NOTES ON TIME SERIE ANALYSIS, ARIMA MODELS AND SIGNAL EXTRACTION

Regina Kaiser and
Agustín Maravall

00-64

Universidad Carlos III de Madrid

NOTES ON TIME SERIE ANALYSIS, ARIMA MODELS AND SIGNAL EXTRACTION
Regina Kaiser and Agustín Maravall*

Abstract _____

Present practice in applied time series work, mostly at economic policy or data producing agencies, relies heavily on using moving average filters to estimate unobserved components (or signals) in time series, such as the seasonally adjusted series, the trend, or the cycle. The purpose of the present paper is to provide an informal introduction to the time series analysis tools and concepts required by the user or analyst to understand the basic methodology behind the application of filters. The paper is aimed at economists, statisticians, and analysts in general, that do applied work in the field, but have not had an advanced course in applied time series analysis. Although the presentation is informal, we hope that careful reading of the paper will provide them with an important tool to understand and improve their work, in an autonomous manner. Emphasis is put on the model-based approach, although much of the material applies to ad-hoc filtering. The basic structure consists of modelling the series as a linear stochastic process, and estimating the components by means of "signal extraction", i.e., by optimal estimation of well-defined components.

Keywords: Times series analysis; ARIMA models; Signal extraction.

*Kaiser, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Getafe, Madrid, Spain. Ph: 34-1-624.96.24, Fax: 34-1-624.98.49, e-mail: kaiser@est-econ.uc3m.es; Maravall, Servicio de Estudios- Banco de España, Alcalá 50 28014 Madrid, e-mail: maravall@bde.es The first author acknowledges support by the Spanish grant PB95-0299 of CICYT.

Contents

1 Introduction

Present practice in applied time series work, mostly at economic policy or data producing agencies, relies heavily on using moving average filters to estimate unobserved components (or signals) in time series. Within the "ad-hoc" filter-design approach, well known examples are the X11 filter for seasonal adjustment, and the Hodrick-Prescott filter (HP) filter to estimate business cycles; see Shiskin et al (1967), and Hodrick and Prescott (1980). Within the "model-based" approach, whereby the filters are derived from statistical models, well known examples are the filters provided by programs STAMP and SEATS; see Koopman et al. (1996), and Gomez and Maravall (1996). (The program X12ARIMA can be seen as a move from ad-hoc filtering towards a partially model-based approach; see Findley et al., 1998). The purpose of the present paper is to provide an informal introduction to the time series analysis tools and concepts required by the user or analyst to understand the basic methodology behind the application of filters. The paper is aimed at economists, statisticians, and analysts in general, that do applied work in the field, but have not had an advanced course in applied time series analysis. Although the presentation is informal, we hope that careful reading of the paper will provide them with an important tool to understand and improve their work, in an autonomous manner. Emphasis is put on the model-based approach, although much of the material applies to the ad-hoc filtering case (in fact, most ad-hoc filters can be seen -at least to a close approximation- as particular cases of the model-based approach.) The basic structure consists of modelling the series as a linear stochastic process, and estimating the component by means of "signal extraction", i.e., by optimal estimation of well-defined components.

A previous word of caution should be said. The standard filtering procedure to estimate business cycles may require some prior corrections to the series, given that otherwise the results can be strongly distorted. An important example is outlier correction, as well as the correction for special effects that can have many different causes (trading day, easter, or holiday effect, legal changes, modifications in the statistical measurement procedure, etc.). This "preadjustement" of the series shall be briefly described in Section 3.3, where references for its methodology and its application in practice will be provided, that also cover the case in which observations are missing. For the rest of the book, we shall assume that the series either has already been preadjusted, or that no preadjustment is needed.

Further, although the discussion and the approach are also valid for other frequencies of observation, in order to simplify, we shall concentrate on quarterly series.

2 Brief review of applied time series analysis

2.1 Some basic concepts

The very basic intuition behind the concept of cyclical or seasonal variation leads to the idea of decomposing a series into "unobserved components", mostly defined by the frequency of the associated variation. If x_t denotes the observed series, the simplest formulation could be

$$x_t = \sum_j x_{jt} + u_t \tag{2.1}$$

where the variables x_{jt} denote the unobserved components, and u_t a residual effect (often referred to as the "irregular component"). In the early days, the components were often specified to follow deterministic models that could be estimated by simple regression. We shall follow the convention: a Deterministic Model denotes a model that yields forecasts with zero error when the model parameters are known. Stochastic Models will provide forecasts with non-zero random errors even when the parameters are known. For example, a deterministic trend component (p_t) could be specified as the linear trend

$$p_t = a + bt, \tag{2.2}$$

and the seasonal component (s_t) could be modelled with dummy variables, as in

$$s_t = \sum_j c_j d_{jt}, \tag{2.3}$$

where $d_{jt} = 1$ when t corresponds to the jth period of the year, and $d_{jt} = 0$ otherwise. An equivalent formulation can be expressed in terms of deterministic sine-cosine functions.

Gradual realization that seasonality evolves in time (an obvious example is the weather, one of the basic causes of seasonality) lead to changes in the estimation procedure. It was found that linear filters could reproduce the moving features of a trend or a seasonal component. A Linear Filter will simply denote a linear combination of the series x_t, as in

$$y_t = c_{-k_1} x_{t-k_1} + \ldots + c_{-1} x_{t-1} + c_0 x_t + c_1 x_{t+1} + \ldots + c_{k_2} x_{t+k_2}, \tag{2.4}$$

and, in so far as y_t is then some sort of moving average of successive stretches of x_t, we shall also use the expression Moving Average (MA) filter. The weights c_j could be found in such a way as to capture the relevant variation associated with the particular component of interest. Thus a filter for the trend would capture the variation associated with the long-term movement of the series, and a filter for a seasonal component would capture variation of a seasonal

4

nature. A filter designed in this way, with an "a priori" choice of the weights, is an "ad-hoc" fixed filter, in the sense that it is independent of the particular series to which it is being applied. Both, the HP and the X11 filters can be seen as "ad-hoc" fixed MA filters (although, strictly speaking, the coefficients as we shall see later, will not be constant.)

Over time, however, application of "ad-hoc" filtering has evidenced some serious limitations. An important one is the fact that, due to its fixed character, spurious results can be obtained, and for some series the component may be overestimated, while for other series, it may be underestimated. To overcome this limitation, and in the context of seasonal adjustment, an alternative approach was suggested (around 1980) whereby the filter adapted to the particular structure of the series, as captured by its ARIMA model. The approach, known as the ARIMA-model-based (AMB) approach, consists of two steps. First, an ARIMA model is obtained for the observed series. Second, signal extraction techniques are used to estimate the components with filters that are, in some well-defined way, optimal.

2.2 Stochastic processes and stationarity

The following summary is an informal review, aimed at providing some basic tools for the posterior analysis, as well as some intuition for their usefulness. More complete treatments of time series analysis are provided in many textbooks; some helpful references are Box and Jenkins (1970), Brockwell and Davis (1987), Granger and Newbold (1986), Harvey (1993), and Mills (1990).

The starting point is the concept of a Stochastic Process. For our purposes, a stochastic process is a real-valued random variable z_t, that follows a distribution $f_t(z_t)$, where t denotes an integer that indexes the period. The T-dimensional variable $(z_{t_1}, z_{t_2}, \ldots, z_{t_T})$ will have a joint distribution that depends on $(t_1, t_2, \ldots, t_T)$. A Time Series $[z_{t_1}, z_{t_2}, \ldots, z_{t_T}]$ will denote a particular realization of the stochastic process. Thus, for each distribution f_t, there is only one observation available. Not much can be learned from this, and more structure and more assumptions need to be added. To simplify notation, we shall consider the joint distribution of $(z_1, z_2, \ldots, z_t)$, for which a time series is available when $t \leq T$.

From an applied perspective, the two most important added assumptions are

Assumption A: The process is stationary;

Assumption B: The joint distribution of $(z_1, z_2, \ldots, z_t)$ is a multivariate normal distribution.

Assumption A implies the following basic condition. For any value of t,

$$f(z_1, z_2, \ldots, z_t) = f(z_{1+k}, z_{2+k}, \ldots, z_{t+k}), \qquad (2.5)$$

where k is a integer; that is, the joint distribution remains unchanged if all time periods are moved a constant number of periods. In particular, letting $t = 1$, for the marginal distribution it has to be that

$$f_t(z_t) = f(z_t)$$

for every t, and hence the marginal distribution remains constant. This implies

$$E z_t = \mu_z; \quad V z_t = V_z \qquad (2.6)$$

where E and V denote the expectation and the variance operators, and μ_z and V_z are constants that do not depend on t.

In practice, thus, stationarity implies a constant mean level and bounded deviations from it. It is a very strong requirement and few actual economic series will satisfy it. Its usefulness comes from the fact that relatively simple transformations of the nonstationary series will render it stationary. For quarterly economic series, it is usually the case that constant variance can be achieved through the log/level transformation combined with proper outlier correction, and constant mean can be achieved by differencing.

The log transformation is "grosso modo" appropriate when the amplitude of the series oscillations increases with the level of the series. As for outliers, several possible types should be considered, the most popular ones being the additive outlier (i.e., a single spike), the level shift (i.e., a step variable), and the transitory change (i.e., an effect that gradually disappears). Formal testing for the log/level transformation and for outliers are available, as well as easy-to-apply automatic procedures for doing it (see, for example, Gómez and Maravall, 2000a). In Section 3.3 we shall come back to this issue; we center our attention now on achieving stationarity in mean.

2.3 Differencing

Denote by B the backward operator, such that

$$B^j z_t = z_{t-j} \qquad (j = 0, 1, 2, \ldots),$$

and let x_t denote a quarterly observed series. We shall use the operators:

- Regular difference: $\nabla = 1 - B$.

- Seasonal difference: $\nabla_4 = 1 - B^4$.

- Annual aggregation: $S = 1 + B + B^2 + B^3$.

6

Thus $\nabla x_t = x_t - x_{t-1}$, $\nabla_4 x_t = x_t - x_{t-4}$, and $Sx_t = x_t + x_{t-1} + x_{t-2} + x_{t-3}$. It is immediately seen that the 3 operators satisfy the identity

$$\nabla_4 = \nabla S \tag{2.7}$$

If x_t is a deterministic linear trend, as in $x_t = a + bt$, then

$$\nabla x_t = b; \tag{2.8}$$
$$\nabla^2 x_t = 0; \tag{2.9}$$

where $\nabla^2 x_t = \nabla(\nabla x_t)$. In general, it can easily be seen that ∇^d will reduce a polynomial of degree d to a constant. Obviously, $\nabla_4 x_t$ will also cancel a constant (or reduce the linear trend to a constant); but it will also cancel other deterministic periodic functions, such as for example, one that repeats itself every 4 quarters. To find the set of functions that are cancelled with the transformations $\nabla_4 x_t$, we have to find the solution of the homogenous difference equation

$$\nabla_4 x_t = (1 - B^4)x_t = x_t - x_{t-4} = 0, \tag{2.10}$$

with characteristic equation $r^4 - 1 = 0$. The solution is given by

$$r = \sqrt[4]{1},$$

that is, the four roots of the unit circle displayed in Figure 2.1. The four roots are

$$r_1 = 1, \quad r_2 = -1, \quad r_3 = i, \quad r_4 = -i. \tag{2.11}$$

The first two roots are real and the last two are complex conjugates, with modulus 1 and, as seen in the figure, frequency $\omega = \pi/2$ (frequencies will always be expressed in radians). Complex conjugate roots generate periodic movements of the type

$$r_t = A^t \cos(\omega t + B) \tag{2.12}$$

where A denotes the amplitude, B denotes the phase (the angle at t=0) and ω the frequency (the number of full circles that are completed in one unit of time.) The period of function (2.12), to be denoted τ, is the number of units of time it takes for a full circle to be completed, and is related to the frequency ω by the expression

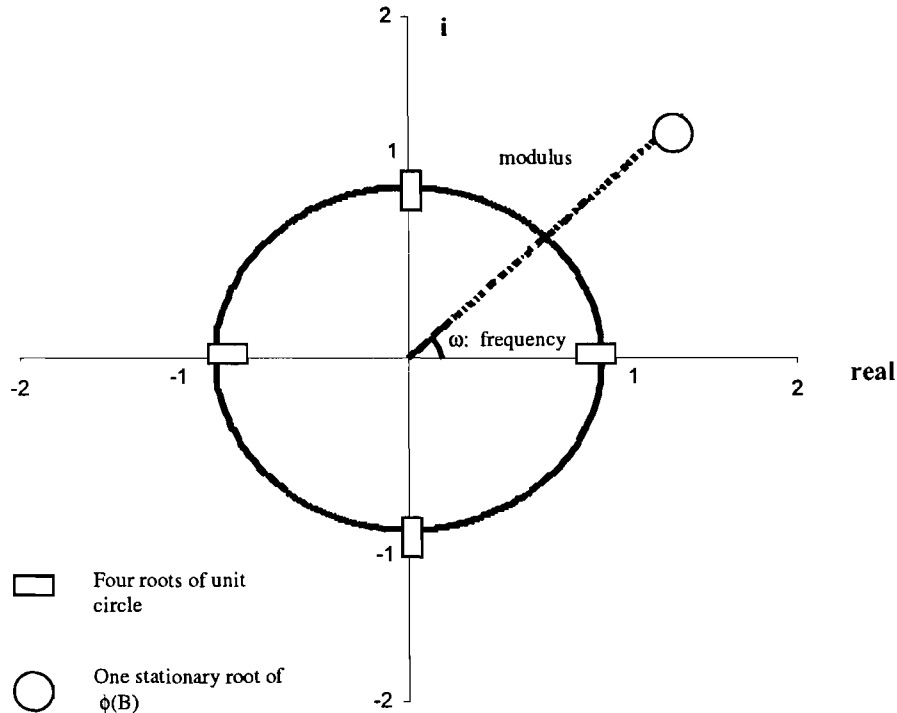$$\tau = \frac{2\pi}{\omega}. \tag{2.13}$$

Fig 2.1 : Roots of Unit Circle

Figure 2.2a illustrates a periodic movement of the type (2.12), with A=1, B=0, and $\omega = \pi/4$. From (2.11), the general solution of $\nabla_4 x_t = 0$ can be expressed as (see for example, Goldberg, 1967)

$$x_t = c_0 + c_1 \cos\left(\frac{\pi}{2}t + d_1\right) + c_2(-1)^t,$$

where c_0, c_1, c_2 and d_1 are constants to be determined from the starting conditions. Realizing that $cos\pi = -1$, the previous expression can also be written as

$$x_t = c_0 + \sum_{j=1}^{2} c_j \cos\left(j\frac{\pi}{2}t + d_j\right),\qquad(2.14)$$

with $d_2 = 0$. Considering (2.13), the first term in the sum of (2.14) will be associated with a period of $\tau = 4$ quarters and will represent thus a seasonal component with a once-a-year frequency; the second term has a period of $\tau = 2$ quarters, and hence will represent a seasonal component with a twice-a-year frequency. The two components are displayed in Figure 2.2b and c. Noticing that the characteristic equation can be rewritten as $(B^{-1})^4 - 1 = 0$, (2.11)

8

implies the factorization

$$\nabla_4 = (1 - B)(1 + B)(1 + B^2).$$

The factor (1-B) is associated with the constant and the zero frequency, the factor (1+B) with the twice-a-year seasonality with frequency $\omega = \pi$, and the factor $(1+B^2)$ with the once-a-year seasonality with frequency $\omega = \pi/2$. The product of these last two factors yields the annual aggregation operator S, in agreement with expression (2.7). Hence the transformation Sx_t will remove seasonal nonstationarity in x_t.

For the most-often-found case in which stationarity is achieved through the differencing $\nabla\nabla_4$, the factorization

$$\nabla\nabla_4 = \nabla^2 S$$

directly shows that the solution to

$$\nabla\nabla_4 x_t = 0$$

will be of the type:

$$x_t = a + bt + \sum_{j=1}^{2} c_j \left[cos(j\frac{\pi}{2}t) + d_j \right], \tag{2.15}$$

with $d_2 = 0$. Thus the differencing will remove the same cosine (seasonal) functions as before, plus the local linear trend (a+bt). For the case $\nabla^2\nabla_4$, the factorization $\nabla^3 S$ shows that the cancelled trend will now be a second order polynomial in t, the rest remaining unchanged. For quarterly series, higher order differencing is never encountered in practice.
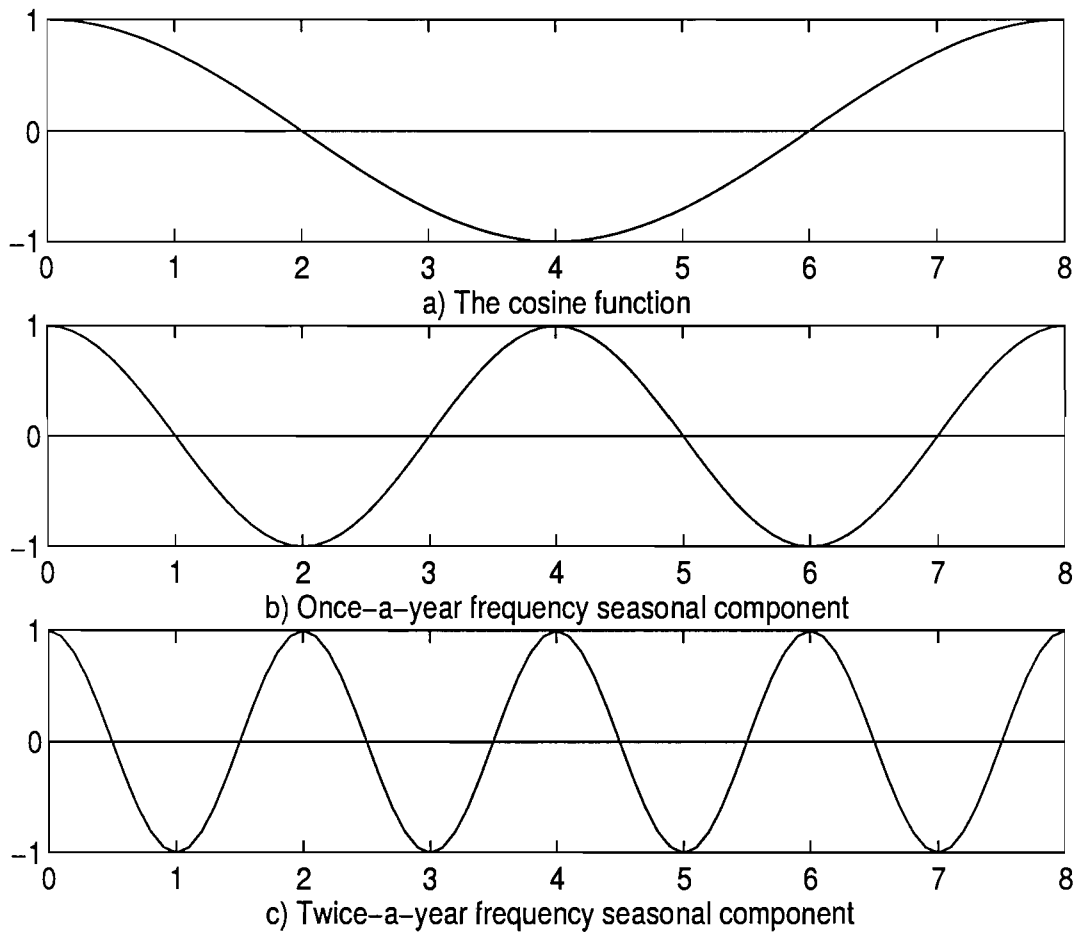
A final and important remark:

- Let D denote, in general, the complete differencing applied to the series x_t so as to achieve stationarity. When specifying the ARIMA model for x_t, we shall not be stating that $Dx_t = 0$ (as, for example, in (2.9),) but that

$$Dx_t = z_t,$$

where z_t is a zero-mean, stationary stochastic process with relatively small variance. Thus every period the solution of $Dx_t = 0$ will be perturbed by the stochastic input z_t (see Box and Jenkins, 1970, Appendix A.4.1). In terms of expression (2.15), what this perturbation implies is that the a,b,c and d coefficients will not be constant but will instead depend on time. This gradual evolution of the coefficients provides the model with an adaptive behavior that will be associated with the "moving"features of the trend and seasonal components.

9

Figure 2.2

a) The cosine function

b) Once-a-year frequency seasonal component

c) Twice-a-year frequency seasonal component

2.4 Linear stationary process, Wold representation, and autocorrelation function

Following the previous notation, if x_t denotes the observed variable and $z_t = Dx_t$ its stationary transformation, under assumptions A and B, the variable $(z_1, z_2, \ldots, z_T)$ will have a proper multivariate normal distribution. One important property of this distribution is that the expectation of some (unobserved) variable linearly related to z_t, conditional on $(z_1, z_2, \ldots, z_T)$, will be a linear function of $z_1, z_2, \ldots, z_T$. Thus conditional expectations will directly provide linear filters. An additional important property is that, because the first two moments fully characterize the distribution, stationarity in mean and variance will imply stationarity of the process. In particular, stationarity will be implied by the constant mean and variance condition (2.6), plus the condition that

$$Cov(z_t, z_{t-k}) = \gamma_k,$$

for $k = 0, \pm 1, \pm 2, \ldots$ Hence the covariance between z_t and z_{t-k} should depend on their relative distance k, not on the value of t. Therefore,

$$(z_1, z_2, \ldots, z_T) \sim N(\mu, \Sigma),$$

where μ is a vector of constant means, and Σ is the variance-covariance matrix

$$\Sigma = \begin{bmatrix} V_z & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ & V_z & \gamma_1 & \cdots & \gamma_{T-2} \\ & & \cdots & \cdots & \\ & & & V_z & \gamma_1 \\ & & & & V_z \end{bmatrix}, \qquad (V_z = \gamma_0),$$

a positive definite symmetric matrix. Let F denote the forward operator, $F = B^{-1}$, such that

$$F^j z_t = z_{t+j}, \quad (j = 0, 1, 2, \ldots),$$

a more parsimonious representation of the 2nd-order moments of the stationary process z_t is given by the Autocovariance Generating Function (AGF)

$$\gamma(B, F) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (B^j + F^j). \tag{2.16}$$

To transform this function into a scale-free function, we divide by the variance γ_0, and obtain the Autocorrelation Generating Function (ACF),

$$\rho(B, F) = \rho_0 + \sum_{j=1}^{\infty} \rho_j (B^j + F^j). \tag{2.17}$$

where $\rho_j = \gamma_j / \gamma_0$. If the following conditions on the AGF:

1. $\rho_0 = 1$;

2. $\rho_j = \rho_{-j}$;

3. $|\rho_j| < 1$ for $j \neq 0$;

4. $\rho_j \to 0$ as $j \to \infty$;

5. $\sum_{j=0}^{\infty} |\rho_k| < \infty$,

are satisfied, then a zero-mean, finite variance, normally distributed process is stationary. Further, under the normality assumption, a complete realization of the stochastic process will be fully characterized by μ_z, V_z and $\rho(B, F)$.

