 22, 241-252.

Kendall, M. (1976), *Time Series*, London: Griffin and Co.

King, W.W.I. (1924), "An Improved Method for Measuring the Seasonal Factor", *Journal of the American Statistical Association*, 19, 301-313.

Kohn, R. and Ansley, C.F (1986), "Estimation, Prediction and Interpolation for ARIMA Models with Missing Data", *Journal of the American Statistical Association*, 81, 751-761.

Kohn, R. and Ansley, C.F. (1987), "Signal Extraction for Finite Nonstationary Time Series", *Biometrika* , 74,411-421.

Kuznets, S. (1933), "Seasonal Variations in Industry and Trade", New York: National Bureau of Economic Research.

Ljung, G. and Box G.E.P. (1978), "On a Measure of Lack of Fit im Time Series Models", *Biometrika*, 65, 297-303.

Lucas, R.E. (1976), "Understanding Business Cycles", in *Stabilization of the Domestic and International Economy*, ed. Brunner, K. and Meltzer, A., Amsterdam: North-Holland.

Macaulay, F.R. (1931), *The smoothing of Time Series*, New York: National Bureau of Economic Research.

McLeod, A.I. and Li, W.K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.

Mann, H.B. and Wald, A. (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, 11, 173-220.

Maravall, A. (1983), "An Application of Non-Linear Time Series Forecasting", *Journal of Business Economics and Statistics*, 1, 66-74.

Maravall, A. (1985), "On Structural Time series Models and the Characterization of Components", *Journal of Business and Economic Statistics*, 3, 4, 350-355.

Maravall, A. (1986), "Revisions in ARIMA Signal Extraction", *Journal of the American Statistical Association*, 81, 736-740.

Maravall, A. (1988a), "A Note on Minimum Mean Squared Error Estimation of Signals with Unit Roots", *Journal of Economic Dynamics and Control*, 12, 589-593.

Maravall, A. (1988b), "The Use of ARIMA Models in Unobserved Components Estimation", in *Dynamic Econometric Modeling*, ed. Barnett, W., Berndt, E. and White, H., Cambridge: Cambridge University Press.

Maravall, A. (1989), "On the Dynamic Structure of a Seasonal Component", *Journal of Economic Dynamics and Control*, 12, 589-593.

Maravall, A. (1993a), "Stochastic Linear Trends: Models and Estimators", *Journal of Econometrics*, 54, 1-33.

Maravall, A. (1993b), "Unobserved Components in Economic Time Series", forthcoming in *Handbook of Applied of Econometrics*, ed. M.H. Pesaran, T. Schmidt, M. Wickens, Oxford Basil Blackwell.

Maravall, A. (1994), "Use and Misuse of Unobserved Components in Economic Forecasting", *Journal of Forecasting*, 13, 157-177.

Maravall, A. and Gómez, V. (1992), "Signal Extraction in Economic Time Series: Program SEATS", EUI Working Paper ECO No. 92/65, Department of Economics, European University Institute.

Maravall, A. and Mathis, A. (1994), "Encompassing Univariate Models in Multivariate Time Series": A Case Study", *Journal of Econometrics*, 61, 197-223.

Maravall, A. and Pierce, D.A. (1983), "Preliminary-Data Error and Monetary Aggregate Targeting", *Journal of Business and Economics Statistics*, 1,3,179-186.

Maravall, A. and Pierce, D.A. (1986), "The Transmission of Data Noise into Policy Noise in U.S. Monetary Control", *Econometrica*, 54, 4, 961-979.

Maravall, A. and Pierce, D.A. (1987), "A prototypical Seasonal Adjustment Model", *Journal of Time Series Analysis*, 8, 177-193.

Maravall, A. and Planas, C. (1996), "Estimation Errors and the Specification of Unobserved Component Models", forthcoming in *Proceedings of the 1996 ASA Meeting,*

Business Economics and Statistics Section.

Nelson, C.R. and Plosser, G.J. (1982), "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10, 139-162.

Nerlove, M. (1964), "Spectral Analysis of Seasonal Adjustment Procedures", *Econometrica*, 32, 241-286.

Nerlove, M., Grether, D.M, and Carvalho, J.L. (1979), *Analysis of Economic Time Series: A Synthesis"*, New York: Academic Press.

Ng, C.N. and Young, P.C. (1990), "Recursive Estimation and Forecasting of Nonstationary Time Series", *Journal of Forecasting*, 9, 173-204.

Osborn, D. (1993), "Moving Average Detrending of Integrated Processes", forthcoming in *Journal of Econometrics.*

Pagan A. (1975), "A Note on the Extraction of Components from Time Series", *Econometrica*, 43, 1, 163-168.