When $\rho_j = 0$ for all $j \neq 0$, the process will be denoted a White Noise process. Therefore, a white noise process is a sequence of normally identically independently distributed random variables.

The AGF (or ACF) is the basic tool in the so-called "Time Domain Analysis" of a time series. The first statistics that we shall compute for a time series $[z_1, \ldots, z_T]$ will be estimates of the autocovariances and autocorrelations using the standard sample estimates

$$\bar{z} = T^{-1} \sum_{t=1}^{T} z_t; \qquad \hat{\gamma}_k = T^{-1} \sum_{t=k+1}^{T} (z_t - \bar{z})(z_{t-k} - \bar{z}); \qquad \hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0.$$

Next, a look at the sample ACF (SACF) will give an idea of the lag dependence in the series: large autocorrelation for low lags will point towards large inertia; large autocorrelation for seasonal lags will, of course, indicate the presence of seasonality. One word of caution should be nevertheless made: the dependence of the autocorrelation estimators on the same time series can induce important spurious correlation between them. These correlations can have serious distorting effects on the visual aspect of the SACF, which may fail to damp out according to expectations (see Box and Jenkins, 1970, section 2.1). Figure 2.3a exhibits the ACF of a quarterly stationary process; figure 2.3b displays the SACF obtained with a sample of 100 observations. As a consequence, care should be taken not to "over-read" SACFs, ignoring large-lag autocorrelations, and focussing only on its most salient features.

To start the modelling procedure, a general result on linear time series processes will provide us with an analytical representation of the process that will prove very useful. This is the so-called Wold (or Fundamental) representation. We present it next.

12

Figure 2.3

a)Theoretical ACF

b)Sample ACF
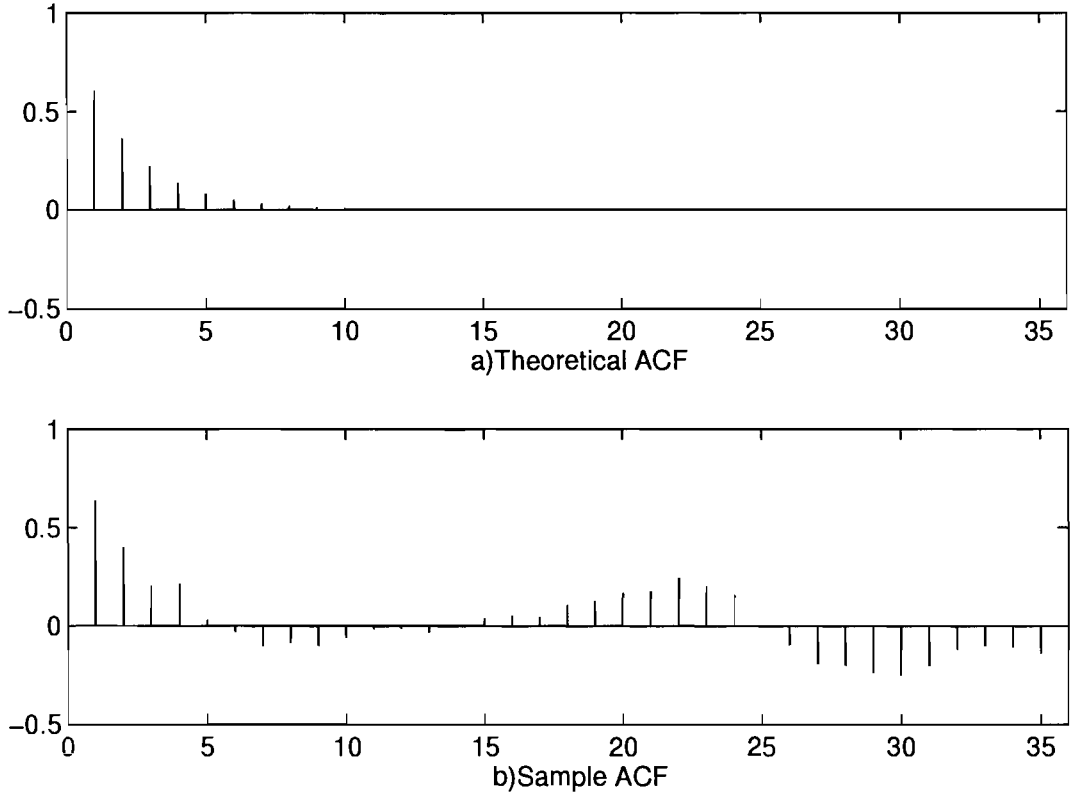
Let z_t denote a linear stationary stochastic process with no deterministic component, then z_t can be expressed as the one-sided moving average

$$z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \ldots =$$
$$= \sum_{j=0}^{\infty} \psi_j a_{t-j} = \Psi(B)a_t,$$
$$\Psi(B) = \sum_{j=0}^{\infty} \psi_j B^j, \qquad (\psi_0 = 1) \tag{2.18}$$

where a_t is a white noise process with zero mean and constant variance V_a, and $\Psi(B)$ is such that

1. $\psi_j \to 0$ as $j \to \infty$;

2. $\sum_{j=0}^{\infty} |\psi_j| < \infty$;

the last condition reflecting a sufficient condition for convergence of the polynomial $\Psi(B)$. Given the ψ_j-coefficients, a_t represents the one-period ahead

13

forecast error of z_t, that is

$$a_t = z_t - \hat{z}_{t|t-1},$$

where $\hat{z}_{t|t-1}$ is the forecast of z_t made at period t-1. Since a_t represents what is new in z_t, that is, what is not contained in its past $[z_{t-1}, z_{t-2}, z_{t-3}, \ldots]$, it will be referred to as the Innovation of the process. The representation of z_t in terms of its innovations, given by (2.18), is unique, and is usually referred to as the Wold representation.

A useful result is the following: If $\gamma(B, F)$ represents the AGF of the process z_t, then

$$\gamma(B, F) = \Psi(B)\Psi(F)V_a. \tag{2.19}$$

In particular, for the variance,

$$V_z = (1 + \psi_1^2 + \psi_2^2 + \ldots)V_a. \tag{2.20}$$

2.5   The spectrum

The spectrum is the basic tool in the so-called "Frequency Domain Approach" to time series analysis. It represents an alternative way to look and interpret the information contained in the second-order moments of the series. The frequency approach is particularly convenient for analyzing unobserved components, such as trends, cycles, or seasonality. Our aim is not to present a complete and rigorous description, but to provide some intuition and basic understanding, that will permit us to use it properly for our purposes. (Two good references for a general presentation are Jenkins and Watts, 1968, and Grenander and Rosenblatt, 1957.)

Consider, first, a time series (i.e., a partial realization of a stochastic process) given by $z_1, z_2, \ldots, z_T$. To simplify the discussion, assume the process has zero mean and that T is even, so that we can write T=2q. In the same way that, as is well known, the T values of z_t can be exactly duplicated ("explained") by a polynomial of order (T-1), they can also be exactly reproduced as the sum of T/2 cosine functions of the type (2.12); this result provides in fact the basis of Fourier analysis.

Figure 2.4a shows, for example, the quarterly time series of 10 observations generated by the five cosine functions of figure 2.4b. To construct this set of functions, we start by defining the Fundamental Frequency $\omega = 2\pi/T$ (i.e., the frequency of one full circle completed in T periods) and its multiples (or Harmonics)
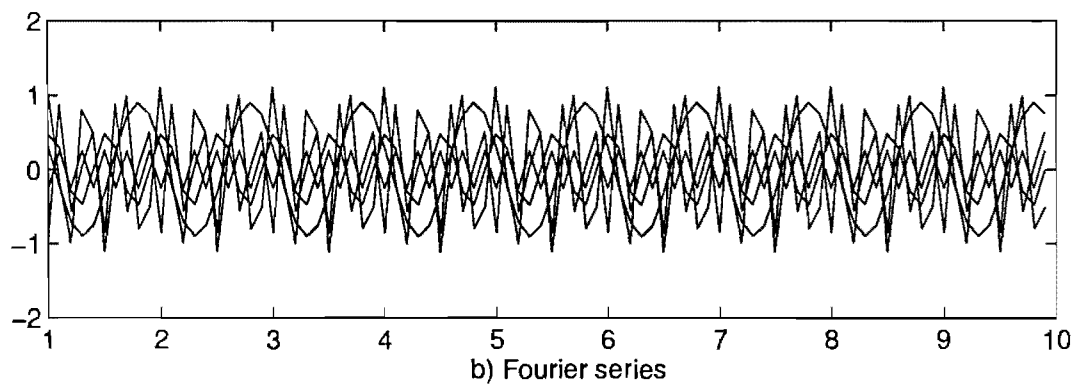
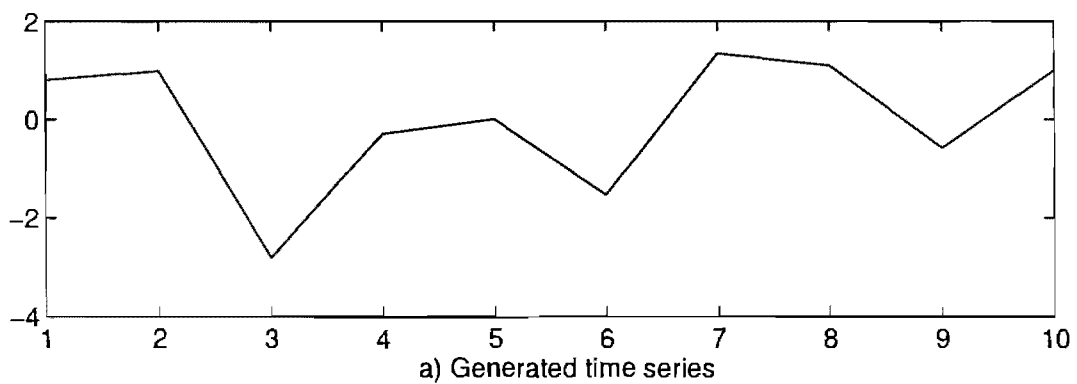$$\omega_j = (2\pi/T)j, \qquad j = 1, 2, \ldots, q.$$

14

Then, express (2.12) as

$$r_{jt} = a_j \cos \omega_j t + b_j \sin \omega_j t, \qquad (2.21)$$

and hence,

$$z_t = \sum_{j=1}^{q} r_{jt}. \qquad (2.22)$$

Figure 2.4



a) Generated time series

b) Fourier series

It is straightforward to check that a_j and b_j are related to the amplitude A_j by

$$A_j^2 = a_j^2 + b_j^2.$$

From (2.21) and (2.22), by plugging in the values of z_t, w_j, and t, a linear system of T equations is obtained in the unknowns a_j's and b_j's, $j = 1, 2, \ldots, q$; a total of T unknowns. Therefore, for each frequency ω_j, we obtain a square amplitude A_j^2. The plot of A_j^2 versus ω_j, $j = 1, \ldots, q$, is the Periodogram of the series.

15

As a consequence, we obtain a set of periodic functions with different frequencies and amplitudes. We can group the functions in intervals of frequency by summing the squared amplitudes of the functions that fall in the same interval. In this way we obtain an histogram of frequencies that shows the contribution of each interval of frequency to the series variation; an example is shown in Figure 2.5a. In the same way that a density function is the model counterpart of the usual histogram, the spectrum will be the model counterpart of the frequency histogram (properly standardized).

Figure 2.5

a) Histogram of frequencies

b) Power spectrum
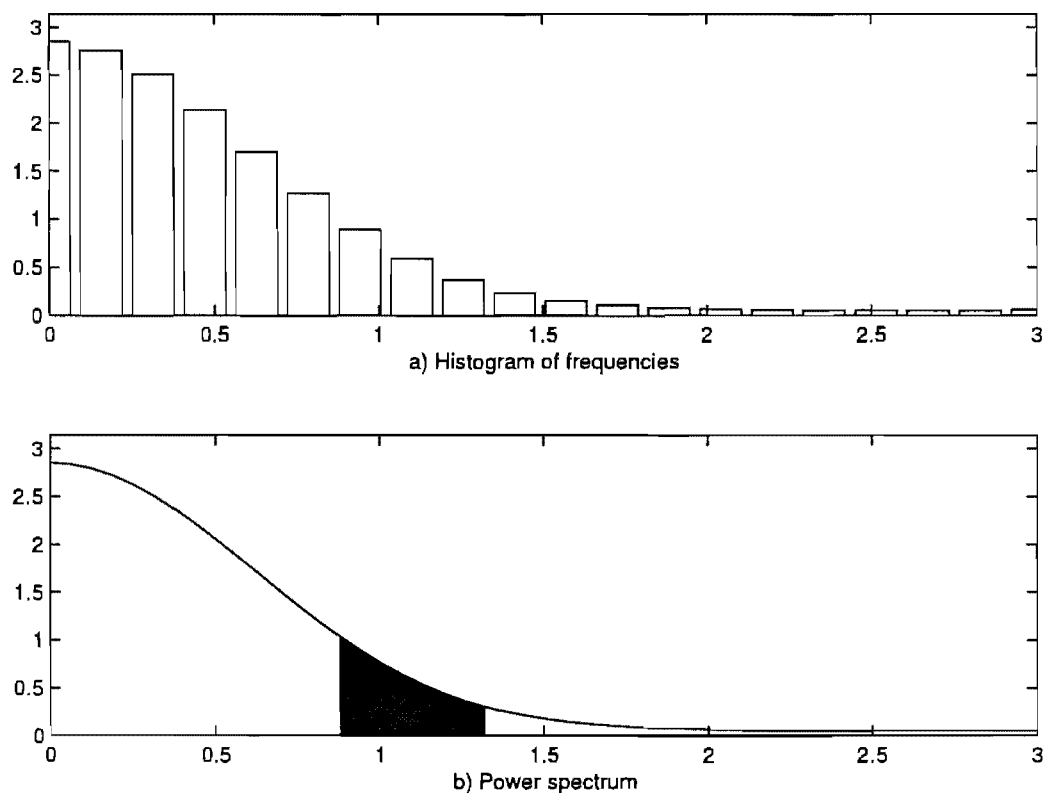
We can now let the interval $\Delta\omega_j$ go to zero, and the frequency histogram will become a continuous function, which is denoted the Sample Spectrum. The area over the differential $d\omega$ represents the contribution of the frequencies in $d\omega$ to the variation of the time series. An important result links the sample spectrum with the SACF (see Box and Jenkins, 1970, Appendix A.2.1). If

16

$H(\omega)$ denotes the sample spectrum, then it is proportional to

$$H(\omega) \propto \left(\hat{\gamma}_0 + 2 \sum_{t=1}^{T-1} \hat{\gamma}_j \cos \omega t \right), \qquad (2.23)$$

where $\hat{\gamma}_j$ denotes the lag-j autocovariance estimator.

The model equivalent of (2.23) provides precisely the definition of power spectrum. Consider the AGF of the stationary process z_t, given by

$$\gamma(B, F) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (B^j + F^j), \qquad (2.24)$$

where B is a complex number of unit modulus, which can be expressed as $e^{i\omega}$. Replacing B and F by their complex representation, (2.24) becomes the function

$$g(\omega) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (e^{-i\omega j} + e^{i\omega j}),$$

or, using the identity

$$e^{-i\omega j} + e^{i\omega j} = 2 \cos(j\omega),$$

and dividing by 2π, one obtains

$$g_1(\omega) = \frac{1}{2\pi} \left[\gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j \cos(j\omega) \right]. \qquad (2.25)$$

The move from (2.24) to (2.25) is the so-called Fourier cosine transform of the AGF $\gamma(B, F)$, and is denoted the Power Spectrum. Replacing the AGF by the ACF (i.e., dividing by the variance γ_0), we obtain the Spectral Density Function

$$g_1^*(\omega) = \frac{1}{2\pi} \left[1 + 2 \sum_{j=1}^{\infty} \rho_j \cos(j\omega) \right]. \qquad (2.26)$$

It is easily seen that $g_1(\omega)$ -or $g_1^*(\omega)$- are periodic functions, and hence the range of frequencies can be restricted to $(-\pi, \pi)$, or $(0, 2\pi)$. Moreover, given that the cosine function is symmetric around zero, we only need to consider the range $(0, \pi)$. It is worth mentioning that the sample spectrum (2.23), divided by 2π, is also the Fourier transform of the sample autocovariance function.

From (2.25), knowing the AGF of a process, the power spectrum is trivially obtained. Alternatively, knowledge of the power spectrum permits us to derive the AGF by means of the inverse Fourier transform, given by

$$\gamma_k = \int_{-\pi}^{\pi} g(\omega) \cos(\omega k) d\omega.$$

Thus, for k=0,

$$\gamma_0 = \int_{-\pi}^{\pi} g(\omega)d\omega, \tag{2.27}$$

which shows that the integral of the power spectrum is the variance of the process. Therefore, the area under the spectrum for the interval $d\omega$ is the contribution to the variance of the series that corresponds to the range of frequencies $d\omega$ (as in Figure 2.5b). Roughly, the power spectrum can be seen as a decomposition of the variance by frequency.

For the rest of the monograph, in order to simplify the notation, power spectra will be expressed in units of 2π, and, because of the symmetry condition, only the range $\omega \in [0, \pi]$ will be considered. We shall refer to this function simply as the Spectrum.

As an example, consider a process z_t, the output of the 2nd-order homogenous difference-equation (deterministic) model

$$z_t + .81z_{t-2} = 0 \tag{2.28}$$

The characteristic equation, $r^2 + .81 = 0$ yields the pair of complex conjugate numbers $r = \pm.9i$, situated in the imaginary axis, they will be associated thus with the frequency $\omega = \pi/2$ (see Figure 2.1). The process follows therefore the deterministic function

$$z_t = .9\cos\left(\frac{\pi}{2}t + \beta\right), \tag{2.29}$$

where we can set $\beta = -\pi/2$. The function (2.29) does not depend on ω and the movements of z_t are all associated with the single frequency $\omega = \pi/2$. This explains the isolated spike for that frequency in Figure 2.6a. To transform the previous model into a stochastic process, we perturb every period the equilibrium (2.28) with a white noise (0,1) variable a_t, so that it is replaced by the stochastic model

$$z_t + .81z_{t-2} = a_t, \text{ or } (1 + .81B^2)z_t = a_t. \tag{2.30}$$

From (2.30), the Wold representation (2.18) is immediately obtained as

$$z_t = \frac{a_t}{1 + .81B^2},$$

with

$$\Psi(B) = 1/(1 + .81B^2).$$

Using (2.19), the AGF of z_t can be obtained through

$$\gamma(B, F) = \frac{V_a}{(1 + .81B^2)(1 + .81F^2)} =$$

18

$$= \frac{V_a}{1.656 + .81(B^2 + F^2)}$$

Replacing $(B^2 + F^2)$ by $2\cos 2\omega$, the spectrum is found to be equal to

$$g(\omega) = \frac{V_a}{1.656 + 1.62\cos 2\omega}; \qquad 0 \le \omega \le \pi.$$

The spike of the previous case, as seen in Figure 2.6b, has now become a hill. If we increase the variance of the stochastic input a_t, as shown in part c of the figure, the width of the hill (i.e., the dispersion of ω around $\pi/2$) increases. Figure 2.7 compares the type of movements generated in the 3 cases. As the variance of the stochastic input becomes larger, the component becomes less stable and more "moving".
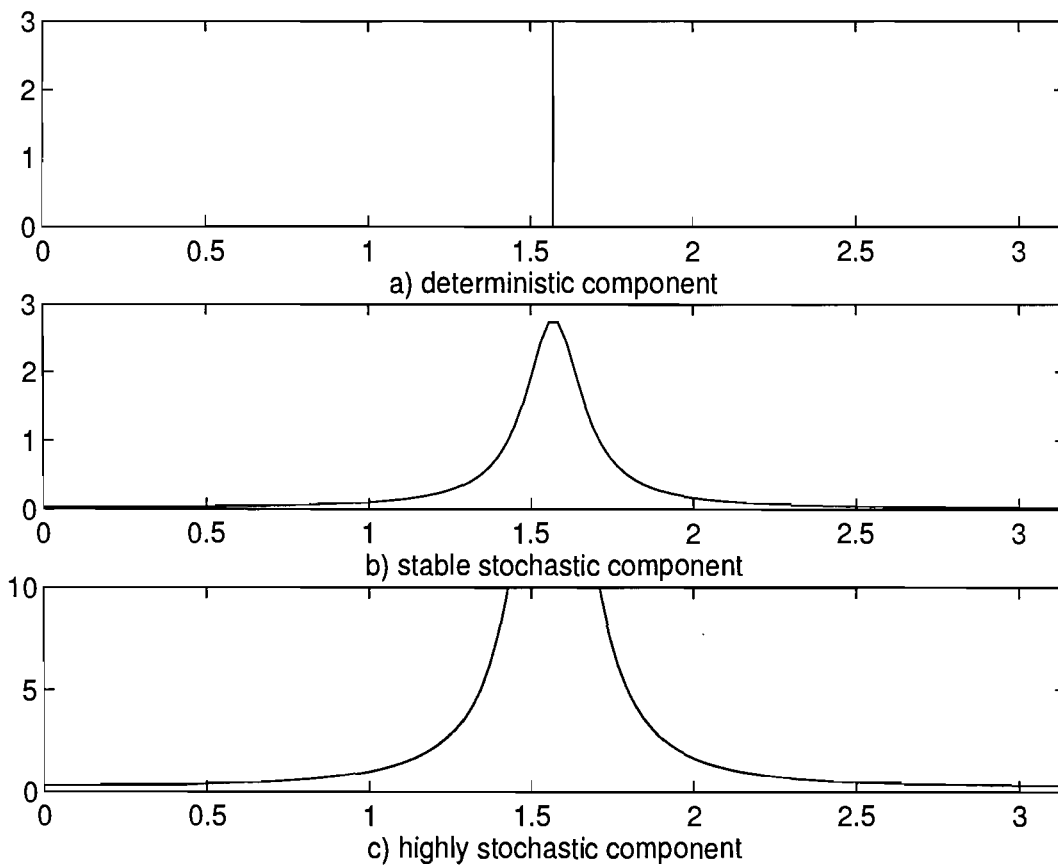
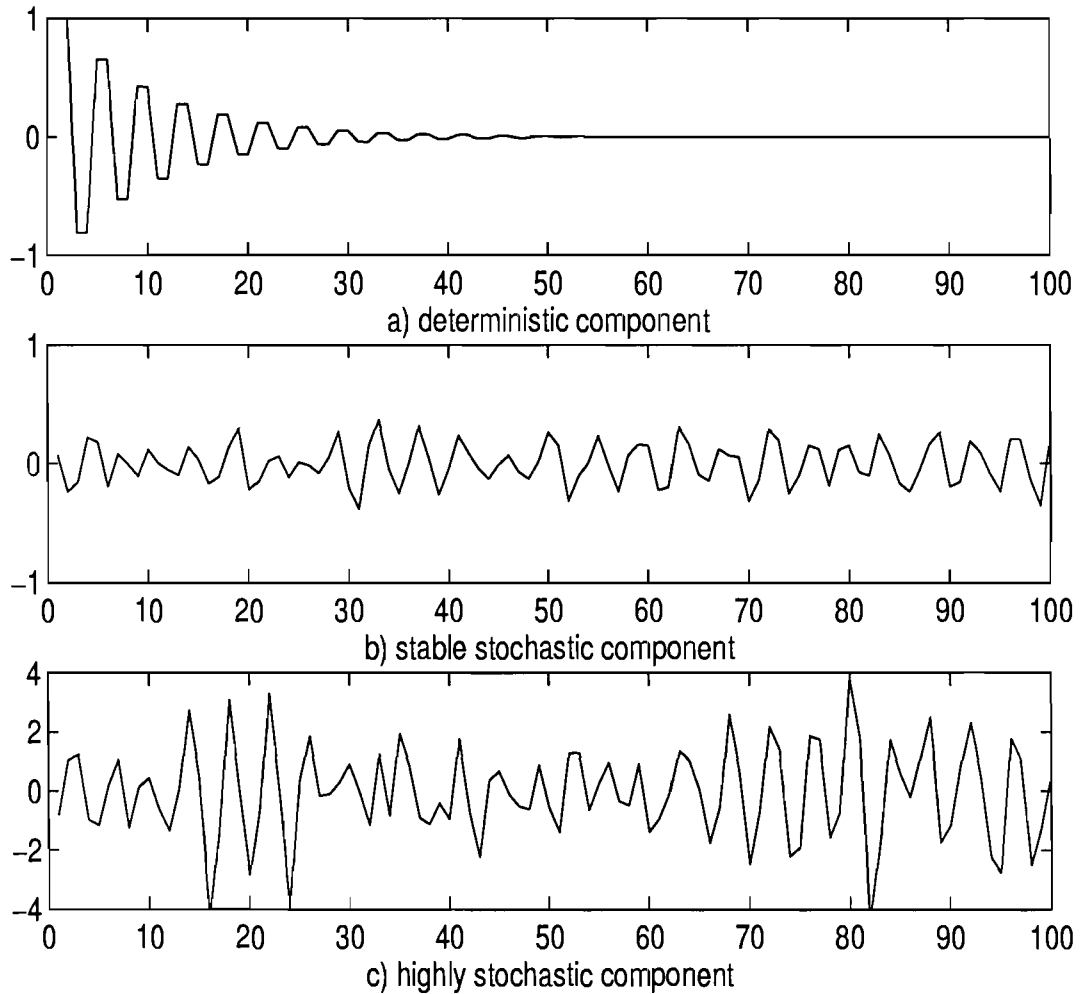## Figure 2.6. Spectra of AR(2) process

a) deterministic component

b) stable stochastic component

c) highly stochastic component

Figure 2.7. Realization of AR(2) process

a) deterministic component

b) stable stochastic component

c) highly stochastic component

In summary, if a series contains an important component for a certain frequency ω_0, its spectrum should reveal a peak around that frequency. Given that a good definition of a trend is a cyclical component with period $\tau = \infty$, the spectral peak in this case should occur at the frequency $\omega = 0$.

To see some examples of spectra for some simple processes, we use the previous result that allows us to move from the Wold representation to the AGF, and from the AGF to the spectrum. The sequence is, in all cases,

$$\begin{aligned} z_t &= \Psi(B)a_t: &&\text{Wold representation ;} \\ \gamma(B,F) &= \Psi(B)\Psi(F)V_a: &&\text{AGF of } z_t \\ &= [\gamma_0 + \sum_j \gamma_j(B^j + F^j)]V_a; \end{aligned}$$

$g(\omega) = [\gamma_0 + 2\sum_j \gamma_j \cos j\omega]V_a$: spectrum.

1. *White noise process.* Then, $\gamma_j = 0$ for $j \neq 0$, and hence

$g(\omega) = $ constant (Figure 2.8a).

2. *Moving Average process of order 1*: MA(1)

$z_t = a_t + \theta_1 a_{t-1}$

$z_t = (1 + \theta_1 B)a_t$, hence $\Psi(B) = (1 + \theta_1 B)$; therefore

$$\begin{aligned} \gamma(B,F) &= \Psi(B)\Psi(F)V_a = (1 + \theta B)(1 + \theta F)V_a = \\ &= [1 + \theta^2 + \theta(B + F)]V_a, \end{aligned}$$

$g(\omega) = [1 + \theta^2 + 2\theta \cos \omega]V_a$

Figure 2.8b shows an example with $\theta < 0$).

3. *Autoregressive process of order 1*: AR(1)

$z_t + \phi_1 z_{t-1} = a_t$; or $(1 + \phi B)z_t = a_t$

$z_t = (1/(1 + \phi B)) a_t$, so that $\Psi(B) = 1/(1 + \phi B)$;

assuming $|\phi| < 1$, it is found that

$$\begin{aligned} \gamma(B,F) &= [(1 + \phi B)(1 + \phi F)]^{-1} V_a = \\ &= \left[1 + \phi^2 + \phi(B + F)\right]^{-1}V_a; \end{aligned}$$

$g(\omega) = [1 + \phi^2 + 2\phi \cos \omega]]^{-1}V_a$.

The case $\phi < 0$ is displayed in Figure 2.8c. The spectrum consists of a peak for $\omega = 0$ that decreases monotonically in the range $[0, \pi]$. Therefore, the AR(1) process in this case reveals a trend-type behavior.

Figure 2.8c also displays (dotted line) the case $\phi > 0$. The resulting spectrum is symmetric to the previous one around the frequency $\omega = \pi/2$, and,

consequently, displays a peak for $\omega = \pi$. The period associated with that peak is, according to (2.13), always 2. Therefore the AR(1) in this case reveals a cyclical behavior with period $\tau = 2$. If the data is monthly, this behavior corresponds to the six-times-a-year seasonal frequency; for a quarterly time series, to the twice-a-year seasonal frequency; for annual data, it would represent a two-year cycle effect.

4. *Autoregressive process of order 2*: AR(2)

$$z_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} = a_t \tag{2.31}$$

or:

$$(1 + \phi_1 B + \phi_2 B^2)z_t = a_t \tag{2.32}$$

Concentrating, as we did earlier, on the homogenous part of (2.31), the characteristic equation associated with that part is precisely the polynomial in B, with $B = r^{-1}$. Thus we can find the dominant behavior of z_t from the solution of $r^2 + \phi_1 r + \phi_2 = 0$. Two cases can happen:

(a) The two roots are real;

(b) The two roots are complex conjugates.

In case (a), if r_1 and r_2 are the two roots (we assume $|r_1|$ and $|r_2|$ are < 1), the polynomial can be factorized as $(1 - r_1 B)(1 - r_2 B)$, and each factor will produce the effect of an AR(1) process. Thus, if both r_1 and r_2 are > 0, the spectrum will display a peak for $\omega = 0$; if one is > 0 and the other < 0, the spectrum will have peaks for $\omega = 0$ and $\omega = \pi$; if both roots are < 0, the spectrum will have a peak for $\omega = \pi$.

In case (b), the complex conjugate roots will generate a cosine-type (cyclical) behavior. The modulus m and the frequency ω can be obtained from the model (2.31) through

$$m = \sqrt{\phi_2}; \quad \omega = \arccos\left(\frac{\phi_1}{2m}\right); \tag{2.33}$$

and the spectrum will display a peak for the frequency ω, as in Figure 2.8d.

In general, a useful way to look at the structure of an autoregressive process of order p, AR(p), a specification very popular in econometrics, is to factorize the full AR polynomial. Real roots will imply spectral peaks of the type 2.8c, while complex conjugate roots will produce peaks of the type 2.8d.
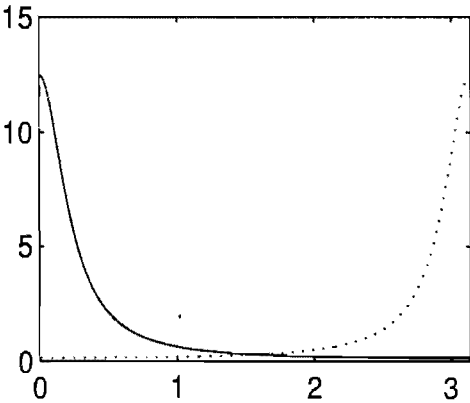
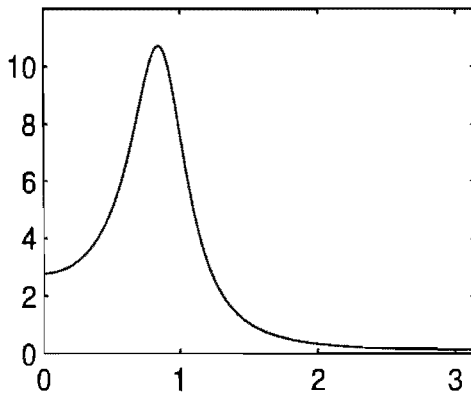## Figure 2.8. Examples of Spectra
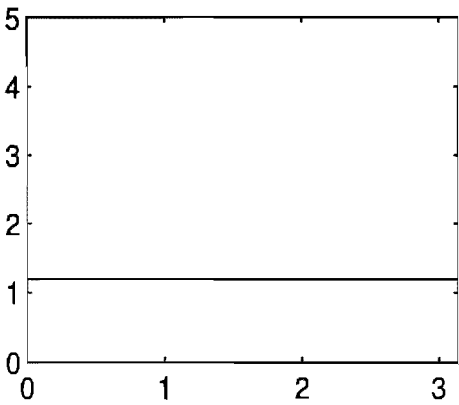
a) Spectrum of white noise
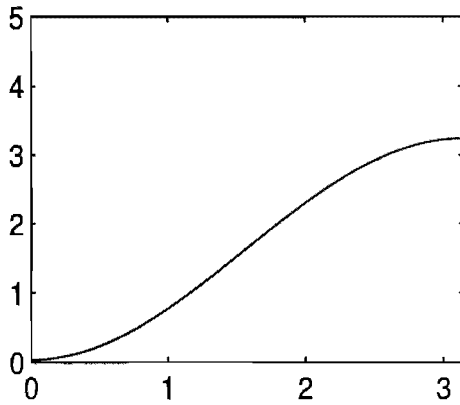
b) Spectrum of MA(1)

c) Spectrum of AR(1)

d) Spectrum of AR(2)

The range of cyclical frequencies

As already mentioned, the periodic and symmetric character of the spectrum permits us to consider only the range of frequencies $[0, \pi]$. When $\omega = 0$, the period $\tau \to \infty$, and the frequency is associated with a trend. When $\omega = \pi/2$, the period equals 4 quarters and the frequency is associated with the first seasonal harmonic (the once-a-year frequency). For a frequency in the range $[0 + \epsilon_1, \pi/2 - \epsilon_2]$, with $\epsilon_1, \epsilon_2 > 0$ and $\epsilon_1 < \pi/2 - \epsilon_2$, the associated period will be longer than a year, and bounded. Economic cycles should thus have a spectrum concentrated in this range. Broadly, we shall refer to this range as the "range of cyclical frequencies".

Frequencies in the range $[\pi/2, \pi]$ are associated with periods between 4 and 2 quarters. Therefore, they imply very short-term movements (with the cycle completed in less than a year) and are of no interest for business-cycle analysis. Given that $\omega = \pi$ is a seasonal frequency (the twice-a-year seasonal harmonic), the open interval of frequencies $(\pi/2, \pi)$, excluding the two seasonal frequencies, will be referred to as the "range of intraseasonal frequencies".

The determination of ϵ_1 and ϵ_2 in order to specify the precise range of cyclical frequencies is fundamentally subjective, and depends on the purpose of the analysis. For quarterly data and business-cycle analysis in the context of short-term economic policy, obviously a cycle of period 100000 years should be included in the trend, not in the business cycle. The same consideration would apply to a 10000 years cycle. As the period decreases (and ϵ_1 becomes bigger), we eventually approach frequencies that can be of interest for business-cycle analysis. For example, if the longest cycle that should be considered is a 10 year cycle (40 quarters), from (2.13), ϵ_1 should be set as $.05\pi$.

At the other extreme of the range, very small values of ϵ_2 can produce cycles with, for example, a period of 1.2 years, too short to be of cyclical interest. If the minimum period for a cycle is set as 1.5 years, then ϵ_2 should be set equal to $.167\pi$, and the range of cyclical frequencies would be $[.05\pi, .33\pi]$. Figure 2.9 shows how, from the decision on what is the relevant interval for the periods in a cyclical component, the range of cyclical frequencies is easily determined (in the figure, the interval for the period goes from 2 to 12 years).

24

Figure 2.9. Cyclical period and frequency
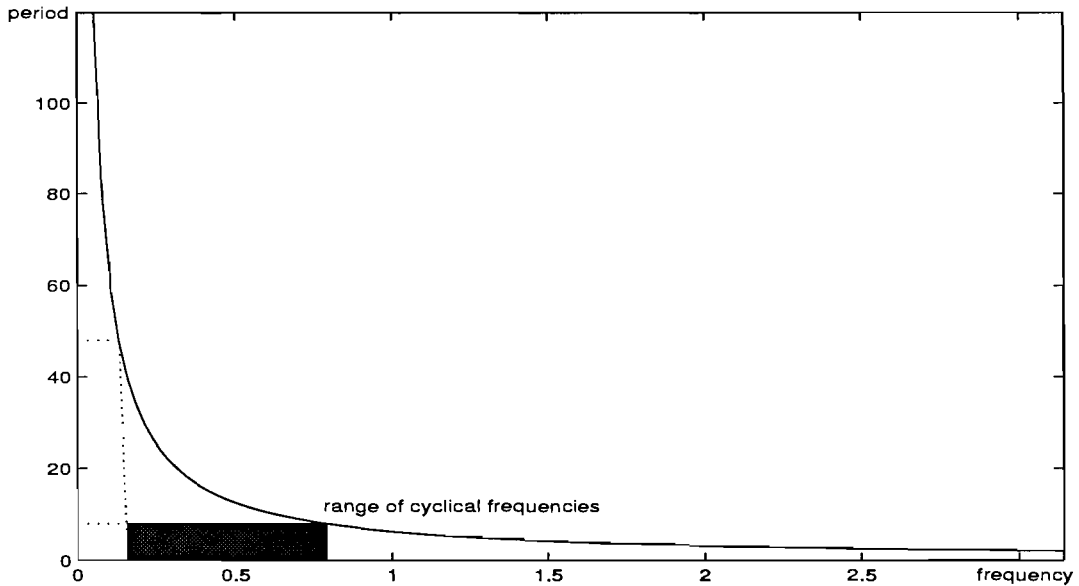
Extension to nonstationary unit roots

In the AR(1) model, we can let ϕ approach the value $\phi = -1$. In the limit we obtain

$$(1 - B)z_t = a_t, \quad \text{or} \quad \nabla z_t = a_t,$$

the popular random-walk model. Proceeding as in case 3. above, one obtains

$$g(\omega) = \frac{1}{2(1 - \cos\omega)} V_a.$$

For $\omega = 0$, $g(\omega) \to \infty$, and hence the integral (2.27) does not converge, which is in agreement with the well-known result that the variance of a random walk is unbounded. The nonstationarity induced by the root $\phi = -1$ in the AR polynomial $(1 + \phi B)$, a unit root associated with the zero frequency, induces a point of infinite in the spectrum of the process for that frequency. This result is general: a unit AR root, associated with a particular frequency ω_0, will produce an ∞ in the spectrum for that particular frequency.

An important example is when the polynomial $S = 1 + B + B^2 + B^3$ is present in the AR polynomial of the series. Given that S factorizes into $(1 + B)(1 + B^2)$, its roots are -1, and $\pm i$, associated with the frequencies π and $\pi/2$, respectively (as seen in Section 2.3). The Fourier transform of S, given by

$$S^* = 4(1 + \cos\omega)(1 + \cos 2\omega),$$

25

displays zeros for $\omega = \pi$ (first factor), and $\omega = \pi/2$ (second factor). Because S^* will appear in the denominator of the spectrum, its zeros will induce points of ∞. Therefore, a model with an AR polynomial including S will have a spectrum with points of ∞ for the frequencies $\omega = \pi/2$, and $\omega = \pi$, i.e., the seasonal frequencies.

It follows that, in the usual case of a seasonal quarterly series, for which a $\nabla\nabla_4$ or a $\nabla^2\nabla_4$ differencing has been used as the stationary transformation, the spectrum of the series would present points of ∞ for the frequencies $\omega = 0$, $\omega = \pi/2$, and $\omega = \pi$. Figure 2.10a exhibits what could be the spectrum of a standard, relatively simple quarterly series.

One final point. Given that a spectrum with points of ∞ has a nonconvergent integral, and that no standardization can provide a proper spectral density, the term spectrum is usually replaced by Pseudo-spectrum (see, for example, Hatanaka and Suzuki, 1967, and Harvey, 1989). For our purposes, however, the points of ∞ pose no serious problem, and the pseudo-spectrum can be used in much the same way as the stationary spectrum (this will become clear throughout the discussion). In particular, if, for the nonstationary series, we use the nonconvergent representation (2.18), compute the function $\gamma(B, F)$ through (2.19) and, in the line of Hatanaka and Suzuki, refer to this function as the "pseudo-AGF", the pseudo-spectrum is the Fourier transform of the pseudo-AGF. Bearing in mind that, when referring to nonstationary series, the term "pseudo-spectrum" would be more appropriate, in order to avoid excess notation, we shall simply use the term spectrum in all cases.

2.6 Linear filters and their squared gain

Back to the linear filter (2.4) of Section 2.1, the filter can be rewritten as

$$y_t = C(B, F)x_t, \tag{2.34}$$

where

$$C(B, F) = \sum_{j=1}^{k_1} c_{-j} B^j + c_0 + \sum_{j=1}^{k_2} c_j F^j.$$

If $k_1 = k_2$ and $c_j = c_{-j}$ for all j values, the filter becomes centered and symmetric, and we can express it as

$$C(B, F) = c_0 + \sum_{j=1}^{k} c_j (B^j + F^j). \tag{2.35}$$

Using the same Fourier transform as with expression (2.24), that is, replacing $(B^j + F^j)$ by $(2\cos j\omega)$, the frequency domain representation of the filter

becomes

$$C^*(\omega) = c_0 + 2\sum_{j=1}^{k} c_j \cos(j\omega). \qquad (2.36)$$

If $k_1 \neq k_2$ or $c_j \neq c_{-j}$, the uncentered or asymmetric filter does not accept an expression of the type (2.36). Additional terms involving imaginary numbers that do not cancel out will be present. This feature will induce a Phase effect in the output, in the sense that there will be a systematic distortion in the timing of events between input and output (for example, in the dating of turning points, of peaks and throughs, etc.). For our purposes, this is a disturbing feature and hence we shall concentrate attention on centered and symmetric filters.