Persons, W. M. (1919), "Indices of Business Conditions", *Review of Economics and Statistics*, 1, 5-107.

Persons, W.M. (1923), "Correlation of Time Series", *Journal of the American Statistical Association*, 18, 713-726.

Pierce, D.A. (1978), "Seasonal Adjustment when Both Deterministic and Stochastic Seasonality are Present", in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census, 242-269.

Pierce, D.A. (1979), "Signal Extraction Eerror in Nonstationary Time Series", *Annals of Statistics*, 7, 1303-1320.

Pierce, D.A. (1980), "Data Revisions in Moving Average Seasonal Adjustment procedures", *Journal of Econometrics*, 14, 1, 95-114.

Pierce, D.A., and MacKenzie, S.K. (1987), "On Concurrent Seasonal Adjustment", *Journal of the American Statistical Association*, 82, 720-732.

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root", *Econometrica*, 55, 277-301.

Phillips, P.C.B. and Perron, P. (1988), "Testing for Unit Roots in Time Series Regression", *Biometrika*, 75, 335-346.

Planas, C. (1995) "Estimation Error in Unobserved Component Models", Ph.D. Dissertation, Department of Economics, European University Institute, Florence, Italy.

Planas (1996) "Linear Signal Extraction with Intervention Techniques in Non Linear Time Series", mimeo, Eurostat, Luxembourg.

Priestley, M.B. (1981) *Spectral Analysis and Time Series*, New York: Academic Press.

Robb, R.A. (1929), "The Variate Difference Method of Seasonal Variation", *Journal of the American Statistical Association*, 24, 250-257.

Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica*, 39, 577-592.

Shiskin, J., Young, A.H. and Musgrave, J.C. (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program", Technical Paper 15, Washington, D.C.: Bureau of the Census.

Slutsky, E. (1937), "The Summation of Random Causes as a Source of Cyclical Processes", *Econometrica*, 5, 105-146.

Sneek, M. (1984), *Modelling Procedures for Economic Time Series*, Amsterdam: Free University Press.

Stock, J.H. and Watson, M.W. (1988), "Variable Trends in Economic Time Series", *Journal of Economic Perspectives*, 2, 147-174.

Stock, J.H. and Watson, M.W. (1989), "New Indexes of Leading and Coincident Economic Indicators", National Bureau of Economic Research, *Macroeconomic Annual*, 351-394.

Stock, J.H. and Watson, M.W. (1991), "A Probability Model for Predicting Re-

cessions with Leading Indicators", in *Leading Economic Indicators: New Approaches and Forecasting Records*, ed. Lahiri, K., and Moore, G.H., Cambridge: Cambridge University Press.

Stock, J.H., and Watson, M.W. (1993), " A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Performances", in *New Business Cycle Indicators and Forecasting*, ed. Stock, J.H. and Watson, M.W., Chicago: Chicago University Press.

Tsay, R. (1984) "Regression Models with Time Series Errors", *Journal of the American Statistical Association*, 79, 118-124.

Walker, A.M. (1962), "Large Sample Estimation of Parameters for Autoregressive Processes with Moving Average Residuals", *Biometrika*, 49,117-131.

Wallis, K.F. (1974), "Seasonal Adjustment and Relations between Variables", *Journal of the American Statistical Association*, 69, 18-31.

Wallis, K.F.(1982), "Seasonal Adjustment and Revisions of Current Data: Linear Filters for the X11 Method", *Journal of the Royal Statistical Society*, Ser. A, 145, 74-85.

Watson, M.W. (1986), "Univariate Detrending Methods with Stochastic Trends", *Journal of the Monetary Economics*, 18,49-75.

Watson, M. W. (1987), "Uncertaincy in model-based-seasonal adjustment procedures and construction of minimax filters", *Journal of the American Statistical Association*, 82, 398, 395-408.

Weber, A.A. (1992), "The Role of Policymakers' Reputation in the EMS Disinflations. An Empirical Evaluation", *European Economic Review*, 36, 1473-1492.

Whittle, P. (1952), "Tests of Fits in Time Series", *Biometrika*, 39, 309-318.

Whittle P. (1963), *Prediction and Regulation using Least-Square Methods*, London: English Universities Press.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series*, Uppsala: Almqvist and Wicksell,

Young, A.H. (1968), "Linear Approximation to the Census and BLS Seasonal Adjustment Methods", *Journal of the Royal Statistical Society*, Ser. A, 145, 74-85.

Yule, G.U. (1927), "On a Method of Investigating Periodicities in Disturbed Series, With Special Reference to Wolfer's Sunspot Numbers", *Philosophical Transactions of the Royal Society*, Ser. A, 226, 267-298.