Being C(B,F) symmetric and x_t stationary, (2.34) directly yields
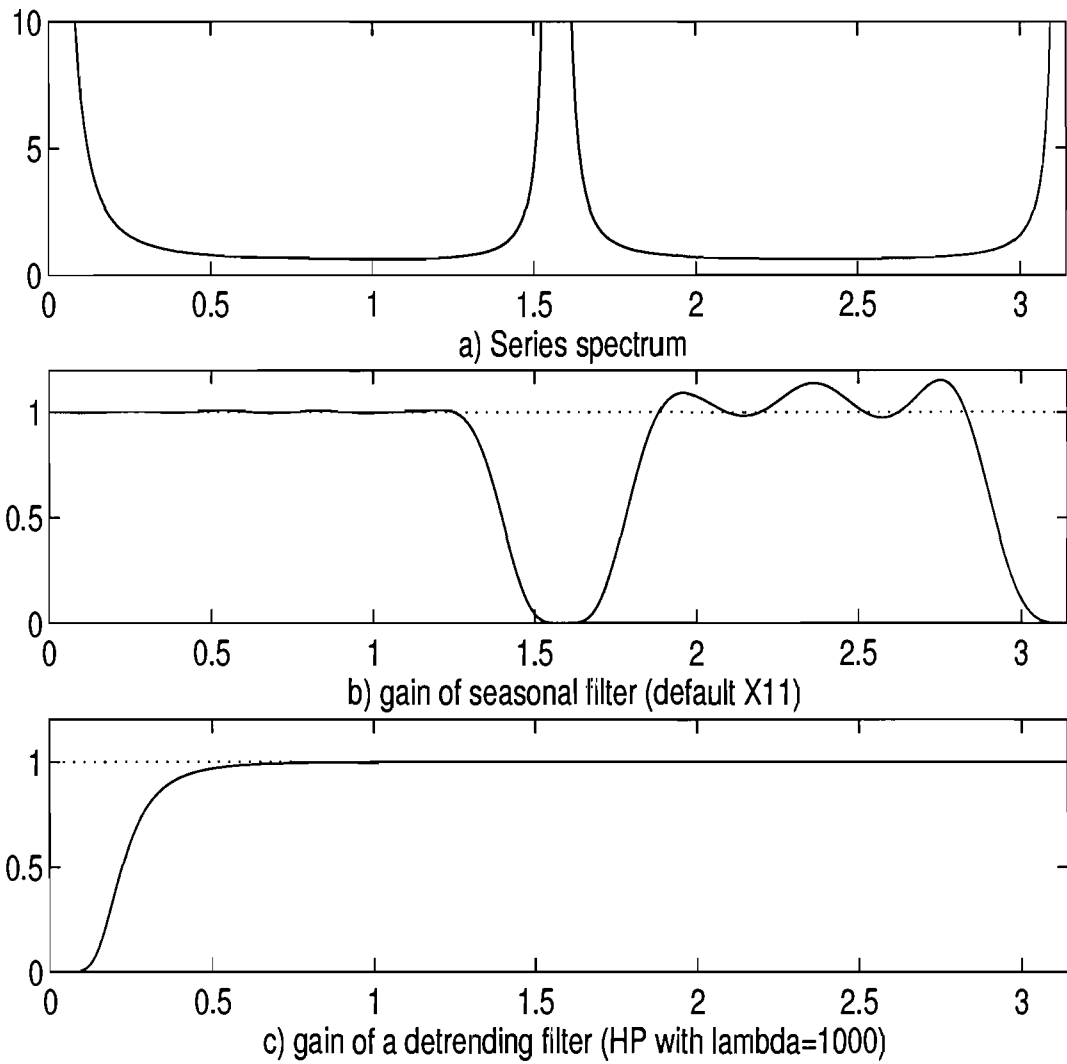
$$AGF(y) = [C(B,F)]^2 ACF(x),$$

so that, applying the Fourier transform, we obtain

$$g_y(\omega) = [G(\omega)]^2 g_x(\omega) \qquad (2.37)$$

where $g_x(\omega)$ and $g_y(\omega)$ are the spectra of the input and output series x_t and y_t and we represent by $G(\omega)$ the Fourier Transform of $C(B,F)$. The function $G(\omega)$ will be denoted the Gain of the filter. From the relationship (2.37), the squared gain determines what is the contribution of the variance of the input in explaining the variance of the output for each different frequency. If $G(\omega) = 1$, the full variation of x for that frequency is passed to y; if $G(\omega) = 0$, the variation of x for that frequency is fully ignored in the computation of y.

When interest centers in the components of a series, where the components are fundamentally characterized by their frequency properties, the squared gain function becomes a fundamental tool, since it tells us which frequencies will contribute to the component and which frequencies will not enter it. As an example, consider a quarterly series with spectrum that of Figure 2.10a. The peaks for $\omega = 0, \pi/2$, and π imply that the series contains a trend component and a seasonal component, associated with the once-and twice-a-year frequencies. A seasonal adjustment filter will be one with a squared gain displaying holes for the seasonal frequencies that will remove the seasonal spectral peaks, leaving the rest basically unchanged (Figure 2.10b displays the squared gain of the default-X11 seasonal adjustment filter). A detrending filter will be one with a squared gain that removes the spectral peak for the zero frequency, and leaves the rest approximately unchanged (Figure 2.10c displays the squared gain of the Hodrick-Prescott detrending filter, for the case of $\lambda = 1000$).

Figure 2.10

a) Series spectrum

b) gain of seasonal filter (default X11)

c) gain of a detrending filter (HP with lambda=1000)

One final important clarification should be made. We said that, in order to avoid phase effects, symmetric and centered filters would be considered. Let one such filter be

$$y_t = c_k x_{t-k} + \ldots + c_1 x_{t-1} + c_0 x_t + c_1 x_{t+1} + \ldots + c_k x_{t+k}. \qquad (2.38)$$

Assume a long series and let T denote the last observed period. When $T \geq t+k$, the filter can be applied to obtain y_t with no problem. However, when $T < t+k$, observations at the end of the series, needed to compute y_t, are not available yet, and hence the filter cannot be applied. As a consequence, the series y_t cannot be obtained for recent enough periods, because unknown future observations of x_t are needed. The fact that interest typically centers on recent periods has lead filter designers to modify the weights of the filters when truncation is needed because a lack of future observations (see, for example, the analysis in Burridge and Wallis, 1984, in the context of the seasonal adjustment filter X11.) Application of these truncated filters yields a preliminary measure of y_t, because new observations will imply changes in the weights, until $T \geq t + k$ and the final (or historical) value of y_t can be obtained. One modification that has become popular is to replace needed future values, not yet observed, by their optimal forecasts, often computed with an ARIMA model for the series x_t. Given that the forecasts are linear functions of present and past values of x_t, the preliminary value of y_t obtained with the forecasts will be a truncated filter applied to the observed series. Naturally, preliminary (truncated) filters will not be centered, nor symmetric. (In particular, the measurement of y_t obtained when the last observed period is t, i.e., when T=t, the so-called "concurrent" estimator, will be a purely one-sided filter). Besides its natural appeal, replacing unknown future values with optimal forecasts has the convenient features of minimizing (within the limitations of the structure of the particular series at hand,) both, the phase effect, and the size of the total revision the preliminary estimator will undergo until it becomes final. To this important issue of preliminary estimation and revisions we shall return in the following sections.

3 ARIMA models and signal extraction

3.1 ARIMA models

Back to the Wold representation (2.18) of a stationary process, $z_t = \Psi(B)a_t$, the representation is of no help from the point of view of fitting a model because, in general, the polynomial $\Psi(B)$ will contain an infinite number of parameters. Therefore we use a rational approximation of the type

$$\Psi(B) \doteq \frac{\theta(B)}{\phi(B)},$$

where $\theta(B)$ and $\phi(B)$ are finite polynomials in B of order q and p, respectively. Then we can write

$$z_t = \frac{\theta(B)}{\phi(B)}a_t, \text{ or}$$

$$\phi(B)z_t = \theta(B)a_t. \tag{3.1}$$

The model

$$(1 + \phi_1 B + \ldots + \phi_p B^p)z_t = (1 + \theta_1 B + \ldots + \theta_q B^q)a_t \tag{3.2}$$

is the Autoregressive Moving-Average process of orders p and q; in brief, the ARMA(p,q) model. For further reference, the Inverse Model of (3.1) is the one that results from interchanging the AR and MA polynomials. Thus

$$\theta(B)y_t = \phi(B)b_t,$$

with b_t white noise, is an inverse model of (3.1). Equation (3.2) can be seen as a non-homogeneous difference equation with forcing function $\theta(B)a_t$, an MA(q) process. Therefore, if both sides of (3.2) are multiplied by z_{t-k}, with $k > q$, and expectations are taken, the right hand side of the equation vanishes, and the left hand side becomes:

$$\gamma_k + \phi_1 \gamma_{k-1} + \ldots + \phi_p \gamma_{k-p} = 0, \tag{3.3}$$

or

$$\phi(B)\gamma_k = 0, \tag{3.4}$$

where B operates on the subindex k. The Eventual Autocorrelation Function (that is, γ_k as a function of k, for $k > q$) is the solution of the homogeneous difference equation (3.3), with characteristic equation

$$r^p + \phi_1 r^{p-1} + \ldots + \phi_p = 0. \tag{3.5}$$

If $r_1, \ldots, r_p$ are the roots of (3.5) the solution of (3.3) can be written as

$$\gamma_k = \sum_{i=1}^{p} r_i^k,$$

and will converge to zero as $k \to \infty$ when $|r_i| < 1, i = 1, \ldots, p$. Comparison of (3.5) with (3.3) shows that $r_1, \ldots, r_p$ are the inverses of the roots $B_1, \ldots, B_p$ of the polynomial

$$\phi(B) = 0$$

that is, $r_i = B_i^{-1}$. Convergence of γ_k implies, thus, that the roots (in B) of the polynomial $\phi(B)$ are all larger than 1 in modulus. This condition can also be stated as follows: the roots of the polynomial $\phi(B)$ have to lie outside the unit circle (of Figure 2.1a). When this happen, we shall say that the polynomial $\phi(B)$ is stable. From the identity

$$\phi(B)^{-1} = \frac{1}{(1 - r_1 B) \ldots (1 - r_p B)},$$

it is seen that stability of $\phi(B)$ implies, in turn, convergence of its inverse $\varphi(B)^{-1}$.

From (2.19), considering that $\Psi(B) = \theta(B)/\phi(B)$, the AGF of z_t is given by

$$\gamma(B, F) = \frac{\theta(B)}{\phi(B)} \frac{\theta(F)}{\phi(F)} V_a. \tag{3.6}$$

and it is straightforward to see that stability of $\phi(B)$ will imply that the stationarity conditions of Section 2.4 are satisfied. The AGF is symmetric and convergent, and the eventual autocorrelation function is the solution of a difference equation, and hence, in general, a mixture of damped polynomials in time and periodic functions. The Fourier transform of (3.6) yields the spectrum of z_t, equal to

$$g_z(\omega) = V_a \frac{\theta(e^{-i\omega})\theta(e^{i\omega})}{\phi(e^{-i\omega})\phi(e^{i\omega})}, \tag{3.7}$$

and the integral of $g_z(\omega)$ over $0 \leq \omega \leq 2\pi$ is equal to $2\pi Var(z_t)$.

A useful result is the following. If two stationary stochastic processes are related through

$$y_t = C(B)x_t,$$

then the AGF of y_t, $\gamma_y(B, F)$, is equal to

$$\gamma_y(B, F) = C(B)C(F)\gamma_x(B, F),$$

where $\gamma_x(B, F)$ is the AGF of x_t. Finally, a function that will prove helpful is the Crosscovariance Generating Function (CGF) between two series, x_t and

y_t, with Wold representation

$$x_t = \alpha(B)a_t$$
$$y_t = \beta(B)a_t.$$

Letting $\gamma_j = E(x_t y_{t-j})$ denote the lag-j crosscovariance between x_t and y_t, $j = 0, \pm 1, \pm 2, \ldots$, the CGF is given by

$$CGF(B, F) = \sum_{-\infty}^{\infty} \gamma_j B^j = \alpha(B)\beta(F)\sigma_a^2.$$

If, in equation (3.2), the subindex t is replaced by t+k (k a positive integer), and expectations are taken at time t, the forecast of z_{t+k} made at time t, namely $\hat{z}_{t+k|t}$, is obtained. Viewed as a function of k (the horizon) and for a fixed origin t, $\hat{z}_{t+k|t}$ is denoted the Forecast Function. (It will be discussed in more detail in subsection 3.2.3). Given that $E_t a_{t+k} = 0$ for $k > 0$, it is found that, for $k > q$, the forecast function satisfies the equation

$$\hat{z}_{t+k|t} + \phi_1 \hat{z}_{t+k-1|t} + \ldots + \phi_p \hat{z}_{t+k-p|t} = 0,$$

where $\hat{z}_{t+j|t} = z_{t+j}$ when $j \leq 0$. Therefore, the Eventual Forecast Function is the solution of

$$\phi(B)\hat{z}_{t+k|t} = 0, \tag{3.8}$$

with B operating on k. Comparing (3.4) and (3.8), the link between autocorrelation for lag k (and longer) and k-period-ahead forecast becomes apparent, the forecast being simply an extrapolation of correlation: what we can forecast is the correlation we have detected. For a zero-mean stationary process the forecast function will converge to zero, following, in general, a mixture of damped exponentials and cosine functions.

In summary, stationarity of an ARMA model, which requires the roots (in B) of the autoregressive polynomial $\phi(B)$ to be larger than 1 in modulus, implies the following model properties: a) its AGF converges; b) its forecast function converges; and c) the polynomial $\phi(B)^{-1}$ converges, so that z_t accepts the convergent (infinite) MA representation

$$z_t = \phi(B)^{-1}\theta(B)a_t = \Psi(B)a_t, \tag{3.9}$$

which is precisely the Wold representation. To see some examples, for the AR(1) model

$$z_t + \phi z_{t-1} = a_t,$$

the root of $1 + \phi B = 0$ is $B_1 = -1/\phi$. Thus stationarity of z_t implies that

$$|B_1| = |\frac{1}{\phi}| > 1,$$

or $|\phi| < 1$.

For the AR(2) model

$$z_t + \phi_1 z_{t-1} + \phi_2 z_{t-2} = a_t,$$

stationarity implies that the two roots, B_1 and B_2 be larger than one in modulus. This requires the coefficients ϕ_1 and ϕ_2 to lie inside the triangular region of Figure 3.1. The parabola inside the triangle separates the region associated with complex roots from the one with real roots (Box and Jenkins, 1970, Section 3.2).

If z_t is the differenced series, for which stationarity can be assumed, that is

$$z_t = Dx_t, \qquad D = \nabla^d, \quad d = 0, 1, 2, \ldots,$$

then the original nonstationary series x_t follows the Autoregressive Integrated Moving-Average process of orders p,d, and q, or ARIMA(p,d,q) model, given by

$$\phi(B)Dx_t = \theta(B)a_t; \tag{3.10}$$

p and q refer to the orders of the AR and MA polynomials, respectively, and d refers to the number of regular differences (i.e., the number of unit roots at the zero frequency). We shall often use abbreviated notation, namely

AR(p): autoregressive process of order p;

MA(q): moving-average process of order q;

ARI(p,d): autoregressive process of order p applied to the dth difference of the series;

IMA(d,q): moving-average process of order q applied to the dth difference of the series.

Further, a series will be denoted I(d) when it requires d regular differences in order to become stationary.

Fig 3.1: Stationarity region for AR(2) parameters
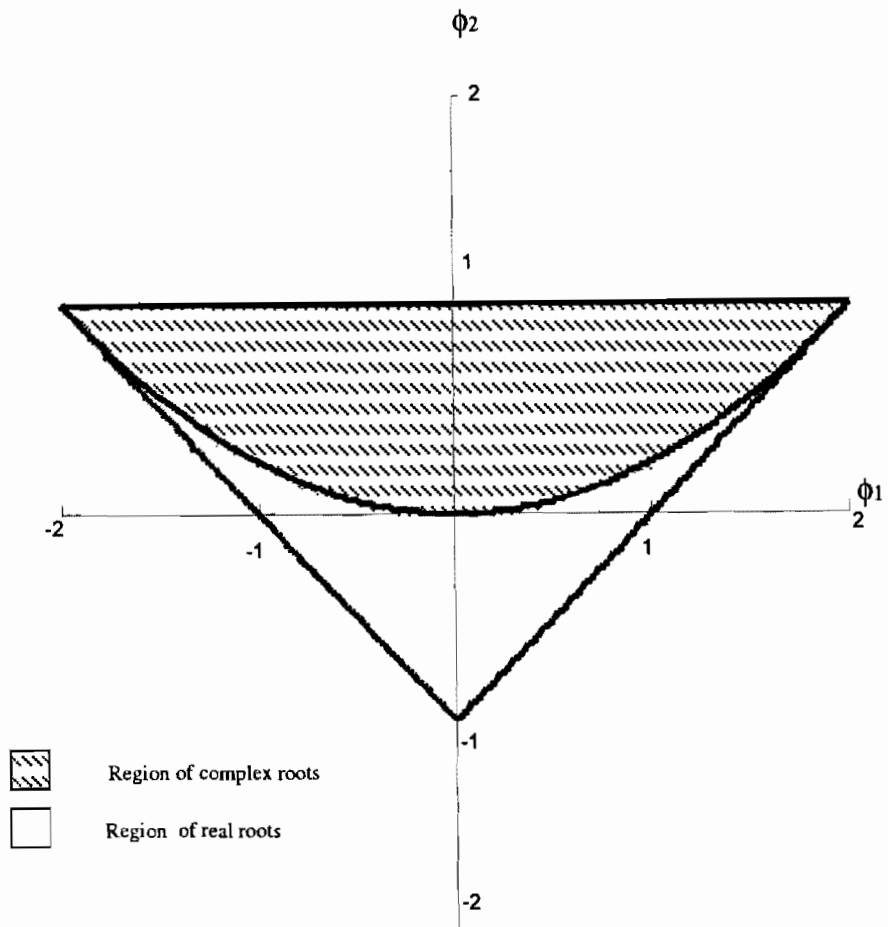
As in the stationary case, taking conditional expectations at time t in both sides of equation (3.10) with t replaced by t+k, where k is a positive integer, it is obtained that

$$\phi(B)D\hat{x}_{t+k|t} = \theta(B)\hat{a}_{t+k|t},$$

where

$$\hat{x}_{t+j|t} = E(x_{t+j} \mid x_t, x_{t-1}, \ldots)$$

is the forecast of x_{t+j} obtained at time t when $j > 0$, and is the observation x_{t+j} when $j \leq 0$; further, $\hat{a}_{t+j|t} = E(a_{t+j} \mid x_t, x_{t-1}, \ldots)$ is equal to 0 when $j > 0$, and is equal to a_{t+j} when $j \leq 0$. As a consequence, the eventual forecast function ($\hat{x}_{t+k|t}$ as a function of k, for $k > q$) will be the solution of the homogenous difference equation

$$\phi(B)D\hat{x}_{t+k|t} = 0,$$

with B operating on k. The roots of D all have unit modulus; if $D = \nabla^d$, then the eventual forecast function will include a deterministic polynomial in t of the type $(a + bt^{d-1})$. If D also includes seasonal differencing ∇_4, then the eventual forecast function will also contain the non-convergent deterministic cosine-type function (2.14), associated with the once and twice-a-year seasonal frequencies, $\omega = \pi/2$ and $\omega = \pi$.

As an example, the forecast function of the model

$$(1 - .7B)\nabla\nabla_4 x_t = (1 + \theta_1 B)(1 + \theta_4 B^4)a_t,$$

will consists of five starting values $\hat{x}_{t+j|t}, j = 1, \ldots, 5$, implied by the MA part with q=5, after which the function will be the solution of the homogenous equation associated with the AR part. Factorizing the AR polynomial as

$$(1 - .7B)(1 - B)^2(1 + B)(1 + B^2),$$

the roots of the characteristic equation are given by

$$r_1 = .7, r_2 = r_3 = 1, r_4 = -1, r_5 = i, r_6 = -i.$$

From Section 2.3, the eventual forecast function can be expressed as

$$\hat{x}_{t+k|t} = c_1^{(t)}(.7)^k + c_2^{(t)} + c_3^{(t)}k + c_4^{(t)}(-1)^k + c_5^{(t)}\cos\left(\frac{\pi}{2}k + c_6^{(t)}\right),$$

where the last two terms reflect the seasonal harmonics (the root $r_4 = -1$ can also be written as $c_4^{(t)}\cos\pi k$). The constants $c_1, \ldots, c_6$ are determined from the starting conditions of the forecast function, and hence will depend on t, the origin of the forecast. This feature gives the ARIMA model its adaptive (or "moving") properties. Notice that, in the nonstationary case, the forecast function (with fixed origin t and increasing horizon k) will not converge.

Concerning the MA polynomial $\theta(B)$, a similar condition of stability will

35

be imposed, namely, the roots $B_1, \ldots, B_q$ of the equation $\theta(B) = 0$ have to be larger than 1 in modulus. This condition is referred to as the Invertibility condition for the process and, unless otherwise specified, we shall assume that the model for the observed series z_t is invertible. This assumption implies that $\theta(B)^{-1}$ converges, so that the model (3.1) can be inverted and expressed as

$$a_t = \theta(B)^{-1}\phi(B)z_t = \Pi(B)z_t, \qquad (3.11)$$

which shows that the series accepts a convergent (infinite) AR expression, and hence can be approximated by a finite AR. Expression (3.11) also shows that, when the process is invertible, the innovations can be recovered from the z_t series.

Some frequency domain implications of nonstationarity and noninvertibility are worth pointing out. Assume that the MA polynomial $\theta(B)$ has a unit root $|B_1| = 1$ -perhaps a complex conjugate pair- associated with the frequency ω_1. Then, $\theta(e^{-i\omega_1}) = 0$, and the spectrum of z_t, given by (3.7), will have a zero for the frequency ω_1. Analogously, if $|B_1| = 1$ is a root of the AR polynomial $\phi(B)$, with associated frequency ω_1, then, $\phi(e^{-i\omega_1}) = 0$, and $g(\omega_1) \to \infty$
. It follows that

- a unit MA root causes a zero in the spectrum;

- a unit AR root causes a point of ∞ in the spectrum;

- an invertible model will have strictly positive spectrum, $g(\omega) > 0$;

- a stationary model has a bounded spectrum, $g(\omega) < \infty$.

To illustrate the spectral implications of unit roots, Figure 3.2a presents the spectrum of the model

$$(1 - B)x_t = (1 + B)a_t.$$

Since the spectrum is proportional to $(1 + \cos\omega)/(1 - \cos\omega)$, the unit AR root $B = 1$ for the zero frequency makes the vertical axis an asymptote. The unit MA root $B = -1$ for $\omega = \pi$ creates a zero for this frequency. The spectrum of the inverse model

$$(1 + B)x_t = (1 - B)a_t$$

is displayed in Figure 3.2b. The unit AR root for $\omega = \pi$ implies that the line $\omega = \pi$ is an asymptote, and the unit MA root for $\omega = 0$ implies a spectral zero at the origin.

For quarterly data with seasonality, the differencing D is likely to contain the seasonal difference ∇_4. A popular specification that increases parsimony

Figure 3.2

a) spectrum of (1−B)x(t)=(1+B)a(t)
b) spectrum of (1−B)x(t)=(1−B)a(t)
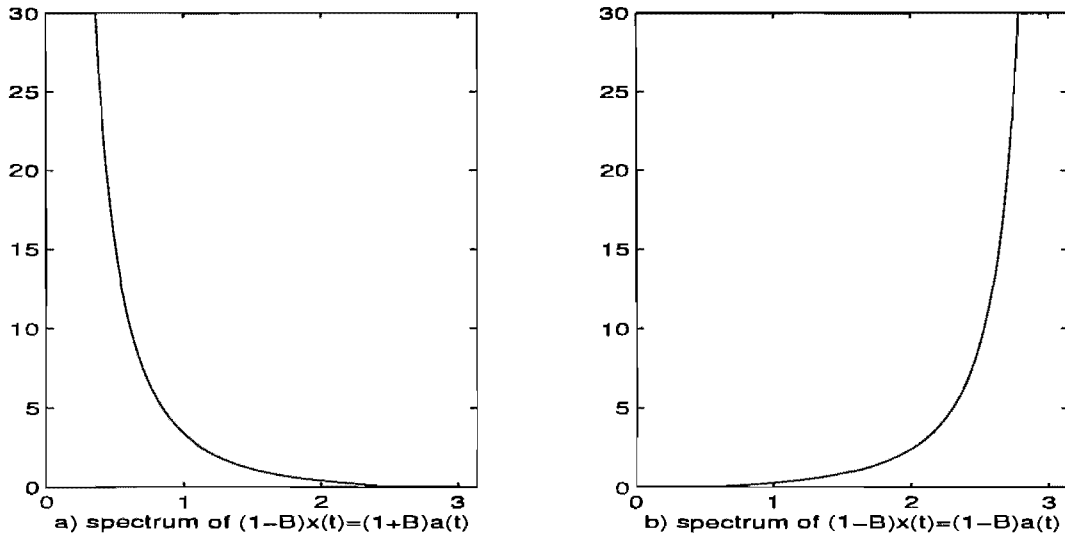
of the model and permits us to capture seasonal effects is the Multiplicative seasonal model

$$\phi(B)\Phi(B^4)\nabla^d\nabla_4^D x_t = \theta(B)\Theta(B^4)a_t \tag{3.12}$$

where the regular AR polynomial in B, $\phi(B)$, is as in (3.2), $\Phi(B^4)$ is the seasonal AR polynomial in B^4, d is the degree of regular differencing, D is the degree of seasonal differencing, $\theta(B)$ is the regular MA polynomial in B, $\Theta(B^4)$ is the seasonal MA polynomial in B^4, and a_t denotes the series white-noise $(0, V_a)$ innovation. The polynomials $\phi(B), \Phi(B^4), \theta(B)$ and $\Theta(B^4)$, are assumed stable, and hence the series ·

$$z_t = \nabla^d\nabla_4^D x_t$$

follows a stationary and invertible process. (To avoid nonsense complications, we assume that the stationary AR and invertible MA polynomials are prime.) If p, P, q, and Q denote the orders of the polynomials $\phi(B), \Phi(\beta), \theta(B)$ and $\Theta(\beta)$, respectively, where $\beta = B^4$, model (3.12) will be referred to as the multiplicative ARIMA $(p, d, q)(P, D, Q)_4$ model. In practice, we can safely restrict the orders to

$$
\begin{aligned}
- \quad p, q &\leq 4; \\
- \quad P &\leq 1; \\
- \quad Q &\leq 2; \\
- \quad d &\leq 2; \\
- \quad D &\leq 1.
\end{aligned}
\tag{3.13}
$$

Two important practical comments (to bear always in mind) are the following:

1. Parsimony (i.e., few parameters) should be a crucial property of ARIMA models used in practice.

2. ARIMA models are a useful tool for relatively short-term analysis. Their flexibility and adaptive behavior contribute to their good short-term forecasting. Long-term extrapolation of this flexibility may imply, however, unstable long-term inference (see Maravall, 1999). As a general rule, short-term analysis favors differencing, while long-term one favors more deterministic trends, that imply less differences.

3.2 Modelling strategy, diagnostics and inference

The so-called Box–Jenkins approach to building ARIMA models consists of the following iterative scheme that contains 4 stages:

3.2.1 Identification

Two features of the series have to be addressed,

- the degree of regular and seasonal differencing;

- the orders of the stationary AR and invertible MA polynomials.

Differencing of the series can employ some of the unit root tests available for possibly seasonal data (see, for example, Hylleberg et al, 1990). Devised to test deterministic seasonals versus seasonal differencing, these test are of little use for our purpose. In our experience, stochastic modelling removes in practice the need for the dilemma: deterministic specification versus differencing. Consider, for example, the two models:

(a) $x_t = \mu + a_t$,

(b) $\nabla x_t = (1 - .99B)a_t$.

For a quarterly series, and realistic series length, it is impossible that the sample information can distinguish between the two specifications. Consequently, the choice is arbitrary. Besides the variance of a_t, Model (a) contains one parameter that needs to be estimated, while Model (b) contains none (although, in this case the first observation is lost by differencing). Model (a) offers, thus, no estimation advantage. If short-term forecasting is the main objective, however, Model (b) will display some advantage because it allows for more flexibility, given that it could be rewritten as $x_t = \mu^{(t)} + a_t$, where $\mu^{(t)}$ is a very slowly adapting mean.

A similar consideration applies to seasonal variations. The model

38

(c) $x_t = \mu + \sum_{j=1}^{3} \beta_j d_{jt} + a_t,$

where d_{jt} denotes a quarterly seasonal dummy variable, is in practice indistinguishable from the direct specification

(d) $\nabla_4 x_t = (1 - .95B^4)a_t.$

The deterministic specification has now 4 parameters; the stochastic one has none, but 4 starting values are lost at the beginning.. The latter can also be expressed as

$$x_t = \mu^{(t)} + \sum_{j=1}^{3} \beta_j^{(t)} d_{jt} + a_t,$$

where $\mu^{(t)}, \beta^{(t)}$ denote slowly adapting coefficients. Within our short-term perspective, there is no reason thus to maintain the deterministic-stochastic dichotomy, and deterministic features can be seen as extremely stable stochastic ones.

Besides the lack of power of unit roots tests to distinguish between models (a) and (b), or (c) and (d), the process of building ARIMA models typically implies estimation of many specifications (if combined with outlier detection and correction, the number may be indeed very large) and the true size of the tests is therefore unknown. In practice, a more efficient and reliable procedure for determining AR unit roots is to use estimation results based on the superconsistency of parameter estimates associated with unit roots, having determined "a priori" how close to one a root has to be in order to be considered a unit root (see Tiao and Tsay, 1983, 1989, and Gómez and Maravall, 2000a).

Once the proper differencing has been established, it remains to determine the orders of the stationary AR and invertible MA polynomials. Here, the basic criterion used to be to try to match the SACF of z_t with the theoretical ACF of a particular ARMA process. In recent years, the efficiency and reliability of automatic identification procedures, based mostly on information criteria, has strongly decreased the importance of the "tentative identification" stage (see Fischer and Planas, 1999, and Gómez and Maravall, 2000a).

3.2.2 Estimation and diagnostics

When $q \neq 0$, the ARIMA residuals are highly nonlinear functions of the model parameters, and hence numerical maximization of the likelihood function, or of some function of the residual sum of squares, can be computationally nontrivial. Within the restrictions in the size of the model given by (3.13), however, maximization is typically well behaved. A standard estimation procedure would cast the model in a state-space format, and use the Kalman filter

to compute the likelihood through the Prediction Error Decomposition. The likelihood is then maximized with some nonlinear procedure. Usually, the V_a parameter, as well as a possible constant mean, are concentrated out of the likelihood. When the series is nonstationary, several solutions have been proposed to overcome the problem of defining a proper likelihood. Relevant references are Bell and Hillmer (1991), Brockwell and Davis (1987), De Jong (1991), Gómez and Maravall (1994), Kohn and Ansley (1986), and Morf, Sidhu and Kailath (1974). Several of these references deal, in fact, with more general models than the straightforward ARIMA one.

Many diagnostics are available for ARIMA models. A crucial one, of course, is the out-of-sample forecast performance. Some test for in-sample model stability are also of interest. Then, there is a large set of test based on the model residuals, assumed to be niid. This implies testing for Normality, for autocorrelation, for homoscedasticity, etc. Besides the ones proposed by Box and Jenkins (1970), additional references can be Newbold (1983), Gourieroux and Monfort (1990), Harvey (1989), and Hendry (1995).

3.2.3 Inference

If the diagnostics are failed, in the light of the results obtained, the model specification should be changed. When the model passes all diagnostics, we may then proceed to inference. We shall look in particular at an application in forecasting, unquestionably the main use of ARIMA models.

Let (3.10) denote, in compact notation, the ARIMA model identified for the series x_t, and, as in Section 3.1, denote by $\hat{x}_{t+j|t}$ the forecast of x_{t+j} made at period t (in Box-Jenkins notation, $\hat{x}_{t+j|t} = \hat{x}_t(j)$.) Under our assumptions, the optimal forecast of x_{t+j}, in a Minimum Mean Square Error (MMSE) sense, is the expectation of x_{t+k} conditional on the observed time series $x_1, \ldots, x_t$ (equal also, to the projection of x_{t+k} onto the observed time series); that is,

$$\hat{x}_{t+j|t} = E(x_{t+k} \mid x_1, \ldots, x_t).$$

This conditional expectation can be obtained with the Kalman filter, or with the Box-Jenkins procedure (for large enough t). Recall that, for known parameters,

$$a_t = x_t - \hat{x}_{t|t-1},$$

that is, the innovations of the process are the sequence of one-period-ahead forecast errors.

The forecast function at time t is $\hat{x}_{t+k|t}$ as a function of k (k a positive integer). In Section 3.1 we saw that for an ARIMA (p,d,q) model, the forecast function consists of q starting conditions, after which it is given by the solution

of the homogenous AR difference equation

$$\phi^*(B)\hat{x}_{t+k|t} = 0, \tag{3.14}$$

where B operates on k, and $\phi^*(B)$ denotes the full AR convolution $\phi^*(B) = \phi(B)D$, and includes thus the unit roots.

A useful way to look at forecasts is directly based on the pure MA representation $\Psi(B)$, even in the nonstationary case of a nonconvergent $\Psi(B)$. Assume the model parameter are known and write

$$x_{t+k} = a_{t+k} + \psi_1 a_{t+k-1} + \ldots + \psi_{k-1}a_{t+1} + \psi_k a_t + \psi_{k+1}a_{t-1} + \ldots. \tag{3.15}$$

Given that, for $k > 0$, $E_t a_{t+k} = 0$ and $E_t a_{t-k} = a_{t-k}$, taking conditional expectations in (3.15) yields

$$\hat{x}_{t+k|t} = E_t x_{t+k} = \sum_{j=0}^{\infty} \psi_{k+j}a_{t-j}; \tag{3.16}$$

so that the forecast is a linear combination of past and present innovations. Substracting (3.16) from (3.15), the k-periods-ahead forecast error is given by the model

$$
\begin{aligned}
e_{t+k|t} &= x_{t+k} - \hat{x}_{t+k|t} \\
&= a_{t+k} + \psi_1 a_{t+k-1} + \ldots + \psi_{k-1}a_{t+1}, \tag{3.17}
\end{aligned}
$$

an MA(k-1) process of "future" innovations. From expression (3.17), the joint, marginal, and conditional distributions of forecast errors can be easily derived, and in particular the standard error of the k-period ahead forecast, equal to

$$SE(k) = (1 + \psi_1^2 + \ldots + \psi_{k-1}^2)^{1/2}\sigma_a. \tag{3.18}$$

Unless the series is relatively short, this standard error, estimated by using ML estimators of the parameters, will provide a good approximation. Figure 3.3 displays the last 3 years of observations and the next 2 years of ARIMA forecasts for a quarterly series. The forecast function is dominated by a linear trend plus seasonal oscillations; the width of the confidence interval increases with the horizon.

41

Figure 3.3 Forecasts and 90% confidence interval
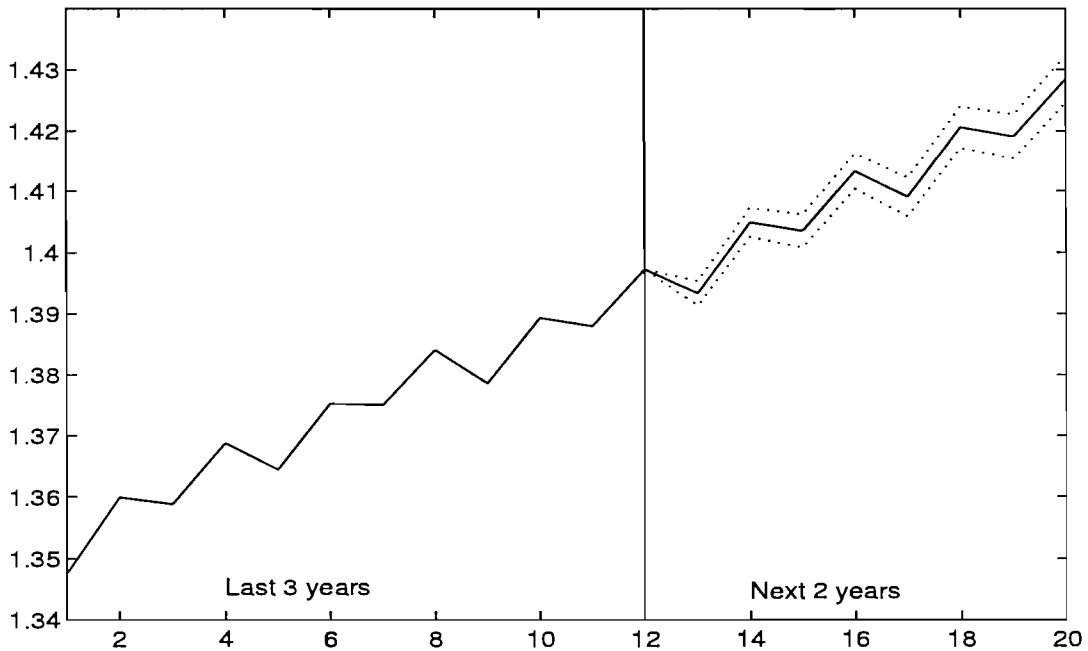
3.2.4 A particular class of models

Box and Jenkins (1970) dedicate a considerable amount of attention to a particular multiplicative model that, for quarterly series, takes the form

$$\nabla\nabla_4 x_t = (1 + \theta_1 B)(1 + \theta_4 B^4)a_t \tag{3.19}$$

(a regular IMA(1,1) structure multiplied by a seasonal IMA(1,1) structure). Given that they identified the model for a series of airline passengers, it has become known as the "Airline model". Often, the model is obtained for the logs, in which case a rough first reading shows that the rate-of-growth of the annual difference is a stationary process.

The model is highly parsimonious, and the 3 parameters can be given a structural interpretation. As seen in Section 3.1, when $\theta_1 \to -1$, the trend behavior generated by the model becomes more and more stable and, when $\theta_4 \to -1$, the same thing happens to the seasonal component. Estimation of MA roots close to the noninvertibility boundary poses no serious problem, and fixing a priori the maximum value of the modulus of a MA root to, for example, .99 produces perfectly behaved invertible models.

If estimation of (3.19) yields, for example, $\hat{\theta}_4 = -.99$, two (mutually exclusive) things can explain the result:

1) seasonality is practically deterministic;

2) there is no seasonality, and the model is overdifferenced.

Determining which of the two is the correct explanation is rather simple by testing for the significance of seasonal dummy variables. When the model has no seasonality, the seasonal filter $\nabla_4 z_t = (1 - .99 B^4) b_t$ would have hardly any effect on the input series. A similar reasoning holds for θ_1 and the possible presence of a deterministic trend. Further, a purely white-noise series filtered with model (3.19) with $\theta_1 = \theta_4 = -.99$ would, very approximately, reproduce the series. Thus the Airline model also encompasses simpler structures with no trend or no seasonality. Adding the empirical fact that it provides reasonably good fits to many actual macroeconomic series (see, for example, Fischer and Planas, 1999, or Maravall, 2000), it is an excellent model for illustration, for benchmark comparison, and for pre-testing.

3.3 Preadjustment

We have introduced the ARIMA model as a practical way of dealing with moving features of series. Still, before considering a series appropriate for ARIMA modelling, several prior corrections or adjustments may be needed. We shall classify them into 3 groups.

1. OUTLIERS

    The series may be subject to abrupt changes, that cannot be explained by the underlying normality of the ARIMA model. Three main types of outlier effects are often distinguished: a) additive outlier, which affects an isolated observation, b) level shift, which implies a step change in the mean level of the series, and c) transitory change, similar to an additive outlier whose effect damps out over a few periods. Chen and Liu (1993) suggested an approach to automatic outlier detection and correction that has lead to reliable and efficient procedures (see Gómez and Maravall, 2000a).

2. CALENDAR EFFECT

    By this term we refer to the effect of calendar dates, such as the number of working days in a period, the location of Easter effect, or holidays. These effects are typically incorporated into the model through regression variables (see, for example, Hillmer, Bell and Tiao, 1983, and Harvey, 1989).

3. INTERVENTION VARIABLES

Often special, unusual events affect the evolution of the series and cannot be accounted for by the ARIMA model. There is thus a need to "intervene" the series in order to correct for the effect of special events. Examples can be strikes, devaluations, change of the base index or of the way a series is constructed, natural disasters, political events, important tax changes, or new regulations, to mention a few. These special effects are entered in the model as regression variables (often called, following Box and Tiao, 1975, intervention variables).

The full model for the observed series can thus be written as

$$y_t = w_t'\beta + C_t'\eta + \sum_{j=1}^{k} \alpha_j \lambda_j(B) I_t(t_j) + x_t \qquad (3.20)$$

where $\beta = (\beta_1, \ldots, \beta_n)'$, is a vector of regression coefficients, $w_t' = (w_{1t}, \ldots, w_{nt})$ denotes n regression or intervention variables, C_t' denotes the matrix with columns the calendar effect variables (trading day, Easter effect, Leap year effect, holidays), and η the vector of associated coefficients, $I_t(t_j)$ is an indicator variable for the possible presence of an outlier at period t_j, $\lambda_j(B)$ captures the transmission of the j-th outlier effect (for additive outliers, $\lambda_j(B) = 1$, for level shifts, $\lambda_j(B) = 1/\nabla$, for transitory changes, $\lambda_j(B) = 1/(1 - \delta B)$, with $0 < \delta < 1$,) and α_j denotes the coefficient of the outlier in the multiple regression model with k outliers. Finally, x_t follows the general (possibly multiplicative) ARIMA model (3.12). As mentioned earlier, there are several procedures for estimation of models of this type, and easily available programs that enforce the procedures (examples are the programs REGARIMA, see Findley et al, 1998, and TRAMO, see Gómez and Maravall, 1996). Noticing that intervention variables, outliers, and calendar effects are regression variables, the full model can be expressed as a regression-ARIMA model. Estimation typically proceeds by iterating as follows: conditional on the regression parameters (β, η, α), exact maximum likelihood estimation of the ARIMA model is performed; then, conditional on the ARIMA model, GLS estimators of the regression parameters are obtained (both steps can be done with the Kalman filter).

Bearing in mind that preadjustment should be a "must" in applied time series work, for the rest of this book, we shall assume that the series do not require preadjustment, or have already been subject to one. The series can be directly seen, then, as the outcome of an ARIMA process.

Figures 3.4 and 3.5 illustrate preadjustment in quarterly (simulated) series. The observed original series is displayed in Figure 3.4a. After removal (through regression) of the outliers automatically identified in the series (2 additive

44

outliers, 1 level shift, and 1 transitory change) whose effect is displayed in Figure 3.5a, of the trading-day effect (captured, in this case, with a variable that counts the number of working days) shown in Figure 3.5b, of the Easter effect, exhibited in Figure 3.5c, and of an intervention variable associated with the introduction of a regulation that affects the seasonal effect for the last two quarters of each year, the remaining series is displayed in Figure 3.4b. This is the preadjusted series, also referred to as the "linearized series", given that it can be assumed the output of a linear stochastic process (modelled in the ARIMA format).

In the final decomposition of the observed series, that we shall be discussing in the following sections, the different regresion effects (outliers, calendar effects, and intervention variables) can be associated with different components. Thus, typically, calendar effects will be associated with the seasonal component, additive and transitory outliers will be assigned to the irregular component, and level shifts to the trend-cycle component. Care should be taken, however, when a separate business-cycle component is being estimated, because it may require a different allocation of the deterministic effects. For example, when annual data is being used, a transitory change that takes 5 or 6 periods to become negligeable should probably be included in the cycle, not in the irregular. Likewise, the correction produced by two level shifts of opposite sign and similar magnitude possibly should be assigned to the cycle, not to the long-term trend.
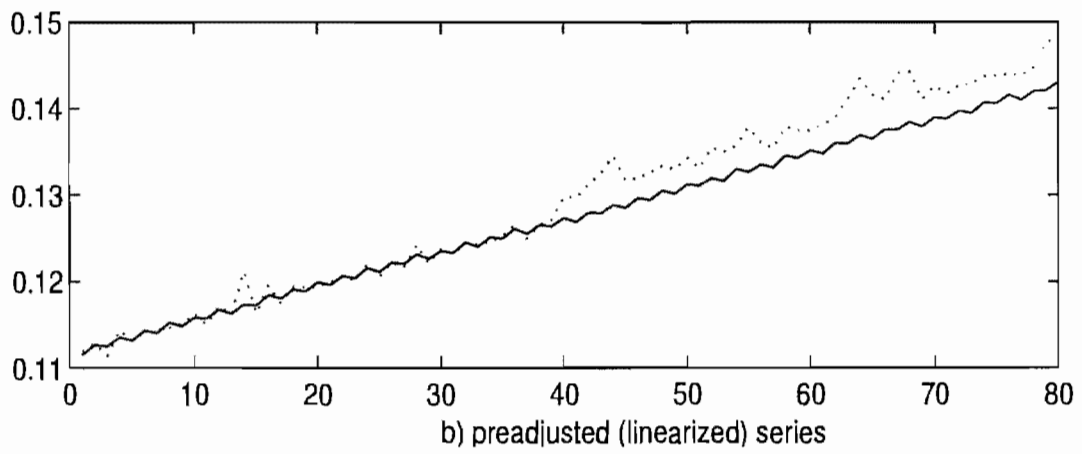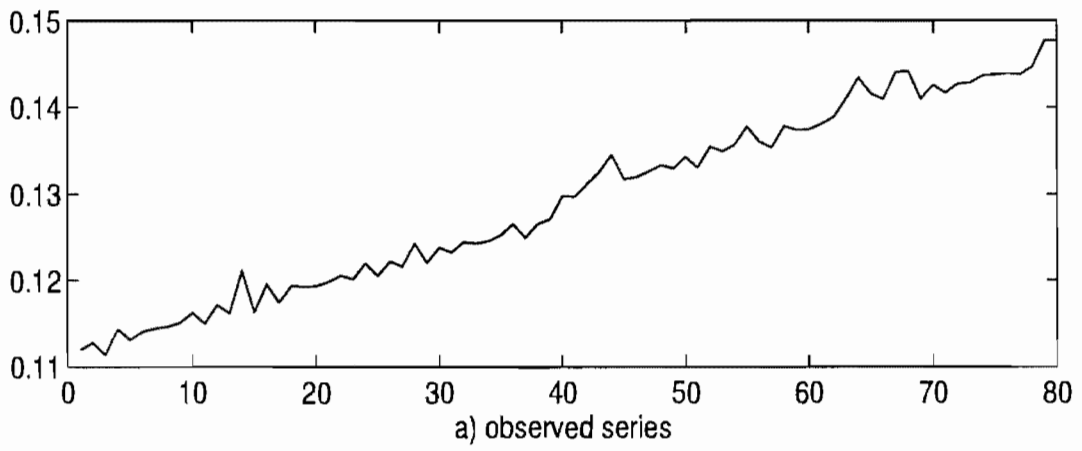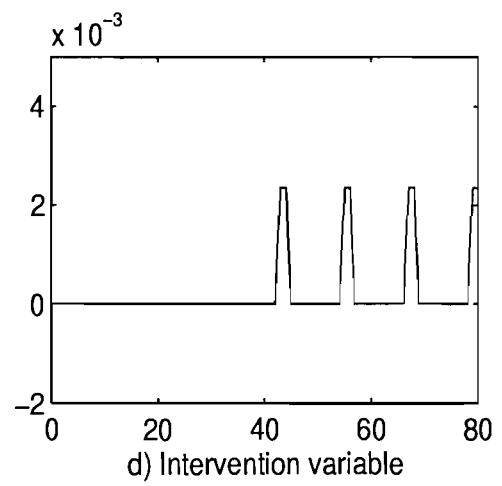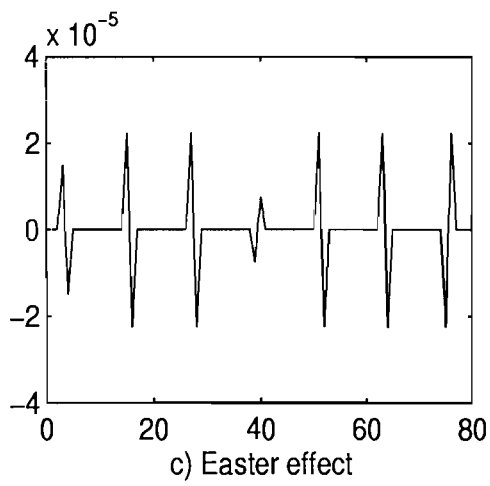
Figure 3.4. Preadjustment

a) observed series

b) preadjusted (linearized) series

46

Figure 3.5. Deterministic Effects

a) Outlier effect

b) Trading day effect

c) Easter effect

d) Intervention variable
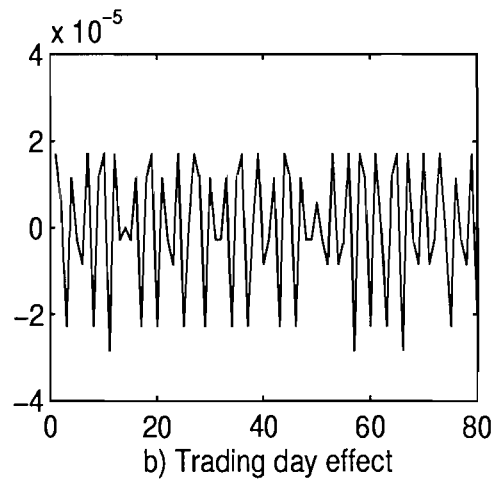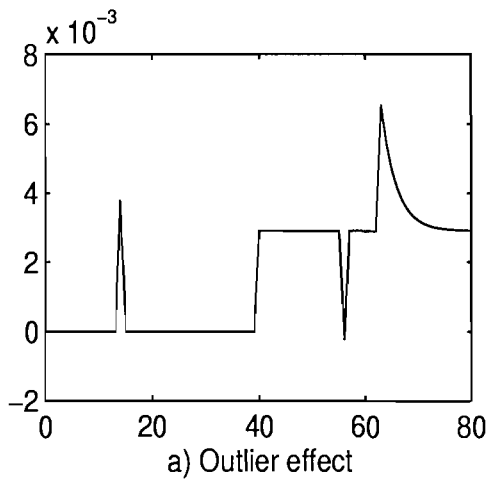
47

3.4 Unobserved components and signal extraction

Assume we are interested in some unobserved component buried in the observed series. Examples can be the seasonally adjusted (SA) series, some underlying short-term trend, or perhaps some cycle. We refer to the component of interest as the Signal, and assume it can be extracted from x_t in an additive manner, as in

$$x_t = s_t + n_t, \tag{3.21}$$

where n_t denoted the non-signal component of the series. (If the signal is the SA series, n_t would be the seasonal component; if the signal is the short-term trend, an additional noise or transitory component may also be included in n_t). The decomposition can also be multiplicative, as in $x_t = s_t n_t$. Taking logs, however, the additive structure is recovered. For the rest of the discussion we shall consider the additive decomposition. (A more complete presentation can be found in Planas, 1997).

We further assume that both components are linear stochastic processes, say

$$\phi_s(B)s_t = \theta_s(B)a_{st} \tag{3.22}$$

$$\phi_n(B)n_t = \theta_n(B)a_{nt}. \tag{3.23}$$

The AR polynomials $\phi_s(B)$ and $\phi_n(B)$ also include possible unit roots; in fact, in the vast majority of applications, at least one of the components will be nonstationary. This is because the very concept of a trend or a seasonal component imply a mean that changes with time, and hence a nonstationary behavior that can be removed by differencing.

Concerning expressions (3.22) and (3.23), the following assumptions will be made:

(A.1) The variables a_{st} and a_{nt} are mutually independent white-noise processes, with zero mean, and variances V_s and V_n, respectively.

(A.2) The polynomials $\phi_s(B)$ and $\phi_n(B)$ are prime.

(A.3) The polynomials $\theta_s(B)$ and $\theta_n(B)$ share no unit root in common.

The first assumption is based on the belief that what causes, for example, seasonality (weather, time of year) is not much related to what may drive a long-term trend (technology, investment), and similarly for other components. Given that different components are associated with different spectral peaks, assumption A.2 seems perfectly sensible. Assumption A.3 is not strictly needed, but in practice it is hardly restrictive and simplifies considerably notation. The assumption states a sufficient condition for invertibility of the x_t series.

Because aggregation of ARIMA models also yields an ARIMA model, the series x_t will follow an ARIMA model, which we write as

$$\phi(B)x_t = \theta(B)a_t, \tag{3.24}$$

where a_t is a white noise variable, $\theta(B)$ is invertible, and $\phi(B)$ is given by

$$\phi(B) = \phi_s(B)\phi_n(B). \tag{3.25}$$

The following identity is implied by (3.22)-(3.24):

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

which shows the relatively complicated relationship between the series innovations and the innovations in the components (see Maravall, 1995)

Having observed a time series $X_T = [x_1, \ldots, x_T]$ our aim is: 1) to obtain Minimum Mean Square Error (MMSE) estimators of $\hat{s}_t$ (and $\hat{n}_t$), as well as forecasts; 2) to obtain the full distribution of these estimators, from which diagnostics can be derived; 3) to obtain standard errors for the estimators and forecasts; and 4) to analyze some important features, such as revisions in preliminary estimators, both in terms of size and speed of convergence to the historical estimators.

1. Known models

For the stationary case, the full distribution of (s_t, X_T) is known. Under some additional assumptions (see, for example, Bell and Hillmer, 1991, and Gómez and Maravall, 1993) an appropriate conditional distribution can also be derived for the nonstationary case. The joint distribution is multivariate normal, so that the conditional expectation of the unobserved s_t, given X_T, is a linear combination of the elements in X_T. This conditional expectation also provides the MMSE estimator, $\hat{s}_t$, which can thus be expressed as the linear filter

$$\hat{s}_t = E(s_t \mid x_1, \ldots, x_T) = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_T x_T.$$

The above conditional expectation can be computed with the Kalman filter (see Harvey, 1989) or with the Wiener-Kolmogorov (WK) filter (see Box, Hillmer and Tiao, 1978). The equivalence of both filters, also when the series is nonstationary, is shown in Gómez (1999). Both filters are efficient; while the Kalman filter has a more flexible format to expand the models, the WK filter is more useful for analysis and interpretation. It will be the one used in the discussion.

We start by considering the case of an infinite realization $(x_{-\infty}, \ldots x_{\infty})$. (In practice, this means that we start with historical estimation for the central years of a long-enough series.) As shown in Whittle (1963), the WK filter that yields the MMSE of s_t when model (3.24) is stationary is given by the ratio of

the AGF of s_t and x_t, namely

$$\hat{s}_t = \left[\frac{AGF(s_t)}{AGF(x_t)}\right] x_t = \left[\frac{V_s \dfrac{\theta_s(B)\theta_s(F)}{\phi_s(B)\phi_s(F)}}{V_a \dfrac{\theta(B)\theta(F)}{\phi(B)\phi(F)}}\right] x_t.$$

(3.26)

Notice that an important feature of the WK filter (enforced in this way) is that it only requires the specification of the model for the signal, once the model for the observed series has been identified. Contrary to other model-based approaches enforced with the Kalman filter, such as the Structural Time Series Model (STSM) approach of Harvey (1989), with the WK filter there is no need to specify the components that aggregate into the non-signal n_t. In view of (3.25), the filter simplifies into

$$\hat{s}_t = \left[k_s \frac{\theta_s(B)\phi_n(B)}{\theta(B)} \frac{\theta_s(F)\phi_n(F)}{\theta(F)}\right] x_t.$$

(3.27)

where $k_s = V_s/V_a$. Direct inspection of (3.27) shows that the filter is the AGF of the stationary model

$$\theta(B)z_t = \theta_s(B)\phi_n(B)b_t,$$

(3.28)

where b_t is white noise with variance (V_s/V_a). The filter is thus convergent in B and F, centered at t, and symmetric.

In order to analyze the properties of the estimated signal, we shall be interested in its spectrum. If $g_s(\omega), g_n(\omega)$ and $g(\omega)$ denote the spectrum of the signal, the non-signal component, and the observed series, respectively, orthogonality of s_t and n_t imply

$$g(\omega) = g_s(\omega) + g_n(\omega),$$

where the two components spectra are nonnegative, and $g(\omega)$ is strictly positive (due to the invertibility condition on the observed series).

The gain of the WK filter, given by the expression in brackets in (3.26), is the Fourier transform of the ratio of two AGFs, so that

$$G(\omega) = g_s(\omega)/g(\omega).$$

Thus, according to (3.26), the spectrum of the MMSE estimator $\hat{s}_t$, denoted $g_{\hat{s}}(\omega)$ is given by

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

$$= G(\omega)g_s(\omega). \tag{3.29}$$

Given that $G(\omega) \leq 1$, it follows that

$$g_{\hat{s}}(\omega) \leq g_s(\omega),$$

and hence the MMSE estimator will underestimate the variance of the theoretical component.

The filter is well defined everywhere when the ϕ-polynomials contain unit roots, and, in fact, extends, in a straightforward manner, to the nonstationary case (see Bell, 1984, and Maravall, 1988). As for the distribution of the estimator $\hat{s}_t$, for the general nonstationary case, assume the polynomial $\phi_s(B)$ can be factorized as

$$\phi_s(B) = \varphi_s(B)D_s,$$

where D_s contains all unit roots, and $\varphi_s(B)$ is a stable polynomial. Multiplying (3.27) by D_s, and replacing $D_s x_t$ by

$$[\theta(B)/\varphi_s(B)\phi_n(B)]a_t,$$

it is obtained that

$$D_s \hat{s}_t = \left[k_s \frac{\theta_s(B)}{\varphi_s(B)} \frac{\theta_s(F)\phi_n(F)}{\theta(F)} \right] a_t, \tag{3.30}$$

which provides the model that generates the stationary transformation of the estimator $\hat{s}_t$. It is seen that MMSE estimation preserves the differencing of the theoretical component, but has an effect on the stationary structure of the model. The part in B of the model generating the estimator is identical to that of the component; the model for the estimator, however, contains a part in F (that gradually converges towards zero), reflecting the contribution of innovations posterior to t to the historical estimator for period t. Theoretical component, given by (3.22), and MMSE estimator will have the same stationary transformation, but the AGF and spectra will differ. Further, it is straightforward to see that the AGF of the historical estimation error,

$$e_t = s_t - \hat{s}_t,$$

is equal to the AGF of the stationary ARMA model

$$\theta(B)z_t = \theta_s(B)\theta_n(B)b_t, \tag{3.31}$$

where b_t is white noise with variance $(V_s V_n)/V_a$ (see Pierce,1979). Stationarity of (3.31) implies that component and estimator are cointegrated.

As was mentioned in Section 2.6, for a finite realization of the x_t process, it will happen that, for periods close enough to both ends of the series, it will not be possible to apply the complete two-sided filter. Denote by $\nu(B, F)$ the

filter in brackets in expression (3.27), namely

$$\nu(B, F) = k_s \frac{\theta_s(B)\phi_n(B)}{\theta(B)} \frac{\theta_s(F)\phi_n(F)}{\theta(F)}, \qquad (3.32)$$

and assume it can be safely truncated after L periods, so that we can write the historical estimator as

$$\hat{s}_t = \nu_0 x_t + \sum_{j=1}^{L} \nu_j (x_{t-j} + x_{t+j}). \qquad (3.33)$$

Let the time series available be $(x_1, \ldots, x_T)$ and, to avoid problems with first observations, let $T > L$. Assume we wish to estimate s_t for $t \leq T$ and $(T - t) > L$, that is for relatively recent periods. According to (3.33), we need $L - (T - t)$ observations at the end of the filter that are not available yet, namely, $x_{T+1}, x_{T+2}, \ldots, x_{T+L-(T-t)}$. Replacing these future values with the ARIMA forecasts computed at time T, we obtain the preliminary estimator. Rewriting (3.33) as

$$
\begin{aligned}
\hat{s}_t &= \nu_L x_{t-L} + \ldots + \nu_0 x_t + \ldots + \nu_{(T-t)} x_T + \\
&+ \nu_{(T-t+1)} x_{T+1} + \nu_{(T-t+2)} x_{T+2} + \\
&+ \ldots + \nu_L x_{t+L},
\end{aligned} \qquad (3.34)
$$

and taking conditional expectations at time T, the preliminary estimator of the signal for time t, when observations end at time T, denoted $\hat{s}_{t|T}$, is given by

$$
\begin{aligned}
\hat{s}_{t|T} &= \nu_L x_{t-L} + \ldots + \nu_0 x_t + \ldots + \nu_{(T-t)} x_T + \\
&+ \nu_{(T-t+1)} \hat{x}_{T+1|T} + \nu_{(T-t+2)} \hat{x}_{T+2|T} + \\
&+ \ldots + \nu_L \hat{x}_{t+L|T}
\end{aligned} \qquad (3.35)
$$

where $\hat{x}_{t_1|t_2}$ denotes the forecasts of x_{t_1} obtained at period t_2. Thus, in compact form, the preliminary estimator can be expressed as

$$\hat{s}_{t|T} = \nu(B, F) x_{t|T}^e \qquad (3.36)$$

where $\nu(B, F)$ is the WK-filter, and $x_{t|T}^e$ is the "extended" series, such that

$$
\begin{aligned}
x_{t|T}^e &= x_t \text{ for } t \leq T \\
x_{t|T}^e &= \hat{x}_{t|T} \text{ for } t > T.
\end{aligned}
$$

The Revision the preliminary estimator will undergo until it becomes the historical one is the difference $(\hat{s}_t - \hat{s}_{t|T})$ or, substracting (3.35) from (3.34),

$$r_{t|T} = \sum_{j=1}^{t+L-T} \nu_{T-t+j} \hat{e}_{T+j|T}, \qquad (3.37)$$

that is, the revision is a linear combination of the forecast errors. Large re-

visions are unquestionably an undesirable feature of a preliminary estimator, and expression (3.37) shows the close relationship between forecast error and revision: the better we can forecast the observed series, the smaller the revision in the preliminary estimator of the signal will be.

Direct application of (3.35), when t is close to the end of the series, may require for models close to noninvertibility (for which $\theta(B)^{-1}$ converges slowly) a very large number of forecasts (perhaps more than 100) in order to complete the filter. The Burman-Wilson algorithm (Burman, 1980), permits us to capture, in a very efficient way, the effect of the infinite forecast function with just a small number of forecasts; for the vast majority of quarterly series, 10 forecasts are indeed enough. A similar procedure can be applied to the first periods of the sample to improve starting values for the signal estimator: one can extend the series at the beginning with backcasts (see Box and Jenkins, 1970), and apply the WK filter to the extended series, using a symmetric Burman-Wilson algorithm.

By combining (3.24) with (3.27), an expression is obtained that relates the final estimator $\hat{s}_t$ to the innovations a_t in the observed series, to be represented by

$$\hat{s}_t = \xi_s(B, F)a_t, \tag{3.38}$$

where $\xi_s(B, F)$ can be obtained from the identity

$$\phi_s(B)\theta(F)\xi_s(B, F) = k_s\theta_s(B)\theta_s(F)\phi_n(F), \tag{3.39}$$

and can be seen to be convergent in F. From (3.38), we can write

$$\hat{s}_t = \xi_s^-(B)a_t + \xi_s^+(F)a_{t+1}. \tag{3.40}$$

When t denotes the last observed period, the first term in (3.40) contains the effect of the starting conditions and of the present and past innovations. The second term captures the effect of future innovations (posterior to t). From (3.40), the concurrent estimator is seen to be equal to

$$\hat{s}_{t|t} = E_t s_t = E_t \hat{s}_t = \xi_s^-(B)a_t,$$

so that the revision

$$r_t = \hat{s}_t - \hat{s}_{t|t}$$

is the (convergent) moving average

$$r_t = \xi_s^+(F)a_{t+1}. \tag{3.41}$$

a zero-mean stationary process. Thus historical and preliminary estimators will also be cointegrated. From expression (3.41) it is possible to compute the relative size of the full revision, as well as the speed at which it vanishes.

The distinction between preliminary estimation and forecasting of a signal

53

is, analytically, inexistent. If we wish to estimate s_t for $t > T$ (i.e., to forecast s_t), expression (3.36) remains unchanged, except that now forecasts will start operating "earlier". For example, if the final estimator is given by (3.34) and the concurrent estimator by

$$\hat{s}_{t|t} = \nu_L x_{t-L} + \ldots + \nu_2 x_{t-2} + \nu_1 x_{t-1} + \nu_0 x_t + \sum_{j=1}^{L} \nu_j \hat{x}_{t+j|t},$$

the one-and two-period-ahead forecasts, $\hat{s}_{t|t-1} = E_{t-1}\hat{s}_t$ and $\hat{s}_{t|t-2} = E_{t-2}\hat{s}_t$, will be given by

$$\hat{s}_{t|t-1} = \nu_L x_{t-L} + \ldots + \nu_2 x_{t-2} + \nu_1 x_{t-1} + \sum_{j=0}^{L} \nu_j \hat{x}_{t+j|t-1};$$

$$\hat{s}_{t|t} = \nu_L x_{t-L} + \ldots + \nu_2 x_{t-2} + \sum_{j=-1}^{L} \nu_j \hat{x}_{t+j|t-2}, \qquad (\nu_1 = \nu_{-1}),$$

and likewise for other horizons. The discussion on revisions in preliminary estimators applies equally to forecasts. A derivation of the estimation errors associated with the different types of estimators can be found in Maravall and Planas (1998).

2. Unknown models

The previous discussion has assumed known models for the unobserved components s_t and n_t. Given that observations are only available on their sum x_t, quite a bit of "a priori" information on the components has to be introduced in order to identify and estimate them. Two approaches to the problem have been followed. One, the so-called "Structural Time Series Model" (STSM) approach, directly specifies models for the components (and ignores the model for the observed series). A trend component, p_t, will typically follow a model of the type

$$\nabla^d p_t = \theta_p(B) a_{pt}, \qquad (3.42)$$

where d=1,2, and $\theta(B)$ is of order ≤ 2; a seasonal component, s_t

$$S s_t = \theta_s(B) a_{st}, \qquad (3.43)$$

with $\theta_s(B)$ also a relatively low order polynomial in B. Irregular components are often assumed white noise or perhaps some highly transitory ARMA model.

A limitation of the STSM approach that has often been pointed out is that the "a priori" structure imposed on the series may not be appropriate for the particular series at hand. This limitation is overcome in the so-called ARIMA Model Based (AMB) approach, where the starting point is the identification of an ARIMA model for the observed series, a relatively well-known problem,

54

and, from that overall model, the appropriate models for the components are derived (there is indeed a close relationship between the STSM and AMB approaches, see Maravall, 1985. The models for the components will be such that their aggregate yields the aggregate model identified for the observations. The models obtained for the trend and seasonal components are also of the type (3.42) and (3.43) and the decomposition may also yield a white noise or a transitory ARMA irregular component. In the applications, we shall use the program SEATS ("Signal Extraction in ARIMA Time Series"; Gómez and Maravall, 1996). The program originated from the work on AMB decomposition of Burman (1980) and Hillmer and Tiao (1982), done in the context of seasonal adjustment, and proceeded along the lines of Maravall (1995) and Gómez and Maravall (2000b).

Although, as we have presented it, the method can be applied to any signal, it has been developed in the context of the basic "trend-cycle + seasonal component + irregular component" decomposition. A summary of this application will prove of help.

3.5   ARIMA-model-based decomposition of a time series

For the type of quarterly series considered in this work ,we briefly summarize the AMB decomposition method.The method starts by identifying an ARIMA model for the observed series. To simplify, assume this model is given by an expression of the type:

$$\nabla \nabla_4 x_t = \theta(B) a_t, \quad a_t \sim niid(0, V_a), \tag{3.44}$$

where we assume that the model is invertible. Next, components are derived, such that they conform to the basic features of a trend, a seasonal, and an irregular component, and that they aggregate into the observed model (3.44). Considering that $\nabla \nabla_4$ factorizes into $\nabla^2 S$, obviously ∇^2 represents the AR $\phi_p(B)$ polynomial for the trend component, and S represents the AR $\phi_s(B)$ polynomial for the seasonal component. The series is seen to contain nonstationary trend (or trend-cycle) and seasonal components, and it can be decomposed into

$$x_t = p_t + s_t + u_t, \tag{3.45}$$

where p_t, s_t, and u_t denote the trend-cycle, seasonal, and irregular components, respectively, the latter being a stationary process. When q (the order of $\theta(B)$) ≤ 5, the following models for the components are obtained

$$\nabla^2 p_t = \theta_p(B) a_{pt}, \quad a_{pt} \sim niid(0, V_p)$$
$$S s_t = \theta_s(B) a_{st}, \quad a_{st} \sim niid(0, V_s) \tag{3.46}$$

$$u_t \sim niid(0, V_u)$$

where a_{pt}, a_{st} and u_t are mutually uncorrelated white noise variables. We refer to (3.46) as the (unobserved component) "structural model" associated with the reduced form model (3.44). Applying the operator $\nabla\nabla_4$ to both sides of (3.45), the identity

$$\theta(B)a_t = S\theta_p(B)a_{pt} + \nabla^2\theta_s(B)a_{st} + \nabla\nabla_4 u_t \qquad (3.47)$$

is obtained. If the l.h.s. of (3.47) is an MA(5) process, setting the order of $\theta_p(B)$, q_p, equal to 2, and that of $\theta_s(B)$, q_s, equal to 3, all terms of the sum in the r.h.s. of (3.47) are also MA(5)'s. Thus we assume, in general, $q_p = 2, q_s = 3$ and equating the AGF of both sides of (3.47), a system of 6 equations is obtained (one equation for each nonzero covariance). The unknowns in the system are the 2 parameters in $\theta_p(B)$, the 3 parameters in $\theta_s(B)$, plus the variances V_p, V_s, and V_u; a total of 8 unknowns. There are not enough equations to identify the parameters, and hence there is, as a consequence, an infinite number of solutions to (3.47). For a more detailed discussion, see Maravall and Pierce(1987).

Denote a solution that implies components as in (3.46) with nonnegative spectra an admissible decomposition. The structural model will not be identified, in general, because an infinite number of admissible decompositions are possible. The AMB method solves this underidentification problem by maximizing the variance of the noise V_u, which implies inducing a zero in the spectra of p_t and s_t in (3.46). The spectral zero translates into a unit root in $\theta_p(B)$ and in $\theta_s(B)$, so that the two components p_t and s_t become noninvertible.

This particular solution to the identification problem is referred to as the "canonical" decomposition (see Box, Hillmer and Tiao, 1978, and Pierce, 1978); from all infinite solutions of the type (3.46), the canonical one maximizes the stability of the trend-cycle and seasonal components that are compatible with the model (3.44) for the observed series. Further, the trend-cycle and seasonal components for any other admissible decomposition can be expressed as the canonical ones perturbated by orthogonal white noise. Also, if the model accepts an admissible decomposition, then it accepts the canonical one (see Hillmer and Tiao, 1982). Notice that, since it should be a decreasing function of ω in the interval $(0, \pi)$, the spectrum of p_t should display the zero at the frequency π. Thus the trend-cycle MA polynomial can be factorized as

$$\theta_p(B) = (1 + \alpha B)(1 + B),$$

where the root B=-1 reflects the spectral zero at π (see Section 2.5). The zero in the spectrum of s_t may occur at $\omega = 0$ or at a frequency roughly halfway between the two seasonal frequencies $\omega = \pi/2$ and $\omega = \pi$.

One simple example may clarify the canonical property. Assume an unob-

served component model for which the trend follows the random-walk model

$$\nabla p_t = a_{pt}, \qquad V_p = 1.$$

This specification is in fact often found in macroeconomic applications of unobserved component models (Stock and Watson, 1988). Part a) of Figure 3.6 displays the spectrum of p_t. It is clear that it does not satisfy the canonical condition because

$$min_\omega g_p(\omega) = g_p(\pi) = .25 > 0.$$

It is straightforward to check that the trend p_t can be decomposed into a canonical trend, p_t^*, plus orthogonal white noise u_t, according to

$$p_t = p_t^* + u_t,$$

where

$$\nabla p_t^* = (1 + B)a_{pt}^*,$$

with $V_{p^*} = .25$. Part b) of Figure 3.6 shows the spectral decomposition of the random walk. The canonical p_t^* is clearly smoother, since it has removed white noise from p_t. The spectral zero for $\omega = \pi$ of the canonical trend is associated with the (1+B) MA polynomial, with unit root B=-1.

Figure 3.6. Canonical Decompostion of a Random Walk

a) Random walk

b) Canonical decomposition
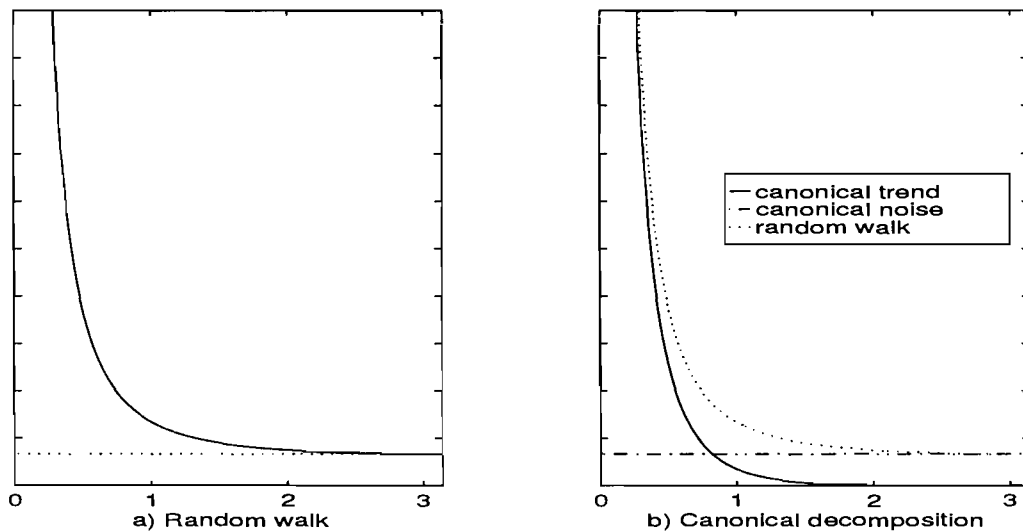
One relevant property of noninvertible series (and hence, of canonical components) is that, due to the spectral zero, no further noise can be extracted from them.

The AMB method computes the trend-cycle, seasonal, and irregular component estimators as the MMSE ("optimal") estimators based on the available

series $X_t = [x_1, \ldots, x_T]$, as described in the previous section . Under our assumptions, these estimators are also conditional expectations of the type E(component | [observed series]), and they are obtained using the WK filter. For a series extending from $t = -\infty$ to $t = \infty$, that follows model (3.44), assume we are interested in estimating a component, which we refer to as the "signal" (the signal will be p_t, then s_t, and finally u_t). Applying result (3.28) to the model (3.46), the WK filter for historical estimation of the trend-cycle component is equal to the AGF of the model

$$\theta(B)z_t = [\theta_p(B)S]b_t, \quad b_t \sim niid(0, V_p/V_a); \tag{3.48}$$

for the seasonal component it is given by the AGF of

$$\theta(B)z_t = [\theta_s(B)\nabla^2]b_t, \quad b_t \sim niid(0, V_s/V_a); \tag{3.49}$$

and for the irregular component, by the AGF of

$$\theta(B)z_t = \nabla\nabla_4 b_t, \quad b_t \sim niid(0, V_u/V_a). \tag{3.50}$$

Notice that this last model is the "inverse" model of (3.44), which is assumed known. Also invertibility of (3.44) guarantees stationarity of the models in (3.48)-(3.50), and hence the three WK filters will converge in B and in F.

For a finite realization, as already mentioned, the optimal estimator of the signal is equal to the WK filter applied to the available series extended with optimal forecasts and backcasts, obtained with (3.44). This is done with the Burman-Wilson algorithm referred to in the previous section.

The following figures illustrate the procedure. Figure 3.7 shows the spectrum of a particular case of model (3.44) and its spectral decomposition into trend, seasonal, and irregular components. The trend captures the peak around $\omega = 0$, and the seasonal component the peaks around the seasonal frequencies. Figure 3.8 displays the WK filters to obtain the historical estimates of the SA series, trend, seasonal and irregular components. From figures 3.8a and b, it is seen, for example, that the concurrent estimator of the SA series requires many more periods to converge to the historical one than that of the trend. Figure 3.9 shows the squared gains of the WK filter (see Section 2.6), that is, which part of the series variation is passed to, or cut-off from, each component. As seen in 3.9c, to estimate the irregular component only the frequencies of no interest for the trend or seasonal component will be employed. Figure 3.10a exhibits a time series of 100 observations generated with the model of Figure 3.7a, and figures 3.10b,c and d the estimates $\hat{n}_{t|100}, \hat{p}_{t|100}$ and $\hat{s}_{t|100}(t = 1, \ldots, 100)$ of the trend, seasonal, and irregular components. Figure 3.11 presents the standard errors of the estimates of Figure 3.10, moving from concurrent to final estimator. The trend estimator converges in a year, while the SA series takes about 3 years for convergence. Finally, Figure 3.12 presents the forecast function of the original series, trend and seasonal components, as well as the

58

associated 90% probability intervals.

Figure 3.7. Spectral AMB Decomposition
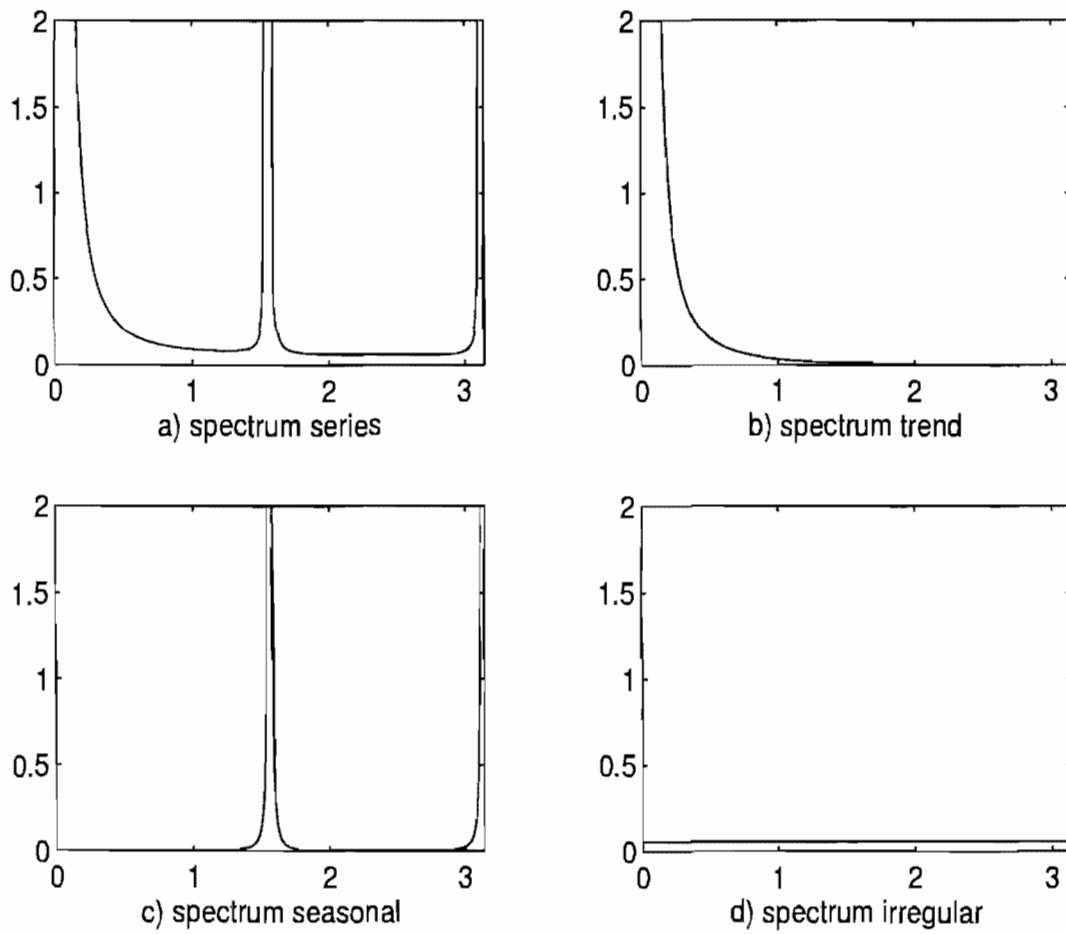

a) spectrum series

b) spectrum trend

c) spectrum seasonal

d) spectrum irregular

Figure 3.8. Wiener–Kolmogorov Filters

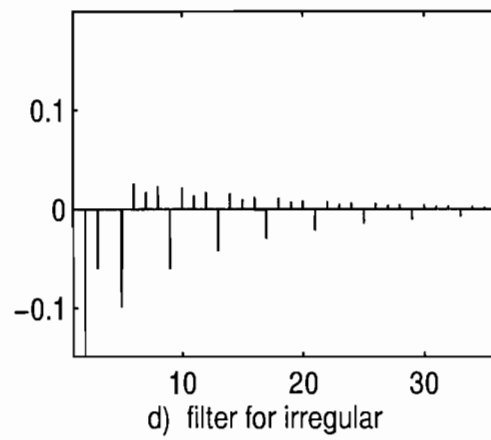a) filter for SA series

b) filter for trend

c) filter for seasonal component

d) filter for irregular

Figure 3.9. Squared Gains

a) SA filter

b) trend–cycle component filter

c) seasonal component filter

d) irregular component filter
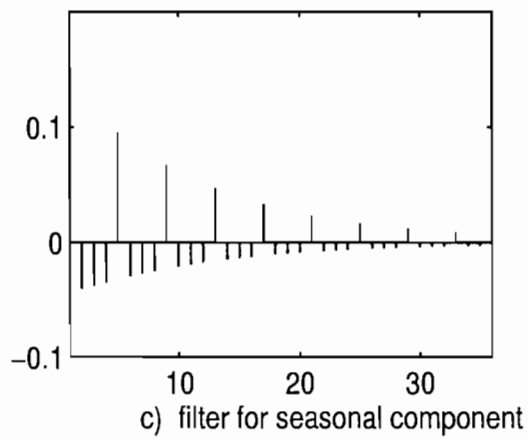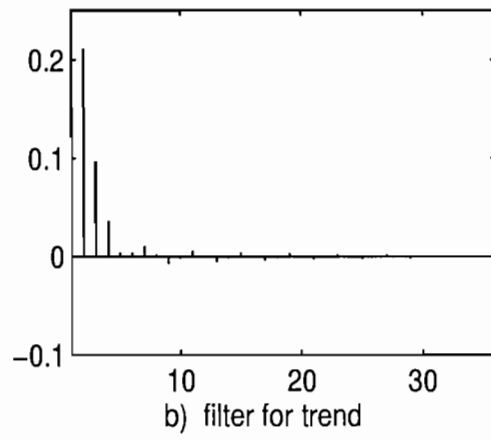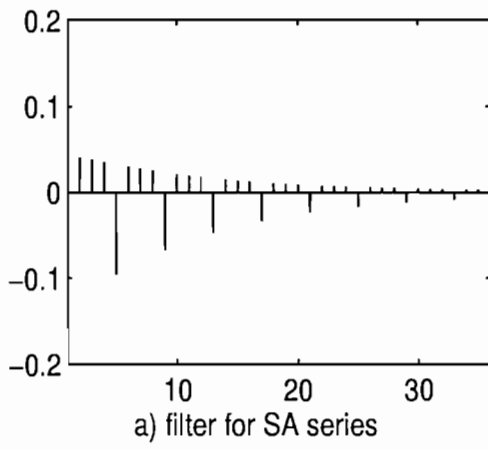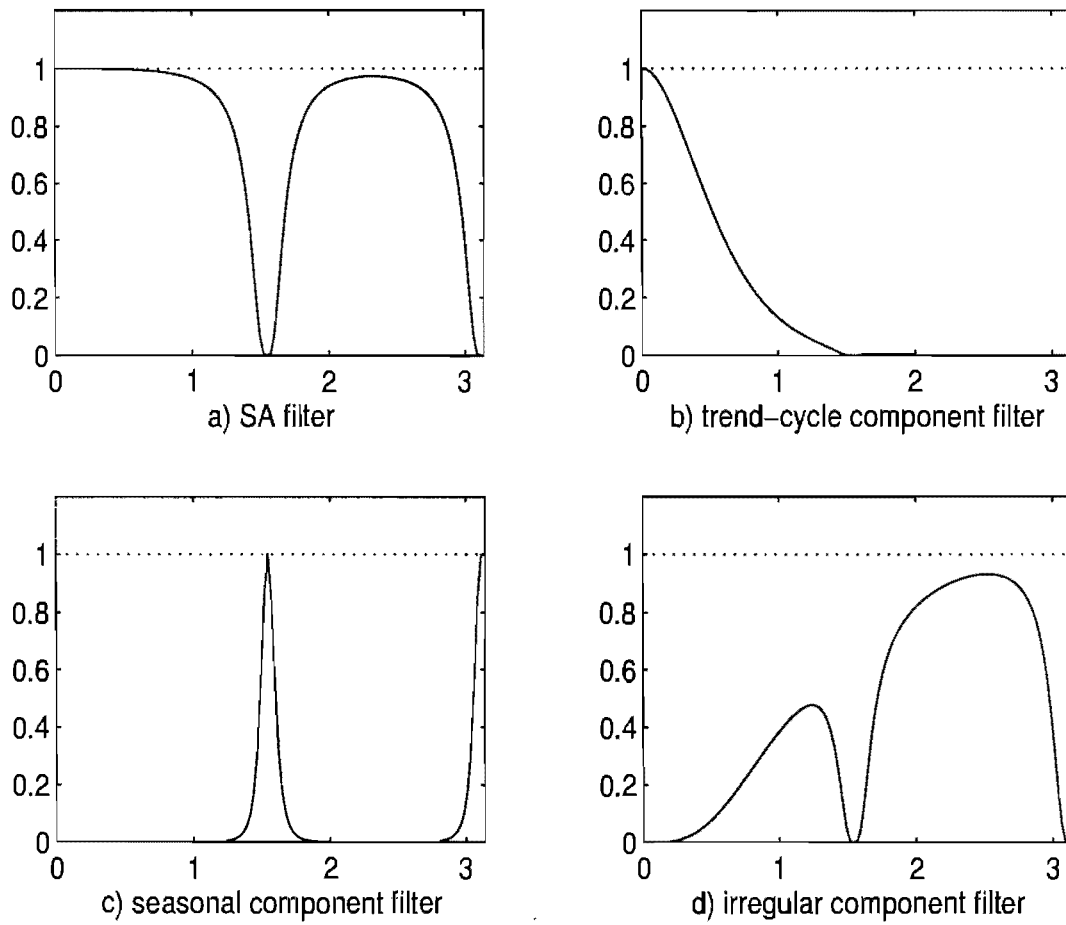
Figure 3.10. Series and Estimated Components

a) Original series
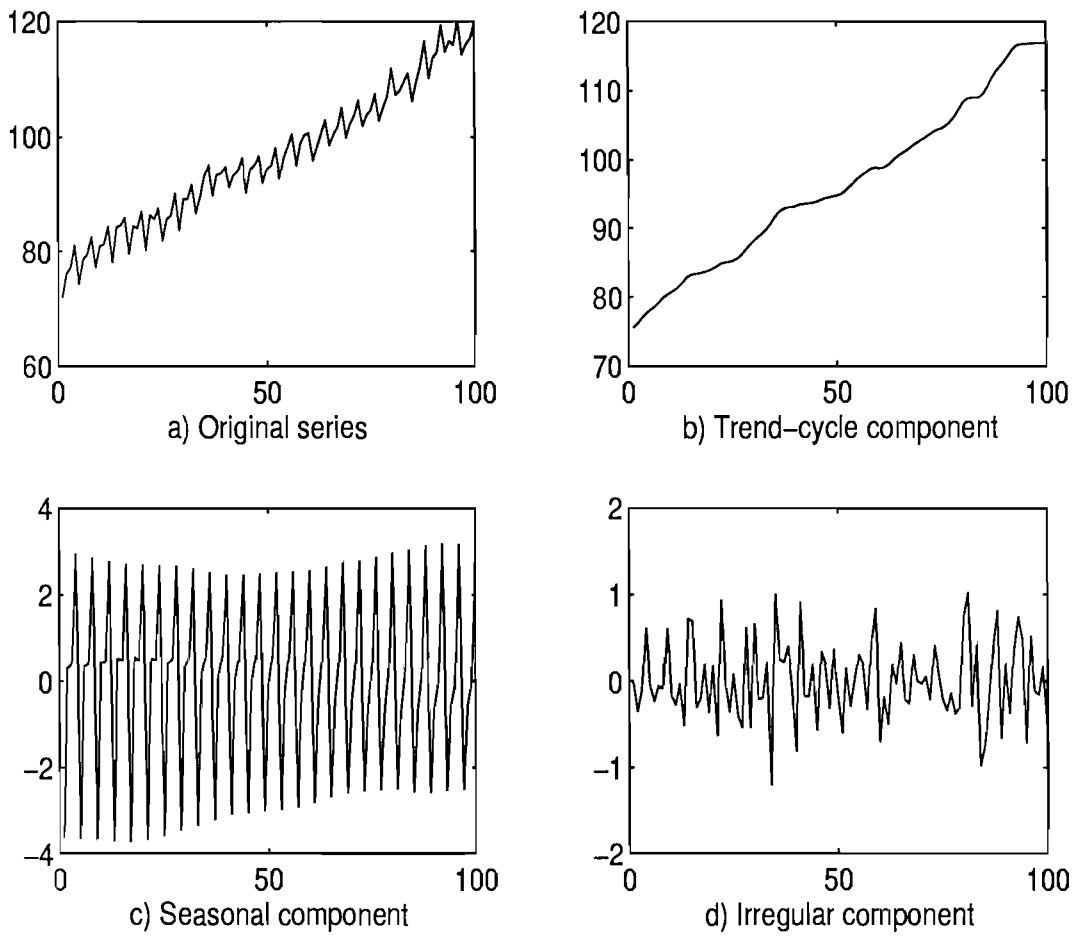
b) Trend-cycle component

c) Seasonal component

d) Irregular component

Figure 3.11. Standard error of estimators

a) SA series

b) Trend-cycle component

c) Seasonal component

Figure 3.12. Forecasts

a) Original series

b) Trend-cycle component

c) Seasonal component
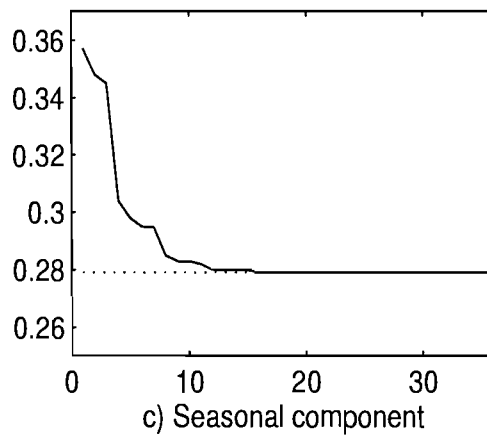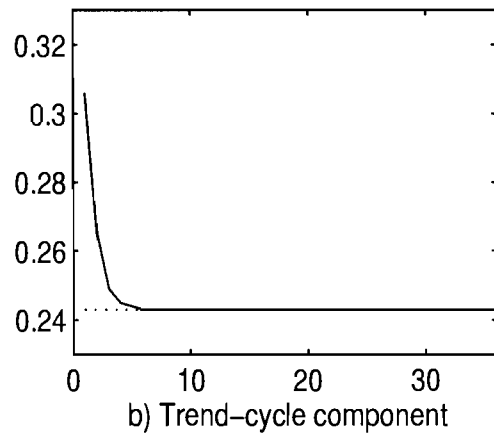
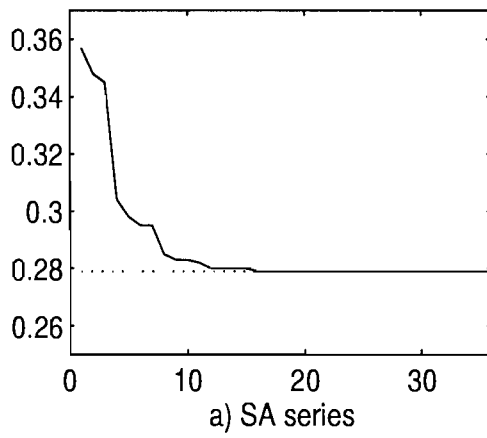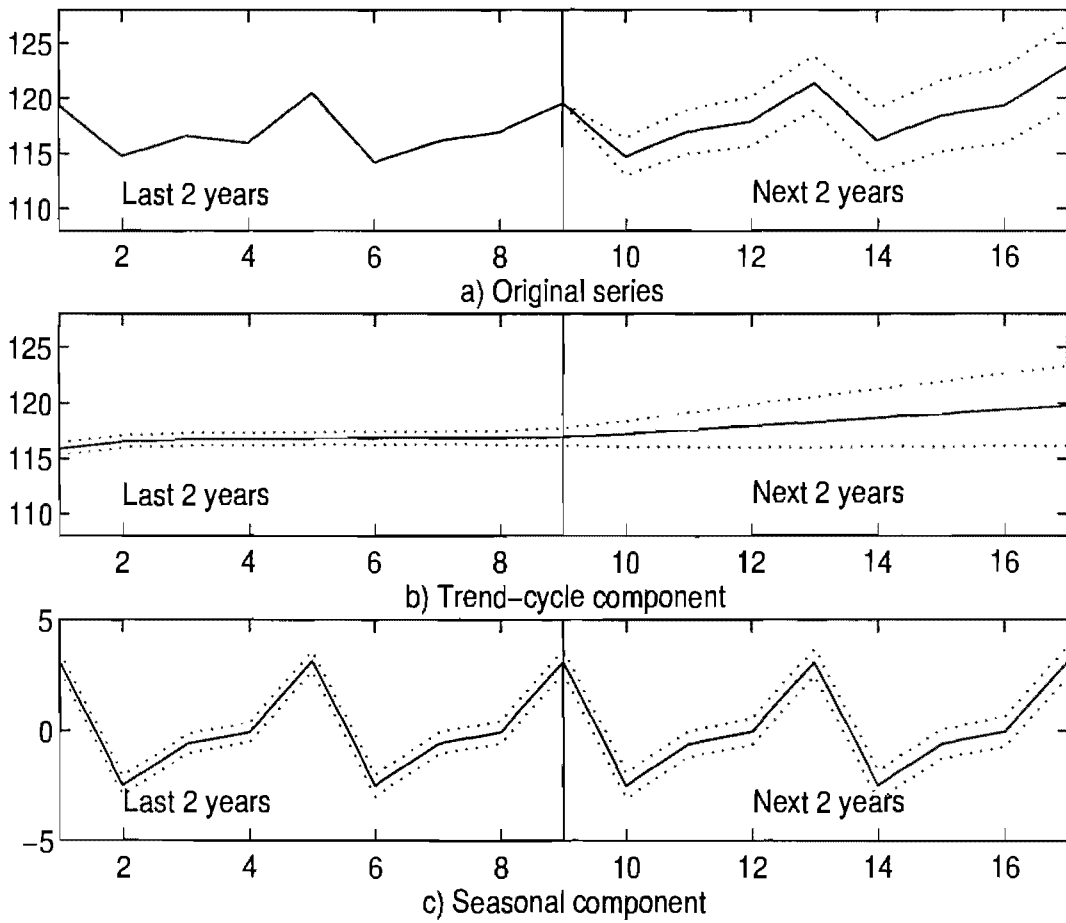## 3.6 Short-term and long-term trends

The previous figures serve also to illustrate an important point, often a source of confusion, namely, the meaning of a trend component. It is a well-known fact that the width of the spectral peak for $\omega = 0$ in parsimonious ARIMA models may vary considerably, so that the same will be true for the squared gain of the trend estimator. Figure 3.13 shows these squared gains for model (3.19), for different combinations of the θ_1 and θ_4 parameters. If the range of cyclical frequencies is broadly defined as starting slightly to the right of $\omega = 0$, and finishing slightly to the left of the fundamental frequency ($\omega = \pi/2$) (so that cycles have periods longer than a year, yet reasonably bounded), then figure 3.13 shows how the squared gain of the trend filter may very well extend over the range of cyclical frequencies, and even exhibit spill-over effects for higher frequencies. This feature is also typical of the squared gains derived from the Structural Time Series Model approach (see Harvey, 1989, and Koopman et al, 1996), and from well-known detrending filter such as the Henderson ones implemented in the X11 family of programs (see Findley et al, 1998).

Figure 3.13

Squared gain for trend—cycle filter: different ARIMA models
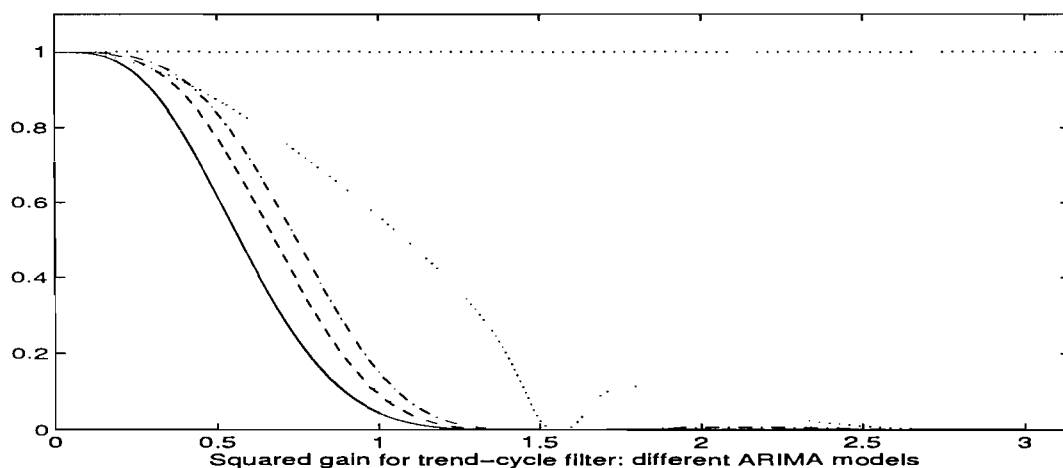
As a consequence, the trend estimators obtained with these procedures may contain a large amount of relatively short-term variation. These short-term trend components should be more properly called trend-cycle components. The contamination of trend with cyclical frequencies is clearly a result of the implicit definition of the trend in the decomposition (3.44). The two components that are removed from the series in order to obtain the trend are the seasonal component and the highly transitory (close to white) noise component. Therefore, the trend is basically defined as the "noise-free SA series", and includes, as a consequence, cyclical frequencies. Its interest rests on the

belief that noise, unrelated to the past and to the future, is more disturbing than helpful in short-term monitoring of the series. (In fact, SA series and trend-cycle components for short-term indicators are both provided at several data-producing agencies; see Eurostat, 1999, and Bank of Spain, 1999.) A discussion of short-term trends is contained in Maravall (1993).

Another important area where trends are used is business-cycle analysis. Here, the trend is also defined as the detrended and SA series, but the concept of detrending is rather different. The aim is to remove a long-term trend that does not include movements with periods shorter than a certain number of years (roughly, the cutting point is set within the range 8 to 10 years). Having defined the band in the frequency range associated with cyclical oscillations (for example, those with period between 2 and 10 years), the issue is to design a "band-pass" filter that permits only the passage of frequencies within that band. Linear filters can only do this job in an approximate manner because the first derivative with respect to ω of their squared gain function is everywhere well defined, and cannot take the form of an exact rectangle, with base the frequency band pass, and height one. The Butterworth family of filters were designed to approximate this band-pass features. One of the members of the family is very well-known in economics, where it is usually called the Hodrick-Prescott (HP) filter (see Hodrick and Prescott, 1980, or Prescott, 1986). Despite the fact that business cycle estimation is basic to the conduct of macroeconomic policy and to monitoring of the economy, many decades of effort have shown that formal modelling of economic cycles is a frustrating issue. As a consequence, applied work and research at economic-policy related institutions has relied (and still relies) heavily on "ad-hoc" band-pass filters and, in particular, in the HP one. One can say that HP filtering of X11-SA series has become the present paradigm for business-cycle estimation in applied work. Figure 3.14 represents, for the example of the previous section, the short-term trend (or trend-cycle component) obtained with the AMB approach, and the long-term trend obtained with the HP-X11 filter. Part a) compares the two squared gains, and part b) the two estimated trends. The short-term character of the AMB trend and the long-term character of the X11-HP trend are clearly discernible.

If business-cycle analysts complain that series detrended with short-term trends, of the type obtained in the AMB approach, contain very little cyclical information, ad-hoc fixed filters to estimate long-term trends are criticized because the trends they yield could be spurious. As seen in Kaiser and Maravall (2000), however, the two types of trends are not in contradiction and can be instead quite complementary. When properly used, their mixture can incorporate the desirable features of the ad-hoc design, with a sensible and complete model-based structure, that fully respects the features of the series

66

at hand. Specifically, the trend-cycle of the AMB decomposition accepts a perfectly sensible model-based decomposition into a long-term trend and a cyclical component, where these two components are closely related to the HP decomposition. The differences, in fact, are those introduced in the Modified HP filter of Kaiser and Maravall (1999), and their aim is to improve end-point estimation, early detection of turning points, and smoothness of the cyclical signal.

Figure 3.14

a) squared gains of trend filters

b) trend estimators
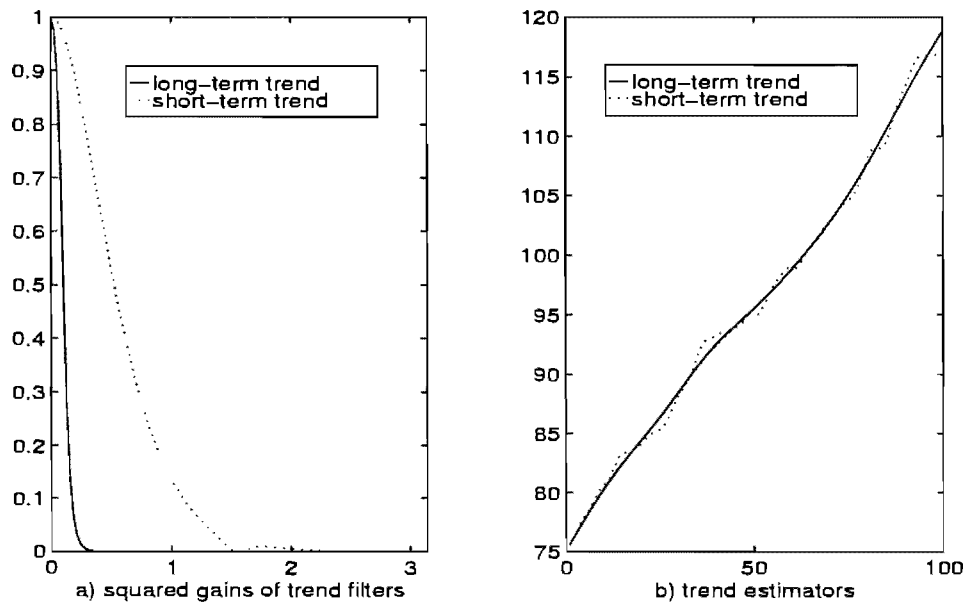
4 References

BANCO DE ESPAÑA (1999), *Boletín Económico*, monthly bulletin, Madrid: Banco de España.

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

BELL, W.R. AND HILLMER, S.C. (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models", *Journal of Time Series Analysis* 12, 283-300.

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 Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series*, Washington, D.C.: U.S. Dept. of Commerce- Bureau of the Census, 309-334.

BOX, G.E.P. AND TIAO, G.C. (1975), " Intervention Analysis with Applications to Economic and Enviromental Problems", *Journal of the American Statistical Association* 70, 71-79.

BROCKWELL,P. AND DAVIS, R. (1987), *Time Series: Theory ans Methods*, Berlin: Springer-Verlag.

BURMAN, J.P. (1980), "Seasonal Adjustment by Signal Extraction", *Journal of the Royal Statistical Society A* 143, 321-337.

BURRIDGE, P. AND WALLIS, K.F. (1984), "Unobserved Component Models for Seasonal Adjustment Filters", *Journal of Business and Economic Statistics* 2, 350-359.

CHEN, C. AND LIU, L.M. (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association* 88, 284-297.

DE JONG, P. (1991), "The Difusse Kalman Filter", *Annals of Statistics* 19, 1073-1083.

EUROSTAT (1999), *Monthly Panorama of European Business*, Luxembourg: Official Publications of the European Communities.

FINDLEY, D.F., MONSELL, B.C., BELL, W.R., OTTO, M.C. AND CHEN, B.C. (1998), " New Capabilities and Methods of the X12 ARIMA Seasonal Adjustment Program" (with discussion), *Journal of Business and Economic Statistics*, 16, 127-177.

FISCHER, B. AND PLANAS, C. (1998) "Large Scale Fitting of ARIMA Models and Stylized Facts of Economic Time Series" Eurostat Working Paper 9/1998/A/8. To appear in *The Journal of Official Statistics.*

GOLDBERG, S. (1967), *Introduction to Difference Equations*, New York: J. Wiley and Sons.

GÓMEZ, V. (1999), "Three Equivalent Methods for Filtering Nonstationary Time Series". *Journal of Business and Economic Statistics* 17, 109-116.

GÓMEZ, V. AND MARAVALL, A. (2000a), "Automatic Modelling Methods for Univariate Series", Working Paper 9808, Research Department, Bank of Spain. To appear as Ch.7 in Peña D., Tiao G.C. and Tsay, R.S. (eds.) *A Course in Advanced Time Series Analysis*, New York: J. Wiley and Sons.

GÓMEZ, V. AND MARAVALL, A. (2000b), "Seasonal Adjustment and Signal Extraction in Economic Time Series". Working Paper 9809, Research Department, Bank of Spain. To appear as Ch.8 in Peña D., Tiao G.C. and Tsay, R.S. (eds.) *A Course in Advanced Time Series Analysis*, New York: J. Wiley and Sons.

GÓMEZ, V. AND MARAVALL, A. (1996), "Programs TRAMO and SEATS; Instructions for the User" (with some updates). Working Paper 9628, Research Department, Banco de España.

GÓMEZ, V. AND MARAVALL, A. (1994), "Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman filter", *Journal of the American Statistical Association* 89, 611-624.

GÓMEZ, V. AND MARAVALL, A. (1993), "Initializing the Kalman filter with Incompletely Specified Initial Conditions" in Chen, R. (ed.) *Approximate Kalman Filter (Series on Approximation and Decomposition)*, London: World Scientific Publ. Co. 39-62.

GOURIEROUX, C. AND MONFORT, A. (1990), *Séries Temporelles et Modéles Dynamiques*, Paris: Economica.

GRANGER, C.W.J. AND NEWBOLD, P. (1986) *Forecasting Economic Time Series*, New York: Academic Press.

GRENANDER, U. AND ROSENBLATT, M. (1957), *Statistical Analysis of Stationary Time Series*, New York: John Wiley.

HARVEY, A.C. (1993), *Time Series Models*, Deddington: Philip Allan.

HARVEY, A.C. (1989), *Forecasting Structural Time Series and the Kalman Filter*, Cambridge: Cambridge University Press.

HATANAKA, M. AND SUZUKI, M. (1967), " A Theory of the Pseudospectrum and its Application to Nonstationary Dynamic Econometric Models", in Shubik, M. (ed.), *Essays in Mathematical Economics in Honor of Oskar Morgenstern*, Princeton, N.J.: Princeton University Press.

HENDRY, D.F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press.

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.

HILLMER, S.C., BELL, W.R. AND TIAO, G.C. (1983), "Modelling Considerations in the Seasonal Adjustment of Economic Time Series", in Zellner, A. (ed.), *Applied Time Series Analysis of Economic Data*, Washington, D.C.: U.S. Department of Commerce-Bureau of the Census, 74-100.

HODRICK, R. AND PRESCOTT, E.C. (1980), "Post-war U.S. Business Cycles: An Empirical Investigation", mimeo, Carnegie-Mellon University, Pittsburgh, P.A.

JENKINS, G.M. AND WATTS, D.G. (1968), *Spectral Analysis and its Applications*, San Francisco: Holden Day.

KAISER, R. AND MARAVALL, A (2000), *Measuring Business Cycles in Economic Time Series*, Lectures Notes in Statistics. Springer Verlag NY. Forthcoming.

KAISER, R. AND MARAVALL, A (1999), "Estimation of the Bussines Cycle: a Modified Hodrick-Prescott filter", *Spanish Economic Review* 1, 175-206.

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.

KOOPMAN, S.J., HARVEY, A.C., DOORNIK, J.A. AND SHEPHARD, N. (1996), *Stamp: Structural Time Series Analyser, Modeller and Predictor*, London: Chapman and Hall.

MARAVALL, A. (2000), "An application of TRAMO and SEATS", Working Paper 9914, Research Department, Bank of Spain. To appear in *Seasonal Adjustment Procedures. Experiences and Perspectives*, Rome, ISTAT.

MARAVALL, A. (1999), "Short-Term Analysis of Macroeconomic Time Series". in A. Gerard-Varet and A Kirman (eds.), *Economics: Beyond the Milennium*, Oxford: Oxford University Press.

MARAVALL, A. (1995), "Unobserved Components in Economic Time Series", in Pesaran, H. and Wickens ,M. (eds.), *The Handbook of Applied Econometrics, vol. 1*, Oxford: Basil Blackwell.

MARAVALL, A. (1993), "Stochastic Linear Trends: Models and Estimators", *Journal of Econometrics* 56, 5-37.

MARAVALL, A. (1988), "A Note on Minimum Mean Squared Error Estimation of Signals with Unit Roots", *Journal of Economic Dynamics and Control* 12, 589-593.

MARAVALL, A. (1985), "On Structural Time Series Models and the Characterization of Components", *Journal of Business and Economic Statistics* 3, 4, 350-355. Reprinted in Harvey, A.C. (ed.), *Time Series*, Cheltenham: Edward Elgar Publ., 1994.

MARAVALL, A. AND PIERCE, D.A. (1987), "A Prototypical Seasonal Adjustment Model", *Journal of Time Series Analysis* 8, 177-193. Reprinted in Hylleberg, S. (ed.), *Modelling Seasonality*, Oxford University Press, 1992.

MARAVALL, A. AND PLANAS, C. (1999), "Estimation Error and the Specification of Unobserved Component Models", *Journal of Econometrics*, 92, 325-353.

MILLS, T.C. (1990), *Time Series Techniques for Economists*, Cambridge: Cambridge University Press.

MORF, M., SIDHU, G.S. AND KAILATH, T. (1974), "Some New Algorithms for Recursive Estimation on Constant, Linear, Discrete-Time Systems", *IEEE Transactions on Automatic Control*, AC 19, 315-323.

NEWBOLD, P. (1983), "Model Checking in Time Series Analysis", in A. Zellner (ed.) *Applied Time Series Analysis of Economic Data*, Washington, D.C.: U.S. Department of Commerce-Bureau of the Census, 133-143.

PIERCE, D.A. (1979), "Signal Extraction Error in Nonstationary Time Series", *Annals of Statistics* 7, 1303-1320.

PIERCE, D.A. (1978), "Seasonal Adjustment when Both Deterministic and Stochastic Seasonality are Present", in Zellner, A. (ed.), *Seasonal Analysis of Economic Time Series*, Washington, D.C.: U.S. Dept. of Commerce-Bureau of the Census, 242-269.

PLANAS, C. (1997), *Applied Time Series Analysis. Modelling, Forecasting, Unobserved Components, and the Wiener-Kolmogorov Filter.* Luxembourg Office for Official Publications of the European Communities.

PRESCOTT, E. (1986), "Theory ahead of Business Cycle Measurement", *Carnegie-Rochester Conference Series on Public Policy* 25, 11-66.

SHISKIN, J., YOUNG, A.H. AND MUSGRAVE, J.C. (1967), "The X11 Variant of the Census Method II Seasonal Adjustment Program", Technical Paper, 15, Washington, D.C.: Bureau of the Census.

STOCK, J.H. AND WATSON, M.W. (1988), " Variable Trends in Economic Time Series", *Journal of Economic Perspectives* 2, 147-174.

TIAO, G.C. AND TSAY, R.S. (1989), "Model Specification in Multivariate Time Series", *Journal of the Royal Statistical Society B*, 51, 132-141.

TIAO, G.C. AND TSAY, R.S. (1983), "Consistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models", *The Annals of Statistics* 11, 856-871.

WHITTLE, P. (1963), *Prediction and Regulation by Linear Least-Squares Methods*, London: English Universities Press.